ISSN Print: 2518-4245
ISSN Online: 2518-4253
Vol. 61(2), June 2024

PROCEEDINGS
OF THE PAKISTAN ACADEMY OF SCIENCES:
A. Physical and Computational Sciences

PAKISTAN ACADEMY OF SCIENCES
ISLAMABAD, PAKISTAN

Proceedings of the Pakistan Academy of Sciences: Part A Physical and Computational Sciences

Proceedings of the Pakistan Academy of Sciences A. Physical and Computational Sciences is the official flagship, the peer-reviewed quarterly journal of the Pakistan Academy of Sciences. This open-access journal publishes original research articles and reviews on current advances in the field of Computer Science (all), Materials Science (all), Physics and Astronomy (all), Engineering Sciences (all), Chemistry, Statistics, Mathematics, Geography, Geology in English. Authors are not required to be Fellows or Members of the Pakistan Academy of Sciences or citizens of Pakistan. The journal is covered by Print and Online ISSN, indexed in Scopus, and distributed to scientific organizations, institutes and universities throughout the country, by subscription and on an exchange basis.

Annual Subscription: **Pakistan:** Institutions, Rupees 8000/-; Individuals, Rupees 4000/- (Delivery Charges: Rupees 300/-)
Other Countries: US$ 200.00 (includes air-lifted overseas delivery)

HEC Recognized; Scopus Indexed

Published by **Pakistan Academy of Sciences**, 3 Constitution Avenue, G-5/2, Islamabad, Pakistan
Email: editor@paspk.org; **Tel:** 92-51-920 7140 & 921 5478; **Websites:** www.paspk.org/proceedings/; www.ppaspk.org

Printed at **Graphics Point.,** Office 3-A, Wasal Plaza, Fazal-e-Haq Road Blue Area Islamabad.
Ph: 051-2806257, **E-mail:** graphicspoint16@gmail.com

PROCEEDINGS OF THE
PAKISTAN ACADEMY OF SCIENCES: PART A
Physical and Computational Sciences

CONTENTS

Volume 61, No. 2, June 2024 **Page**

Review Articles

Research Articles

Instructions for Authors

Submission of Manuscripts: Manuscripts may be submitted as an e-mail attachment at editor@paspk.org or submit online at http://ppaspk.org/index.php/PPASA/about/submissions. Authors must consult the ***Instructions for Authors*** at the end of this issue or at the Website: www.paspk.org/proceedings/ or www.ppaspk.org.

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 99-112 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)676

Pakistan Academy of Sciences

Review Article

Analytical Techniques for Elemental Analysis: LIBS, LA-TOF-MS, EDX, PIXE, and XRF: A Review

Muhammad Aslam Baig[1*], Amir Fayyaz[1], Rizwan Ahmed[2], Zeshan Adeel Umar[2], Haroon Asghar[2], Usman Liaqat[3], Rinda Hedwig[4], and Koo Hendrik Kurniawan[4]

[1]Atomic and Molecular Physics Laboratory, Quaid-i-Azam University, Islamabad, Pakistan
[2]National Centre for Physics, Quaid-i-Azam University Campus, 45320 Islamabad, Pakistan
[3]School of Chemical and Materials Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan
[4]Research Centre of Maju Makmur Mandiri Foundation, 40/80 Srengseng Raya, Jakarta 11630, Indonesia

Abstract: This contribution presents the applications of laser-induced breakdown spectroscopy (LIBS) for qualitative and quantitative elemental analysis of several samples such as geological ores, cement, and industrial materials. A combination of LIBS with the laser ablation time–of–flight mass spectrometer (LA-TOF-MS) assists in elemental quantification and isotopic abundance analysis. Different elemental components such as aluminum (Al), calcium (Ca), iron (Fe), potassium (K), lithium (Li), magnesium (Mg), sodium (Na), rubidium (Rb), silicon (Si), strontium (Sr), and titanium (Ti) in cement varieties, as well as rare earth metals, lanthanum (La), cerium (Ce), and neodymium (Nd) in geological ore samples, and cobalt (Co), chromium (Cr), nickel (Ni), iron (Fe), and molybdenum (Mo) in industrial standard material samples, have been precisely identified and their relative weight percentages estimated using CF-LIBS and LA-TOF-MS diagnostic techniques. To ensure the reliability of LIBS for the elemental analysis of industrial materials, a comparative study with other compatible techniques, including EDX, PIXE, and XRF, utilizing different excitation sources is presented. The results suggest that LIBS offers unique advantages, enabling rapid elemental analysis in circumstances where other techniques may face certain limitations.

Keywords: Calibration Free-LIBS, LA-TOF-MS, PIXE, XRF, EDX.

1. INTRODUCTION

Laser-induced breakdown spectroscopy (LIBS) is an analytical atomic emission spectroscopy technique. It is rapid, reliable, efficient, micro-destructive, and compatible with the elemental analysis of industrial [1, 2], geological [2], and pharmaceutical samples [3]. LIBS is based on the laws of quantum mechanics describing the discrete nature of atomic levels and the generation of a micro-plasma on the surface of the target sample [3-10]. LIBS employs high-power short laser pulses to create a high-temperature plasma on the target surface. Essentially, the laser beam is focused on the target surface, causing a small portion of the sample to be ablated. This ablated material then interacts with the latter part of the laser pulse, leading to the formation of a highly energetic plasma comprising free electrons, ions, and neutral particles. As the plasma undergoes cooling, electrons in the excited states of atoms and ions make transitions to their characteristic ground states. This process results in the emission of light characterized by distinct spectral lines, bands, and an overarching continuum. These unique lines, which describe the material, have two primary characteristics; spectral lines at specific wavelengths and relative intensities. These factors depend on both the composition of the emitting

Received: February 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Muhammad Aslam Baig <baig@qau.edu.pk>

atoms and their atmosphere. The emitted light from the plasma is collected through a CCD/ICCD detection system covering a particular wavelength range. The LIBS data are then drawn as an emission spectrum using commercially available software. Characteristic LIBS spectral emission lines are associated with each element in the Periodic Table. The elemental composition of the trace and major elements in the sample can be quickly determined using the LIBS spectrum [11-15].

For the last couple of years, various LIBS research articles on elemental quantification analysis have been published using different LIBS techniques, such as Calibration Free-LIBS (CF-LIBS) including the Boltzmann plot technique, electron number-density conservation method, and Calibration Curve-LIBS (CC-LIBS) with ANN, univariate analysis, multiple linear-regression, PCA, PLS, and SVM [9, 10, 16-21]. CF-LIBS was first proposed by Ciucci *et al*. [11] for the elemental study of metallic alloys and the estimation of the chemical composition of the atmosphere. They proposed a mathematical model explaining the physical conditions of the laser-produced plasmas to estimate chemical composition. With CF-LIBS, referenced samples or calibration curves are unnecessary, and effective avoidance of matrix effects is possible. The application of the CF-LIBS technique involves taking into account the following key assumptions: ensuring identical chemical composition between plasmas and samples (i.e., stoichiometric ablation), expecting particles within the plasma, when elevated in energy levels, to adhere to the Boltzmann distribution (i.e., locally in thermal equilibrium), ensuring that spectral lines are devoid of self-absorption (optically thin plasma), and recording spectra that include all elemental species. The CF-LIBS method has been applied to examine various samples, including alloys containing Al, C-Fe, Fe, and Cu, as well as diverse geological and raw materials such as sedimentary rocks, pigments found on Roman frescoes and parchments, meteorites, glass, soil, kidney stones, cement, coral skeletons, and human hair. Nevertheless, the precision and accuracy of the CF-LIBS technique still need improvement [12-15, 22, 23].

In this review, analyses of various samples including geological ores containing rare earth metals, cement, and industrial alloys are presented. The investigation involves qualitative studies through LIBS and quantitative analysis utilizing a combination of CF-LIBS, laser ablation time–of–flight mass spectrometry (LA-TOF-MS), energy dispersive X-ray (EDX), X-ray fluorescence (XRF), and proton-induced X-ray emission (PIXE) diagnostic analytical methods. The results obtained using CF-LIBS and LA-TOF-MS are compared and cross-validated with those obtained from EDX, XRF, and PIXE, focusing on major, minor, and trace constituent elements within different samples. The primary aim of this review is to explore the capabilities of LIBS for the chemical quantification of novel materials.

2. MATERIALS AND METHODS

2.1. Calibration Free LIBS (CF-LIBS) Method

In the Calibration-Free LIBS (CF-LIBS) approach, the initial step for elemental analysis involves constructing Boltzmann plots using various spectroscopic parameters, such as the relative intensity of observed spectral lines, transition probability, upper-level energy, statistical weight, and central wavelength corresponding to the constituent elements. This process assumes that the plasma is in local-thermal equilibrium (LTE). The Boltzmann distribution function for the population density of the excited state can be expressed as follows [24-28]:

$$n_k^s = n^s \frac{g_k}{P(T)} \exp\left(-\frac{E_k}{k_B T}\right) \quad (1)$$

Here, n_k^s represents the population concentration for the excited level k of the species s, n^s denotes the total concentration for elements s in an LTE plasma, g_k stands for the statistical mass for a transition of the upper-state, and $P(T)$ is the quantum statistical sum. The partition function, defined as $P(T) = \sum_k g_k \exp(-E_k/k_B T)$, involves the summation of statistical weights up to the level k where the energy E_k is associated with the upper level, k_B signifies the Boltzmann constant, and T is the plasma temperature. In the context of an optically-thin plasma state, the intensity of an emission line I_{ki} is expressed as proportional to n_k^s. Incorporating this intensity relation in Equation (1), we get:

$$I_{ki} = A_{ki} n^s \frac{hc}{\lambda_{ki}} \frac{g_k}{P(T)} \exp\left(-\frac{E_k}{k_B T}\right) \quad (2)$$

The measured intensity is directly influenced by the efficiency of the collecting method; hence, Equation (2) is formulated as:

$$\overline{I_{ki}} = FC^S A_{ki} \frac{g_k}{P(T)} \exp\left(-\frac{E_k}{k_B T}\right) \quad (3)$$

Where, $\overline{I_{ki}}$ is the net integrated emission intensity of the spectral line, F is the experimental factor, i.e., depending on the efficiency of the collecting method and size of the LTE plasma, and C^S denotes the chemical weight of the emitted constituents. By ensuring a natural log on both sides of Equation (3) [20-27], the following relationship is obtained:

$$\ln\left[\frac{\lambda_{ki}\overline{I_{ki}}}{hcA_{ki}g_k}\right] = -\frac{E_k}{k_B T} + \ln\left[\frac{FC^S}{P(T)}\right] \quad (4)$$

Equation (4) is a straight-line equation of the type, y = bx + q_S.

$$y = \ln\left[\frac{\lambda_{ki}\overline{I_{ki}}}{hcA_{ki}g_k}\right]; x = E_k; q_S = \ln\left[\frac{FC^S}{P(T)}\right] \quad (5)$$

A Boltzmann plot can be characterized for each species by taking $\ln[\lambda_{ki}\overline{I_{ki}}/hcA_{ki}g_k]$ on the y-axis and E_k on the x-axis. In the Boltzmann plot, the reciprocal of the slope ($b = -1/k_B T$) yields the plasma temperature and q_S is determined from the intercept along the y-axis. Finally, the chemical weight of each element in the material is calculated from the resulting relationship:

$$C^S = \frac{1}{F} P(T) \exp(q_S) \quad (6)$$

The Sah-Boltzmann equation can be applied to estimate the chemical weight of the ionic components present in the sample plasma. In brief, the following equation is used to calculate the chemical weight of each ionic constituent in the sample [28-32].

$$\frac{n^{\alpha,z}}{n^{\beta,z+1}} = \frac{I_{\alpha,z}}{I_{\beta,z+1}} \times \frac{P_{\alpha,z}}{P_{\beta,z+1}} \exp\left(-\frac{E_{k,\beta,z+1} + E_{k,\alpha,z}}{k_B T}\right) \quad (7)$$

where, $E_{k,\alpha,z}$ and $E_{k,\alpha,z+1}$ represent the ionization energies of integral component α in charge states z and z+1, respectively. Similarly, $n^{\alpha,z}$, and $n^{\beta,z+1}$ denote the electron plasma number-density of element α in charge states z and z+1, respectively. In the same way, $P_{\alpha,z}$ and $P_{\beta,z+1}$stand for the quantum statistical sum of the higher charge-state (z) and lower charge-state (z+1). The concentration contribution of any element in the sample is determined by summing up both neutral and ionic contributions. The individual contribution to the total chemical weight ratio of any element is then computed using the following relationship:

$$n^{\alpha} = n^{\alpha,z} + n^{\alpha,z+1} \quad (8)$$

$$C^{\alpha}(\%) = \frac{n^{\alpha}}{\sum n^{\alpha}} \times 100 \quad (9)$$

2.2. Materials

In this study, we investigated 5 geological samples containing rare earth elements, 7 samples of cement, and 4 molybdenum (Mo) based alloy samples for elemental analysis. The rare earth ore samples were collected from a mineralogical and geological site in the northern sectors of Gilgit-Baltistan, Pakistan. To prepare the samples for LIBS, we subjected them to a 40-minute ultrasonic bath with acetone, followed by drying in a furnace at 100 °C for 90 minutes to eliminate moisture. Subsequently, the target samples were excellently ground as a powder form to achieve reliable uniformity, and pellets having a surface area of ~0.95 cm^2 and a thickness of 0.5 cm were formed by applying a pressure of about 7 tons. Similarly, a total of seven cement samples labeled HS-1, HS-2, HS-3, HS-4, HS-5, HS-6, and HS-7 were obtained from various production facilities for cement in Pakistan. Pellets having a thickness of about 0.5 cm and a surface area of ~1.77 cm^2 were formed using a hydraulic press without the addition of any adhesive substance. For the Mo-based alloy samples with different compositions (purity > 99.95%), we employed arc melting in an environment enriched with argon gas within a copper hearth cooled by water. Before the insertion of pure argon, a vacuum of 10^{-6} mbar was established. A Ti–getter underwent melting for a duration of one minute to absorb any oxygen remnants prior to the real sample melting process. Each sample underwent five re-melting cycles, with flipping after each cycle to confirm chemical consistency. Although multiple samples were examined, this contribution focuses on presenting a detailed analysis of one representative sample from each category.

3. ANALYTICAL TECHNIQUES

3.1. Laser Induced Breakdown Spectroscopy (LIBS)

The custom-built tabletop LIBS experiments have

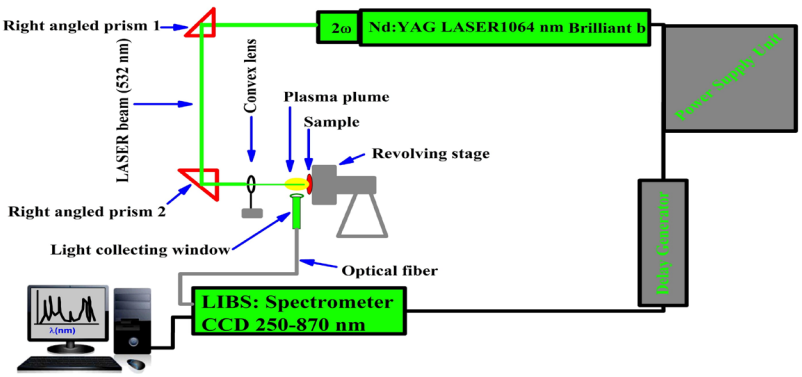

Fig. 1. A schematic diagram of the experimental setup of the LIBS system.

been performed by using a Q–switched Nd:YAG (Brilliant–B, Quantel (France)) laser beam (532 nm wavelength) having a pulse energy of 110 mJ, 10 Hz repetition frequency, and pulse interval of 5 ns that was concentrated by a converging lens of 10 cm focal length on the target surface to generate the plasma plume. Figure 1 shows an experimental arrangement for the LIBS analyses.

The sample was placed on a rotating stand, continuously revolving around an axis. The utilization of the rotating stand was preferred due to several advantageous effects. These included providing a clean area of the target substance to each single pulse, enhancing the replicability of mass ablation, averting the growth of intense trash, and avoiding non-uniformity of the target. The focused laser beam at the sample side had a spot size of around 0.5 mm, corresponding to a laser fluence of around 55 Jcm^{-2}. To capture the plasma optical spectrum, an optical fiber with a focusing lens (positioned vertically to the laser beam) featuring high–OH and a core diameter of approximately 600 µm was employed. Multiple compact spectrometers, each outfitted with a 10 µm slit-size in the detection system from Avantes, Holland, were utilized to capture the optical spectrum across the wavelength from 250 to 870 nm. A delay-time of 2 µs among the laser-beam emission and the reading of the detection setup was implemented, with a gate width of about 1 ms.

The time-resolved spectra of several elements were also recorded using a setup at the Maju Makmur Mandiri Research Center, Indonesia. In this setup, the light emission signal by the laser-produced plasma was captured using an optical fiber with a numerical aperture of 0.27 which was attached to a 1.0 m Czerny Turner type spectrograph outfitted with a 5.0 µm wide entrance slit and an 1800 grooves/mm grating. The resolution of this spectrograph at 313.1 nm is about 0.012 nm. The exit slit of the spectrograph was attached with a time-gated ICCD (intensified (Andor i–Star, UK) CCD, (1024) × (256) pixels,). To synchronize the delay time between the laser pulse and the ICCD, the

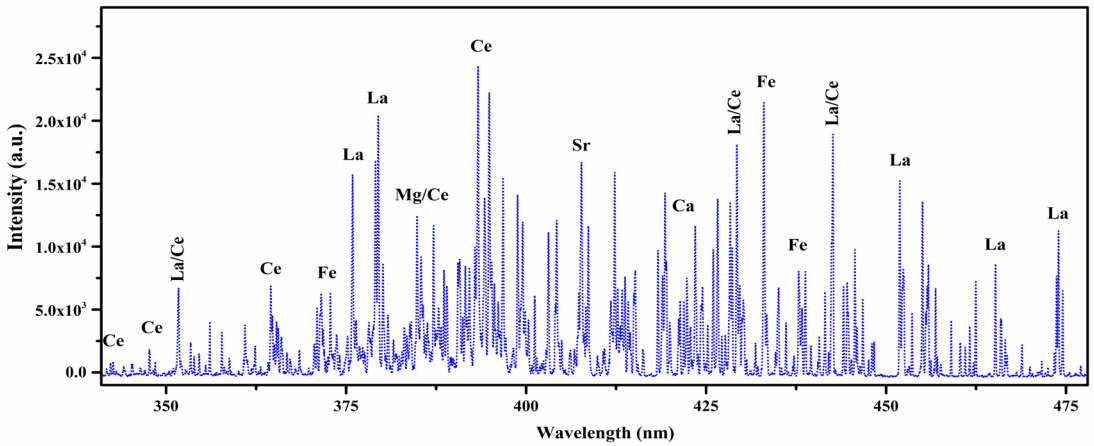

Fig. 2. LIBS emission spectrum of the rare earth geological sample from 335 nm to 480 nm wavelength region.

ICCD was triggered using an output signal from the laser system. The delay-time of the ICCD was set at 1 µs and the corresponding gate-delay was adjusted at 30 µs. Acquisition and analysis of the data were carried out with an individual computer system. All the experiments were conducted under atmospheric air pressure and the spectra were registered through the averaging of ten single-shot datasets under similar investigational circumstances, aiming to enhance the signal-to-noise ratio.

In Figure 2, we display the emission spectrum of the rare earth geological sample having a wavelength range from 335 nm to 480 nm. The emission spectrum shows the dominating emission lines of rare earths along with other elements such as Ca, Fe, Mg, Na, Si, and Ti. In Figure 3, the emission spectrum (covering the spectral range of 420-562 nm) of a cement sample is presented. The National Institute of Standards and Technology (NIST) database system [33] enabled the identification of each spectral line in the spectrum. In the cement test sample, the emission lines of Ca (I) and Ca (II) predominate, followed by Fe, K, Li, Mg, Na, Sr, and Ti. The signal height is indicative of the component concentration in the sample. A calibration-free LIBS procedure was employed for quantitative analysis, depending on plasma parameters derived from Boltzmann plots of the essential elements. The calculated weight percentage (wt.%) of the Ca in the various cement samples is around (71-76 %). The wt.% of the other elements was estimated to be in the range of; Al (1-3%), Fe (3-8%), K (2-5%), Li (0.3-1.3%), Mg (1-3.9%), Na (1-4%), Si (6-14%), Sr (0.1-0.4%), and Ti (0.35-0.95%). Figure 4 displays the emission spectrum of the Mo-based alloy sample, containing the wavelength scale from 257 nm to 392 nm. The spectrum reveals the principal presence of Mo emission lines, with cobalt (Co), nickel (Ni), chromium (Cr), and iron (Fe) lines.

To perform quantitative analysis, the number-density of electron plasma (N_e) was determined by calculating the full width at half–maximum (FWHM) of the broadening of the Stark line using the well-recognized equation [12, 14, 21].

$$N_e = \frac{\Delta\lambda_{1/2}}{2\omega_s} \times 10^{16}\ (cm^{-3}) \qquad (10)$$

Where the contribution of Stark broadening to the overall line profile is denoted by $\Delta\lambda^{s}_{FWHM}(nm)$ and the Stark broadening parameter by ω_s (nm).

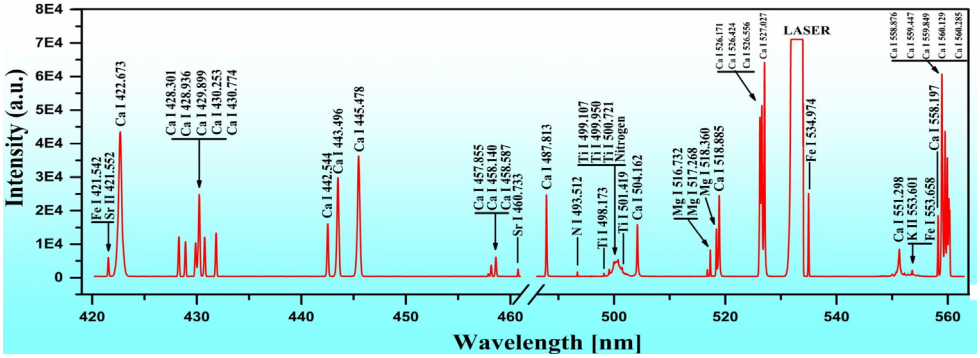

Fig. 3. LIBS emission spectrum of the cement sample from wavelength section 420-562 nm at a laser wavelength of 532 nm, delay-time of 2 µs, and laser energy ~110 mJ.

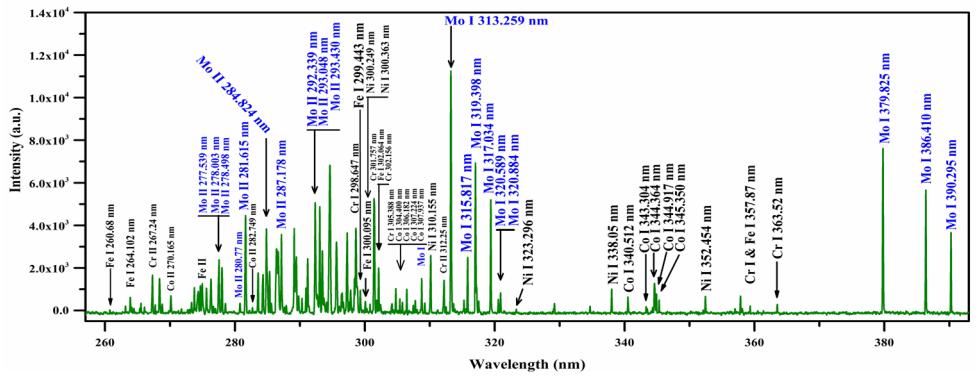

Fig. 4. The LIBS spectrum of the alloy sample within the wavelength region from 257 nm to 392 nm.

We have used the optically–thin emission line of neutral calcium (Ca I) at 610.27 nm to estimate the electron number density for all the cement samples. The broadening parameter of the specific Stark emission line (ω_s = 0.00698 nm) was sourced from existing literature [34]. A Voigt fitting considering the instrumental width factor (0.06 ± 0.01 nm) was used to determine the number-density of the electron plasma. The estimated N_e for the cement sample was about (1.1 ± 0.6) × 10^{17} cm^{-3}. The McWhirter's criterion in which the collisional mechanisms are dominating the radiative processes was used to check whether the plasma satisfies the LTE condition or not. We have used the Ca (I) emission line at 422.67 nm to check the LTE criterion [12, 14, 27]:

$$N_e(cm^{-3}) \geq 1.6 \times 10^{12}(\Delta E_{(eV)})^3[T_{(k)}]^{1/2} \quad (11)$$

The density of the electron plasma for the cement sample is estimated to be ~10^{15} cm^{-3}, determined through the electron temperature (T_e of about 9000 K) and energy difference (ΔE ~ 2.93 eV). Similarly, electron number densities computed for several cement varieties using the same Ca (I) line are around 10^{15} cm^{-3}. This observation indicates that the N_e (1.1 ± 0.6) × 10^{17} cm^{-3} derived from broadened emission lines due to the Stark effect for all samples significantly exceeds the electron number-density calculated using McWhirter's criterion. Consequently, it suggests that the plasma can be observed as being in LTE.

3.2. Energy Dispersive X-ray (EDX)

The target samples that were examined by LIBS were studied using the energy-dispersive X-ray (EDX) system. Before EDX spectroscopical analysis, a layer of carbon was applied to the samples. This precaution was taken because carbon possesses a low atomic number, ensuring that its X-ray graph peak does not disrupt the peaks of other elements. To coat the samples, a carbon film containing vaporized carbon fiber was deposited. Carbon vaporized from the fiber due to the passage of electric current through it and deposited onto the sample. Optimal thicknesses for sample preparation for EDX analysis range from 5 nm to 20 nm.

EDX is an analytical technique that is frequently exploited in the elemental investigation of materials in the form of solid thin films, solid powder, liquid samples, or even a pellet, etc. [2, 34]. The production of X-rays in EDX connected with scanning-electron-microscopy (SEM) involves a two-step process. Initially, a high-energy electron beam impacts the sample surface, transferring some of its energy to the nuclear electrons of the target substance. This energy excites the electrons, causing them to be ejected from their shells, leaving behind vacancies (holes). Subsequently, electrons from shells at elevated energy levels fill these vacancies in shells with lower energy levels, emitting the energy difference (ΔE) in the form of X-rays. By employing this methodology, the energies of the emitted X-rays, which are distinctive to each element, and the corresponding signal intensities are utilized to quantify the elements present in the analyzed sample. The X-rays produced from the target are then captured using a 30 mm^2 Si (Li) detector. The compositional analysis of cement, Al-based alloys, and rare earth geological samples was carried out using an Oxford ($X\text{-}MAX^{N\text{-}20}$) EDX instrument, coupled with an SEM instrument functioning at a 20 keV level [34, 35].

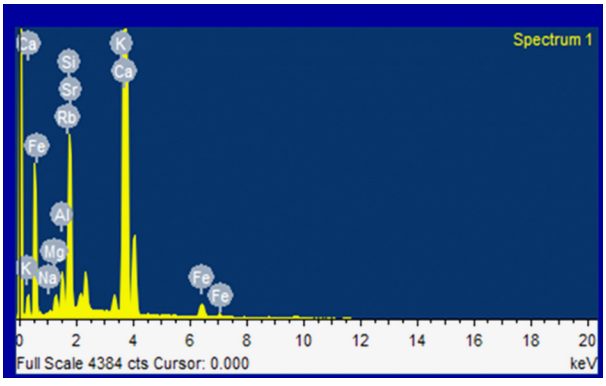

Fig. 5. The K_α K_β, L_α, and L_γ emission peaks detected from the cement target sample using EDX.

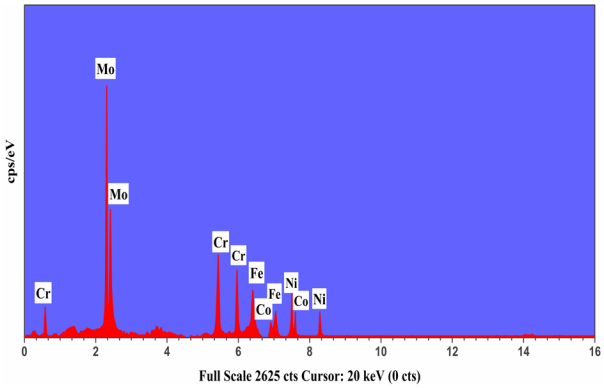

Fig. 6. Emission peaks: K_α and K_β in alloy sample with major molybdenum.

In this study, the results obtained using EDX for Mo-based alloys, cement, and rare earth geological samples are presented. The EDX spectra were collected from various spots on a homogeneous sample surface with an over-scan area of about ~100 μm^2 at an accelerating voltage of the electron beam of 20 kV. The EDX spectra illustrate the distribution of the energy spectrum of X-rays emanating from the specified point (micro-area) within the sample. Figure 5 reveals the detection of evident characteristic X-ray peaks, including K_α, K_β, L_α, and L_γ belong to various elements such as Al, Ca, Fe, K, Mg, Na, Si, and Sr in the cement sample. Similarly, the X-ray emission peaks of various elements are detected in the EDX spectra of the Mo-based alloy samples as presented in Figure 6.

3.3. Laser Ablation Time–of–Flight Mass Spectrometer (LA-TOF-MS)

Laser Ablation Time of Flight Mass Spectrometer (LA–TOF-MS) was also applied to observe the chemical weight of the rare earth geological, cement, and molybdenum-based alloy samples. There is no need for special pretreatment of the samples in the mass spectrometric analysis. We used the same pallets of the samples which were already used in the CF-LIBS analysis. The experimental configuration for LA–TOF-MS involves a stainless steel vacuum chamber with a 30 cm diameter, assisting ionization and extraction regions. This chamber is connected to a two-meter-long field-free flight tube, as shown in Figure 7. The extraction region is composed of three electrodes (3 cm × 3 cm), with two of them featuring 1 cm openings in the center covered by a fine tungsten mesh. To maintain a vacuum of ~1×10^{-6} mbar during the experiment, the chamber is linked to a turbo-molecular pump. Ion detection is carried out at the terminus of the flight tube using a channeltron electron multiplier detector (CEM-D), which is connected to an oscilloscope (Tektronix) having 500 MHz digital storage capacity [35-38].

In Figure 8, we show the mass spectrum of the Uraninite sample obtained using LA-TOF-MS to determine its relative chemical wt.%. The spectrum represents the signal corresponding to the ionic element along the y-axis and the arrival time of the ions at the detector is presented along the x-axis. An enlarged section is presented as an inset showing the peaks of the ingredient components present in the Uraninite sample. The peaks corresponding to the ions of the elements for instance Si^+, Ti^+, Mn^+, Na^+, Al^+, Ca^+, Fe^+, Ni^+, O+, and Zn^+ along with the ionic oxides such as LiO^+, SiO^+, FeO^+, MnO^+ and ZnO^+ are identified. A calibration curve is also presented as an inset to complement the elemental identifications. Interestingly, the signal corresponding to uranium ion is detected nearly at 88 microseconds [39].

Figure 9 illustrates a comparative analysis of individual elemental peaks from a rare earth ore sample, utilizing both the mass spectrum from the LA-TOF-MS system and the emission spectrum from the EDX system. Figure 9(a) displays the mass spectrum obtained from the LA-TOF-MS system, while Figure 9(b) illustrates the distinctive X-ray emission peaks representing different elements detected in the EDX spectrum of the geological mineral sample containing rare earth metals. In both spectra, the y-axis represents the relative intensity of the corresponding signals. The x-axis

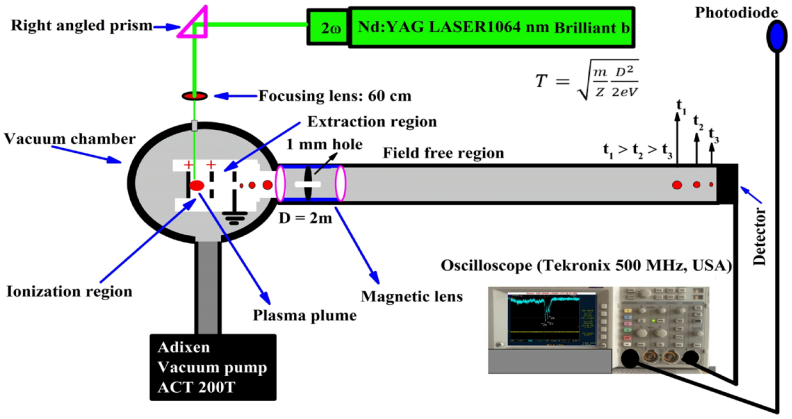

Fig. 7. A visual representation illustrating the configuration of the experimental arrangement for the LA-TOF-MS system, focusing on laser ablation/ionization processes.

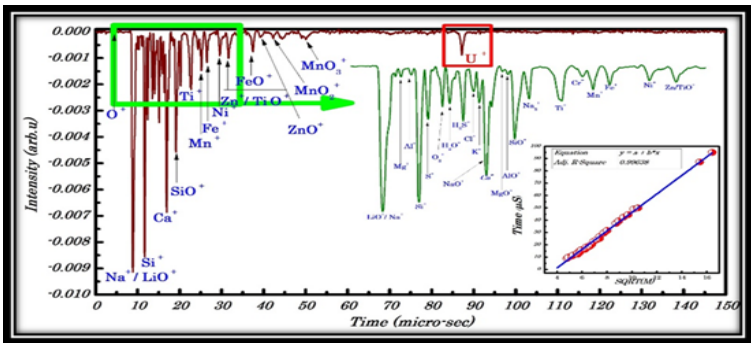

Fig. 8. A LA-TOF-MS spectrum of the Uraninite sample, inset shows the calibration curve.

of the LA-TOF-MS spectrum corresponds to time in microseconds (µs), while the x-axis of the EDX spectrum represents energy in kilo-electron volt (keV). Remarkably, the spectra obtained through EDX and LA-TOF-MS exhibit consistent element identification, except for oxygen. It is noteworthy, that these two methodologies are entirely distinct: the EDX data depict electron-induced emissions (of K_α, K_β, L_α, and L_γ, etc.) lines (X-ray region) from constituent elements along the x-axis, while the LA-TOF-MS mass spectrum illustrates the arrival times of ionic masses at the detector, where lighter ions arrive earlier than heavier ones. To enable comparison, we have aligned the x-axis to correlate the LA-TOF-MS ionic mass peaks with the corresponding EDX energy peaks of silicon and lanthanum. The comparison is highly favorable, underscoring a close correlation in the compositional results acquired from both techniques. The element silicon emerges as the predominant component in both approaches, followed by C, Ca, Ti, Na, Mg, Fe, and rare earth metals such as lanthanum, cerium, and neodymium.

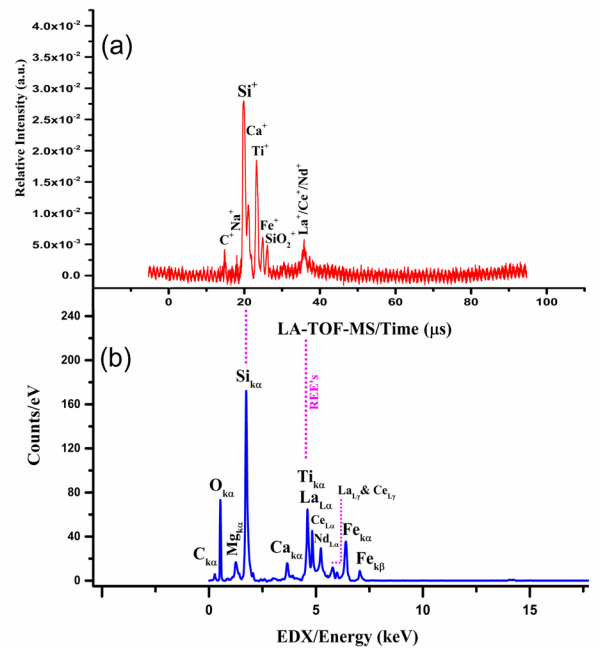

Fig. 9. (a) The quantitative comparison of LA-TOF-MS (laser energy fluence per pulse ~ 3 J/cm^2) and (b) EDX (operated at 20 KeV) for analyzing the rare earth geological sample.

In Figure 10, we show the mass spectrum of the Mo-based alloy sample with the predominant peak corresponding to molybdenum, followed by peaks representing chromium (Cr), iron (Fe), cobalt (Co), and nickel (Ni). The y-axis of the graph illustrates the relative intensity of the ion signals, while the x-axis represents the arrival time of the ions. Inset (a) provides a well-resolved view of the peaks corresponding to Cr, Fe, Co, and Ni. This inset is particularly valuable as it showcases the individual peaks within the high-resolution time range of 10 to 12 microseconds, allowing for more precise identification and analysis of each element

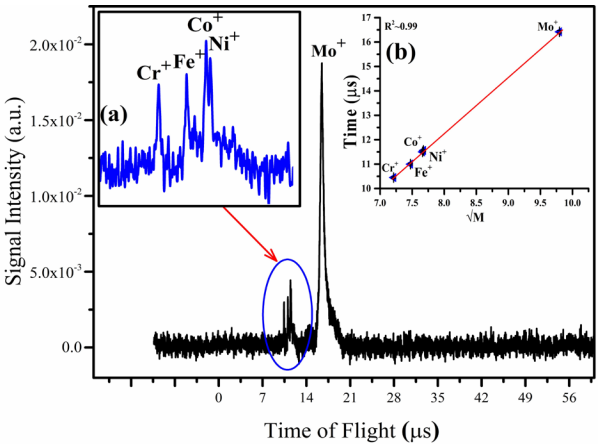

Fig. 10. The mass spectrum of the molybdenum-based alloy sample obtained using the LA-TOF-MS system. Inset (a) shows high-resolution peaks within a time scale from 10 to 12 µs, while (b) represents a linear regression calibration curve.

in the alloy. For a more detailed analysis, a linear regression calibration curve that correlates the square root of the mass with the time–of–flight of the ions is present as an inset (b). This calibration curve enhances the accuracy of mass determination in the analysis.

3.4. Proton–Induced X-ray Emission (PIXE)

In PIXE analysis, X-rays are generated by the excitation of the target atoms caused by a high-energy incident proton beam or alpha particles (^4_2He). The incident ions may experience further elastic or inelastic scattering during the interaction. The unstable excited target atoms decay to a stable energy state by restoring to their original electron configuration. The de-excitation of atoms may cause emission of the electromagnetic radiation in the form of X-rays (at a certain angle with the normal) which are characteristic of the excited atom as shown in Figure 11. The emission spectrum consists of various emission lines such as K, L, M, etc., generated by the electronic transitions. The elemental composition is estimated by registering the X-rays produced from the target sample [40, 41].

The PIXE analysis was conducted utilizing a high-power proton beam with an energy of around 3 MeV at a beam current of 5 nA. This proton beam was sourced from a 5 MV Pelletron-Tandem Accelerator (PTA) facility that is installed at the National Center for Physics in Islamabad, Pakistan. The facility serves as a crucial resource for providing the necessary particle acceleration for PIXE investigations in the experiments, allowing for precise and detailed elemental analysis [41, 42]. In Figure 12, the comparison between the two methodical techniques, for instance, LA-TOF-MS and PIXE for the cement target sample is presented. In Figure 12(a), the PIXE spectrum detects major peaks of elements such as Al, Ca, Si, Fe, Ti, and K. In Figure 12(b), we present the mass spectra from LA-TOF-MS, where a dominating ion peak of Ca is followed by K, Ti, Li, Na, Al, Si, Mg, Sr, and Fe, showing results consistent with the PIXE analysis of the same cement sample. The inset of the PIXE spectrum displays the same spectrum without fitting on a logarithmic scale. However, elements of light atomic numbers such as Li, Na, and Mg, and heavy atomic numbers such as Sr, remained undetected in the PIXE analysis due to the PIXE detection limit. The PIXE investigation of the cement target sample is carried out using the commercially available GUPIXWIN software. Interestingly, Ca in PIXE analysis is also a dominating constituent of the cement sample followed by Si and Fe.

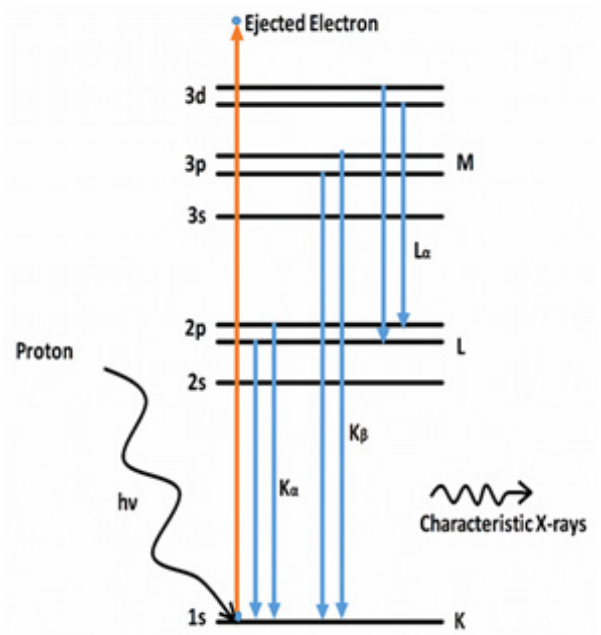

Fig. 11. The schematic diagram of the PIXE experiment for the cement sample with a small thickness Δx.s.

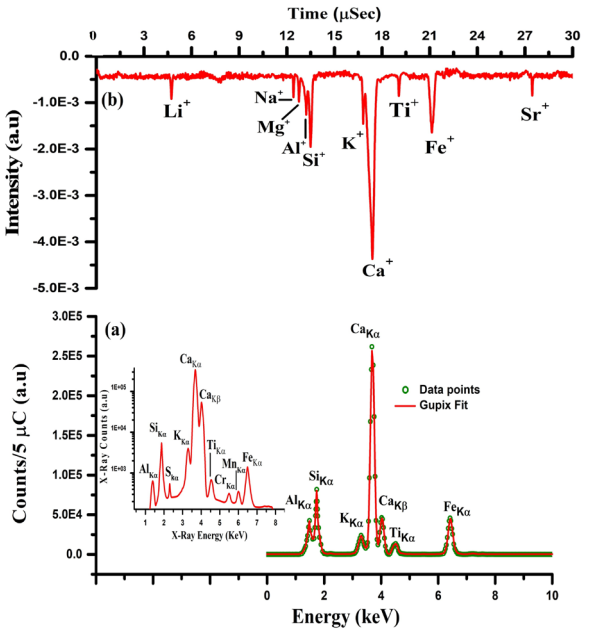

Fig. 12. A comprehensive spectral study of the cement sample through two analytical techniques: (a) PIXE spectrum with an inset showing spectrum without fitting and (b) LA-TOF-MS mass spectrum. The PIXE spectrum is supplemented by nonlinear fitting conducted with GUPIXWIN software. The PIXE data was obtained using a proton ion beam with an energy of around 3 MeV and a beam current of ~5 nA. These ions were sourced using a 5 MV PTA system.

3.5. X-ray Fluorescence (XRF)

In the X-ray fluorescence (XRF) technique, an incoming X-ray knocks out an electron from one of the orbitals near the nucleus of an atom of the material. A positively charged hole (+) is generated in the orbital, which produces a stable state for another upper high-energy atomic electron. To reinstate the neutral or equilibrium state of an atom, an electron from a higher energy state falls into the lower state and the excess energy is emitted in the form of a fluorescent X-ray (K_α and K_β lines). The energy or wavelength, i.e., ($\lambda = 1240$ eVnm/E) corresponding to these emitted X-rays is specific for an element in the material to be analyzed. Hence, this emitted wavelength or energy of the X-ray photon makes XRF a fast analytical instrument for the analysis of elemental composition. Generally, the energy or wavelength of the emitted X-ray photon for a specific element is independent of the chemistry of the target sample [34, 35].

In various fields such as scientific, military, and forensic laboratories, metallurgical, and pharmaceutical industries, the XRF method is widely used for chemical weight analysis as the detection limits of this technique are much improved than SEM-EDX. However, this technique has limitations in the sample size but it yields excellent precision on samples having a flat surface of dimension greater than 1 mm^2. This analytical technique is also comparable with other techniques such as LIBS and LA-TOF-MS. These techniques can provide excellent elemental as well as mass isotopic analysis and can be employed as a primary reference to highlight the limitations and precision of other techniques. The elemental composition of the cement samples was investigated using the XRF technique. A JSX model-(3202M) setup, operating at (5 to 50) kV with a current range of 0.01–1.0 mA and an energy resolution of 149 eV at 5.9 keV [34, 35], was employed for this purpose. The XRF energy spectrum shown in Figure 13 reveals major elements like Fe, Ca, and Si, while also indicating the presence of trace elements such as Sr, K, Ti, and Al in the cement sample. Although this technique can easily detect elements such as Na-U (Concentration > 100 ppm or 10-50 mg/kg), Na and Mg remain undetected, maybe due to the lower concentrations which are about 0.22% and 1.40%, respectively. Interestingly, the results are quite complementary in the domain of Al to Sr showing good compatibility between these two techniques.

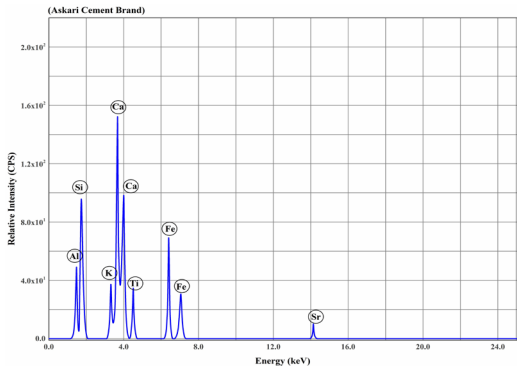

Fig. 13. XRF spectrum of cement under high vacuum with 50 kV anode voltage and 1 mA anode current.

To compare the analytical performance of the techniques, we performed a comprehensive analysis by comparing quantitative analysis for the cement sample. The quantitative results obtained from various techniques, including CF-LIBS, PIXE, EDX, XRF, and LA-TOF-MS are illustrated in Figure 14. Mainly, the quality control of Portland cement hinges on its major components, namely Al_2O_3, CaO, Fe_2O_3, and SiO_2. Following the standards set by the British authorities, the acceptable ranges for Ca, Si, Al, and Fe in Portland cement are (60% – 67%), (21% – 22%), (6%), and (3.5%), respectively [43-45]. Figure 14 illustrates that the cement sample falls within the specified limits for calcium, silicon, and iron. Remarkably, LIBS, EDX, XRF, and LA-TOF-MS methods consistently identify main elements such as Fe, K, Al, Ca, Ti, and Si. These results demonstrate an agreement in the composition obtained through the various analytical methods. However, PIXE analysis deviates notably for elements with atomic weights less than 20 (Li, Mg, Na) due to its lower detection limits. Additionally, the XRF technique's constrained detection capabilities result in the non-detection of Li and Na. This variance in PIXE and XRF results underlines the importance of considering the specific detection capabilities of each analytical technique in the comprehensive evaluation of cement composition.

4. RESULTS AND DISCUSSION

The application of EDX in combination with scanning electron microscopy (SEM) is commonly used for the qualitative and quantitative assessment of materials. The EDX depth profile usually reaches around 5 microns for the minimum

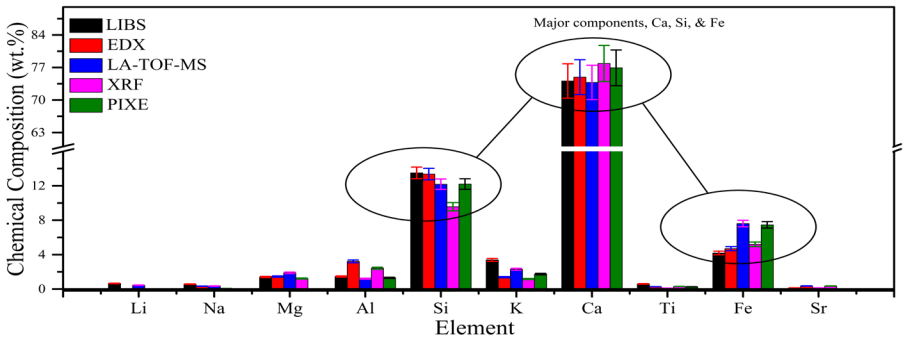

Fig. 14. A comparative analysis of the chemical weight % of the cement sample through various diagnostic techniques, including CF-LIBS, EDX, LA-TOF-MS, XRF, and PIXE.

detectable mass, with a detection limit of 0.01 wt.% according to theoretical calculations. For bulk materials, the limit is typically 0.1 wt.%, while for light elements, usually less than 10 atomic number, it's approximately 1-2 wt.%. Complex materials tend to have a detection limit greater than 5 wt.%. Therefore, under the optimized conditions for multielement analysis, the EDX analysis can only identify elements present in concentrations exceeding 0.1 wt.%. In this work, major elements with concentrations greater than 10 wt.%, such as Si and Ca, are consistently identified as predominant components with all analytical techniques. This highlights the reliability of EDX in identifying elements at relatively higher concentrations and reveals its role in both qualitative and quantitative material analysis, particularly for major components with significant presence.

In this study, the XRF technique was applied to examine the same set of samples. XRF analysis is suitable for multi-elemental analyses of several materials, containing soils, rocks, sediment, paint, herbal medicines, and cement, and in the construction industry for verifying the metallic composition. The XRF technique typically provides a depth profile reaching ~100 microns, with detection limits for most elements ranging from 2 to 20 ng/cm^2 for micro, and thin solid samples, aerosols, and liquids. The results of this study confirm that when the chemical wt.% of an ingredient in the sample exceeds 0.01 wt.% and the size range of the target sample is in millimeters in diameter or larger, more precise analysis and quantification of elements can be achieved. It's important to note that certain elements, specifically Mg and those with relatively lower concentrations compared to Mg, remained undetected in this analysis. This is due to the smaller concentrations of those elements in the sample, as well as a limitation characteristic of the XRF technique. However, the technique successfully identified major elements with concentrations exceeding 10 wt.%, demonstrating its effectiveness in analyzing and quantifying elements present in significant amounts.

PIXE employs proton beams for analysis owing to their small energy loss in the target sample. The limit of detection in PIXE investigation is influenced by the X-ray spectrum from the environment, primarily generated by proton collisions of less than 1.5 to 2 MeV. An optimal proton energy for PIXE analysis falls in the range of 3 to 5 MeV. Remarkably, PIXE analysis exhibits low sensitivity to elements for instance C, H, Li, Mg, N, Na, and O due to their absorption in the X-ray detector's detection window. Conversely, its sensitivity is high when it comes to metallic elements and heavy elements in alloys or metallic samples. As a result, PIXE proves particularly suitable for studying biological, metallic alloy, or geological samples containing elements like Ca, Fe, Si, Sr, and Ti, etc. In the context of this study, elements such as Al, Ca, Fe, K, Si, and Ti were efficiently detected and quantified by PIXE, demonstrating its robust compatibility with LIBS and other diagnostic methods.

In LA-TOF-MS, there is no specific limit on the sample size or dimension. It needs a few mJ energy to ablate the sample surface. Once the sample is ablated and plasma is formed the only thing which needs to be done is the accelerating voltages across the metallic plates under a high vacuum. The mass spectra of LA-TOF-MS, (as shown in Figures 8 to 10), demonstrate that all the elements detected by LIBS are also present in these mass spectra. These results also reveal that both the

analytical techniques such as EDX and PIXE are in good agreement with the LA-TOF-MS. However, in the case of light elements, PIXE is unable to detect the lower Z elements but the LA-TOF-MS technique shows a good detection sensitivity.

In this work, a limited number of LIBS analyses were presented using geological, cement, and alloy samples. From these results, it is evident, that it is possible to differentiate between the samples of various natures by comparing the LIBS spectra and there is a good reproducibility of the LIBS analysis. Both qualitative and quantitative investigations of a sample can be conducted without producing scratches to the material's surface. The LIBS technique provides a depth profile extending to approximately hundreds of microns with a limit of detection varying from 10–50 ppm. In the current study, LIBS results consistently demonstrated a reliable detection limit, specifically showcasing the ability to detect elements such as Li to Sr within the cement analysis and C to Nd within the analysis of rare earth mineralogical samples. This highlights the use of LIBS in achieving both qualitative and quantitative insights into distinct sample compositions. Furthermore, the results obtained through LIBS exhibited reliability with those achieved using the LA-TOF-MS technique. Additionally, the analysis of major elements in the samples, such as Si, Ca, Ti, Fe, and Sr, conducted through EDX, PIXE, and XRF techniques also aligns well with the findings from LIBS and LA-TOF-MS. However, it is important to highlight an inconsistency observed when analyzing light elements, specifically Li, Na, and Mg, using EDX, PIXE, and XRF. In these examples, the results from the three techniques diverge, indicating a variation in their ability to accurately detect and quantify such lighter elements.

5. CONCLUSIONS

Different analytical techniques, including LIBS, LA-TOF-MS, EDX, PIXE, and XRF, have been employed to perform both qualitative and quantitative analyses on different varieties of samples. The LIBS qualitative analysis involved capturing emission spectra from a micro-plasma generated at 532 nm, with a pulse energy of 110 mJ, laser irradiance around ~2.8 GW/cm^2, and an integration time of 1 ms. The findings from LIBS demonstrate its rapid growth as a technique for chemically quantifying elements in diverse materials. Furthermore, the combination of LIBS with an LA-TOF-MS proved to be an effective and complementary approach for both isotopic and elemental analysis. Comparative analyses of geological, cement, and metallic alloy samples using LIBS were performed in sequence with results obtained from LA-TOF-MS, EDX, PIXE, and XRF. LIBS emerged as a rapid and micro-destructive technique, offering a relatively cost-effective solution. The incorporation of LIBS with LA-TOF-MS has been identified as the optimal analytical method for the quantitative elemental analysis of geological ores containing rare earths, cement, nephrite, and metallic alloy samples.

6. ACKNOWLEDGMENTS

We are grateful to the Pakistan Academy of Sciences for providing financial support to establish the LIBS setup and for the fabrication of the LA-TOF-MS equipment at the National Centre for Physics, Islamabad.

7. CONFLICT OF INTEREST

The authors declare no conflict of interest.

8. REFERENCES

1. R. Noll, C. Fricke-Begemann, S. Connemann, C. Meinhardt, and V. Sturm. LIBS analyses for industrial applications–an overview of developments from 2014 to 2018. *Journal of Analytical Atomic Spectrometry* 33(6): 945-956 (2018).
2. A. Fayyaz, U. Liaqat, K. Yaqoob, R. Ahmed, Z.A. Umar, and M.A. Baig. Combination of laser-induced breakdown spectroscopy, and time–of–flight mass spectrometry for the quantification of CoCrFeNiMo high entropy alloys. *Spectrochimica Acta Part B: Atomic Spectroscopy* 198: 106562 (2022).
3. G.S. Senesi. Laser-Induced Breakdown Spectroscopy (LIBS) applied to terrestrial and extraterrestrial analogue geomaterials with emphasis to minerals and rocks. *Earth-Science Reviews* 139: 231-267 (2014).
4. V. Palleschi. Laser-induced breakdown spectroscopy: Principles of the technique and future trends. *ChemTexts* 6: 1-16 (2020).
5. S.S. Harilal, B. O'Shay, M. Tillack, and M.V. Mathew. Spectroscopic characterization of laser-induced tin plasma. *Journal of Applied Physics* 98(1): 013306-7 (2005).

6. E. Tognoni, G. Cristoforetti, S. Legnaioli, V. Palleschi, A. Salvetti, M. Müller, U. Panne, and I. Gornushkin. A numerical study of expected accuracy and precision in calibration-free laser-induced breakdown spectroscopy in the assumption of ideal analytical plasma. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62(12): 1287-1302 (2007).
7. D.A. Cremers, and L.J. Radziemski (Eds.). Handbook of laser-induced breakdown spectroscopy. *John Wiley & Sons* (2013).
8. Z. Wang, M.S. Afgan, W. Gu, Y. Song, Y. Wang, Z. Hou, W. Song, and Z. Li. Recent advances in laser-induced breakdown spectroscopy quantification: From fundamental understanding to data processing. *TrAC Trends in Analytical Chemistry* 143: 116385 (2021).
9. G. Guo, G. Niu, Q. Shi, Q. Lin, D. Tian, and Y. Duan. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods* 11(23): 3006-3013 (2019).
10. S. Zhang, X. Wang, M. He, Y. Jiang, B. Zhang, W. Hang, and B. Huang. Laser-induced plasma temperature. *Spectrochimica Acta Part B: Atomic Spectroscopy* 97: 13-33 (2014).
11. A. Ciucci, M. Corsi, V. Palleschi, S. Rastelli, A. Salvetti, and E. Tognoni. New procedure for quantitative elemental analysis by laser-induced plasma spectroscopy. *Applied spectroscopy* 53(8): 960-964 (1999).
12. Q. Abbass, N. Ahmed, R. Ahmed, and M.A. Baig. A comparative study of calibration free methods for the elemental analysis by laser induced breakdown spectroscopy. *Plasma Chemistry and Plasma Processing* 36: 1287-1299 (2016).
13. K.S. Singh, and A.K. Sharma. Spatially resolved behavior of laser-produced copper plasma along expansion direction in the presence of static uniform magnetic field. *Physics of Plasmas* 23(12): 122104 (2016).
14. S.S. Harilal, B.E. Brumfield, N.L. LaHaye, K.C. Hartig, and M.C. Phillips. Optical spectroscopy of laser-produced plasmas for standoff isotopic analysis. *Applied Physics Reviews* 5(2): 021301 (2018).
15. Z. Wang, C. Yan, J. Dong, T. Zhang, J. Wei, and H. Li. Acidity analysis of iron ore based on calibration-free laser-induced breakdown spectroscopy (CF-LIBS) combined with a binary search algorithm (BSA). *RSC Advances* 6(80): 76813-76823 (2016).
16. J. Peng, W. Xie, J. Jiang, Z. Zhao, F. Zhou, and F. Liu. Fast quantification of honey adulteration with laser-induced breakdown spectroscopy and chemometric methods. *Foods* 9(3): 341 (2020).
17. A. Safi, B. Campanella, E. Grifoni, S. Legnaioli, G. Lorenzetti, S. Pagnotta, F. Poggialini, L. Ripoll-Seguer, M. Hidalgo, and V. Palleschi. Multivariate calibration in Laser-Induced Breakdown Spectroscopy quantitative analysis: The dangers of a 'black box'approach and how to avoid them. *Spectrochimica Acta Part B: Atomic Spectroscopy* 144: 46-54 (2018).
18. D. Diaz, A. Molina, and D.W. Hahn. Laser-induced breakdown spectroscopy and principal component analysis for the classification of spectra from gold-bearing ores. *Applied Spectroscopy* 74(1): 42-54 (2020).
19. P. Inakollu, T. Philip, A.K. Rai, F.Y. Yueh, and J.P. Singh. A comparative study of laser induced breakdown spectroscopy analysis for element concentrations in aluminum alloy using artificial neural networks and calibration methods. *Spectrochimica Acta Part B: Atomic Spectroscopy* 64(1): 99-104 (2009).
20. J.B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, and I.L. Hecho. Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis. *Analytical and Bioanalytical Chemistry* 385: 256-262 (2006).
21. Z.A. Umar, U. Liaqat, R. Ahmed, and M.A. Baig. Detection of lead in soil implying sample heating and laser-induced breakdown spectroscopy. *Applied Optics* 60(2): 452-458 (2021).
22. T. Takahashi, B. Thornton, T. Sato, T. Ohki, K. Ohki, and T. Sakka. Partial least squares regression calculation for quantitative analysis of metals submerged in water measured using laser-induced breakdown spectroscopy. *Applied Optics* 57(20): 5872-5883 (2018).
23. T. Zhang, S. Wu, J. Dong, J. Wei, K. Wang, H. Tang, X. Yang, and H. Li. Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (LIBS) coupled with support vector machine (SVM) and partial least square (PLS) methods. *Journal of Analytical Atomic Spectrometry* 30(2): 368-374 (2015).
24. E. Mal, R. Junjuri, M.K. Gundawar, and A. Khare. Optimization of temporal window for application of calibration free-laser induced breakdown spectroscopy (CF-LIBS) on copper alloys in air employing a single line. *Journal of Analytical Atomic Spectrometry* 34(2): 319-330 (2019).

25. F.O. Borges, J.U. Ospina, G.H. Cavalcanti, E.E. Farias, A.A. Rocha, P.I. Ferreira, G.C. Gomes, and A. Mello. CF-LIBS analysis of frozen aqueous solution samples by using a standard internal reference and correcting the self-absorption effect. *Journal of Analytical Atomic Spectrometry* 33(4): 629-641 (2018).
26. J.A. Aguilera, and C. Aragón. Multi-element Saha–Boltzmann and Boltzmann plots in laser-induced plasmas. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62(4): 378-385 (2007).
27. A. Mansoori, B. Roshanzadeh, M. Khalaji, and S.H. Tavassoli. Quantitative analysis of cement powder by laser induced breakdown spectroscopy. *Optics and Lasers in Engineering* 49(3): 318-323 (2011).
28. J.A. Aguilera, and C. Aragón. Characterization of a laser-induced plasma by spatially resolved spectroscopy of neutral atom and ion emissions.: Comparison of local and spatially integrated measurements. *Spectrochimica Acta Part B: Atomic Spectroscopy* 59(12): 1861-1876 (2004).
29. O. Samek, D.C. Beddows, J. Kaiser, S.V. Kukhlevsky, M. Liska, H.H. Telle, and A.J. Whitehouse. Application of laser-induced breakdown spectroscopy to in situ analysis of liquid samples. *Optical Engineering* 39(8): 2248-2262 (2000).
30. A. De Giacomo, M. Dell'Aglio, O. De Pascale, S. Longo, and M. Capitelli. Laser induced breakdown spectroscopy on meteorites. *Spectrochimica Acta Part B: Atomic Spectroscopy* 62(12): 1606-1611 (2007).
31. D.D. Pace, R.E. Miguel, H.O. Di Rocco, F.A. García, L. Pardini, S. Legnaioli, G. Lorenzetti, and V. Palleschi. Quantitative analysis of metals in waste foundry sands by calibration free-laser induced breakdown spectroscopy. *Spectrochimica acta part B: Atomic Spectroscopy* 131: 58-65 (2017).
32. J.M. Gomba, C. D'Angelo, D. Bertuccelli, and G. Bertuccelli. Spectroscopic characterization of laser induced breakdown in aluminium–lithium alloy samples for quantitative determination of traces. *Spectrochimica Acta Part B: Atomic Spectroscopy* 56(6): 695-705 (2001).
33. Atomic spectra database, NIST (2024). https://dx.doi.org/10.18434/T4W30F.
34. A. Fayyaz, U. Liaqat, Z.A. Umar, R. Ahmed, and M.A. Baig. Elemental analysis of cement by calibration-free laser induced breakdown spectroscopy (CF-LIBS) and Comparison with laser ablation–time-of-flight–mass spectrometry (LA-TOF-MS), energy dispersive x-ray spectrometry (EDX), x-ray fluorescence spectroscopy (XRF), and proton induced x-ray emission spectrometry (PIXE). *Analytical Letters* 52(12): 1951-1965 (2019).
35. N. Ahmed, R. Ahmed, Z.A. Umar, and M.A. Baig. Laser ionization time of flight mass spectrometer for isotope mass detection and elemental analysis of materials. *Laser Physics* 27(8): 086001 (2017).
36. Z.A. Umar, U. Liaqat, R. Ahmed, and M.A. Baig. Classification of nephrite using calibration-free laser induced breakdown spectroscopy (CF–LIBS) with comparison to laser ablation–time-of-flight–mass spectrometry (LA–TOF–MS). *Analytical Letters* 53(2): 203-216 (2020).
37. A. Jabbar, M. Akhtar, S. Mehmood, N. Ahmed, Z.A. Umar, R. Ahmed, and M.A. Baig. On the detection of heavy elements in the Euphorbia indica plant using laser-induced breakdown spectroscopy and laser ablation time of flight mass spectrometry. *Journal of Analytical Atomic Spectrometry* 34(5): 954-962 (2019).
38. N. Ahmed, U. Liaqat, M. Rafique, M.A. Baig, and W. Tawfik. Detection of toxicity in some oral antidiabetic drugs using LIBS and LA-TOF-MS. *Microchemical Journal* 155: 104679 (2020).
39. J.O. Thompson, S.T. Alavi, J.R. Walensky, and A.G. Suits. Time of flight mass spectrometry with direct extraction of a uranium plasma. *International Journal of Mass Spectrometry* 445: 116190 (2019).
40. K. Ishii. PIXE and its applications to elemental analysis. *Quantum Beam Science* 3(2): 12 (2019).
41. W. Akram, K. Shahzad, A. Awais, I. Ahmad, M. Arif, I. Ahmad, and M. Madhuku. Roadside dust contamination with toxic metals along industrial area in Islamabad, Pakistan. *Nuclear Science and Techniques* 25(3): 1-6 (2014).
42. A. Awais, J. Hussain, M. Usman, W. Akram, K. Shahzad, T. Ali, I. Ahmad, and M. Maaza. The charge state distribution of B, C, Si, Ni, Cu and Au ions on 5 MV Pelletron accelerator. *Nuclear Science and Techniques* 28: 1-5 (2017).
43. G.N. Pandey, and S.D. Shukla (Eds.). A Textbook of Chemical Technology. *Vikas Publishing House, New Delhi* (1978).
44. N. Furman (Ed.). Standard Method of Chemical Analysis (6th ed). *Van Nostrand Reinhold Co. New York* (1962).
45. G.H. Jeffery, J. Bassett, J. Mendham, and R.C. Denney, (Eds.). Vogel's Textbook of Quantitative Chemical Analysis, 5th Edition. *Longman Scientific and Technical, Harlow, UK* (1989).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 113-128 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)850

Pakistan Academy of Sciences

Review Article

Blockchain-Based Verifiable Computation: A Review

Maham Zara[1*], Shuzhen Wang[1], and Hasan ul Moin[2]

[1]School of Computer Science and Technology Xidian University, Xian, Shaanxi, 710071, China

[2]Shaheed Zulfikar Ali Bhutto Institute of Science & Technology, Hyderabad Campus, Hyderabad, Sindh, 71000, Pakistan

Abstract: Verifiable computation has been studied as a way to verify the outcomes of an outsourced computation. It is usually seen from the view of a user who wishes to outsource computation to a centralized third party but wants to ensure that the party provides correct results. With the said scheme, the verifier requests the prover to perform the computational task and then verifies the outcome by checking the output and the proof obtained from the prover. However, there are several security challenges within a centralized third party to execute verification tasks. Recently, the advancement in blockchain technology has offered an opportunity to solve these security challenges. Blockchain is a distributed ledger and decentralized technology that eliminates the need for third-party verification. In recent years, the emergence of innovative applications of verifiable computing techniques within blockchain technology has been witnessed. These applications focus on ensuring secure key management, enhancing smart contracts, and fortifying sybil-resistance. The use of blockchain in the realm of verifiable computing has drawn the attention of many researchers. However, our research into relevant papers revealed a notable lack of comprehensive surveys on blockchain-based verifiable computing in the literature. To overcome this gap, we conducted a comprehensive survey on blockchain-based verifiable computation. First, we address fundamental concepts related to blockchain-based verifiable computation. Afterwards, we offer a series of criteria to evaluate existing blockchain-based verifiable computation techniques. Finally, based on our comprehensive review and evaluation metrics, we explore various open challenges and potential research prospects. These include zero-knowledge proofs (ZKP) integration, addressing privacy preservation, scalability, and traceability. Future research should focus on robust privacy-preserving methods, using ZKP for enhanced security, off-chain computations for scalability, and decentralized file systems like Interplanetary File System (IPFS) to improve traceability.

Keywords: Blockchain, Verifiable Computation, Smart Contract, Privacy, Security, Ethereum.

1. INTRODUCTION

As cutting-edge computer technology has become more ubiquitous in recent years, the internet sector has grown and thrived, and the size of data that needs to be processed is increasing exponentially. However, due to computing power limitations and equipment costs, ordinary users are often unable to complete massive computing tasks. To solve this problem, outsourcing computation comes to the forefront, which allows users to assign computing tasks to one or more powerful servers through the Internet in an efficient and cost-effective way. However, how to securely and reliably process and compute outsourced data has become a critical security issue with significant challenges [1]. Verifiable computation has been studied to confirm the result of an outsourced computation. It is seen from the perspective of a user who wishes to perform outsourced computation to a centralized third party while also wanting to confirm the validation of results. However, there are many security issues with the centralized third party to perform verification tasks. Recently, blockchain technology has provided an opportunity to solve security issues. Blockchain has emerged as a popular technology, and its features, such as transparency, decentralization, and immutability, make it a suitable choice for setting up a trustless platform by eliminating the centralized third party

Received: April 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Maham Zara <mahamzaara@gmail.com>

and replacing it with a system of publicly verifiable. Blockchain was proposed through Bitcoin for peer-to-peer financial transactions [2]. The blockchain rapidly gained interest in academics and industry. It works as a widely used public transaction ledger or distributed database. Instead of relying on a trusted third party, the underlying time stamping, data encryption, incentive mechanism and distributed consensus establish the core of blockchain security [3]. It can address the issue of establishing trust among each node in a decentralized system by employing a consensus and verification method, allowing distrusted users to exchange data or execute transactions without engaging a trusted third party. In recent years, new applications of verifiable computation techniques for secure key management, smart contracts, and Sybil resistance have evolved in blockchain technology while ensuring desired performance and privacy assurances. Many studies have endeavoured to use blockchain for verified computation. According to our evaluation of relevant papers, we observed that there is no thorough survey of blockchain-based verifiable computation in the available literature. To bridge this gap, we conducted a comprehensive survey on verifiable computation based on blockchain. In this paper, we first cover the core concepts of verifiable computation and blockchain technology. Following that, we present several criteria to evaluate existing blockchain-based verifiable computation techniques. In addition, we examine existing works and evaluate them using the metrics we propose. Lastly, we highlight open research issues and suggest future directions for research, drawing insights from a comprehensive literature review of existing works. More precisely, the contributions of our survey can be outlined as follows.

- A brief introduction to blockchain is given, along with highlighting its essential features and the core architectural framework of blockchain systems.
- An overview of verifiable computation is provided with different techniques for ensuring the accuracy of the computed output.
- To illustrate how blockchain-based verifiable computing works, we explain the general framework of blockchain-based verifiable computation, drawing insights from the literature.
- We explore existing blockchain-based verifiable computation techniques and use the specified criteria to evaluate and compare their benefits and drawbacks.
- We pinpoint various open issues and suggest potential research directions for future research in order to motivate future research efforts.

The rest of the paper is laid out in the following manner. In Section 2, we discuss the core ideas of blockchain, an introduction to verifiable computing, and an examination of verifiable computation utilizing blockchain technology. Section 3 presents a set of standards that can serve as the criteria for evaluating existing blockchain-based verifiable computation schemes. In Section 4, we provide an extensive examination of current research, along with an evaluation using the previously mentioned criteria. Section 5 addresses the existing challenges in Blockchain-based verifiable computing and outlines potential research paths to encourage further investigation. The paper culminates with a conclusion in Section 6.

1.1. Comparison with Existing Surveys

Much research has been performed to explore the applications of Blockchain technology, verifiable computation, and their intersection. However, there is still a significant lack of comprehensive review focusing on Blockchain-based verifiable computation. This serves as our motivation to conduct a thorough review in this domain. Table 1 displays a comparative analysis between our survey and other previously conducted surveys in the field. It gives a clear summary of the distinctions and contributions of our work in comparison to existing works. Yu *et al*. [1] focused their attention on verifiable computation and illustrated numerous application scenarios and use cases to underscore its practical significance. On the other hand, Šimunić *et al*. [4] offered a comprehensive overview of verifiable computing techniques employed in widely used blockchain applications. However, Šimunić *et al*. [4] did not include a comprehensive review of blockchain technology, a gap that our paper addresses. Dorsala *et al*. [5] conducted a thorough survey on existing cloud services based on blockchain, which was undoubtedly valuable. However, they ignored a review of smart contracts and Ethereum, which are thoroughly covered in the present study. Gamage *et al*. [6], Soni and Bhushan [7] and Shi *et al*. [8] mainly focused on blockchain,

Table 1. Comparison of our survey with Literature.

Topic	[1]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	Our Survey
Give a brief review on blockchain	✗	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓
Comparison among public, consortium and private blockchains	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✓
Give a review on Smart contracts	✗	✓	✗	✓	✗	✗	✓	✓	✗	✓	✓
Give a review on Ethereum	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Give a review of verifiable computation	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓
Focus on blockchain-based verifiable computation	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓

Note: ✓: Discussed; ✗: Not discussed

applications and issues. Shi *et al*. [8] conducted a systematic literature review of blockchain approaches designed for Electronic Health Record (EHR) systems, concentrating solely on security and privacy.

However, a comparison among public, consortium, and private blockchains and a review of Ethereum is not discussed in this work. Peng *et al*. [9] analyzed privacy concerns within the context of permissionless blockchains and offered a summary of potential privacy threats. This paper highlights the important aspects of blockchain technology that is related to privacy and security. Ahmed *et al*. [10] provided a detailed review of the blockchain-based Identity Management System (IDMS) and self-sovereign identity ecosystem. These authors conducted a survey that covered various adversarial attack types that could potentially damage the blockchain-based IDMS. Furthermore, Li *et al*. [11] presented a comprehensive review of blockchain-based trust approaches in cloud computing systems. This paper highlights the application of blockchain from the perspective of trust. In this survey paper, a double-blockchain structure-based cloud transaction model and a cloud-edge trust management framework are discussed. Saleh [12] did a detailed review of an emerging area that integrates blockchain technology and decentralized AI within the context of cybersecurity. The authors highlight the potential research directions for blockchain-enabled decentralized AI in cybersecurity to improve security, privacy, and trust in AI systems. However, a comparison among public, consortium, and private blockchains and a review of Ethereum are not discussed in this paper.

2. BASIC KNOWLEDGE

This section provides the fundamental concepts of blockchain and its unique features. It also provides a basic understanding of verifiable computation, including several techniques utilized in the verifiable computation field. Furthermore, this section gives a thorough overview of verifiable computing applications in blockchain technology, as well as real-world scenarios.

2.1. Blockchain

Blockchain technology is an immutable and distributed digital record system that is decentralized and secured with advanced cryptography. Because of these core attributes, blockchain has been positioned as a revolutionary technology in the domain of financial technology (fintech) [13]. This system is cloned across multiple nodes in a peer-to-peer network, with consensus methods applied to generate agreement on transaction histories [6]. A blockchain serves as a secure database, where encrypted data blocks are connected to form a chronological and reliable single source of truth for the information stored. It is a distributed, decentralized ledger designed to securely store some simple, hierarchical, and verifiable data.

Although the primary objective of blockchain technology was to allow peer-to-peer financial transactions without depending on third party. The core principles of blockchain technology are currently being employed to create a wide range of decentralized applications across realms of digital assets [14], Internet of Things [15], smart contracts [16], cloud computing [17], and 5G networks [18].

Figure 1 illustrates the fundamental technological architectural model.

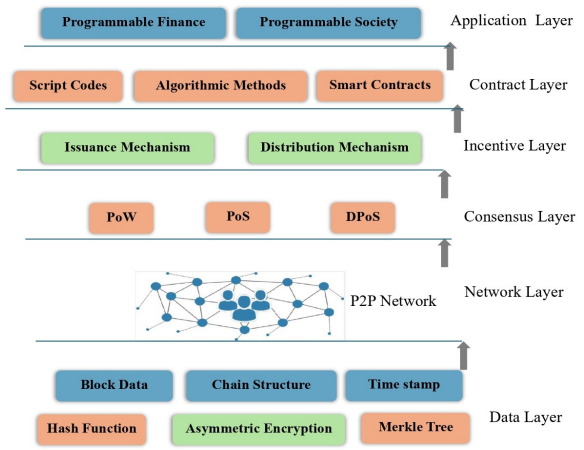

Fig. 1. Blockchain basic technology architecture model.

Application Layer: Various blockchain application scenarios, such as programmable finance and programmable society, are encapsulated in this layer.
Contract Layer: Fundamentally, it encompasses a range of algorithmic techniques, script codes and smart contracts, among others, that serve as the programming base for the upper application layer.
Incentive Layer: The incentive layer combines financial characteristics with blockchain technology, specifically the distribution mechanism and issuance mechanism of economic incentives. The main goal of incentives is to gain the attention of participants to contribute to computing power.
Consensus Layer: To build mutual confidence, several consensus algorithms, such as Delegated Proof-of-Stake (DPoS), Proof of Stake (PoS), and Proof of Work (PoW) are used to encapsulate network nodes.
Network Layer: This layer, which serves as the network backbone of blockchain, primarily covers data authentication methods, data transfer methods, and peer-to-peer network technology.
Data Layer: The data layer maintains all information records and records of transaction data, as well as the underlying timestamps and data blocks, in blockchain form.

2.1.1. Main characteristics of blockchain

To summarize, blockchain includes the following main characteristics:

Decentralization: In terms of blockchain, decentralization is defined as the lack of a central authority managing aspects such as databases, the execution of code, accounts, identities, and balances. It is the core principle of blockchain since nodes record all transaction data, eliminating the necessity for a central authority. This reduces the risk of a single point of vulnerability. Consensus mechanisms like PoS, PoW, and others play a vital role in safeguarding the security of the blockchain, even in the absence of a trustworthy authority and no service costs.

Persistency: Before being added to the Blockchain, every block and all transactions are verified for more benefits. Blocks containing invalid transactions could be detected promptly. Due to public verification, any malevolent action to undermine the system by implementing malicious transactions is impossible. Once data is stored on a blockchain, it becomes persistent and cannot be altered.

Anonymity: Any user can interact with the blockchain using a randomly generated address that masks the true identity of the user. However, the members can see the details of the encoded transaction.

Transparency: Every member has free access to all transactions or interactions logged in the blockchain. Furthermore, many parties (miners) offer their computing resources to create new blocks and verify recently generated blocks and transactions. These methods ensure a high level of transparency, which enhances the integrity of data recorded within the blockchain. The transparency and trust of cloud computing are improved if blockchain technology is employed and the meta-data of interactions between cloud services and users are stored within the blockchain.

Auditability: Since the data recorded on a blockchain is openly accessible, it is sensitive to public auditing. This feature is vital for cloud computing because the majority of existing methods for cloud data integrity and auditing rely on complicated cryptographic primitives and third-party auditors [19]. The concept of auditability in blockchain technology minimizes the expenses associated with auditing and removes the necessity for reliable third parties, which is a feature of conventional cloud auditing methods.

Security and privacy: Every transaction on the blockchain is cryptographically hashed. In simple terms, the information on the network conceals the true nature of the data. In blockchain, each participating entity is required to generate a set of asymmetric keys by using public-key cryptography in order to initiate transactions. The sender's private key is used for signing each transaction before it is sent. The signature is verified using the public key of sender during transaction verification. The ownership, non-repudiation, and confidentiality of the data are guaranteed by the asymmetric key. In blockchain technology, encoding access control policies into smart contracts eliminates the need for a centralized authority. However, public blockchain systems have the limitation of being unable to ensure the confidentiality of data stored on the blockchain.

Despite the fact that public blockchain schemes include decentralization, transparency, immutability, and trust, they offer privacy and scalability issues. In order to address these problems, consortium blockchains and private blockchains are being introduced. Nevertheless, private blockchains or consortium blockchains sacrifice transparency and decentralization in the trade of privacy and scalability. The comparison among public, private and consortium blockchains is listed in Table 2.

2.1.2. Smart contracts

A blockchain-based program that can be executed and stored is known as a smart contract (SC). The logic of contractual contracts between parties is captured in the program code. The consensus peers evaluate the code, and the consensus protocol of the blockchain confirms the integrity of execution. Assuming that the blockchain's underlying consensus mechanism is reliable, smart contracts are performed by a trustworthy global machine that faithfully executes every command [20]. A detailed survey on blockchain, blockchain functionality, blockchain security analysis, blockchain vulnerabilities, and prospective blockchain applications, is presented by Soni and Bhushan [7].

2.1.3. Ethereum

Ethereum is a publicly accessible source blockchain that was designed in 2013 and made publicly available in 2015 [21]. It has two distinct characteristics:

- Ethereum enables developers to execute distributed apps on the Ethereum blockchain.
- Distributed apps are consensus-based applications that are robust to network outages.

In this network, developers have the benefit of writing smart contracts.

2.2. Overview of Verifiable Computation

Verifiable computation, also known as verifiable computing, enables a client, acting as a verifier, to outsource computational tasks to clients who may not be fully trustworthy. Despite this, the verifier retains the ability to verify the accuracy of the outcomes. This approach effectively eliminates the risk of untrustworthy clients delivering incorrect results without actually accomplishing the task. The main goal of verifiable computation is to securely outsource computational tasks. Verifiable computation generally involves two key parties: a client and a trusted third party such as Prover

Table 2: Comparison among public, consortium and private blockchains.

S. No.	Characteristic	Public	Consortium	Private
1	Decentralization	Yes	No (selected set of nodes spread across multiple organizations)	No (single organization)
2	Immutability	Tamper-roof	Could be tampered	Could be tampered
3	Transparency	Yes	Could be public or restricted	Could be public or restricted
4	Persistency	Yes	No	No
5	Public auditability	Yes	No	No
6	Privacy	No	Partial	Yes
7	Smart contracts	Yes	Yes	Yes

(cloud). The client forwards an input x to the cloud, which serves as the host for the outsourced function F(x). The cloud executes the outsourced function on the given input and subsequently sends back the resulting output y to the client with mathematical proof. Subsequently, the client has the capability to validate that the output provided by the cloud corresponds to the genuine output obtained from the computation of the function on the given input, i.e., y = F(x). The mathematical proof should be able to verify the accuracy of the output for the specific input and function that was outsourced, as well as confirm the legitimacy of the input, the output, and the function used in the computation.

However, If the prover is malevolent, then it may mislead the verifier by sending false outcomes. Additionally, the malevolent prover can use a different input x or function F. Therefore, it is important for a verifier to be able to identify these issues without the need to recompute y.

2.2.1. Techniques of verifiable computation

Verifiable computation employs various techniques to ensure the accuracy of the result computed by the prover (cloud), such as Proof-based Verifiable Computation (PBVC), Replication-based Verifiable Computation (RBVC), and Challenge-based Verifiable Computation (CBVC).

i. Proof-based Verifiable Computation (PBVC): In a proof-based method, the verifier (client) can efficiently and logically check the outcomes when the prover (cloud) delivers the results. The client will accept the results if the overall computation has been performed correctly. On the contrary, if there is any sign of inaccuracy or errors in the computation, the client is prone to confidently reject the results. This technique guarantees that inaccurate results are reliably detected and rejected by the client.

ii. Replication-based Verifiable Computation (RBVC): Replication-based methods utilize multiple cloud services for computation. The results obtained from these multiple sources are then compared to each other to estimate their similarity and ensure accuracy. If the comparison of results from the multiple clouds is successful and aligned, the client accepts the outcome; otherwise, a dispute resolution protocol is initiated to identify the malicious cloud that provides incorrect results. A significant drawback of RBVC is that the client is burdened with the cost of computation for each cloud provider involved.

iii. Challenge-based Verifiable Computation (CBVC): Challenge-based methods involve the outsourcing of computation to a single cloud provider. In this approach, any public party can challenge the outcome provided by the cloud. If there are no objections to the outcome, the client readily accepts it. Otherwise, a dispute resolution method is used.

For a more in-depth understanding of verifiable computation and the techniques involved, one may refer to Yu *et al*. [1].

2.3. Overview of Blockchain-based Verifiable Computation

Verifiable computation has been extensively researched as a method to verify the outcome of an outsourced computation, eliminating the risk of dishonest clients providing inaccurate results without accomplishing the task. However, a centralized third party introduces security challenges in performing verification tasks. However, adopting blockchain technology in verifiable computation can eliminate the centralized third-party requirement. The concept of blockchain-based verifiable computation is proposed by Kumaresan and Bentov [22]. Much research on blockchain-based verifiable computation has been suggested; Figure 2 depicts the basic model of blockchain-based verifiable computation. The model includes four main components: a blockchain system, a worker, a client (Bob), and the Interplanetary File System (IPFS).

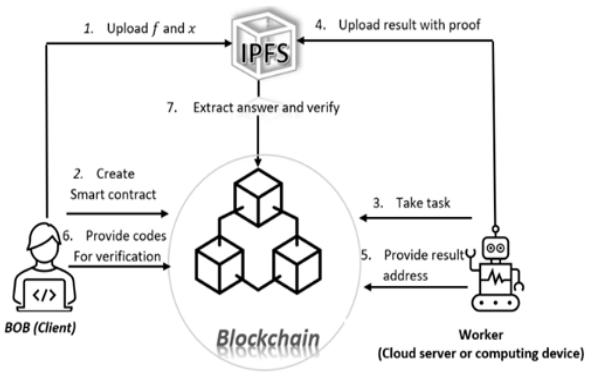

Fig. 2. The basic model of blockchain-based verifiable computation.

- **Client:** In this model, the scenario assumes that Bob requires the evaluation of a function (f) using a specific input x. However, Bob is unable to do computing tasks due to limited computation and power resources. As a result, he decides to delegate the computational task to compute the value of f(x) with logical proof.
- **Worker:** The worker in this scenario may either be a cloud server or a personal computing device, aiming to leverage its processing power for conducting computational tasks on behalf of others. These tasks can be sourced from the blockchain, where they are published by Bob. To access these tasks, the worker might use some explorer, such as Bitcoin Explorer, which facilitates the retrieval of information from the blockchain.

Blockchain: A blockchain system that supports smart contracts, such as Ethereum, to ensure fairness. As illustrated in Figure 2, the blockchain rests between Bob and the worker to monitor their interactions and verify the computation result. Furthermore, it monitors Bob's earnings and the worker's deposit to make sure that anyone who violates protocol pays penalties.

IPFS: In this model, due to the blockchain's limited capacity for data handling, the IPFS is employed to store the evidence of accuracy provided by the worker.

In this situation, both Bob and the worker are motivated by their desired profits. Bob prioritizes ensuring the precision and correctness of computing tasks, while the worker is mainly interested in being rewarded for finishing these tasks. Moreover, section 3 discusses the essential set of standards that can serve as the criteria for evaluating existing blockchain-based verifiable computation schemes.

3. EVALUATION CRITERIA

This section provides a set of criteria that will serve as the evaluation criteria for existing blockchain-based verifiable computation schemes.

- **Verification Correctness (VC):** The term verification correctness fundamentally implies that the assessment of the accuracy of the computed outcome meets the defined standards of the specific verifiable computation scheme. During the verification process, if a blockchain-based verification computation scheme achieves this criterion, it reduces the risk of an incorrect result.
- **Public Verifiability (PV):** As the verifiable computation field is rapidly growing, it is no longer sufficient to allow interested parties to confirm the computational outcomes. It is essential to enable anyone to verify the computational outcomes. In an electronic voting (e-voting) system, a voter can review and confirm any vote given by other voters, to ensure that the voting process is fair and effective. A client can verify the outcomes of data processing carried out by others and rate the quality of the service provider to help in the decision of whether to use their services or not. Moreover, public verifiability schemes can help users verify computational outcomes through public servers by potentially reducing the computational burden on users. Thus, public verifiability has become an essential feature in many application scenarios.
- **Privacy Preservation (PP):** Privacy preservation helps to protect the confidential information of both the client and the service provider. Storing access policies in smart contracts and frequently interacting with them can expose the service provider's sensitive data to the public. To address the privacy issue of access policies, as proposed by Yang *et al*. [23], encrypted access policies can be preserved in smart contracts. However, the ability to provide access permission remains with the cloud, which is not a trusted party in this scenario.
- **Traceability (TR):** Traceability means that selfish mining and malicious clients/workers can be detected quickly. By employing this approach, clients have the ability to monitor and trace any changes made to outsourced data by reviewing the records. When a blockchain-based verifiable computation scheme satisfies the traceability criterion, clients can identify hostile computational tasks by examining the activities. As a result, traceability must be taken into consideration.
- **Zero-knowledge Proof (ZKP):** The term Zero-Knowledge Proof (ZKP) refers to a scenario involving two parties in the system: a prover and a verifier. The prover makes every possible effort to persuade the verifier that it has certain information without revealing

the original information. Simultaneously, the verifier is unable to provide evidence to third parties that the prover has certain information. With this method, the verifier gains only the knowledge that the computation was executed accurately without obtaining any additional information. ZKP is classified into two types: interactive and non-interactive. In an interactive ZKP, the prover performs a series of activities to persuade the verifier that the outcome is accurate, with a certain probability p. The greater the number of rounds of interaction, the greater the probability p. A Non-Interactive Zero-Knowledge Proof (NIZKP) needs no contact between the prover and the verifier. This provides additional flexibility, as the verifier can independently verify the proof at their convenience without the necessity for real-time interaction. If a blockchain-based verifiable computation scheme satisfies ZKP, the service provider could be able to prove to anyone that it performs computational tasks accurately without revealing any sensitive information. In other words, ZKP enhances security.

- **Scalability:** Scalability refers to a platform's ability to handle higher transaction loads while simultaneously accommodating a growing number of nodes within the network. A network may be scaled in two ways: vertically by updating hardware to boost individual machine capacity and horizontally by dividing the workload across multiple machines to satisfy growing needs. If a blockchain-based verifiable computation scheme satisfies the scalability metric, the scheme can process a large amount of transaction throughput efficiently. As a result, scalability must be taken into consideration.
- **Memory and Communication Costs:** Scalability is a common challenge in blockchain-based solutions. Large records that contain all of the data generated in the network must be in local storage (memory) and require significant amounts of peer-to-peer communication. These basic needs lead to significant memory and communication costs, respectively. Therefore, the cost of employing the method for large-scale computational processes over significant multi-agent networks is expensive. During the computation task, if a blockchain-based verification computation scheme satisfies the memory and communication costs criteria, it reduces the scalability problem. As a result, memory and communication costs must be taken into consideration.
- **Efficiency:** The efficiency of any system is the ability to attain the desired outcome with minimum requirements in terms of effort, energy, materials, time, and cost. In general, blockchain systems typically include a distributed ledger for storing blocks of data and a consensus mechanism for generating these blocks. Despite each block being relatively small in size (approximately 2 MB in the case of Bitcoin), the ledger's overall size will expand over time as new blocks are added to the system (typically every 10 minutes). On the other hand, in order to generate blocks, the system will use energy, effort, financial resources, and time to obtain consensus among distributed system nodes. It is necessary to consider efficiency as an important metric for an efficient scheme.

4. METHODOLOGICAL APPROACHES AND IN-DEPTH REVIEW OF BLOCKCHAIN-BASED VERIFIABLE COMPUTATION APPLICATIONS

In this section, related literature is discussed and it covers the method for selecting the surveyed papers. We searched for relevant literature in the leading academic databases, including the IEEE, Elsevier, ACM Digital Library, Springer, and Wiley online library. We employed a two-step literature search technique. In the first stage, the terms "verifiable computation" and "blockchain" were used to search titles, abstracts, and keywords. Considering that in certain publications, "verifiable computation" may be referred to as "verifiable computing" or "authenticated computation," we combined these words with "blockchain" in order to find further similar articles. Finally, we selected 22 research papers that focused on blockchain-based verifiable computation applications. 61% of the publications were published in journals, and 39% were featured in the proceedings of international conferences. Since blockchain technology permits parties to build reliable interactions even if they previously distrusted each other. As a result, it is logical to utilize blockchain to outsource computing, and numerous efforts have been undertaken in industry and academics.

The concept of blockchain-based verifiable computation is introduced by Kumaresan and

Bentov [22]. In the paper, it details how a user creates a Bitcoin output script that specifies an exact payment amount in advance. The computational task is subsequently transferred to cloud processing. The validation of this script can be achieved by either supplying accurate outcomes from the outsourced computation or by revealing specific pre-shared secrets. To summarize, they presented two protocols that encourage verifiable computation schemes. The first scheme compiles any public verification scheme and ensures pay on computation while not protecting client privacy. The second scheme compiles the designated verifier scheme, such as ZKP, preserves client privacy, and punishes hostile workers who provide incorrect proof but do not guarantee pay on computation. However, traceability and scalability are not considered in this work.

Zhang *et al*. [24] and Zhang *et al*. [25] define outsourcing computation using Bitcoin scripts BCPay and Bpay, respectively. The authors developed a robust, fair payment approach, similar to Kumaresan and Bentov [22], in which a cloud provider is paid for computing and only then if it offers valid evidence of correctness. In these works, a checking-proof protocol is provided, and a fair payment system with soundness and robust fairness without reliance on a third party is presented. In these papers, traceability is achieved as relevant operational data is safely stored within the blockchain. In terms of the blockchain, the contents are publicly accessible, and both the server and client can confirm the authenticity of data on the blockchain, ensuring public verifiability. Despite the benefits listed above, their work has several restrictions. In these schemes, the metric of scalability is not considered. BPay and BCPay are potentially vulnerable to malleability attacks, as neither scheme employs private channels. Consequently, these papers do not effectively address the issue of privacy preservation. Furthermore, these approaches did not take ZKP into account; hence, they only achieved minimal verification confidentiality. However, the efficiency metric is considered in the BCPay scheme.

Wang *et al*. [26] presented a fair payment scheme for cloud storage based on the Ethereum network. Blockchain technology facilitates decentralized payment, while payment fairness is guaranteed through a smart contract that includes a pre-existing penalty. In their proposed scheme, there is no trustworthy third party. Their proposed scheme operates in the following manner: When Alice intends to purchase items from an online store, first she needs to register herself as a user, then select her desired item, add it to the shopping cart, and proceed with the payment. Meanwhile, Alice receives a receipt for her order. The seller dispatches the items associated with this order to Alice. The purchased goods are then delivered to Alice. A normal payment procedure has been completed at this point. However, if the seller does not give the products to Alice after she pays, Alice might appeal to the seller at time A and request that the things be sent. The payment procedure is finished if Alice gets the purchase. Otherwise, Alice initiates a penalty transaction at time B to claim the seller's pre-set penalty on the blockchain network. In this way, malicious parties can be traced, and public verifiability and verification correctness can be satisfied. Scalability and efficiency metrics are also considered in their scheme. However, privacy preservation can be compromised due to malicious attacks. The limitation of the scheme presented in this paper lies in its incomplete decentralized structure. The framework is based on a cloud storage platform, which impacts its decentralized structure.

A secure payment for outsourced computations (SPOC) is presented by Krol and Psaras [27]. SPOC is a distributed payment system that permits payments to be transferred between requesting and executing nodes that do not necessarily trust each other. In their proposed system, the requestor (client) uploads computational tasks on a blockchain node, allowing any node within the network to claim and execute them. The blockchain operates independently without the need of any central authority. The integrity of blockchain is maintained by hundreds of miners who charge a minimum fee for their services. Once the computational task is done, the outcomes are sent to the client, and the node that performed the computational task receives payment for its services. trusted execution environments smart contracts, and deposit techniques are employed to guarantee that all involved parties act appropriately. However, privacy preservation is at risk as the information stored in the contract is publicly accessible. Furthermore, ZKP, traceability, scalability and efficiency metrics are not considered in their proposed scheme.

A blockchain-based electronic voting (e-voting) system is presented by Hjlmarsson *et al*. [28]. Their proposed e-voting system is based on smart contracts to guarantee a secure and cost-effective election while protecting the privacy of a voter. However, public verifiability, traceability, scalability, and efficiency are not considered because their proposed system is based on a private blockchain. Additionally, the system also achieved minimum confidentiality as ZKP is not taken into consideration. A scalable method based on smart contracts is presented by Eberhardt and Tai [29]. The authors introduced an off-chain processing model based on NIZKP. This model is proposed to enhance the privacy and transaction performance of blockchain systems. The cloud and the user go through a one-time setup process. The user then creates a contract for verification and uploads it on the Ethereum network. After, the cloud performs the computational task and provides proof of its accuracy. Off-chain processing models have the potential to enhance the scalability and privacy of blockchain networks. Due to the zero-knowledge feature, confidential data used in off-chain computing does not need to be published publicly to verify accuracy and privacy is maintained. However, traceability and public verification metrics are compromised.

A blockchain-based search framework is introduced by Jiang *et al*. [30]. Their proposed scheme allows public verification of outsourced encrypted data. Their suggested scheme relies on Ethereum's smart contract to guarantee that the user receives accurate search results. Furthermore, a stealth authorization scheme is developed to provide privacy-preserving and secure access authorization shipping. Despite those benefits above, traceability, zero-knowledge proof, scalability, and efficiency metrics are compromised. Dorsala *et al*. [31] introduced a fair payment model employing Yang *et al*. [23] as a verification contract. It indicates that when both the cloud provider and the user are authentic, the expense of executing a fair and verifiable process on the Ethereum platform is minimal. In this research, the authors have developed unbiased protocols through the use of smart contracts for two categories of verifiable computations such as replication-based verifiable computation and proof-based verifiable computation. In a proof-based verifiable computation scheme, they employed NIZKP to check the sustainability of the computation. Whereas, replication-based verifiable computing outsources the same calculation to several workers, and the output accuracy is checked by comparing outcomes provided by numerous workers. However, their protocols do not guarantee privacy. Moreover, scalability and efficiency metrics are not fulfilled in their scheme.

Zhou *et al*. [32] introduced MIStore, a blockchain-based system designed for storing threshold-based medical insurance data. In this work, by integrating blockchain technology, the system acquires several unique advantages, such as tamper-resistance, decentralization, and record-nodes, which allow clients to authenticate publicly verifiable data. The tamper-resistance attribute of blockchain provides users with great confidence. Furthermore, because of decentralization, users may interact with each other without the involvement of third parties. Despite the benefits listed above, their work faces some restrictions. MIStore could be vulnerable to malleability attacks due to the lack of private channels, resulting in privacy concerns that are not fully addressed in this paper. Moreover, this scheme did not take ZKP into account; hence, they only achieved minimal verification confidentiality. Although efficiency is considered in their work, the efficiency of MIStore primarily relies on the blockchain platform and the performance of cryptographic schemes employed. As a result, efficiency is notably restricted by the Ethereum blockchain.

Dong *et al*. [20] proposed a solution based on smart contracts with the goal of achieving verifiability and cost efficiency. In their proposed solution, the client outsourced the identical task to two different cloud providers and proposed a prisoner's dilemma scenario between them to prevent collusion. They established three contracts to obtain accurate results from the two reliable clouds. Firstly, the system rewards the honest cloud and penalizes the hostile one. Secondly, clouds can collaborate and use a colluder's contract to solve the prisoner's dilemma. Lastly, the traitor's contract includes an additional reward for the honest cloud in order to counter collusion. Dong *et al*. [20] assume that the client is reliable and a middleman is necessary to settle conflicts when the cloud outcomes do not match. As the blockchain is publicly accessible and data is irreversibly stored

on the blockchain, the privacy of computation input/output is a major concern. To overcome this issue, the authors adopted an appropriate collision-resistant hash function and two other cryptographic techniques: agreements and NIZKP. However, due to NIZK, public verification is not possible in this work.

Avizheh *et al*. [33] explored the concept of verifiable outsourcing through the application of a smart contract on a cryptocurrency blockchain. They implemented the Canetti, Riva, and Rothbulm (CRR) protocol to facilitate verifiable computing between two cloud platforms. The CRR protocol functions in the following manner: The client requests both clouds to perform a specific function f on the input x. Each cloud, identified as $Cloud_m i$ where $m \epsilon 1, 2$, returns its result $y_m = f(x)$ to the client. In the process, if the outcomes from both clouds align, the result is considered valid. Conversely, if there is a mismatch in the results, the client then implements the malicious cloud identification protocol, a mechanism specifically designed to pinpoint which cloud is behaving maliciously by generating inaccurate data. The smart contract is designed to function as a trusted third party, managing interactions between the involved parties and facilitating the required transfer of payments between them. However, a smart contract is incapable of maintaining confidentiality; using it as a trusted third party signifies that the communication channels associated with it are both authenticated and public. This setup ensures transparency and verifiability in the transactions, but it also means that the privacy of the communicated data is limited. Besides, they did not take ZKP into account. Hence, they only achieved minimal verification confidentiality. Furthermore, scalability and efficiency metrics are not fulfilled.

Teutsch and Reitwießner [34] proposed a system called TrueBit. In this work, the computational activities are assigned to a single cloud provider, which then sends the results to a smart contract. Then, challengers are asked to analyze and challenge the accuracy of the results. This step plays an essential role as it enables the verification of the result's accuracy, assuring the authenticity of the process. When a challenger initiates the challenge, it triggers the verification game. This game involves progressively examining smaller segments of the computation in each round, allowing for a detailed scrutiny of the process. In this approach, challengers are encouraged: they receive rewards for identifying errors and face penalties for raising false alarms. To encourage challengers to take an active role in the verification process, it periodically pushes the honest cloud to deliberately submit incorrect results (forced mistakes) and rewards verifiers for detecting errors. TrueBit can safely access and utilize bits of enormous data sets as long as the data is publicly and permanently stored somewhere.

Yang *et al*. [35] presented a new data deletion method based on blockchain that can strengthen the transparency in the deletion process. In their scheme, the data owner may validate the deletion outcome regardless of how maliciously the cloud server behaves. Furthermore, by employing blockchain, the proposed method can accomplish public verification without the use of a trusted third party. Moreover, their approach supports traceability, and for privacy preservation of the data, the owner should encrypt the file before submitting it. However, the scalability metric is not considered in this scheme.

Li *et al*. [36] proposed a novel decentralized framework RepChain. This framework is a blockchain-based system designed to preserve privacy in reputation management, specifically for E-commerce platforms. The presented framework enhances accessibility by allowing access to reputation ratings across many platforms. It computes the total reputation of an entity based on the aggregated ratings it receives. The system is also privacy-preserving and is immune to various rating and anomalous rating assaults. To fight against aberrant rating assaults, they have used ZKP. Besides, the authors employed a consortium blockchain to properly track all rating transactions. In this setup, the reputation of suppliers is traceable through a chain of transactions. However, the scalability metric is not considered in this scheme

In the industry, there are three notable practical systems, including Golem [37], iExec [38], and SONM [39], all leveraging Ethereum for the purpose of outsourcing computing tasks to large-scale computational systems. Golem only provides result verifiability if the user's equipment has more than 8G of memory, and it cannot guarantee that the worker will get the promised award. Based on

its reputation system, SONM can achieve a level of fairness in its operations. iExec maintains system fairness by implementing a combination of proof of contribution, a reputation score mechanism, and majority voting. However, it is important to highlight that, despite these measures, both SONM and iExec do not facilitate result verifiability. Table 3 presents a summary and comparison of existing works on blockchain-based verifiable computing methods.

Recent research has focused on enhancing security and privacy in distributed systems using blockchain and Generative Adversarial Network (GAN) technologies [40-41]. Ghani *et al*. [40] proposed a technique to address significant concerns regarding data integrity and privacy in facial recognition applications by combining Blockchain technology, micro-batch aggregation, and GANs. Furthermore, Ghani *et al*. [41] introduced a novel framework that addresses privacy concerns while maintaining accurate face recognition. The proposed framework combines cutting-edge methods such as distributed computing, blockchain, and GANs. Accurate face recognition and the preservation of the integrity of personal data are balanced in this system by utilizing tools like Dlib for face analysis, Ray Cluster for distributed computing, and Blockchain for decentralized identity verification. Moreover, Magsi *et al*. [42] presented a method for detecting and preventing Content Poisoning Attack (CPA) in Vehicular Named Data Networks (VNDN) by integrating a threshold-based content-caching mechanism with blockchain technology. Although blockchain-based verifiable computation systems have the potential to ensure fairness, traceability, and transparency, challenges with scalability, efficiency, and privacy still need to be addressed. Recent research combining cutting-edge technologies such as off-chain processing and GANs suggests a promising direction, but practical implementations need to balance these factors to achieve trustworthy, scalable, and secure methods.

5. OPEN ISSUES AND FUTURE RESEARCH DIRECTIONS

In light of the thorough literature review of existing blockchain-based verifiable computation schemes, several open issues have been highlighted to

Table 3. Summary and Comparison of Blockchain-Based Verifiable Computation Methods.

S. No.	Paper	System / Scheme	VC	PV	PP	TR	ZKP	Scalability	Memory and Communication Costs	Efficiency
1	[20]	—	✓	×	✓	✓	✓	×	×	✓
2	[22]	—	✓	✓	✓	×	✓	×	×	✓
3	[24]	BCPay	✓	✓	×	✓	×	×	×	✓
4	[25]	BPay	✓	✓	×	✓	×	×	×	×
5	[26]	—	✓	✓	×	✓	×	✓	✓	✓
6	[27]	SPOC	✓	✓	×	×	×	×	×	×
7	[28]	E-Voting	✓	×	✓	×	×	×	×	×
8	[29]	ZoKrates	✓	×	✓	×	✓	✓	✓	✓
9	[30]	—	✓	✓	✓	×	×	×	×	×
10	[31]	—	✓	✓	×	✓	✓	×	×	×
11	[32]	MIStore	✓	✓	×	✓	×	×	×	✓
12	[33]	—	✓	✓	×	✓	×	×	×	×
13	[34]	ZoKrates	✓	×	✓	×	✓	✓	✓	✓
14	[35]	Data deletion scheme	✓	✓	✓	✓	×	×	×	✓
15	[36]	RepChain	✓	✓	✓	✓	✓	×	×	✓

Note: —: Not available; ✓: Support this requirement; ×: Without consideration

encourage future blockchain-based verifiable computing research.

5.1. Open Issues

Drawing from the review and comparative study of the existing paper, we have observed the following open issues in the research of blockchain-based verifiable computation schemes.

Firstly, the aspect of privacy preservation appears to be neglected in most existing literature. In the realm of blockchain-based verifiable computation schemes, it is insufficient to merely obtain results that are correct and can be publicly verified. It is necessary to preserve the privacy of data owners and service providers. Therefore, the challenge of ensuring privacy preservation in blockchain-based verifiable computation schemes stands as an open issue that should be addressed. Secondly, it's noteworthy that the existing research and implementations seldom consider the metric of ZKP. So, high confidentiality cannot be achieved, sensitive information can be exposed, and User privacy can be in danger. Thirdly, the blockchain-based verifiable computation scheme should be scalable. validated and processed by every node in the network, every blockchain system is limited in terms of scalability. However, most of the existing works do not consider scalability metrics. Therefore, scalability is also a prominent issue that should be addressed. Fourthly, traceability stands as a key metric in the realm of blockchain-based verifiable computation methods. It is essential for service providers to furnish not only the most recent data but also past records (or data operation logs). These logs enable clients to identify any illegal activities by the service provider. However, most existing works do not adequately address traceability metrics. Although operation logs on the blockchain can improve traceability, this leads to an increased volume of operation logs, which in turn can slow down the processing of transactions. As a result, developing an effective approach to guarantee traceability remains a challenge that needs to be addressed with caution.

5.2. Future Research Directions

After reviewing the literature and examining the open challenges, it becomes clear that blockchain technology still has significant progress to make before it can be effectively utilized for verifiable computation. This subsection discusses various possible future research directions in the realm of verifiable computation schemes based on blockchain technology. Firstly, privacy preservation is expected in blockchain-based verifiable computation schemes. As proposed by Yang *et al*. [23], one direct solution to address privacy issues related to access policies is to store these policies in encrypted form within smart contracts. Yet, in such a situation, the authority to approve access permits remains with the cloud, an entity that is not considered as trustworthy. Therefore, designing a trusted party blockchain-based verifiable computation solution that can support privacy preservation is a significant future research topic. Secondly, ZKP enhances security, privacy, and safety. If a blockchain-based verifiable computation scheme satisfies zero-knowledge proof, the service provider can achieve the ability to prove to anyone that it performs accurate tasks without revealing any private information. Therefore, for future work, there is a need for additional exploration into the ZKP aspect to extend its application to a broader range of blockchain-based verifiable computation scenarios. Thirdly, Scalability is also a prominent issue in blockchain-based verifiable computation schemes. A direct approach to address this issue is that researchers can utilize off-chain computations. However, a scalable blockchain-based verifiable computation approach is a potential direction. Fourthly, traceability is a crucial metric in blockchain-based verifiable computation methods. Owing to the limited performance of distributed nodes, maintaining extensive blockchain operation records suffers significant storage expense and slows transaction processing. One possibility is to implement the IPFS, which can be seen in Figure 2. A decentralized file system like this is secure and highly efficient for preserving large-scale operation records. It is achieved by addressing each file uniquely and taking maximum advantage of the storage space of each node in the network. As a result, developing an affordable mechanism for preserving operational records is a promising field of research.

6. CONCLUSIONS

Over the last decade, verifiable computation has become immensely popular. However, verifiable computation is limited by centralization, lack of

transparency and lack of trust. On the contrary, blockchain is a new technology that is being embraced in a wide range of engineering sectors, including verifiable computing. In this paper, we provided a detailed review of blockchain-based verifiable computation techniques. Initially, we covered the fundamentals of blockchain technology, verifiable computation, and blockchain-based verifiable computation. We highlighted the key components of the blockchain that are essential to reengineering verifiable computing. Next, we presented a series of evaluation criteria for existing verifiable computation techniques based on a distributed ledger. Using the proposed criteria, we thoroughly evaluated, examined, and compared existing works. The literature review of existing blockchain-based verifiable computation techniques has revealed several open issues and potential areas for future research. A significant issue, which is often neglected in existing research, is the requirement for robust privacy preservation methods to guarantee the confidentiality of both data owners and service providers. By applying ZKP, user privacy and confidentiality can be improved. Moreover, scalability is a common challenge in blockchain-based solutions. As blockchain transactions must be validated and processed by every node in the network, every blockchain system is limited in terms of scalability. However, innovative approaches, such as off-chain computations can enhance performance and manage the growing demand for computational resources. Traceability is essential in blockchain-based verifiable computation schemes yet unexplored metrics. It is essential to ensure the accountability and integrity of service providers by allowing clients to check the validity and history of data transfers. Utilizing decentralized file systems like IPFS can improve traceability. Future research should be focused on robust privacy-preserving methods, using ZKP for enhanced security, off-chain computations for scalability, and using decentralized file systems like IPFS to improve traceability.

7. CONFLICT OF INTEREST

The authors declare no conflicts of interest.

8. REFERENCES

1. X. Yu, Z. Yan, and A.V. Vasilakos. A Survey of Verifiable Computation. *Mobile Networks and Application* 22: 438–453 (2017).
2. S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system (2008). https://bitcoin.org/bitcoin.pdf (accessed 10 February 2024).
3. M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman. Blockchain technology: Beyond bitcoin. *Applied Innovation* 2: 6-10 (2016).
4. S. Šimunić, D. Bernaca, and K. Lenac. Verifiable computing applications in blockchain. *IEEE Access* 9: 156729-156745 (2021).
5. M.R. Dorsala, V.N. Sastry, and S. Chapram. Blockchain-based solutions for cloud computing: A survey. *Journal of Network and Computer Applications* 196: 103246 (2021).
6. H. Gamage, H. Weerasinghe, and N. Dias. A survey on blockchain technology concepts, applications, and issues. *SN Computer Science* 1: 1–15 (2020).
7. S. Soni, and B. Bhushan. A comprehensive survey on blockchain: working, security analysis, privacy threats and potential applications. *2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kerala, India* (July 5-6, 2019) 1: 922–926 (2019).
8. S. Shi, D. He, L. Li, N. Kumar, M.K. Khan, and K.K. R. Choo. Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. *Computers & Security* 97: 101966 (2020).
9. L. Peng, W. Feng, Z. Yan, Y. Li, X. Zhou, and S. Shimizu. Privacy preservation in permissionless blockchain: A survey. *Digital Communications and Networks* 7: 295–307 (2021).
10. M.R. Ahmed, A.K.M.M. Islam, S. Shatabda, and S. Islam. Blockchain-Based Identity Management System and Self-Sovereign Identity Ecosystem: A Comprehensive Survey. *IEEE Access* 10: 113436-113481 (2022).
11. W. Li, J. Wu, J. Cao, N. Chen, Q. Zhang, and R. Buyya. Blockchain-Based Trust Management in Cloud Computing Systems: A Taxonomy, Review and Future Directions. *Journal of Cloud Computing* 10(1): 35 (2021).
12. A.M.S. Saleh. Blockchain for Secure and Decentralized Artificial Intelligence in Cybersecurity: A Comprehensive Review. *Blockchain: Research and Applications* 5: 100193 (2024).
13. M.M. Memon, M.A. Hashmani, F.T. Simpao, A.C. Sales, N.Q. Santillan, and D. Khan. Blockchain in Healthcare: A Comprehensive Survey of

Implementations and a Secure Model Proposal. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences* 60(3): 1-13 (2023).

14. M. Rosenfeld. Overview of colored coins (2012). https://bitcoil.co.il/BitcoinX.pdf (accessed 10 February 2024).
15. M. Conoscenti, A. Vetro, and J.C. De Martin. Blockchain for the internet of things: A systematic literature review. *IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco* (November 29 - December 2, 2016) pp. 1–6 (2016).
16. S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.Y. Wang. Blockchain-enabled smart contracts: Architecture, applications, and future trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49: 2266–2277 (2019).
17. K. Gai, J. Guo, L. Zhu, and S. Yu. Blockchain meets cloud computing: A survey. *IEEE Communications Surveys Tutorials* 22: 2009–2030 (2020).
18. D.C. Nguyen, P.N. Pathirana, M. Ding, and A. Seneviratne. Blockchain for 5g and beyond networks: A state of the art survey. *Journal of Network and Computer Applications* 166: 102693 (2020).
19. F. Zafar, A. Khan, S.U.R. Malik, M. Ahmed, A. Anjum, M.I. Khan, N. Javed, M. Alam, and F. Jamil. A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends. *Computers & Security* 65: 29–49 (2017).
20. C. Dong, Y. Wang, A. Aldweesh, P. McCorry, and A. van Moorsel. Betrayal, distrust, and rationality: Smart counter-collusion contracts for verifiable cloud computing. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, *Dallas, Texas, USA* (30 October – 3 November 2017) pp. 211–227 (2017).
21. N.Z. Benisi, M. Aminian, and B. Javadi, Blockchain-based decentralized storage networks: A survey *Journal of Network and Computer Applications* 162: 102656 (2020).
22. R. Kumaresan, and I. Bentov. How to use bitcoin to incentivize correct computations. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, *Scottsdale, Arizona, USA* (November 3-7, 2014) pp. 30–41 (2014).
23. C. Yang, L. Tan, N. Shi, B. Xu, Y. Cao, and K. Yu. AuthPrivacyChain: A blockchain-based access control framework with privacy protection in cloud. *IEEE Access* 8: 70604–70615 (2020).
24. Y. Zhang, R.H. Deng, X. Liu, and D. Zheng. Blockchain based efficient and robust fair payment for outsourcing services in cloud computing. *Information Sciences* 462: 262–277 (2018).
25. Y. Zhang, R.H. Deng, X. Liu, and D. Zheng. Outsourcing service fair payment based on blockchain and its applications in cloud computing. *IEEE Transactions on Services Computing* 14: 1152–1166 (2018).
26. S. Wang, Y. Wang, and Y. Zhang. Blockchain-based fair payment protocol for deduplication cloud storage system. *IEEE Access* 7: 127652–127668 (2019).
27. M. Krol, and I. Psaras. Spoc: Secure payments for outsourced computations. *arXiv preprint arXiv*: 1807.06462 (2018).
28. F. Hjlmarsson, G.K. Hreiarsson, M. Hamdaqa, and G. Hjlmtsson. Blockchain-based e-voting system. *In 2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, *San Francisco, CA, USA* (July 2-7, 2018) pp. 983–986 (2018).
29. J. Eberhardt, and S. Tai. Zokrates-scalable privacy-preserving off-chain computations. *IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, *Halifax, NS, Canada* (July 30 – August 3, 2018) pp. 1084–1091 (2018).
30. S. Jiang, J. Liu, L. Wang, and S.M. Yoo. Verifiable search meets blockchain: A privacy-preserving framework for outsourced encrypted data. In ICC 2019-2019 *IEEE International Conference on Communications (ICC)*, *Shanghai, China* (May 20-24, 2019) pp. 1–6 (2019).
31. M.R. Dorsala, V. Sastry, and S. Chapram. Fair payments for verifiable cloud services using smart contracts. *Computers & Security* 90: 101712 (2020).
32. L. Zhou, L. Wang, and Y. Sun. Mistore: a blockchain-based medical insurance storage system. *Journal of Medical Systems* 42: 1–17 (2018).
33. S. Avizheh, M. Nabi, R. Safavi-Naini, and M. Venkateswarlu K. Verifiable computation using smart contracts. *Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop*, *London, UK* (November 11, 2019) pp. 17–28 (2019).
34. J. Teutsch, and C. Reitwießner. A scalable verification solution for blockchains. *arXiv preprint arXiv*:1908.04756 (2019).
35. C. Yang, X. Chen, and Y. Xiang. Blockchain-based publicly verifiable data deletion scheme for cloud storage. *Journal of Network and Computer*

Applications 103: 185–193 (2018).
36. M. Li, L. Zhu, Z. Zhang, C. Lal, M. Conti, and M. Alazab. Anonymous and verifiable reputation system for e-commerce platforms based on blockchain. *IEEE Transactions on Network and Service Management* 18: 4434–4449 (2021).
37. J. Zawistowski, P. Janiuk, A. Regulski, A. Skrzypczak, A. Leverington, P. Bylica, M. Franciszkiewicz, P. Peregud, A. Banasiak, M. Stasiewicz, and R. Zagórowicz. The Golem Project. *Golem Whitepaper* (2016). https://assets.website-files.com/62446d07873fde065cbcb8d5/62446d07873fdeb626bcb927_Golemwhitepaper.pdf (accessed 30 December 2023).
38. G. Fedak, H. He, O. Lodygensky, and E. Alves. iExec Blockchain-based decentralized cloud computing v3.0 (2018). https://iex.ec/wp-content/uploads/2022/09/iexec_whitepaper.pdf (accessed 30 December 2023).
39. SONM. Supercomputer organized by network mining (2017). https://whitepaper.io/document/326/sonm-whitepaper (accessed 30 December 2023).
40. M.A.N.U. Ghani, K. She, M.A. Rauf, M. Alajmi, Y.Y. Ghadi, and A. Algarni. Securing Synthetic Faces: A GAN-Blockchain Approach to Privacy-Enhanced Facial Recognition. *Journal of King Saud University-Computer and Information Sciences* 36(4): 102036 (2024).
41. M.A.N.U. Ghani, K. She, M.A. Rauf, S. Khan, J.A. Khan, E.A. Aldakheel, and D.S. Khafaga. Enhancing Security and Privacy in Distributed Face Recognition Systems through Blockchain and GAN Technologies. *Computers, Materials & Continua* 79(2): 2610 (2024).
42. A.H. Magsi, L.V. Yovita, G. Ali, G. Muhammad, and Z. Ali. A Content Poisoning Attack Detection and Prevention System in Vehicular Named Data Networking. *Sustainability* 15(14): 10931 (2023).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 129-144 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)863

Pakistan Academy of Sciences

Review Article

Machine Learning, Deep Learning, and Hybrid Approaches in Real Estate Price Prediction: A Comprehensive Systematic Literature Review

Rabia Naz[1], Bushra Jamil[2], and Humaira Ijaz[2*]

[1]Department of Software Engineering, University of Sargodha, Sargodha, Pakistan
[2]Department of Information Technology, University of Sargodha, Sargodha, Pakistan

Abstract: The real estate refers to an extensive field that deals with the purchase, selling, or management of properties, and it stands out as an influential industry in the economic development process, indicating that the precise determination of the price is one of the most effective tools for decision-making among different subjects of the market and authorities. Price prediction improves investment plans, risk management, fair price transactions, and provides key inputs to economic and urban planning. This systematic review categorizes the existing approaches into three groups: machine learning, deep learning, and hybrid models, based on the selected literature from a broad search of numerous databases and using rigorous criteria. The review indicates that the traditional and current machine learning models have relatively high levels of predictive accuracy for small datasets. However, deep learning techniques are preferable for handling large and complex data, while hybrid models have even more potential to increase prediction accuracy. The present study indicate that these sophisticated techniques can enhance and enrich price forecasting models, which can be insightful to various industrial decision-makers and informative for future research endeavours.

Keywords: Real Estate, Machine Learning, Deep Learning, Price Prediction, Hybrid Approach.

1. INTRODUCTION

In recent economic research and policy debates, considerable attention has been devoted to examining the impact of asset prices on macroeconomic policy. One such asset class that has played a significant role is real estate [1]. Real estate refers to land and all permanent structures or developments affixed to it whether natural or manmade. It encompasses houses, buildings, infrastructure, land and other precious resources like minerals, trees or water availability on the land. During the pandemic, the real estate market experienced significant shifts and continued expansion. The Real Estate Association estimates that in 2019 the GDP was boosted by the real estate sector by 7.62% [2]. The global economy is greatly impacted by the thriving real estate sector. The Asia-Pacific area had a 6% increase in transactions, while the European markets saw an 8% growth [3]. The entire spectrum of real estate investing and economic activity is included in the real estate market including transactions, investments, and ancillary services such as brokerage, appraisal, real estate consultancy, and management. It involves interactions among various stakeholders, such as real estate developers, sales representatives, prospective users, and intermediaries like brokers and appraisers. Real estate transactions are grounded in commodity and monetary relations, unfolding within specific temporal and spatial dimensions [4].

Real estate prices undergo analysis for rental assessments but their primary focus centers on property valuation often facilitated through automated valuation models. These statistically dependent automated valuation models provide current estimates of market values for particular properties based on real estate data including age, number of rooms, and comparable transactions in addition to pricing patterns. Valuations, a ubiquitous requirement, are frequently conducted by various market participants who work in real estate such

Received: March 2024; Revised: April 2024; Accepted: June 2024
* Corresponding Author: Humaira Ijaz <humaira.bilalrasul@uos.edu.pk>

as brokers, mortgage lenders, appraisers, investors, fund managers, market researchers, and analysts [5]. The real estate software market was anticipated to be valued at $9.73 billion in 2021 with an estimated growth rate of 9.7% from 2021 to 2028 [6].

The real estate market is important to the country's economy, as it plays a crucial role in many sectors lie urban planning, investment decision-making and formulating government policies. Thus, it is advantageous for investors, homeowners, government and regulatory bodies, and monetary authorities to keep an eye on market trends and make precise predictions on real estate prices. Moreover, the price prediction models help in risk evaluation, investment management, and preventing financial crises due to real estate market downturns. However, because there are so many direct and indirect variables that affect prediction accuracy [7]. The rising number of customer-reported post-purchase or post-rental regrets is causing industry anxiety. According to Trulia, 44% of real estate buyers regret their choices either purchases or rentals. This remorse is mostly attributable to a lack of knowledge about properties and the intricate nature of the acquisition procedure which leaves important information like fees hidden [8]. The current real estate management system cannot provide predictive insights into property prices for users.

Due to its multifaceted importance, many researchers have applied various techniques from traditional statistical models to advanced machine learning techniques. These traditional models include econometric models, and autoregressive integrated moving average models which have been used for real estate price prediction. However, recently machine learning techniques have radically changed the field, fostering the analysis of massive datasets and the recognition of complex patterns that traditional methods could not identify. Therefore, the use of machine learning to estimate real estate values is one of the most promising approaches [9]. This research aims to conduct a thorough examination of papers that are specifically focused on real estate price prediction. Our attention goes beyond the investigation of predictive algorithms to the technology that supports real estate price forecasting. Under close examination, we clarify the critical importance, diverse functions, and uses of high-tech methods—machine learning, deep learning, and hybrid approaches in real estate price prediction. An analysis of the previous studies reveals not only the complex workings of predictive models but also the revolutionary power and priceless insights provided by machine learning, deep learning, and hybrid approaches. This adds to the growing conversation about the nexus between technology and real estate. Furthermore, this study aims to reveal trends, difficulties, and possibilities within the existing literature that can serve as valuable information for further research and application of findings. By identifying the pros and cons of each strategy and showcasing how these technologies can be further employed strategically, we seek to enhance awareness of how these technologies can help address the emerging needs of the real estate industry.

Zulkifley *et al.* [10] examined machine learning algorithms for predicting home prices. Aspects classified as locational, structural, neighborhood, and economic factors were used in the investigations. XGBoost, multiple linear regression, artificial neural network, and support vector regression were utilized. Particularly, locational characteristics played a critical role in support vector regression, artificial neural network, and XGBoost's house value prediction. However, this study had limitations, such as a lack of a unified evaluation metric and a focus on specific attribute categories. Yalgudkar *et al*. [11] performed a survey on housing price prediction. This survey's strengths lie in the application of these diverse algorithms to predict housing prices, with notable mentions of accuracy improvements using random forest and gradient boosting. The survey provided insights into the challenges faced and suggests future research directions, accentuating the coupling effect of multiple regression models, exploring machine learning and deep learning methods, and finding efficient ways to apply complex models. Tekouabou *et al.* [12] explored machine learning applications in real estate prediction using SCOPUS-indexed papers from 2008 to 2022. With a peak in 2021, seventy-two articles were analyzed, revealing a preference for simple machine learning algorithms over deep learning. Dominant countries included China, the United States, and India. Common methods included decision trees, random forests, neural networks, boosting trees, support vector machine, and linear regression. Geerts *et al.* [13] analyzed 93 papers on residential property valuation,

spanning from 1992 to 2021, categorizing them based on model and data novelty scores. The most widely utilized hedonic model types were structural equation modelling and multiple regression analysis. Utilizing random forest, gradient-boosted trees, neural networks, deep learning and other sophisticated machine learning techniques gained popularity. Conspicuously, the most recent papers (2020–2021) explored advanced machine learning and deep learning techniques with advanced spatial data, images, graphs, and text, indicating a shift toward more innovative approaches.

In Table 1, we present a summary comparing various surveys on price prediction in real estate, with an emphasis on the different machine learning and deep learning techniques used. Each reference is associated with its objective, the key models used in the study, and the limitations identified in the research. While existing studies have contributed valuable insights, there is a discernible gap in the comprehensive analysis of recent methodologies and their applicability to diverse real estate markets. Our goal is to synthesize findings from myriad sources, identify common trends, methodologies, algorithms, attributes, and features, address specific limitations in current approaches and highlight future development areas. By taking this approach, we hope to offer a more nuanced view of the current situation and offer fresh insights that will improve real estate price prediction models' precision and applicability.

The contribution of our paper is four folds given as follows:

- Based on comprehensive literature synthesis the existing approaches used for real estate price prediction are divided into three categories: machine learning, deep learning and hybrid model.
- This review provides a comprehensive analysis by integrating both qualitative and quantitative factors used to get more accurate real estate price prediction.
- In-depth analysis of algorithms and methodologies used in existing studies to highlight their strengths and address potential limitations.
- Providing valuable insights with practical implications for researchers, practitioners, and policymakers in real estate.

2. METHODOLOGY

This systematic literature review implements the preferred reporting items for systematic reviews and meta-analyses framework to assess the actual value of the real estate forecast, as shown in Figure 1. Systematic reviews and meta-analyses performed using the methodological approach of random search, selection, assessment, and result aggregation allowed the authors to maintain transparency, consistency, and repeatability at the screening stage of the review process. This explication of methodology covers the sequential phases of our data collating plan i.e. identification, filtering, inclusion screening, and eventually, inclusion, in full.

2.1. Identification

The first step includes a thorough survey of the relevant resources spanning a wide range of search tools in major academic databases. Google Scholar and Scopus were the two major databases

Table 1. Comparative Analysis.

Ref.	Objective	Key Models	Limitations
[10]	Evaluate house price prediction	MLR, SVR, ANN, XGBoost	Inconsistent RMSE reporting, limited focus on structural attributes.
[11]	Housing price prediction using ML/DL	MLR, Lasso, Ridge, SVM, RF, ANN, XG Boost	Need for larger datasets, consideration of additional features, and inclusion of global factors like inflation and GDP
[12]	Analyze ML in real estate prediction	DT, boosting trees, NN, RF, SVM, LR	Limited exploration of developing countries, bias towards structured data, and challenges in explainability.
[13]	Explore ML/DL trends in property valuation	MRA, SEM, RF, GBT, NN, DT, DL	Limited availability of large, high-quality datasets, challenges in handling diverse feature sets, potential gap between academia and industry practices

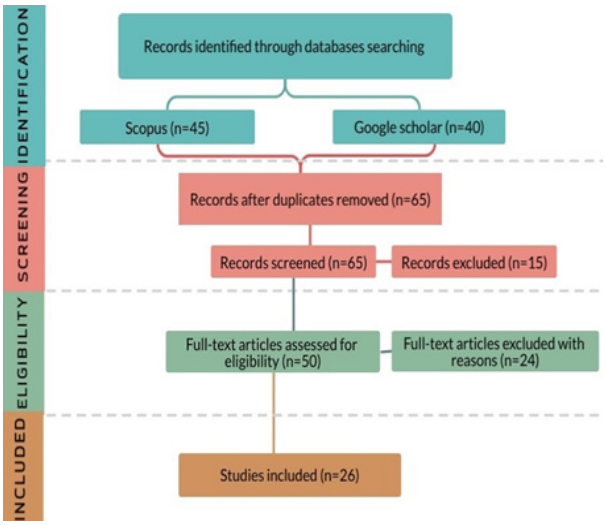

Fig. 1. PRISMA Framework.

that were used for the research investigation. In the interest of getting the most relevant papers out there on the prediction of real estate prices, a careful search strategy was designed. Searches with important words such as real estate, and price prediction, and their synonyms were used in the search engine. A total of 40 records were screened from Google Scholar and 45 records from Scopus were the initial search. These documents were then placed in the referencing software while maintaining the quality and gatekeeping role by eliminating duplicates. Table 2 outlines databases as well the search keywords that we have used in researching real estate price prediction. The data collection databases consist of Scopus and Google Scholar while the keywords are in real estate, price prediction, machine learning, and deep learning.

2.2. Screening

After removing duplicates, 65 unique records remained for further evaluation. During the screening phase, each record's title and abstract were studied to determine their relevance to our research question. Records that did not pertain to real estate price prediction were excluded at this stage. The screening process aimed to refine the list of potentially eligible studies.

2.3. Eligibility and Exclusion

Following the screening phase, the remaining 50 records were assessed for eligibility in more depth. We gathered full-text articles and carefully reviewed them to see if they satisfied the predetermined inclusion requirements. Criteria were set for the research including how good the work was, the possibility of the study to predict real estate prices and which method was used in the research. A total of 24 full-text publications were not included in the analyses because they were not qualified according to the rigidly set requirements. Thorough explanations were given for their exclusion and reasons for rejecting, which injected the decision-making process with transparency and fairness. Then, only twenty-six fulfilling articles that are resistant to the next phase are to be identified.

2.4. Inclusion

In the last step, we ended up with twenty-six studies which were based on certain specified parameters that we had predefined. The formation of our extensive research on real estate price projection partly depends on the studies we conducted. Therefore, the selected studies were incorporated into this systemic review based on the evidence of their methodology, the importance of their conclusions, and the strategies they used to address the main research issues.

3. FACTORS AFFECTING REAL ESTATE PRICE PREDICTION

It is typical to deal with qualitative or-quantitative features when discussing real estate and its attributes in an analysis of the real estate market. Qualitative characteristics are usually expressed orally. This provides information about the feature's status (such as its use as residential, recreational, agricultural, or industrial property) or allows for ranking variations (such as neighborhood characteristics, from most vulnerable to most compelling or from lowest to highest) [14]. The subjective preferences of

Table 2. Databases and Search Keywords.

No.	Databases	Keywords
1	Scopus	("real estate" AND "price prediction") OR ("real estate" AND "machine learning") OR ("real estate" AND "deep learning") OR ("real estate" AND "hybrid")
2	Google Scholar	("real estate" AND "price prediction") OR ("real estate" AND "machine learning") OR ("real estate" AND "deep learning") OR ("real estate" AND "hybrid") OR "land price"

decision-makers regarding viewpoints, architectural styles, and living situations are included in these qualitative components. However, a lack of accurate measurements can occasionally affect the qualitative results for these factors. Conversely, real estate characteristics, business cycles, and macroeconomic factors can all be considered quantitative factors. Metrics such as industrial production, gross domestic product, share indexes, unemployment rates, and a nation's current account are examples of macroeconomic factors. Past selling prices, land acreage, building years, floor area, surface area, number of stories, and building conditions are all considered real estate features [6]. Figure 2 shows a summary of the attributes involved in real estate price prediction.

3.1. Quantitative Attributes

Quantitative attributes in real estate provide measurable, numeric insights for objective analysis. Macroeconomic indicators like unemployment rates, share indexes, current accounts, industrial production, and gross domestic product offer a broad economic context. Attributes like past sales prices, land area, construction year, floor space, surface area, number of floors, and building conditions that accompany the characteristics of the property in question may form property-specific quantitative attributes. This criterion lets the stakeholders acquire analytical algorithms and data on the way to make sound decisions and conduct a full assessment of the economic viability and investment avenues.

Fig. 2. Attributes involved in the real estate.

3.2. Qualitative Attributes

The qualitative factors that are inseparable from real estate include subjective and non-numeric elements which give beauty to a property. The use of the property, either for housing, recreation, farming or industry, determines its respective role. Classification of a neighborhood as unfavorable, average, or favorable is according to certain features. Cultural peculiarities in architecture appear in the form of view preferences, building types, and the living environment. Although these subjective aspects are not measured with exactness, these qualitative factors have a considerable impact on the perceived desirability adding to a broad comprehension of a property's market value. A comprehensive analysis of the real estate property typically involves qualitative and quantitative parameters which is important for a better understanding.

4. TECHNIQUES USED FOR REAL ESTATE PRICE PREDICTION

This section scrutinizes methodologies outlined in the targeted scholarly articles, classifying them into three distinct categories: machine learning, deep learning, and hybrid-based techniques. This part of the discussion deals with the main ideas of each academic paper which is in the main target. Figure 3 provides an extensive examination of machine learning, deep learning, and hybrid approaches used in the selected studies. This visual aid provides a synopsis of the many methodologies employed in the research, exhibiting the range of models and techniques employed for different purposes, including real estate analysis and price forecast. This figure contributes to a comprehensive knowledge of the approaches used in the area by offering a consolidated perspective of the methodological landscape utilized in the analyzed publications.

4.1. Machine Learning

Artificial intelligence's machine learning discipline enables computers to learn from large, complex datasets without the need for explicit programming [15]. Through the application of mathematical and statistical tools, machine learning endows

Fig. 3. All the methodologies mentioned in targeted studies.

machines with the ability to autonomously execute intellectual tasks that were traditionally the domain of human beings [16]. Essentially focused on deriving models from data, machine learning, often employed for prediction purposes, is intrinsically linked with the notion of uncertainty [17]. Since the 1990s, academic research has investigated machine learning use in predicting residential prices. Machine learning-based models are a good option for stakeholders looking to forecast home prices because of their effectiveness in solving problems that are beyond the scope of human capacity.

The capability of machine learning techniques to find functional connections in historical records has been demonstrated by the widespread use of these algorithms to anticipate real estate values in recent years [18]. Within machine learning, there are three categories:

- Supervised Learning: It uses example input-output pairs to build a function that draws inputs to outputs based on labelled training records. This method uses tagged samples to infer the intended function based on a predefined set of objectives [19].
- Unsupervised Learning: By evaluating unlabeled datasets without the need for human interaction, it highlights a data-driven approach [19].
- Semi-supervised Learning: It is a hybrid technique that may be used for both labelled and unlabeled data. It blends aspects of supervised and unsupervised approaches. Semi-supervised learning, which falls in between "without supervision" and "with supervision," is especially useful in real-world situations when there is a dearth of labelled data and a surplus of unlabeled data. Predictions produced with a semi-supervised learning model should ideally perform better than those made with only tagged input [19].

Gampala *et al.* [20] used a methodology that utilized supervised learning algorithms within the domain of machine learning. The researchers used a variety of techniques, including random forest classification [21], decision trees [22], naive bayes [23], and linear regression [24]. The researchers concluded that linear regression proved to be the most effective in predicting house values based on the provided dataset.

Tchuente *et al.* [25] used machine learning algorithms, covering neural networks (MLP), random forest, adaboost [26], gradient boosting [27], and KNN [28]. For neural networks, specific hyperparameters include variations in network architecture, activation functions, learning rates, and optimizers. The best-performing ensemble learning techniques - random forest, AdaBoost, and gradient boosting are emphasized. The researchers highlighted the crucial role of geographic coordinates, introduced through geocoding, in enhancing the predictive accuracy of these models.

Uzut *et al.* [29] discussed data mining methods and focused on three primary methodologies: linear regression, random forest, and gradient boosting. 414 real estate properties are included in the dataset, which was gained from the University of California, Irvine. The outcomes of their research divulged that the gradient-boosting algorithm attains the highest accuracy.

Al Kurdi *et al.* [30] used many classification methods for modelling [31] and forecasting resale home values, including decision tree, random forest, AdaBoost, naïve bayes, and logistic regression. Decision tree C5.0 showed remarkable accuracy, a TNR over 92%, and a TPR above 92% when the study evaluated the performance of many algorithms.

Three machine learning techniques were applied by Ho *et al.* [32]: SGD-based SVR [33], random forest, and gradient boosting. The researchers compared these algorithms to estimate housing prices, emphasizing their accuracy and prediction capabilities. While support vector machine was highlighted for its computational efficiency, random forest and gradient boosting demonstrated superior predictive accuracy with lower errors. The study concluded that machine learning, specifically random forest and gradient boosting, holds promise for accurate property price predictions.

To assess the prices of real estate transactions in Taichung, Taiwan, Pai *et al.* [34] employed BPNN [35], CART [36], GRNN [37], and LSSVR [38] to gather, purify, and restructure real estate attribute data. Genetic algorithms [39] were utilized for model parameter optimization, and a 5-fold cross-validation assessed model robustness. As a consequence, three machine learning models provided extremely accurate predictions, while one performed well. The LSSVR model standed out as the most accurate, surpassing previous studies in MAPE measurements.

Truong *et al.* [40] focused on random forest, XGBoost, and lightGBM for housing price prediction. The researchers introduced hybrid regression and stacked generalization regression [41]. Data analysis highlighted location, age, and various features' impact on prices. The evaluation measure was RMSLE. Random forest exhibited low overfitting, while hybrid regression outperformed the training set. Stacked generalization regression exceled in generalization on the test set. Further research was suggested on factors influencing tree-based model performance and combining machine learning, deep learning methods.

Dalal *et al.* [42] discussed the application of support vector regression for predicting real estate prices in China. It reviewed various studies on real estate market determinants. support vector regression was contrasted with BPNN. The results of the investigations revealed that the SVR model outperformed the BPNN model in terms of MAE, MAPE, and RMSE.

Sanjar *et al.* [43] investigated real estate price variation in Taipei, Taiwan, focusing on the Cathay House Price Index and Sinyi Home Price Index. BPFNN and RBFNN were applied with 11 macroeconomic parameters as input for predicting price variations. The study compared the prediction performance using MAE and RMSE. For the Cathay index, RBFNN outperformed BPFNN, achieving lower MAE and RMSE compared to BPFNN. Conversely, for the Sinyi index, BPFNN exhibited better performance with lower MAE and RMSE compared to RBFNN.

Ziweritin *et al.* [44] discussed three main methodologies: linear regression using square feet as a feature, multivariate regression models using multiple features, and polynomial regression with features raised to different powers. The study evaluated the performance using RMS value. Among the models, the multivariate regression model with features of square feet, bedrooms, and bathrooms harvested the best result.

Haque *et al.* [45] used various regression techniques, comprising multiple linear regression, ridge regression, lasso regression, elastic net regression, adaptive regression, and gradient boosting regression, focusing on Vijayawada, A.P. datasets. It compared algorithms based on scores, MSE, and RMSE. Gradient boosting regression achieved the highest accuracy.

Lee *et al.* [46] investigated the use of machine learning, particularly linear regression, support vector regression, k-nearest neighbors, and random forest, in real estate price prediction. With the lowest prediction error of 0.3713 among all techniques, linear regression produced the best outcomes.

Adetunji *et al.* [47] utilized random forest. The authors used the UCI Machine Learning repository Boston housing dataset with 506 entries and 14 features. This study used MAE, R^2, and RMSE to weigh the model's enactment. The consequences indicated that the random forest predicted house prices with an acceptable difference of ±5 compared to actual values.

Li *et al.* [48] involved the use of the LightGBM framework. The study compared different approaches, including neural networks, and concluded that the LightGBM model, augmented with logarithmic transformation, geo data, and apartment brand information, produced the best MAPE results.

Matey *et al.* [49] used linear regression, lasso regression, and decision tree. Among these, linear regression attained the preeminent result by an accuracy of 83.54%. This study focused on data collection from various sources such as Kaggle, Magicbricks, 99acres, and government websites.

Rizun *et al.* [50] explored various methodologies including hybrid models, fuzzy logic, artificial neural networks, k-nearest neighbors, and machine learning techniques such as decision tree, random forest, naive bayes, logistic regression, and AdaBoost. Among these methods, the paper identified the decision tree using C5.0 and AdaBoost as particularly effective for the dataset, with the decision tree concentrating on rule generation and providing the best accuracy results accuracy, TNR, and TPR, all above 92%.

4.2. Deep Learning

Over the past few decades, advances in sophisticated learning algorithms and efficient pre-processing techniques have led to considerable advancement in machine learning. The evolution of ANNs into increasingly intricate structures, which gave rise to what is now known as deep learning, has been a noteworthy milestone in this trajectory [51]. Deep learning, a specific category within machine learning, comprises multiple layers of ANNs, offering a high-level abstraction for data modelling [52]. In several domains, including natural language processing, gaming, and image processing, it has demonstrated the ability to provide superior prediction outcomes [53]. Deep learning is the most well-liked and generally acknowledged use of artificial intelligence. Well-known tech companies like Google, Microsoft, Facebook, and Amazon have made substantial financial commitments to the study and development of this technology. As of 2016, Google alone claimed to have contributed to over 1,000 deep-learning projects. Currently, a variety of activities require these systems, such as text translation, speech recognition, photo tagging, finding new exoplanets, playing strategic games, evaluating fMRI data, and enabling autonomous driving of automobiles [54].

Zhan *et al.* [55] used backpropagation neural network and convolutional neural network [56] using dataset that included macroeconomic factors and home features from Taiwanese real estate transactions between January 2013 and December 2018. The study contrasted two scenarios: one that incorporated macroeconomic data into its prediction, and the other that solely used home qualities. PCA and normalization were used in the data preprocessing to improve model performance. The assessment measures consisted of r-square, adjusted r-square, MAE, MAPE, RMSLE, and RMSE. With an emphasis on a 5-month historical data span, the results showed that convolutional neural network performed better than backpropagation neural network in predicting house values.

Xu *et al.* [57] anticipated housing prices using deep learning techniques, including convolutional neural network. This study's convolutional neural network model had two convolutional layers, a modified loss function for continuous value regression, and a dropout structure to prevent overfitting. The kind of home, the building area, the location, and macroeconomic variables, for example, GDP, property asset, and consumption level are among the features that have been chosen for prediction. The convolutional neural network model's efficacy was validated by the experimental results, which showed a mean square error of 0.01057 and an accuracy of 98.68%.

4.3. Hybrid Approach

Nouriani *et al.* [58] adopted a hybrid approach, combining deep learning and time series forecasting methods for predicting house prices. The researchers initially employed a deep learning model to predict individual house prices. The model in use consisted of four hidden layers, one output layer with projected house values, and one input layer with property features. The deep learning model was trained by forward and backpropagation with an Adam optimizer. This study forecasted the trajectory of property prices using the prevalent time series projecting method. This model accounted for the temporal aspects of fluctuations in housing values. Due to the researchers' emphasis on the intricate nature of the interaction between influencing elements and housing prices, a dual technique for an all-encompassing prediction approach was adopted.

Chou *et al.* [59] anticipated housing prices using a hybrid strategy that incorporated aspects of ensemble and optimization techniques. The

researchers employed machine learning techniques such as support vector regression, multilayer perception (ANNs), CART, and linear regression as baseline models. Furthermore, it improved prediction performance by employing ensemble techniques, particularly bagging ANNs. A hybrid model known as PSO-Bagging-ANNs combined the use of particle swarm optimization (PSO) for ANN parameter optimization with bagging for aggregation.

Kabir *et al.* [60] used a hybrid approach combining feature engineering and multiple regression algorithms. The methodology involved creative feature engineering, including transformations, category changes, and the introduction of new variables. The study utilized ridge and lasso regressions, along with gradient boosting, as the main regression algorithms. The feature selection was performed using lasso, and a hybrid regression model was proposed, combining lasso and gradient boosting with different weightings. The results indicated that this hybrid method outperformed individual regression algorithms.

Das *et al.* [61] used a hybrid approach, combining machine learning techniques with a specialized embedding model called geospatial network embedding (GSNE). In the hybrid method, the GSNE model which was intended to gather and incorporate neighborhood information based on housing's proximity to different points of interest (POIs) such as areas, schools, and train stations was used in concurrence with machine learning regression models.

Wang *et al.* [62] utilized disparate data from real estate transactions, public facilities, and satellite maps. Based on the input characteristics, the research did a comprehensive preprocessing and separated the data into 13 attribute groups. When metrics for assessment like MAPE were used to compare the numerous machine learning, deep learning models, XGBoost performed better than its machine learning counterparts, and deep learning models that integrated data from public facilities and satellite maps performed better than their machine learning counterparts. Attention mechanisms, particularly the joint self-attention model, were introduced and proven to enhance model flexibility and accuracy. The Joint Self-Attention model performed the best.

Krishnasamy *et al.* [63] used a hybrid approach for live guideline value (GV) prediction in land pricing across Chennai's metropolitan area. The methodology involved utilitarian association rule mining, utilizing 30 customer land-buying pattern attributes obtained through questionnaires. Spatial parameters were measured using GIS, and models like ANN [64] and associative multilayer perceptron [65] were proposed for GV prediction. Potrawa *et al*. [66] explored automated real estate valuation, accentuating visual impact on house market values. It introduced a framework combining ConvNets and crowdsourcing for luxury-level estimation using Zillow data. The proposed method surpassed Zillow's estimates, achieving a 5.8% median error rate compared to Zestimate's 7.9% which showcased the effectiveness of incorporating visual features in real estate valuation.Yousif *et al.* [67] examined many studies that used different approaches, including regression models, hybrid models, and machine learning models, to forecast home prices. They suggested creating a brand-new benchmark dataset called REPD-3000, which included textual and visual data for 3000 dwellings. With the lowest MAE values of 14.38 and 16.60 for REPD-2000 and REPD-3000, respectively, their suggested multi-kernel deep learning regression model beated alternative techniques, such as a multi-kernel SVR.

Table 3 presents a representation of machine learning, deep learning, and hybrid approaches, showcasing different algorithms, metrics and their corresponding accuracies.

5. ANALYTICAL DISCUSSION

Real estate price prediction is a challenging task that needs analysis of various qualitative and quantitative factors. Initially, we have selected 193 attributes but after thorough study, several qualities were unnecessary and hence removed. The remaining 141 attributes were then methodically classified into two categories: qualitative (37) and quantitative attributes (104). By leveraging this mix of attributes, a rich picture of the real estate environment for the sake of understanding key variables that influence property dynamics is provided. Therefore, these attributes compound the focus and are good for anyone interested in understanding how various inputs drive the real estate dynamics.

Table 3. Summary of machine learning, deep learning, and hybrid approaches analysis: algorithms, metrics, and accuracy.

Ref.	ML Algorithms	Metrics									Accuracy
		Qualitative						Quantitative			
		Macro-economic					Others	PP	N	SP	
		UR	SI	CA	IP	GDP					
[20]	RF, DT, NB, LR	✓	✓	✓	✓						-
[25]	MLP, RF, AdaBoost, GB, KNN						✓				-
[29]	LR, RF, GB										LR: 63-68%, RF: 74-78%, GB: 77-78%
[30]	DT, RF, AdaBoost, NB, Logistic Regression						✓				Logistic Regression: 81.5%, DT: 92%, RF: 86.5%, NB: 88%, AdaBoost: 96%
[32]	SGD based SVR, RF, GB										SVM: 82.7%, RF: 90.3%, GB: 89.4% (base model); 90.4% (after hyperpara-meter tuning)
[34]	BPNN, CART, GRNN, LSSVR						✓	✓			LSSVR: MAPE - 0.228%, NMAE - 8.11 x 10^(-4)
[40]	RF, XGBoost, light-GBM, HR, SGR										RMSE - 7671, MRE - 0.22
[42]	SVR, BPNN										SVR (MAE: 1.363, MAPE: 0.01, RMSE: 1.893) BPNN (MAE: 1.788, MAPE: 0.017, RMSE: 2.481)
[43]	BPFNN, RBFNN					✓					-
[44]	LR, MVR, PR										-
[45]	MLR, RR, LR, ER, AR, GBR								✓		MLR (MSE: 391875744, RMSE: 197958) RR (MSE: 391740496, RMSE: 197924) LR (MSE: 391875537, RMSE: 197958) ER (MSE: 489642930, RMSE: 221278) AR (MSE: 32161481079, RMSE: 179336) GBR (MSE: 12037006088, RMSE: 109713).
[46]	LR, SVR, KNN, RF Regression										-
[47]	RF										-
[48]	LightGBM						✓				MAPE values: NN - 10.832%, LightGBM - 9.603%, LightGBM + log - 8.598%, LightGBM + log + Geo Data - 8.412%, LightGBM + log + Geo Data + Brand - 8.349%.
[49]	LR, Lasso Regression, DT						✓				Linear Regression (83.54%), Lasso Regression (82.92%), DT(77.88%)

[50]	DT, AdaBoost					-
[55]	BPNN, CNN	✓				-
[57]	CNN					GM: ~90%, XGBoost: ~96.5%, CNN: ~98.68%
[58]	DL (Forward and Backpropagation with Adam optimizer)				✓	-
[59]	SVR, MLP, CART, Linear Regression + PSO-Bagging-ANNs (Particle Swarm Optimization + Bagging)		✓		✓	-
[60]	Ridge Regression, Lasso Regression, GB (Hybrid Regression)				✓	-
[61]	GSNE + ML Regression Models		✓	✓		-
[62]	ML + DL (XGBoost, Joint Self-Attention Model)					-
[63]	ML + DL (ANN, Associative MLP)		✓			-
[66]	ConvNets + Crowdsourcing		✓			-
[67]	Multi-Kernel DL Regression		✓			-

Figure 4 illustrates the allocation of research papers across various categories, highlighting the predominance of studies in machine learning, deep learning, and combinations of both methodologies. Among the examined studies, the greatest emphasis is on machine learning (61.5%), trailed by hybrid methodologies (30.8%), with a lesser portion devoted to deep learning (7.7%). The distinct colors in the chart emphasize each category and provide a clear and concise overview of the research distribution in the specified domains. Given the widespread adoption of machine learning in the majority of the targeted research papers, a more in-depth examination of the distribution of machine learning categories was conducted. Figure 5 provides a visual breakdown, showcasing the prevalence of specific machine learning categories. Among the various machine learning categories explored, regression emerges as the most prominent, representing 30.8% of the total distribution and is followed by classification and data mining, each constituting 15.4%. Other categories include supervised learning, gradient boosting, and various machine learning approaches. This visually appealing representation aids in understanding the distribution and emphasis of machine learning methodologies in the examined studies.

Table 4 combines and extends the insights gained from Figure 6, illustrating the percentage distribution of methodologies used in predictive modelling across various research papers. The methodologies, namely machine learning, deep learning, and hybrid approaches are presented alongside their respective percentages of representation in the papers. According to analysis machine learning is recognized for its versatility in diverse applications and well-established algorithms, albeit with potential challenges in handling complex patterns and unstructured data. Deep learning is highlighted for its excellence in managing complex data relationships and effectiveness in image and speech recognition. However, it demands substantial computational resources. The hybrid strategy increases predicted accuracy by conjoining the benefits of deep learning and machine learning. Nevertheless, this approach introduces increased complexity in model design and potential challenges in integration. Together, these findings offer a comprehensive understanding

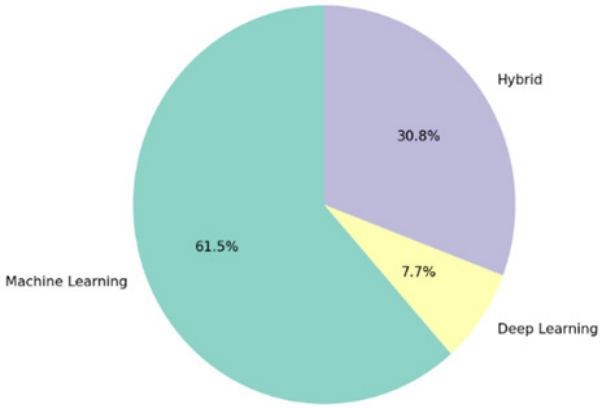

Fig. 4. Distribution of papers by methodology.

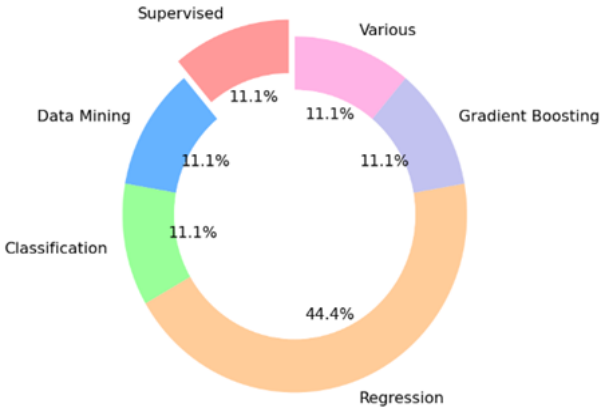

Fig. 5. Distribution of machine learning categories.

of the methods that are frequently employed in real estate predictive modelling, along with the benefits and drawbacks of each.

Figure 6 illustrates the distribution of various machine learning algorithms across the targeted studies. The data reveals a diverse machine learning techniques employed. Linear regression and decision trees stand out as the most frequently utilized algorithms, each appearing in 6 studies. Random forest, AdaBoost, and gradient boosting also demonstrate notable usage with counts of 6, 3, and 4, respectively. This visualization provides insights into the popularity of machine learning algorithms, helping in understanding the methodological preferences within the examined research studies. The color-coded bars enhance the visual appeal, making it easier to discern the distribution and relative prevalence of each algorithm. Figure 7 illustrates the distribution of the number of variables reported across the studies reviewed in this paper. The x-axis represents the number of variables used in each study, while the y-axis lists the corresponding references. For example, reference [67] utilized 15 variables, [63] used 8 variables, [62] utilized 13 variables, and [61] used 43 variables. Some references, such as [20, 45, 49, 50, and 66] are marked as "unknown", indicating that the number of variables was not explicitly mentioned in these studies. This distribution highlights the inconsistency in the number of variables considered across the targeted research studies.

6. CONCLUSIONS

In summary, this review represents the applications of machine learning, deep learning, and hybrid models making substantial advancements with enhanced accuracy and deeper insights for real estate price prediction. The comprehensive literature review categorized existing approaches into three primary categories: machine learning, deep learning and hybrid approaches. Machine

Table 4. Methodologies in real estate predictive modelling.

Tec.	Papers %age	Merits	Demerits
ML	61.5%	• Versatility in diverse applications • Well-established algorithms and models	• May struggle with complex patterns • Limited in handling unstructured data
DL	7.7%	• Excellent for complex data relationships • Effective in image and speech recognition	• Requires substantial computational resources
Hybrid Approach	30.8%	• Combines the strengths of ML and DL • Improved predictive accuracy	• Increased complexity in model design • Potential challenges in integration

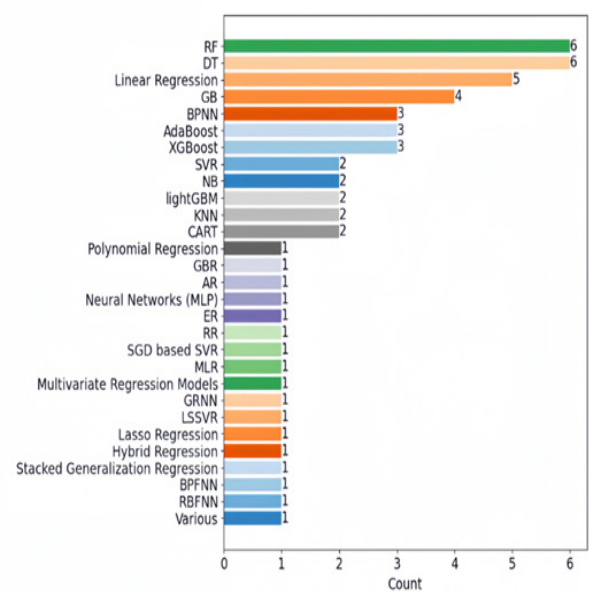

Fig. 6. Distribution of ML algorithms.

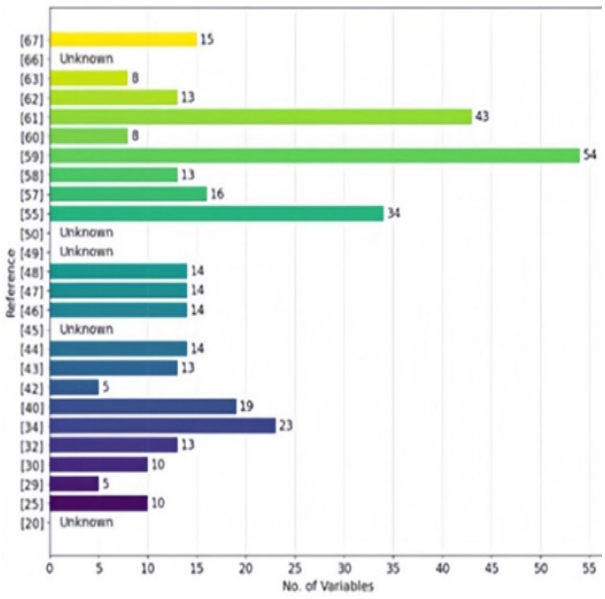

Fig. 7. Distribution of number of variables.

learning is used in 61.5% studies due to its flexibility and well-established methodologies, but it has difficulties in recognizing diverse patterns and working with unstructured information. Deep learning is applied in 7.7% studies, highlighting its potential of managing complex data relations, but has heavy computational penalties. As highlighted in one-third of the studies, hybrid solutions were used as being highly effective since they integrated both ML and DL to improve the prediction capability. According to the findings of this review, machine learning is the most prevalent technique in real estate price prediction, although deep learning and hybrid techniques are gaining increasing attention. By comparing and contrasting these three approaches, this analysis provides insight into the strengths and limitations of each, which is beneficial for both researchers and practitioners. However, there were some research gaps and limitations noted in the current study. Specifically, there is challenge of computational resource requirements in deep learning and the increased demand for improving the data processing of unstructured data in the machine learning domain. Also, the number of studies concerning deep learning is also significantly smaller, which points to the fact that this field has been researched less despite the ability to use this approach for the real estate price prediction. The review suggests that future research should be directed at filling these gaps by identifying better performing algorithms, and examining the possibility and scope of using multiple forms of data, under different market conditions. In doing so, academicians, researchers and professionals in the real estate will be better placed in identifying patterns that will aid in obtainment of better accuracy in their valuation and prediction models.

7. CONFLICT OF INTEREST

The authors declare no conflict of interest.

8. REFERENCES

1. R. Cellmer, A. Cichulska, and M. Bełej. Spatial analysis of housing prices and market activity with the geographically weighted regression. *ISPRS International Journal of Geo-Information* 9(6): 380 (2020).
2. F. Roy. Housing Financing at the Crossroads: Access and Affordability in an Aging Society. In: Understanding China's Real Estate Markets. Management for Professionals. B. Wang, and T. Just (eds). *Springer, Cham*, pp. 159-178 (2021).
3. F. Ullah, and S.M.E. Sepasgozar. Key factors influencing purchase or rent decisions in smart real estate investments: A system dynamics approach using online forum thread data. *Sustainability* 12(11): 4382 (2020).
4. P. Gepner, N.H. Tien, M.T.H. Dao, and D.T. Minh. Analysis of business strategy of leading Vietnamese real estate developers using SWOT matrix. *International Journal of Multidisciplinary Research*

and Growth Evaluation 3(1): 181-187 (2022).
5. Y. Casali, N.A. Yonca, T. Comes, and Y. Casali. Machine learning for spatial analyses in urban areas: a scoping. *Sustainable Cities and Society* 85: 104050 (2022).
6. A. Wandhe, L. Sehgal, H. Sumra, A. Choudhary, and M. Dhone. Real Estate Prediction System Using ML. *11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing, IEEE Publishers* (28-29 April 2023) pp. 1-4 (2023).
7. P.F. Pai, and W.C. Wang. Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences* 10(17): 5832 (2020).
8. F. Ullah, and F. Al-Turjman. A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. *Neural Computing and Applications* 35(7): 5033-5054 (2023).
9. J.I. Pérez-Rave, R.F. Guerrero, A.S. Vallina, and F.G. Echavarría. A measurement model of dynamic capabilities of the continuous improvement project and its role in the renewal of the company's products/ services. *Operations Management Research* 16(1): 126-140 (2023).
10. N.H. Zulkifley, S.A. Rahman, N.H. Ubaidullah, and I. Ibrahim. House Price Prediction using a Machine Learning Model: A Survey of Literature. *International Journal of Modern Education & Computer Science* 12(6): 46-54 (2020).
11. S.S. Yalgudkar, and N.V. Dharwadkar. A Literature Survey on Housing Price Prediction. *Journal of Computer Science & Computational Mathematics* 12(3): 41-45 (2022).
12. S.C.K. Tekouabou, Ş.C. Gherghina, E.D. Kameni, Y. Filali, and K.I. Gartoumi. AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey. *Archives of Computational Methods in Engineering* 31(2): 1079–1095 (2023).
13. M. Geerts, S.V. Broucke, and J. De Weerdt. A Survey of Methods and Input Data Types for House Price Prediction. *ISPRS International Journal of Geo-Information* 12(5): 200 (2023).
14. A. Gdakowicz, and E. Putek-Szeląg. The Use of Statistical Methods for Determining Attribute Weights and the Influence of Attributes on Property Value. *Real Estate Management and Valuation* 28(4): 33-47 (2020).
15. Q.D. Buchlak, N. Esmaili, J.C. Leveque, F. Farrokhi, C. Bennett, M. Piccardi, and R.K. Sethi. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurgical Review* 43(5): 1235-1253 (2020).
16. Y. Liu, F.R. Yu, X. Li, H. Ji, and V.C.M. Leung. Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials* 22(2): 1392-1431 (2020).
17. E. Hüllermeier, and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning* 110(3): 457-506 (2021).
18. J.S. Chou, D.B. Fleshman, and D.N. Truong. Comparison of machine learning models to provide preliminary forecasts of real estate prices. *Journal of Housing and the Built Environment* 37(4): 2079-2114 (2022).
19. I.H. Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science* 2(3): 160 (2021).
20. V. Gampala, N.Y. Sai, and T.N.S. Bhavya. Real-estate price prediction system using machine learning. *International Conference on Applied Artificial Intelligence and Computing,* pp. 533-538 (2022).
21. A. Ghosh, and A. Senthilrajan. Comparison of machine learning techniques for spam detection. *Multimedia Tools and Applications* 82(19): 29227-29254 (2023).
22. N. Borah, U. Baruah, T.R. Mahesh, V.V. Kumar, D.R. Dorai, and J.R. Annad. Efficient Assamese Word Recognition for Societal Empowerment: A Comparative Feature-Based Analysis. *IEEE Access* 11: 82302-82326 (2023).
23. T. Biswas, S.C. Pal, I. Chowdhuri, D. Ruidas, A. Saha, A.R.M.T. Islam, and M. Shit. Effects of elevated arsenic and nitrate concentrations on groundwater resources in deltaic region of Sundarban Ramsar site, Indo-Bangladesh region. *Marine Pollution Bulletin* 188: 114618 (2023).
24. D. Maulud, and A.M. Abdulazeez. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends* 1(2): 140-147 (2020).
25. D. Tchuente, and S. Nyawa. Real estate price estimation in French cities using geocoding and machine learning. *Annals of Operations Research* 308(1): 571-608 (2022).
26. S. Demir, and E.K. Sahin. An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. *Neural Computing and Applications* 35(4): 3173-

3190 (2023).
27. C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review* 54: 1937-1967 (2021).
28. S. Zhang, and J. Li. Knn classification with one-step computation. *IEEE Transactions on Knowledge and Data Engineering* 35(3): 2711-2723 (2021).
29. O.G. Uzut, and S. Buyrukoglu. Prediction of real estate prices with data mining algorithms. *Euroasia Journal of Mathematics, Engineering, Natural and Medical Sciences* 8(9): 77-84 (2020).
30. B.A. Kurdi, H. Raza, S. Muneer, M.B. Alvi, N. Abid, and M.T. Alshurideh. Estate Price Predictor for Multan City Townships Using Marching Learning. *International Conference on Cyber Resilience, Dubai, UAE* (6-7 October 2022) pp. 1-5 (2022).
31. L.M. Halman, and M.J. Alenazi. MCAD: A Machine learning based cyberattacks detector in Software-Defined Networking (SDN) for healthcare systems. *IEEE Access* 11: 37052-37067 (2023).
32. W.K.O. Ho, B.S. Tang, and S.W. Wong. Predicting property prices with machine learning algorithms. *Journal of Property Research* 38(1): 48-70 (2021).
33. Q. Klopfenstein, and S. Vaiter. Linear support vector regression with linear constraints. *Machine Learning* 110(7): 1939-1974 (2021).
34. P.F. Pai, and W.C. Wang. Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences* 10(17): 5832 (2020).
35. F. Miao, X. Xie, Y. Wu, and F. Zhao. Data Mining and deep learning for predicting the displacement of "Step-like" landslides. *Sensors* 22(2): 481 (2022).
36. N. Jalal, A. Mehmood, G.S. Choi, and I. Ashraf. A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University-Computer and Information Sciences* 34(6): 2733-2742 (2022).
37. C.S. Tu, W.C. Tsai, C.M. Hong, and W.M. Lin. Short-term solar power forecasting via general regression neural network with grey wolf optimization. *Energies* 15(18): 6624 (2022).
38. Z. Ren, F. Fang, N. Yan, and Y. Wu. State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology* 9(2): 661-691 (2022).
39. K. Drachal, and M. Pawłowski. A review of the applications of genetic algorithms to forecasting prices of commodities. *Economies* 9(1): 6 (2021).
40. Q. Truong, M. Nguyen, H. Dang, and B. Mei. Housing price prediction via improved machine learning techniques. *Procedia Computer Science* 174: 433-442 (2020).
41. P. Yadav, S.C. Sharma, R. Mahadeva, and S.P. Patole. Exploring Hyper-Parameters and Feature Selection for Predicting Non-Communicable Chronic Disease Using Stacking Classifier. *IEEE Access* 11: 80030-80055 (2023).
42. A. Dalal, M. Bagherimehrab, and B.C. Sanders. Quantum-assisted support vector regression for detecting facial landmarks. *Bulletin of the American Physical Society* 67(3): 1-20 (2022).
43. K. Sanjar, O. Bekhzod, J. Kim, A. Paul, and J. Kim. Missing data imputation for geolocation-based price prediction using KNN–MCF method. *ISPRS International Journal of Geo-Information* 9(4): 227 (2020).
44. S. Ziweritin, C.C. Ukegbu, T.A. Oyeniran, and I.O. Ulu. A Recommendation Engine to Estimate Housing Values in Real Estate Property Market. *Inernational Journal of Scientific Research in Computer Science and Engineering* 9(1): 1-7 (2021).
45. M.A. Haque, N. Sarker, N.S.S. Singh, M.A. Rahman, M.N. Hasan, M. Islam, M.A. Zakariya, L.C. Paul, A.H. Sharker, G.E.M. Abro, M. Hannan, and R. Pk. Dual band antenna design and prediction of resonance frequency using machine learning approaches. *Applied Sciences* 12(20): 10505 (2022).
46. S.H. Lee, J.H. Kim, and J.H. Huh. Land Price Forecasting Research by Macro and Micro Factors and Real Estate Market Utilization Plan Research by Landscape Factors: Big Data Analysis Approach. *Symmetry* 13(4): 616 (2021).
47. A.B. Adetunji, O.N. Akande, F.A. Ajala, O. Oyewo, Y.F. Akande, and G. Oluwadara. House price prediction using random forest machine learning technique. *Procedia Computer Science* 199: 806-813 (2022).
48. T. Li, T. Akiyama, and L. Wei. Constructing a highly accurate price prediction model in real estate investment using LightGBM. *IEEE 4th International Conference on Multimedia Information Processing and Retrieval* 273-276 (2021).
49. V. Matey, N. Chauhan, A. Mahale, V. Bhistannavar, and A. Shitole. Real Estate Price Prediction using Supervised Learning. *IEEE Pune Section International Conference, PUNE, India* (15-17 December 2022) pp. 11-15 (2022).
50. N. Rizun, and A. Baj-Rogowska. Can web search queries predict prices change on the real estate market? *IEEE Access* 9: 70095-70117 (2021).
51. C. Janiesch, P. Zschech, and K. Heinrich. Machine

learning and deep learning. *Electronic Markets* 31(3): 685-695 (2021).
52. F. Sigrist and N. Leuenberger. Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. *European Journal of Operational Research* 305(3): 1390-1406 (2023).
53. F.J. Boge. Two dimensions of opacity and the deep learning predicament. *Minds and Machines* 32(1): 43-75 (2022).
54. A.M. Ozbayoglu, M.U. Gudelek, and O.B. Sezer. Deep learning for financial applications: A survey. *Applied Soft Computing* 93: 106384 (2020).
55. C. Zhan, Z. Wu, Y. Liu, Z. Xie, and W. Chen. Housing prices prediction with deep learning: an application for the real estate market in Taiwan. *IEEE 18th International Conference on Industrial Informatics* 1: 719-724 (2020).
56. Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems* 33(12): 6999-7019 (2021).
57. X. Xu, and Y. Zhang. Residential housing price index forecasting via neural networks. *Neural Computing and Applications* 34(17): 14763-14776 (2022).
58. A. Nouriani, and L. Lemke. Vision-based housing price estimation using interior, exterior & satellite images. *Intelligent Systems with Applications* 14: 200081 (2022).
59. J.S. Chou, D.B. Fleshman, and D.N. Truong. Comparison of machine learning models to provide preliminary forecasts of real estate prices. *Journal of Housing and the Built Environment* 37(4): 2079-2114 (2022).
60. H.M.D. Kabir, M. Abdar, A. Khosravi, D. Nahavandi, S.K. Mondal, S. Khanam, S. Mohamed, D. Srinivasan, S. Nahavandi, and P.N. Suganthan. Synthetic Datasets for Numeric Uncertainty Quantification: Proposing Datasets for Future Researchers. *IEEE Systems, Man, and Cybernetics Magazine* 9(2): 39-48 (2023).
61. S.S.S. Das, M.E. Ali, Y.F. Li, Y.B. Kang, and T. Sellis. Boosting house price predictions using geo-spatial network embedding. *Data Mining and Knowledge Discovery* 35: 2221-2250 (2021).
62. P.Y. Wang, C.T. Chen, J.W. Su, T.Y. Wang, and S.H. Huang. Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access* 9: 55244-55259 (2021).
63. S. Krishnasamy, M. Rajiah, K.K. SenthilKumar, and S.N. Rajendiran. Association rule-based multilevel regression pricing and artificial neural networks based land selling price prediction based on market value. *Concurrency and Computation: Practice and Experience* 35(5): e7550 (2023).
64. M. Alkasassbeh, and S.A.H Baddar. Intrusion detection systems: A state-of-the-art taxonomy and survey. *Arabian Journal for Science and Engineering* 48(8): 10021-10064 (2023).
65. A. Ahrari, S. Elsayed, R. Sarker, D. Essam, and C.A.C. Coello. Adaptive multilevel prediction method for dynamic multimodal optimization. *IEEE Transactions on Evolutionary Computation* 25(3): 463-477 (2021).
66. T. Potrawa, and A. Tetereva. How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research* 144: 50-65 (2022).
67. A. Yousif, S. Baraheem, S.S. Vaddi, V.S. Patel, J. Shen, and T.V. Nguyen. Real estate pricing prediction via textual and visual features. *Machine Vision and Applications* 34(6): 126 (2023).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 145-152 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)867

Pakistan Academy of Sciences

Research Article

Determination of Critical Slip Surface in Loose Rock Slope Stability Analysis

Zulkifl Ahmed[1], Sumra Yousuf[2*], Mahwish Zahra[3], Anum Aleha[3], Abid Latif[4], Tahir Sultan[4], Tanveer Ahmad Khan[2], and Muhammad Yousaf Raza Taseer[5]

[1]School of Resource and Civil Engineering, Northeastern University, Shenyang, China
[2]Department of Building and Architectural Engineering, Faculty of Engineering and Technology, Bahauddin Zakariya University, 60000 Multan, Pakistan
[3]Department of Architecture Design, National Fertilizer Corporation Institute of Engineering and Technology (NFC-IET), 60000 Multan, Pakistan
[4]Department of Civil Engineering, Faculty of Engineering and Technology, Bahauddin Zakariya University, 60000 Multan, Pakistan
[5]Department of Structure and Materials, Faculty of Civil Engineering, Universiti Teknologi Malaysia 81310, UTM, Johar, Malaysia

Abstract: Determination of representative Critical Slip Surface (CSS) for loose rock slope is one of the most important topics for slope reinforcement design. In this paper, a SLOPE/W software is used to analyze the failure characteristics of CSS considering the spatial variability effect of strength parameters. Initially, the Morgenstern-Price limiting equilibrium method was selected within the framework of SLOPE/W software to examine the failure mechanism of CSS and the corresponding Factor of Safety (FS) for a loose rock slope comprised of two different materials. Also, the variability effect of shear strength parameters (cohesion, friction angle) on minimum FS, the maximum depth (D), sliding arc length (L), distribution range of slip surfaces and slip surface entry point distance (D_e) were investigated through software. The results showed that all slip surfaces are mostly parallel and the local failure can happen at the top of the slope. Statistically, local failure has entry and exit points situated at the crest and near the toe of the slope, respectively. Shear strength parameters have a remarkable effect on FS, D, L and D_e of critical slip surface. The distribution range of CSS decreased with an increasing the amount of cohesion and friction angle. These findings can help to locate the actual position of CSS and slip surface entry point distance in case of loose rock slope.

Keywords: Morgenstern-Price Method, Failure Mechanism, Critical Slip Surface, FS, Entry Point Distance, Loose Rock.

1. INTRODUCTION

Engineers mostly simulate rock slope stability to precisely classify the most distinct critical slip surface. Engineers also use several analytical methods to evaluate the stability factor of natural or artificial slopes [1]. Limit equilibrium techniques are leading among these methods, in which the empirically based established site-stability charts may be applicable for slope stability calculation in certain conditions. In general, nevertheless, laboratory tests, field assessments and more precise numerical solutions are normally suggested. The outcomes of the above studies are usually reported as the probability of failure and Factor of Safety (FS) for the geotechnical structures [2, 3]. The shape of the sliding surface has been also widely used previously to determine the stability of various types of rock slopes. Such as circular arc shape was used to evaluate the stability of slopes [4]. The spiral slip surface was estimated for rock slope stability analysis [5], and the circular slip surface was determined for heterogeneous rock slope stability factor computation [6]. Other scholars

Received: March 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Sumra Yousuf <sumra.yousafrm@gmail.com>

have studied the stability of natural and or open pit slopes by ellipsoidal, revolution, cylindrical and critical circular slip surfaces [7-9].

In static analysis, the Critical Slip Surface (CSS) is the surface that produces the minimum FS during slope stability and the smallest value of FS can calculated by studying circular arc with cohesion (c) and friction (ϕ) parameters [10]. When a circular slip surface is studied, repeated trials can be done numerically. The greater difficulties involved in automating the search for non-circular surfaces. Most available computer programs that can be used to analyze and locate the most critical slip surface with repeated trials are more reliable for CSS determination. In most cases, only a circular slip surface was analyzed for general slope types when a computer program was used to perform the search automatically. In some cases, the most critical slip surface may not be approximated accurately by a circular arc [11]. In the cut slope case, the appropriate assumption is that the slip surface is repeated circular will result in an unknown. The method of judgment of the critical slip surface is directly connected to the method of determining the minimum Factor of Safety (FS). Some scholars have used limit equilibrium techniques to evaluate FS and critical slip surfaces statically [12-16], or numerically [17, 18]. To determine the minimum FS for a sliding surface, a general limiting equilibrium technique may allow a precise and accurate evaluation method during large-scale stability investigations. Therefore, commonly used approaches are limiting equilibrium techniques and cannot be used to locate CSS with general constrain under composite conditions. On the other hand, optimization techniques are considered an effective tool to estimate FS for single-slip surface [19]. On the other hand, dynamic programming is used to allocate the non-circular slip surface [18]. Monte-Carlo techniques can evaluate the FS of the critical slip surface of a slope [20]. Other researchers proposed some conventional solutions for rock slope stability analyses [21, 22].

In the current research, a simple geometrical method and GeoStudio software code are used to calculate the CSS and corresponding minimum FS. The failure mechanics for CSS are analyzed by varying the values of cohesion and friction. In this research, the CSS allocation provides concrete guidelines for choosing a typical sliding surface in risk assessments and system reliability analyses of heterogeneous rock slopes beside the road.

2. METHODOLOGY

In geotechnical engineering investigating the stability of slopes is also one of the oldest kinds of numerical study. In the mid of 20th century, [12-16] presented the knowledge of dividing the probable sliding mass into several number of slices. In the 1960 the dawn of computer technology made it conceivable to voluntarily switch the iterative measures essential in the scheme which controlled statistically more laborious equations for example those established previously [12-16]. In this research, SLOPE/W was selected which is a product of GeoStudio. One of the more commanding topographies of this joined method is that it provides access to sorts of investigations of a much more complex and broader field of difficulties, together with the use of stresses and calculated pore-water pressures in a stability study. It is possible in limit equilibrium (LE) computer programs such as SLOPE/W to deal with extremely irregular pore-water pressure conditions, complex stratigraphy, a variety of nonlinear and linear shear strength models, material heterogeneity, concentrated loads, structural reinforcement and practically any type of sliding surface shape.

2.1. General Limit Equilibrium Formulation

Figure 1 shows a representative sliding surface (AB), with arc length L, slices discretization scheme and the possible forces on a slice of a slope. Normal forces (E) and shear forces (X) are acting on the slice base and the slice sides respectively (Figure 1). Limit equilibrium (LE) calculation consists of two FS equations which permit the boundary of inter-slice normal-shear force conditions. One equation gives the FS regarding horizontal force equilibrium (F_f) while the other equation gives the

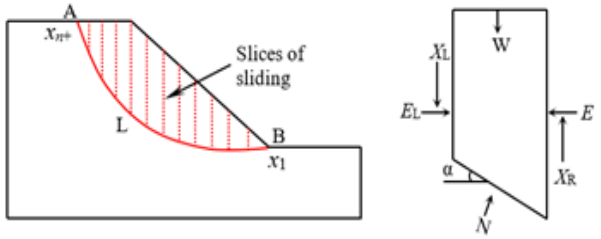

Fig. 1. Slice discretization and slice forces in a sliding mass.

FS about moment equilibrium (F_m). Morgenstern and Price [13] proposed an equation to handle inter-slice shear forces as:

$$X = E\lambda f(x) \quad (1)$$

Where, X = shear force, E = normal force, λ = dimensionless function and $f(x)$ = a function. In the current analysis a distinctive half-sine function was chosen and the *FS* equation relating to moment equilibrium is:

$$Fm = \frac{\sum(\acute{c}\beta R + (N - u\beta)\, R\tan\emptyset')}{\sum wx - \sum Nf \pm \sum Dd} \quad (2)$$

The FS equation regarding horizontal force equilibrium is:

$$Ff = \frac{\sum(\acute{c}\beta\cos\alpha + (N - u\beta)\tan\emptyset'\cos\alpha)}{\sum N\sin\alpha - \sum D\cos\omega} \quad (3)$$

And base normal is defined as:

$$N = \frac{W + (XR - XL) - \frac{(\acute{c}\beta\sin\alpha + u\beta\sin\alpha\tan\emptyset')}{F}}{\cos\alpha + \frac{\sin\alpha\tan\emptyset'}{F}} \quad (4)$$

In above equations ϕ' = effective internal angle of friction, c' = effective cohesion, N = base normal force, u = water pressure, W = slice weight, α = slice base inclination, D = point load and β, R, x, f, d, ω are geometric parameters.

Equation (4) is acquired with the summation of vertical forces. $F = Fm$ when base normal force N is substituted in Equation (2) and $F = F_f$ when N is substituted in Equation (3). The slice base normal force (N) is reliant on the inter-slice shear forces X_R and X_L on both sides of a slice. Limit equilibrium formulation calculates F_f and F_m as a choice of dimensionless function (λ) values. A plot is shown in Figure 2, computed with these values, which illustrates how F_f and F_m vary with lambda (λ). In the end, for the whole slope, there is only one *FS*. F_f and F_m were similar when both force and moment equilibrium were satisfied and also *FS* was the same in each slice.

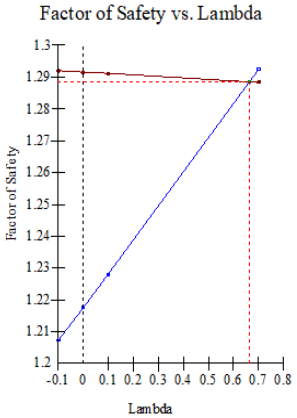

Fig. 2. Relationship between FS and λ.

3. NUMERICAL SIMULATION

3.1. Illustrative Example

A loose rock slope is located beside National Highway N70 in Pakistan considered as a case study. The slope material is composed of boulders as shown in Figure 3. It consists of 20-1000 mm diameter rounded rock pieces. This formation is valley floor sediment and is supposed to consist of two sandy layers. The total station was used to measure the boundaries and coordinates of the slope. Water pressure was measured with a piezometer. The coordinates of the water line are presented in Table 1.

3.2. Model Parameters

The width and height of the slope are 70 m and 35 m respectively as shown in Figure 4(a). A slope angle of 45° was recorded. The slope is composed of two different layers. The shear strength properties of slope material were estimated via direct shear test apparatus. The shear strength parameters were

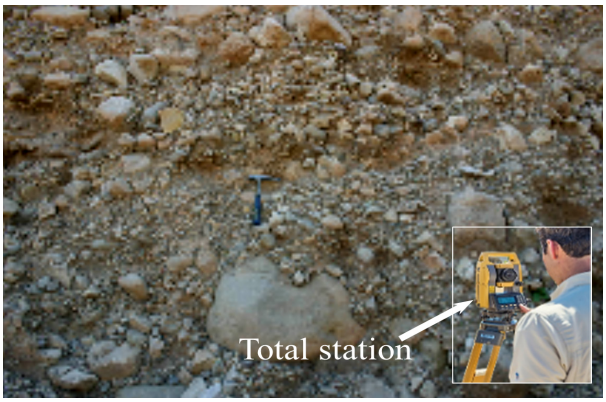

Fig. 3. Topography of loose rock slope and total station.

Table 1. Coordinates of piezo-metric line.

Coordinates	X(m)	Y(m)
1	0	22
2	30	17
3	52	13
4	70	13

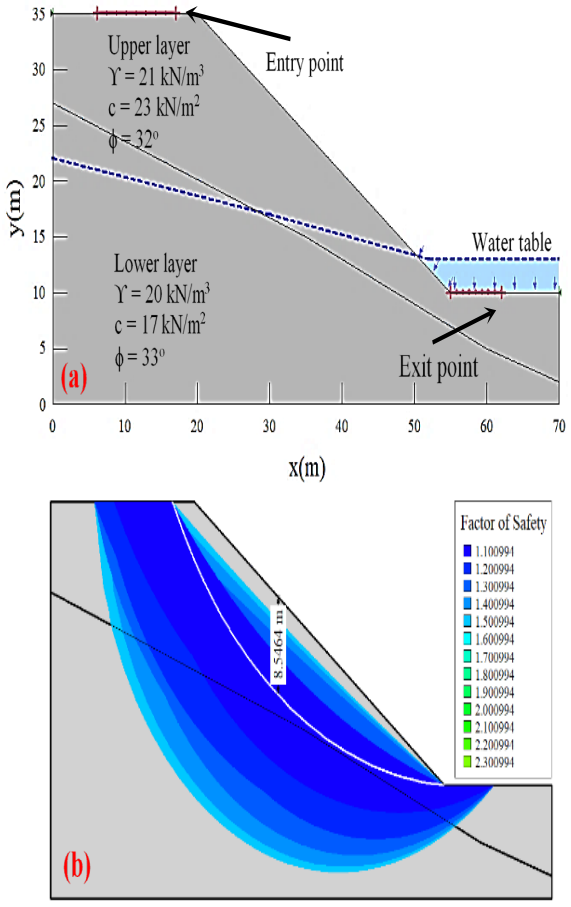

Fig. 4. (a) Slope model and (b) CSS in SLOPE/W analysis.

assigned to the corresponding region. The half-sin function and Mohr-Coulomb failure criterion are selected in a computer program (SLOPE/W) for further analysis. The trial slip surface entry and exit point were located according to Figure 4(a). The stability factor and the location of the CSS were then simulated by SLOPE/W software. Another effective way of graphically viewing a summary of the CSS is a safety map. The valid trial critical failure surface is shown with the factor of safety. This type of presentation clearly shows the location and distribution of the trial failure surface. Figure 4b shows that the critical failure surface initiates from the crest and cuts the slope near the toe.

3.3. Failure Characteristics of CSS

In this sub-section, the failure mechanism and the location of the critical slip surface are analyzed at various values of cohesion and friction angle. Failure mechanism of a critical failure surface can be acquired from a deterministic analysis and generating the different values of cohesion, and internal friction angle within the network of SLOPE/W software. Slope material is very loose and composed of two layers (upper layer and lower layer) of different materials. Three different cases are considered based on strength properties, namely, upper layer c = 16 (kN/m^2), ϕ = 21^o and γ = 17 (kN/m^3); c = 21 (kN/m^2), ϕ = 25^o and γ = 18 (kN/m^3) and c = 25 (kN/m^2), ϕ = 30^o and γ = 19 (kN/m^3) and lower layer c = 10 (kN/m^2), ϕ = 25^o and γ = 20 (kN/m^3); c = 15 (kN/m^2), ϕ = 30^o and γ = 21 (kN/m^3) and c = 20 (kN/m^2), ϕ = 35^o and γ = 21 (kN/m^3). Failure surface location and its failure mechanism for the three sets of input parameters (c, ϕ, γ) are plotted in Figure 5. The typical understandings have various aspects of CSS location by varying strength parameters. All CSSs are located between the crest and the toe of the slope and the failure of the slope is a local failure at an average value of strength parameters. CSS is plotted with the entry and exit method. The representative insights have different appearances of CSS failure and location and are categorized into three parts follows;

a. In the first case, the CSS was shallow and cut the toe of the slope as shown in Figure 5(a). The weak zone is comparatively small and is situated around the surface of the slope. The location of the CSS, in this case, is well dependable on the position of the weak zone and the overall failure style is a local failure. Because the factor of safety (FS) for the CSS passing through the weak zone is lesser than that for other slip surfaces representing an overall slope failure. When the failure of a slope surface happens in such a manner that the sliding surface cuts the slope surface or passes from the slope toe, it is called slope failure as shown in Figure 5(a, b). In this case, the slip surface entry and exit points are located near the top and on the toe of the slope respectively.
b. When the strength parameters were kept relatively large the CSS was deep. In the second case, the weak zones are relatively large and situated at the uppermost portion of the slope. The associated failure mechanism for a typical CSS is a narrow failure as shown in Figure 5(b). The judgment is that the FS is smaller for large sliding surfaces and running through the weak zone representing an overall failure.
c. When the large value of strength parameters is selected the CSS is located at a small distance from the toe of the slope and the failure mode is an overall failure. Because the FS for a CSS

is bigger as compared two other all valid slip surfaces in the weak areas of slope. In this case, the failure of the slope is a deep-seated failure and the location of CSS is not consistent with the location of the weak zone. The reason is that the load of a slope is a component of the gravity of the slope, which is proportional to the volume of the slide mass. The shear resistance of the slope consists of two parts, cohesion part and friction angle part. As the size of a sliding mass decreases, the decreasing rate of the associated CSS is smaller than that of the sliding volume. In other words, if the shear strength parameter remains constant and the size of the sliding mass decreases, the load of a slope decreases faster than the resistance of the slope. For a comparatively large repeated CSS as shown in Figure 5(c), the cohesion in the weak zone is small enough to generate FS smaller than the FS for an overall CSS.

It can be seen from the above analysis, that the failure mode for a heterogeneous slope with a relatively small value of strength parameters is an overall failure in most cases. Although local failures may happen in certain cases, the associated CSS always have entry points (the uppermost points of slip surfaces) located at the top of the slope and exit points (the lowermost points of slip surfaces) located near the toe of the slope.

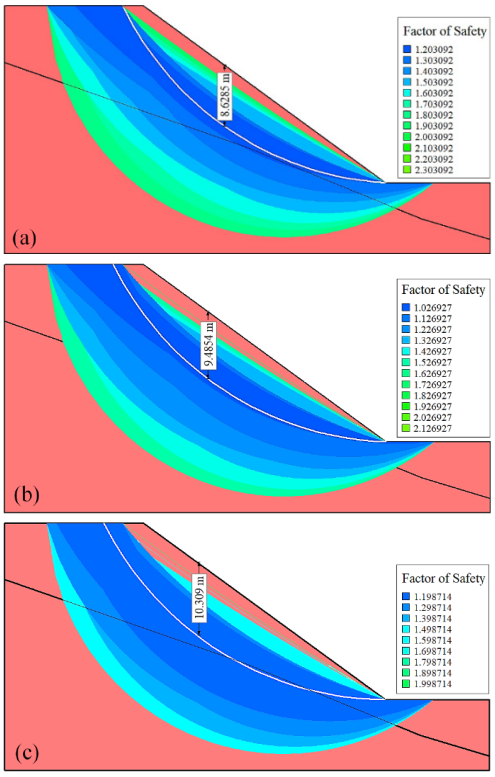

Fig. 5. Failure characteristics of critical slip surfaces, (a) small value of strength parameters, (b) comparatively large value of strength parameters, and (c) large value of strength parameters.

4. DISCUSSION

In this section, the distribution range of critical slip surfaces and corresponding FS are discussed by varying the values of c, and ϕ. The critical slip surface with an exit or an entry point on the face of the slope were assign as horizontal coordinates of the CSS uppermost part and lowermost part [17]. The distribution of CSS shows similar phenomena, and the results verified by Zhang *et al*. [23] study.

The influence of cohesion (c) and angle of internal friction (ϕ) on FS, depth (D), arc length (L) and its distribution are discussed in this subsection. First, the influence of cohesion is analyzed for the same slope model configuration. For this purpose, the unit weight (γ) and angle of internal friction (ϕ) are fixed to 20 kN/m^3 and 30^o, respectively. The value of cohesion (c) ranging from 10 kN/m^2 to 25 kN/m^2 is assigned in SLOPE/W. Four different slope models are tested by varying the values of cohesion as 10, 15, 20 and 25 (kN/m^2), respectively. The summary of computed safety factors and the depth (D) of CSS are graphically portrayed in Figure 6, which shows that all valid CSSs fall inside the range of trial slips.

The results exhibit that the FS value increases as increasing the value of the internal friction angle. Cohesion and friction have a significant influence on the FS of cut slope especially for greater values. The variability effect of c on FS, maximum depth (D) of CSS and distribution range of all valid CSSs were analyzed with different combinations of c, for both layers, which are plotted in Figure 6. For this purpose, three different cases based on c parameters are generated in SLOPE/W. The distribution range of SSs is consistent with the safety map.

As clearly shown in Figure 6(a–c), when the cohesion increased, there were big differences among the distributions of valid CSSs. In addition, the FS and maximum D increases linearly from 8. 55 m to 12.05 m. The reason is that the cohesion is a strength parameter increasing this value increase

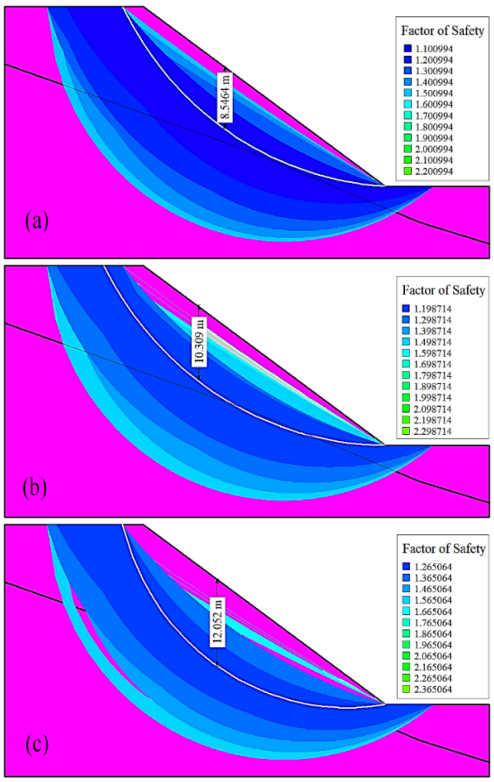

Fig. 6. Display of multiple trial slip surfaces, (a) small value of strength parameters, (b) comparatively large value of strength parameters, and (c) large value of strength parameters.

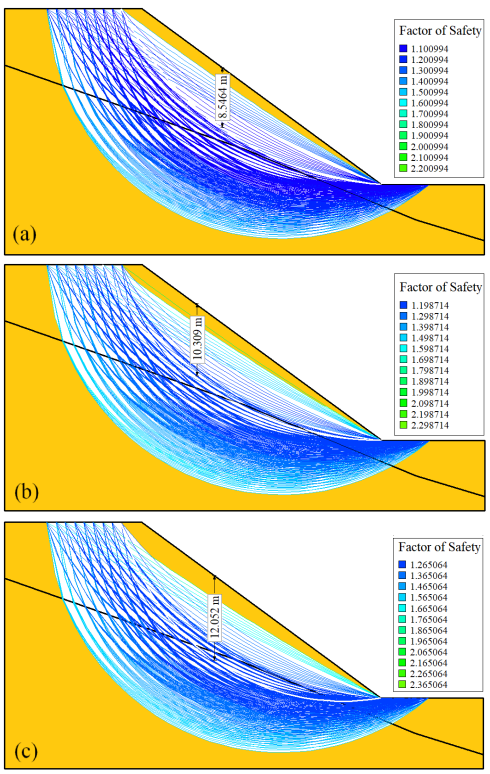

Fig. 7. Display of CSSs. (a) Small value of strength parameters, (b) comparatively large value of strength parameters and (c) large value of strength parameters.

in FS and depth of slip surface. In other words, very large or local CSSs are more likely to occur as the cohesion is large. Because cohesion is a strength parameter.

This paragraph examines the changeability influence of the angle of internal friction (ϕ) on the FS, maximum D and distribution range of CSSs. The same procedure for cohesion is adopted to investigate the effect of friction. By varying the value of friction angle, for both layers, three cases are investigated via SLOPE/W software. The variation of friction angle was set to be 20°, 25°, 30° and 25°, 30° and 35° for the upper and lower layers. The reason is that the value of ϕ varies from 20° and 35° for this particular case. Unit weight and cohesion are kept constant and are set to be 20 kN/m^3 and 15 kN/m^2, respectively. The effect of angle of internal friction on the distribution of CSS, D and FS is plotted in Figure 7. As the value of ϕ increased, the FS and depth and the distribution range of CSS also changed significantly. Local failures happened in the first two cases as ϕ = 20° and 25°, and 25° and 30° for both layers respectively (Figure 7 (a, b)). When the value of ϕ was 30° and 35° chosen, no local failure occurs as shown in Figure 7(c). The reason is that ϕ is a strength parameter increasing this increase in resistance force, which produced greater FS [1]. In other words, when the value of ϕ was large, All CSSs were extremely likely to enter the slope from the crest (entry point) and cut the slope near the toe. By contrast, when the value of ϕ was comparatively small, local failures have entry and exit points located on the crest and around the toe of the slope.

Slip surface entry point distance is an important parameter to define the reinforcement area of the slope. This is the distance from the crest to the failure point of slope as shown in Figure 8. The results of GeoStudio 2D software were used to determine slip surface entry point distance (D_e) and the length of failure arc (L) in the case of loose rock slope. The relationship between D_e and L and FS is plotted in Figure 9, where this can be established that L has an important influence on FS and D_e. It can be seen that entry point distance (D_e) and FS increased significantly as the L increased.

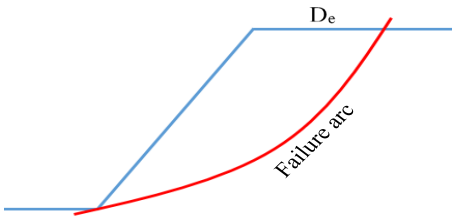

Fig. 8. Schematic diagram of the entry point distance (D_e) and length of failure surface.

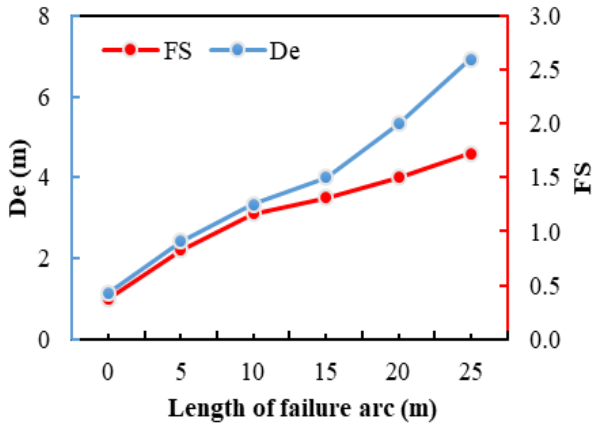

Fig. 9. Effect of entry point distance on the length of failure arc with corresponding FS.

Figure 9 is used to find the relation between entry point distance (D_e) and L, FS, slope height (h), γ, c and ϕ. The entry point distance (D_e) and L can be estimated using the following relations.

$$D_e = 0.89 \ln\left(\frac{c}{\gamma h \tan(\phi)}\right) + 3.2 \quad (5)$$

$$L = 0.81 \ln\left(\frac{c}{\gamma h \tan(\phi)}\right) + 6.19 \quad (6)$$

Where D_e is entry point distance of a CSS, c is cohesion in kN/m^2), γ is density of material, h is slope height (m), ϕ is internal friction angle of slope material and L is the length of slip surface.

5. CONCLUSIONS

In this study, the location of critical slip surfaces (CSS) in loose rock slope stability analysis was determined. The failure characteristics of CSS and the corresponding FS were also analyzed with the help of SLOPE/W software. The effect of cohesion (c), and internal friction (ϕ) on the failure mechanism of CSS and its depth (D) is also simulated. The following conclusions are drawn;

A loose rock slope can only show an overall failure by ignoring the shear strength properties. When the values of c and ϕ were small the circular failure did not occur and the CSS was a local one in case of loose rock slope. The location of CSS is consistent with the weak zone. Cohesion (c) and internal friction angle (ϕ) significantly affect the depth and length of CSS. Decreeing the value of c and ϕ results in a decrease in the length of the sliding arc (L). The shear strength parameters (c, ϕ) have a significant effect on the location of the critical slip surface and corresponding FS. Failure surface entry point distance (D_e) and the length of the sliding arc (L) have a logarithmic relationship with shear strength parameters. In future the effect of water pressure and earthquake loads must be considered for better understanding.

6. ACKNOWLEDGEMENTS

This work was conducted with the support of Bahauddin Zakariya University, Multan, Pakistan.

7. CONFLICT OF INTEREST

The authors declare no conflict of interest.

8. REFERENCES

1. M. Kinde, E. Getahun, and M. Jothimani. Geotechnical and slope stability analysis in the landslide-prone area: A case study in Sawla–Laska road sector, Southern Ethiopia. *Scientific African* 23: e02071 (2024).
2. X. Tang, C. Chen, D. Shan, P. Zhang, and J. Xue. Slope reliability assessment using an innovative critical failure path approach. *Frontiers in Earth Science* 12: 1428309 (2024).
3. A.S. Al-Jawadi. Predicting Slip Surfaces for Slope Stability Assessment Along Highway 80 in Mosul, Northern Iraq. *Geotechnical and Geological Engineering* 42(5): 2997–3008 (2024).
4. I. Khan, A. Afayou, N. Abbas, A. Khan, N. Alam, and S. Shah. Enhanced Geo-technical Methods for Evaluating Slope Stability in Unconsolidated Strata: A Comprehensive Analysis. *Journal of Mining and Environment* 15: 991-1010 (2024).
5. Z. Li, R. Wu, T. Hu, S. Xiao, L. Zhang, and D. Zhang. Stability analysis of an unstable slope in chongqing based on multiple analysis methods.

Processes 11: 2178 (2023).
6. M. Kassa, and M. Meten. Slope stability analysis using kinematic, limit equilibrium, and finite element models at Dessie town, northern Ethiopia. *Arabian Journal of Geosciences* 16: 675 (2023).
7. L. Li, Y. Wang, L. Zhang, C. Choi, and C. Ng. Evaluation of critical slip surface in limit equilibrium analysis of slope stability by smoothed particle hydrodynamics. *International Journal of Geomechanics* 19: 04019032 (2019).
8. J. Zhang, and H. Huang. Risk assessment of slope failure considering multiple slip surfaces. *JC Geotechnics* 74: 188-95 (2016).
9. P.R. Kumar, K. Muthukkumaran, and C. Sharma. Technological Advancements and Sustainable Practices in Rock Slope Stability–Critical Review. *Chemistry of the Earth* 136: 103699 (2024).
10. M. Hajiazizi, and H. Tavana. Determining three-dimensional non-spherical critical slip surface in earth slopes using an optimization method. *Engineering Geology* 153: 114-124 (2013).
11. R. Baker. Determination of the critical slip surface in slope stability computations. *International Journal for Numerical and Analytical Methods in Geomechanics* 4: 333-59 (1980).
12. A.W. Bishop. The use of the slip circle in the stability analysis of slopes. *Geotechnique* 10: 129-150 (1955).
13. N. Morgenstern, and V.E. Price. The analysis of the stability of general slip surfaces. *Geotechnique* 15: 79-93 (1965).
14. E. Spencer. A method of analysis of the stability of embankments assuming parallel inter-slice forces. *Geotechnique* 17: 11-26 (1967).
15. A. Bishop, and N. Morgenstern. Stability coefficients for earth slopes. *Geotechnique* 10: 129-153 (1960).
16. O. Hungr, F. Salgado, and P. Byrne. Evaluation of a three-dimensional method of slope stability analysis. *Canadian Geotechnical Journal* 26: 679-86 (1989).
17. X. Qi, and D. Li. Effect of spatial variability of shear strength parameters on critical slip surfaces of slopes. *Engineering Geology* 239: 41-9 (2018).
18. P. Regmi, and K. Jung. Application of dynamic programming to locate the critical failure surface in a rainfall induced slope failure problem. *KSCE Journal of Civil Engineering* 20: 452-62 (2016).
19. H. Pham, and D. Fredlund. The application of dynamic programming to slope stability analysis. *Canadian Geotechnical Journal* 40: 830-47 (2003).
20. A. Malkawi, W. Hassan, and S. Sarma. Global search method for locating general slip surface using Monte Carlo techniques. *Journal of Geotechnical and Geoenvironmental Engineering* 127: 688-98 (2001).
21. H. Zheng, D. Liu, and C. Li. Slope stability analysis based on elasto-plastic finite element method. *International Journal for Numerical Methods in Engineering* 64: 1871-88 (2005).
22. Y. Cheng, T. Lansivaara, and W. Wei. Two-dimensional slope stability analysis by limit equilibrium and strength reduction methods. *Computers and Geotechnics* 34: 137-50 (2007).
23. W. Zhang, Q. Zhao, J. Chen, R. Huang, and X. Yuan. Determining the critical slip surface of a fractured rock slope considering preexisting fractures and statistical methodology. *Landslides* 14: 1253-63 (2017).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 153-165 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)870

Pakistan Academy of Sciences

Research Article

Sentiment Analysis using Bidirectional Encoder Representations from Transformers

Adil Rehman[1], Khushal Das[2], Kamlish[3*], and Fazeel Abid[4]

[1]Department of Information Systems, Dr. Hasan Murad School of Management, University of Management and Technology, Lahore, Pakistan
[2]Department of Computer Science, Modelling, Electronics and Systems Engineering, University of Calabria, Rende, Italy
[3]Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad, 44000, Pakistan
[4]Department of Information Technology and Computer Science, University of Lahore, Pakistan

Abstract: In the contemporary digital landscape, a significant volume of data is generated through social networks such as Twitter, Facebook, and Instagram. This study presents a method for extracting sentiments from Twitter, focusing on two sentiment-based datasets: the Twitter and emotional sentiments datasets. After extraction and preprocessing, we employed three deep learning models: Recurrent Neural Networks (RNNs), Bidirectional Long Short-Term Memory (BiLSTM), and a pre-trained Bidirectional Encoder Representations from Transformers (BERT) model. We introduced Se-BERT, a model designed for emotional sentiment analysis. Our experiments showed that Se-BERT achieved accuracy levels of 97.29% for tweet sentiments (positive and negative) and 86.77% for emotional sentiments (joy, sadness, love, fear, anger, surprise). These results demonstrate that Se-BERT outperforms RNN and BiLSTM in terms of accuracy for sentiment analysis, thereby significantly enhancing information retrieval and providing a deeper understanding of user behaviour.

Keywords: Sentiment Analysis, Tweets Dataset, BiLSTM, BERT Model.

1. INTRODUCTION

The internet is crucial in modern life, enabling information retrieval, communication, and business activities. It features human-created content—like blogs, stories, and tweets, which provide expressive platforms for sharing thoughts—and machine-generated content, such as automated news and chatbots [1]. While human-generated content fosters personal expression and connections, machine-generated content is often seen as impersonal, lacking the human touch, and unable to comprehend emotions [2]. Differentiating between these two types of content can be challenging, especially in online interactions requiring vigilance. Additionally, managing online content to ensure safety while protecting freedom of expression presents ongoing challenges, as it involves finding the right balance between oversight and openness. Despite challenges, the internet remains a powerful connection, information-sharing, and access tool. It's essential to recognise the limitations of machine-generated content and approach online interactions with caution and critical thinking, especially as technology becomes increasingly embedded in our daily lives [3].

Understanding people's feelings is essential for fostering positive outcomes and mitigating negative ones. Sentiment analysis, a branch of Natural Language Processing (NLP), automates the detection of emotions in text, categorising them as positive or negative and even identifying specific emotions like joy or sadness [4]. This process involves converting text into a numeric vector format, using American Standard

Received: March 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Kamlish <kamlish@outlook.com>

Code for Information Interchange (ASCII) or Unicode, which encapsulates the expressive information contained within the text. However, the challenge lies in the imperfections of models due to insufficient training with diverse datasets, impacting optimal performance [5]. Our study employs neural networks, including Recurrent Neural Networks (RNN) and advanced models like Bidirectional Long Short-Term Memory (BiLSTM) and Sentiment-Bidirectional Encoder Representations from Transformers (Se-BERT), to analyse textual data and extract insights on how people feel about various topics or products [2]. This research introduces two primary approaches for sentiment analysis: emotion detection and NLP-based techniques.

Further, sentiment analysis has garnered significant attention due to its diverse applications in financial markets, social media, and human-computer interaction. Financial Sentiment Analysis (FSA) has evolved with two main research streams focusing on developing advanced techniques and applying them to market applications, highlighting the importance of understanding market sentiment and its relationship with investor sentiment [6]. Similarly, the advent of Web 4.0 has further expanded the applications of sentiment analysis across various domains, utilising machine learning and deep learning approaches to analyse users' emotions and generate insights for businesses, governments, and researchers [7]. Moreover, emotion detection has become crucial for enhancing human-computer interactions, with machine learning techniques such as SVM and Naive Bayes being widely used to analyse textual data and evaluate performance based on accuracy [8]. Our study introduces Se-BERT, a model designed for sentiment analysis using BERT. By providing significant insights into user behaviours and enhancing information retrieval capabilities, Se-BERT performs better in analysing tweets and emotional sentiments. By integrating the methodologies and insights from these seminal works, our research contributes to the growing body of knowledge in sentiment analysis and its practical applications in various fields.

Emotion Detection: This sentiment analysis method uses machine learning algorithms to identify and analyse emotions in text, categorising them into emotions like happiness, sadness, anger, or fear. It's beneficial for analysing unfiltered expressions on social media [9]. Emotion detection in texts, like tweets about a brand, identifies sentiment—positive, negative, or neutral—and guides marketing and product strategies.

Sentiment Analysis using NLP: This approach uses NLP techniques to categorise and assess text sentiment. NLP, a subfield of AI, analyses interactions between computers and human language [10]. Part-of-speech tagging, dependency parsing, and named entity recognition help identify critical phrases and sentiments in texts. For instance, analysing a product review could reveal critical features and associated feelings, which are helpful for marketing insights.

Advanced methods like Long Short-Term Memory (LSTM) with attention mechanisms enhance complex data analysis, such as noisy commercial time-series data [11]. These models focus on important textual information and integrate emotional data, helping analyse user personality and sentiment on platforms like social media [12]. This approach is crucial across various fields, from market analysis to customer feedback on social media, where understanding sentiment can guide business strategies and product development [13]. Like all machine learning models, sentiment categorisation models require data to be transformed into a standardised vector format, using ASCII or Unicode, to encapsulate textual meaning. This transformation is crucial for models like BERT, which rely on pre-trained language models for fine-grained sentiment analysis [14]. These NLP models and deep learning architectures have advanced significantly, enabling text conversion into machine-readable vector forms. These vectors represent text in an n-dimensional space, facilitating machine understanding through word embeddings [15]. In the stock market, sentiment analysis is increasingly used to gauge investor sentiment through online reviews and communications, with studies showing the BERT model outperforming the LSTM and SVM models [11]. During the COVID-19 lockdowns, sentiment analysis on Twitter became crucial for understanding political and social dynamics. In Nepal, sentiment analysis has been applied using multiple languages, with LSTM models providing the best results for analysing text and user personality from various social media sources [16]. Further development in this field includes an attention-based LSTM model that integrates sentiment and attention data to

predict the personality traits of users based on their online interactions [12].

In a study, Jin *et al*. [11] predicted the final stock market price by incorporating stockholders' sentiments and utilizing empirical modal decomposition (EMD). They also enhanced the LSTM model with an attention mechanism to improve performance and reduce time delays in stock market predictions. The primary contribution lies in predicting prices, which directly influences the stock market by analyzing the sequence of stock prices and closing prices, and improving the LSTM mechanism for better accuracy. Chen *et al*. [17] introduced a novel approach to extracting target expressions from opinionated phrases. They employed a Bidirectional LSTM combined with a Conditional Random Field (CRF) layer (BiLSTM-CRF) for classification. The key research challenge is developing a deep learning model for document-level classification. Rhanoui *et al*. [18] addressed this by implementing a BiLSTM model that combines convolutional and bidirectional recurrent neural networks, using Doc2vec for document-level sentiment analysis. Basiri *et al*. [19] further explored deep learning sentiment analysis by investigating both BiLSTM and Bi-GRU architectures, achieving a maximum accuracy of 90%. Building on these efforts, Li *et al*. [20] advanced the field by leveraging the BERT pre-trained language model for End-to-End Aspect-Based Sentiment Analysis (E2E-ABSA), where BERT outperformed other state-of-the-art-models. Additionally, integrating Word2Vec and Glove embeddings with an LSTM-CRF model, alongside BERT, resulted in even better performance.

BERT, a pre-trained bidirectional and unsupervised language model developed by Google, serves as the foundation for contextual word representations in ABSA tasks. Hoang *et al*. [21] demonstrated that BERT outperformed previous single-sentence classification models. BERT classifies sentiments as positive, negative, neutral, or conflict using the Fine-grained Stanford Sentiment Treebank dataset [14]. Its transformer-based architecture and fixed-size word vector representations enable fine-grained sentiment classification. In Chinese stock reviews, BERT has been applied to predict sentiments from various online stockholder evaluations [22]. Despite the challenges of expanding labelled datasets for Chinese stock classifications, BERT has demonstrated improvements through fine-tuning with a BERT + FC model, which shows high efficacy and potential for broad application. BERT has also been employed to detect bots on social media by analysing sentiment features [23]. Its integration with glove for word embedding and comparison against models like Random Forest, SVM, and logistic regression has yielded superior results in identifying bot-driven content. The use of BERT extends to analysing tweets related to COVID-19 and identifying sentiments from global and India-specific Twitter data [15]. This application has shown enhanced performance in sentiment analysis from tweets. Additional points on sentiment analysis include its applications in financial predictions, social media monitoring, customer service, and healthcare. Challenges involve dealing with sarcasm and figurative language, requiring advanced detection techniques. The growth of multilingual sentiment analysis necessitates language-specific training [24]. Ethical concerns about privacy, data usage, and biases in machine learning models are critical considerations in sentiment analysis deployment.

Understanding how people feel in written text, known as sentiment analysis, doesn't always work the same for every type of writing. To improve it, we can try different ways to use different sets of writing. It's not just about the techniques we use; it's also about getting the text ready for analysis. We should test different models and methods to make sentiment analysis work well. We should also clean up the text, like removing common words and making everything lowercase. Other tricks, like picking out the most important words or creating new features from the text, can also help. Different types of writing, like social media posts or news articles, might need special treatment. Words can mean different things in different places, so we might need unique models for each type. Ultimately, improving sentiment analysis means trying many things and seeing what works best for different kinds of writing. We implemented three baseline models, RNN, BiLSTM, and BERT, and did a comparative analysis using two sentiment analysis-based datasets of 800000 and 16000 rows simultaneously. Later, we Proposed a Se- BERT model based on CNN layer and BERT embeddings and achieved the best results compared to baseline models.

Despite numerous studies focusing on sentiment analysis using various models, there remains a gap in achieving optimal accuracy and efficiency, particularly with large and diverse datasets. Se-BERT, derived from BERT, outperforms RNN and BiLSTM using advanced deep learning. Se-BERT excels in analysing tweet and emotional sentiment datasets, with impressive accuracy of 97.29% and 86.77%, effectively capturing nuanced emotional expressions in social media data.

2. METHODOLOGY

We have divided the proposed methodology into two parts: data description and proposed architecture. The data description and architecture are described below.

I. **Data Description:** It's all about datasets, along with the diagrams and tables related to them.
II. **Se-BERT Model - Architecture:** It's all about BERT and the parameters used in code and architecture.

BERT is a Bidirectional Encoder Representation from Transformers and is based on transformers. Transformers are designed to proceed with sequential input data such as translation or text summarisation, and transformers are encoder and decoder stacks. BERT, which Google's pre-trained model introduces. BERT is an encoder stack of transformers with two variations: BERT base (12 layers) and BERT large (24 layers). Since its beginnings, it has completed various NLP tasks with state-of-the-art accuracy. BERT's architecture can better understand the language, learn variation in data forms, and perform well crossways a variation of NLP tasks because it was trained on a vast text corpus. Due to its bidirectional nature, BERT collects data from both the left and right sides of a token's context through the training phase. BERT pre-training model based on transformer encoders is a series of layers used in high dimensional representation of an input sequence. The main feature of the Transformer encoder is the attention mechanism, which allows the model to attend to different parts of the input data when generating the representation. The attention mechanism allows the model to capture the relationships between other parts of the input data and to use this information to develop a more accurate picture. Each Transformer encoder layer has a self-attention mechanism to assess input relationships and a feed-forward network to process and produce the final representation.The output of one layer can be used as input to the next layer, allowing the model to create deeper and more complex expressions. BERT tokenizers are used in the preprocessing steps of transformer encoders. This tokenisation process helps to confirm that the input data is constant with the pre-training and fine-tuning of the BERT model, as discussed in Figure 1.

Due to the large volume of text it was trained on, BERT has shown the ability to perform well on NLP tasks. For tasks such as Se-BERT, performance has demonstrated the ability to increase with the assistance of advanced training on Twitter text, termed post-training. To accomplish the Se-BERT challenge, the post-training created Se-BERT as a question-and-answer exercise using the machine reading comprehension technique known as review reading comprehension. Solving Se-BERT as a classification task performed sequentially using the

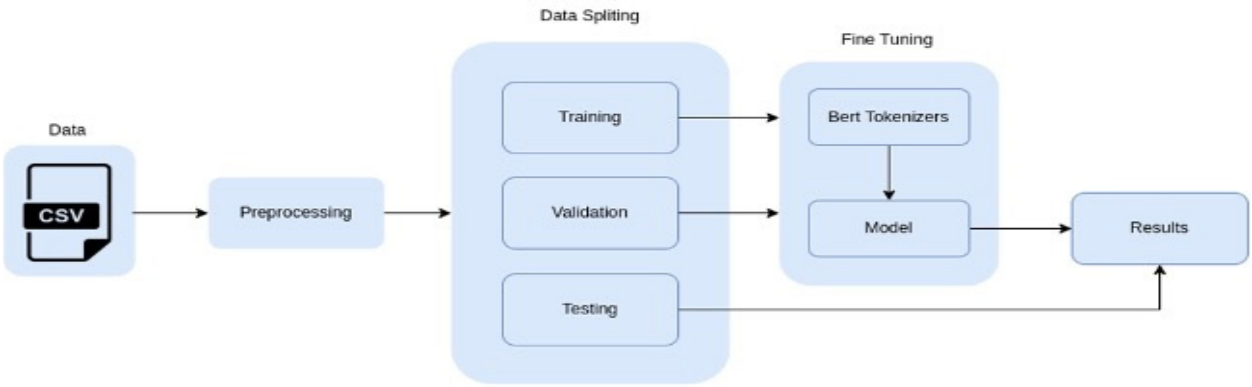

Fig. 1. Illustration of the Se-BERT model methodology, outlining data preprocessing, model architecture, and training phases.

BERT model by building auxiliary sentences has improved the results of the state-of-the-art model initialised as single-sentence classification. The architecture of transformers is based on attention mechanics; it determines the essential classifications for each computing step. In addition to projecting the input to a higher dimensional space vector, the encoder uses significant keywords as extra input to the decoder. It improves the decoder since it better understands the crucial sequences and the keywords that offer the context for the phrase. The BERT Tokenizer, a word tokenizer that turns words into tokens, is used first to handle the BERT model's text input. It yields a set of tokens, each of which denotes a word, along with two extra tokens: the set's opening addition is the classifier token [CLS]. When two groups of sentences are compared using BERT. Three distinct embedding layers with the exact dimensions are used to process this collection of tokens later. The sum of these layers is then sent to the encoder layer. Layers (Token Embedding Layer, Segment Embedding Layer, and Position Embedding) As shown in Figure 2, this is the complete architecture of the BERT Model. Still, using the BERT model, we need to apply for the single sentence sentiment analysis.

We have categorised the methodology into the data description and the proposed model architecture. Below is an explanation of the data definition and model architecture.

2.1. Data Description

We utilized two distinct sentiment analysis datasets for our implementation. The first dataset, the Sentiment Tweets dataset, contains 800,000 tweet records. The second dataset, the Emotional Sentiment dataset, includes 16,000 records categorized into six different emotions: fear, anger, surprise, sadness, love, and joy, as shown in Figure 3. By selecting the datasets provided by KanAnova [25] and Praveen [26], we implemented three models and analysed the results using the Se-BERT model.

Fig. 2. Representation of BERT model inputs, including token, segment, and positional embeddings.

2.2. Baseline Methods

2.2.1. Recurrent neural networks (RNN)

RNNs are an essential subset of neural networks usually utilised in natural language processing. They belong to a type of neural network that, while maintaining hidden states, permits the use of prior outputs as inputs. RNN is a conceptual part of memory and stores all textual-related data. RNNs are appropriately titled because they constantly complete the same task for every input in a sequence, with the results dependent on earlier calculations. In the RNN diagram, the input layer' x' is sent to the neural network and passed to the middle layer 'h', which contains multiple hidden layers. Then, it gets a response in the output layer 'y'. In this diagram, A, B, and C are the parameters used to improve the result of the RNN model, and 't' is the current state used in hidden layers to fetch back to improve results on behalf of recent developments, as shown in Figure 4. We implemented an RNN using two datasets. For the first sentiment analysis dataset, which contains 800,000 sentiment records labelled as positive or negative, we pre-processed the data by removing stop words from the sentences. We then analysed the labelled data using a word cloud. After this analysis, we implemented the Word2Vec

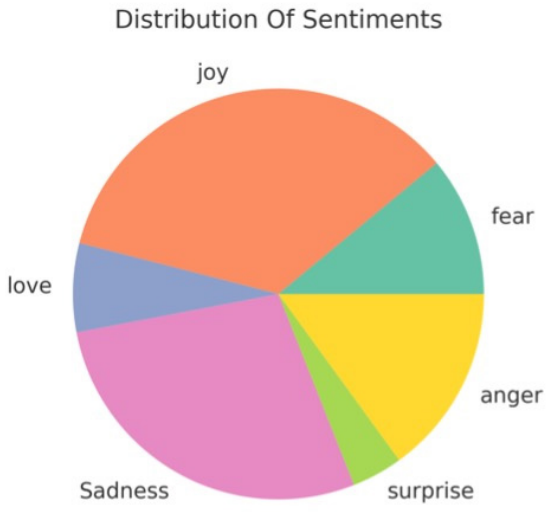

Fig. 3. Distribution of different emotional sentiments in the emotions' dataset.

model with 100 embedding layers. We used various hyperparameters, including the TensorFlow tokenizer with a vocabulary length of 60,000 and a maximum sentence length of 60. For training the RNN model, we used the Adam optimizer, which is known for its effectiveness, and for the loss function, we used the mean squared error method.

2.2.2. *BiLSTM model*

A bidirectional Long Short-Term Memory (BiLSTM) is a type of RNN and classification processing model that includes two LSTMs, one receiving input data forward and the other receiving it backwards. By successfully expanding the network's information pool, BiLSTM improves the contextual availability of the algorithm. It works with two instructions since it also works with two hidden layers, so this is a central point of disagreement of LSTM [16], but BiLSTM has proved to be of good result in NLP. However, according to BiLSTM, it contains two directions: one takes the input to the forward direction, and the other takes the backward direction, as shown in Figure 5. BiLSTM [28] improves the availability of information and context availability to the algorithms. We implemented BiLSTM with two datasets of sentiment analysis and emotional sentiment analysis. In the first dataset of sentiment analysis, we analysed the data with sentiments labels of positive and negative. After conducting research, we implemented the Word2Vec model and initialised the embedding matrix with 100 embedding dimensions. We used different hyperparameters, including the TensorFlow tokenizer with a vocabulary length of 60,000, a maximum sentence length of 60, a batch size of 1024, and ten epochs to train BiLSTM model. we used Adam optimiser, known for its effectiveness, during the training process. For the loss function, we use the binary cross-entropy method.

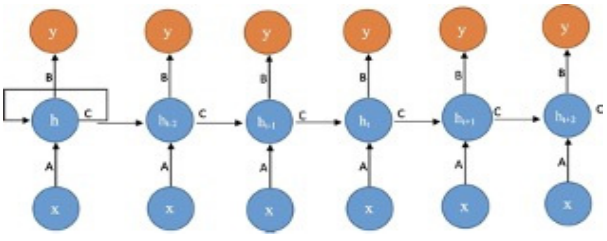

Fig. 4. The RNN model's structure shows input, hidden, and output layers.

2.3. Se-BERT Model Architecture

In the Se-BERT model, we used BERT transformer encoders and a 1D convolution layer (Conv1D) to analyse the best results. BERT is a Bidirectional Encoder Representation from Transformers and is based on transformers. Transformers are designed to proceed with sequential input data such as translation or text summarization. BERT pre-trained model, which is introduced by Google BERT Large. The model usages 24 layers, the hidden size is 1024, BERT's overall constraint size is 110 MB, Adam's optimiser is employed, and 2E-5 is used as the learning rate., the batch size of training and test set is 200. The epochs for training data are used as 6. BERT tokenizer is used to convert words into tokens. Then, we set the maximum length of the sentence to 60 according to our dataset sentence full length, and for the loss, we applied the entropy Loss function. Eight hours were used to train our data using Kaggle GPU, and we achieved 94. A Conv-1D, known as 1D CNN, is a convolutional neural network that operates on one-dimensional input data, such as a time series or a sequence of words. It is like a 2D CNN, commonly used for image processing tasks, but the main difference is that a 1D CNN linearly processes data. In a 1D CNN, the input data is run through several convolutional layers, each applying different filters. These filters are designed to detect specific patterns or features in the input data, such as changes in amplitude or frequency. The output of each convolutional layer is then passed through a pooling layer, which reduces the data's dimensionality and helps extract the most essential features. 1D CNNs can be used for various tasks, including natural language processing, speech recognition, and time series forecasting. They are particularly well suited to working with sequential data, where the order of the data is essential.

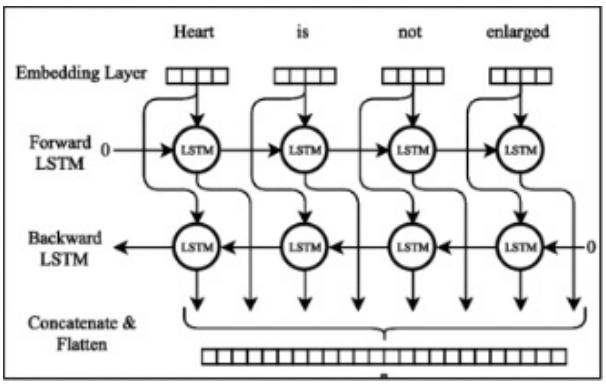

Fig. 5. BiLSTM model architecture [27].

The goal of an emotional dataset is to analyse a user's emotions based on their communication and tweets, particularly in the context of their interactions with other users. To achieve this, we used a BERT model combined with an Adam optimiser, which helps improve the training process's efficiency. One of the critical parameters we set was the maximum length of the input sequence, which we limited to 130 characters. It helped ensure that the model could process the input data quickly and accurately. We also set the learning rate to $5e^{-5}$, which helped ensure the model could learn from the input data effectively. To train the model, we used a loss function called categorical cross-entropy. This loss function is commonly used in multiclass classification tasks like this one, where we try to predict one of several emotional categories. We could ensure the model made accurate predictions by minimising the cross-entropy loss. Overall, the model's performance was excellent, with an accuracy of 97.29%. It means that the model could correctly predict the emotional category of the input data in nearly all cases. This level of accuracy is awe-inspiring, given the complexity of the dynamic dataset and the wide range of emotions it covers. One of the key advantages of using a BERT model for this type of task is its ability to analyse the context of the input data. Unlike other models that analyse individual words or phrases, BERT can consider the entire context of a sentence or communication, which helps improve the accuracy of predictions.

In addition to its accuracy, the BERT model is also highly flexible and can be adapted to work with a wide range of input data types. It is ideal for analysing emotions in various contexts, including social media, email communications, and more. Overall, the results of our analysis demonstrate the power and flexibility of the BERT model for analysing emotions in complex datasets. We accurately predicted emotional categories from user communications and tweets using this model with effective optimisation techniques and loss functions. Using the e-BERT model implementation, deep learning layers are used to achieve better performance; the architecture of my created model is defined below, as shown in Figure 6. One of the significant benefits of using a particular model for sentiment categorisation is that it can utilise both pre-labelled and unlabelled data, thanks to the capabilities of a tool called BERT. It is essential for social media networks, which generate vast text data every second. However, manually labelling this data is an incredibly time-consuming task that requires a significant amount of labour. To break it down further, sentiment categorisation is a task in which we try to determine the emotional tone of a text - for example, whether it expresses positive or negative sentiment. This task is becoming increasingly important as the amount of text data generated on social media platforms grows. One of the most popular tools used for sentiment categorisation is a model called BERT (Bidirectional Encoder Representations from Transformers). One of the key advantages of using BERT is that it can use pre-labelled and unlabelled data. This is significant because labelled data is data humans manually tag to indicate their sentiment, whereas unlabelled data has yet to be classified.

The ability to use unlabelled data is essential for social media networks, which generate vast amounts of data every second. Manually tagging this data is an incredibly labour-intensive operation that would be nearly impossible to keep up with. However, using BERT makes it possible to categorise this data automatically without manual labelling. The reason BERT can do this is because it is what is known as a pre-trained model. It means

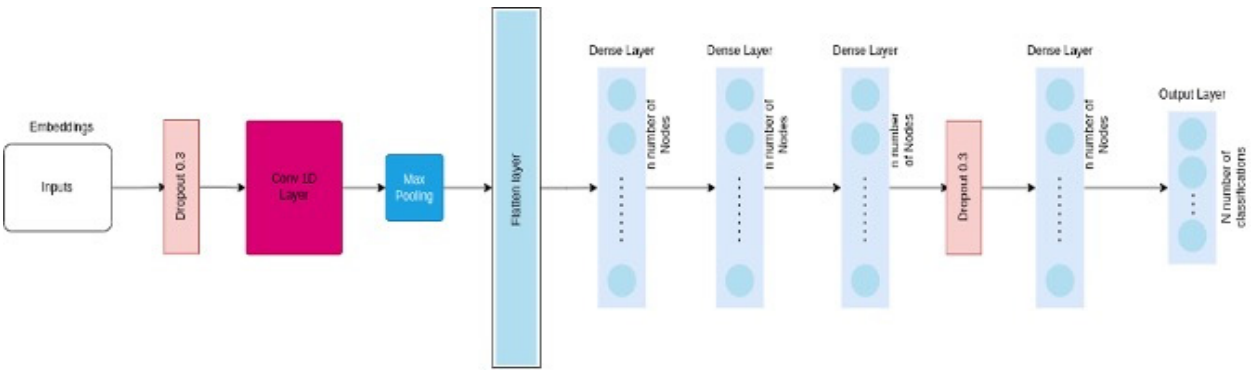

Fig. 6. Architecture of the Se-BERT model, combining BERT transformer encoders with a Conv1D layer.

it has already been trained on a large amount of text data and has learned to recognise specific patterns and features. When it encounters new, unlabelled data, it can use this existing knowledge to categorise it based on how it sees it. Pre-labelled data remains invaluable, as it provides a clear signal for the model to learn from. The ability to use unlabelled data sets BERT apart, making it a powerful tool for sentiment categorization. Automatically categorising sentiment in real-time text data is increasingly important for social media networks. BERT's ability to use pre-labelled and unlabelled data is a handy tool for this task. While manual data labelling is still essential, unlabelled data can significantly speed up the process and make it more efficient.

3. RESULTS AND DISCUSSION

In our sentiment analysis, we used two datasets with three deep learning models: RNN, BiLSTM, and BERT's pre-trained model. Among these, we achieved the best performance with the BERT base model, which is a part of NLP (Natural Language Processing). We used RNN, BiLSTM, and BERT pre-trained models as benchmarks and then initialized the Se-BERT model. For sentiment classification, the model takes a text input from natural language processing and outputs labels in an encoded numerical format (e.g., 0, 1). To achieve optimal performance, we applied multiple models and preprocessing techniques. We utilized the RNN and BiLSTM [16] models, with BiLSTM being an advanced version of the RNN deep learning model, and also employed the pre-trained BERT model (Bidirectional Encoder Representations from Transformers) for sentiment predictions. For the preprocessing and execution of baseline models, we evaluated the results using classification reports that included metrics such as Precision, Recall, and F1-score.

Precision is a metric that measures the accuracy of the model's predictions. It is defined as the ratio of correct sentiment predictions (True Positives, TP) to the total number of sentiment predictions (True Positives + False Positives, TP + FP) [29].

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall computes the completeness of a model, which is defined as the ratio of correct sentiment prediction (TP) and the total number of actual sentiment predictions (TP + FN).

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

The F-score (the F1-score or F-measure) represents the harmonic average of precision and recall and is usually used to optimise a model towards either precision or recall.

$$Recall = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

3.1. Baseline Methods

3.1.1. Recurrent neural networks (RNN)

After implementing the model, we achieved 82% accuracy on the tweet sentiment dataset. The sentiment analysis classification report for the RNN model's precision, recall, and F1-score is presented in Figure 7, and all results are displayed in Table 1. Precision measures the accuracy of the model, defined as the proportion of correct sentiment predictions (True Positives, TP) to the total number of sentiment predictions (True Positives + False Positives, TP + FP). The second dataset belongs to users' emotions, which depend on their behaviour.

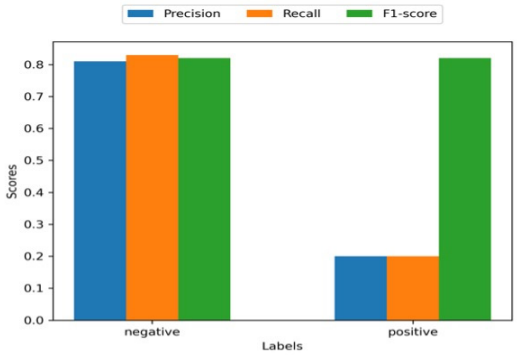

Fig. 7. Classification report for RNN on tweets sentiment dataset, showing precision, recall, and F1-score

Table 1. RNN classification report on tweets sentiments.

Parameters	Precision	Recall	F1-score	Support
Negative	0.81	0.83	0.82	39989
Positive	0.20	0.20	0.82	40011
Accuracy	---	---	0.82	80000
Macro avg	0.07	0.17	0.82	80000
Weighted avg	0.13	0.34	0.82	80000

We analysed this emotional dataset. We received six emotions as a label for the user's sentences. We used glove embedding to initialise the embedding metrics with 300 embedding dimensions. We implemented a categorical cross-entropy loss function with Adam optimiser using this method. After implementing the RNN model, we achieved 73% accuracy in the emotional sentiment analysis dataset. The emotional sentiment analysis classification report for the RNN model's precision, recall, and F1-score is presented in Figure 8, with the complete details shown in Table 2.

3.1.2. BiLSTM model

BiLSTM, an advanced form of RNN, proceeds bidirectionally forward and backwards and performs well, then generates results that we achieve better than RNN, with 84.52% accuracy in the tweets sentiments dataset. We show sentiment analysis results of the BiLSTM model (precision, recall, and F1-score) in Figure 9 diagram and Table 3. In the second dataset, which belongs to users' emotions and depends on their behaviour, we analysed the BiLSTM model with the emotional sentiment dataset. We received six emotions as a label for the user's sentences. We used glove embedding to initialise the embedding metrics with 300 embedding dimensions. This method utilised the Adam optimiser with a 0.001 learning rate and categorical cross-entropy loss function. BiLSTM, which is an advanced form of RNN, proceeds bidirectional forward and backward and performs well, then generates good results that we achieve better than RNN, that is, 92.15% accuracy in the dataset of emotional sentiments, and we show the impact of BiLSTM model precision, recall, and F1-score in Figure 10 and Table 4. For contextual word representations, we implemented word embedding to translate text into the machine-readable vectorised format using Word2Vec embeddings; we used bidirectional LSTMs and Word2Vec embeddings and pre-trained glove

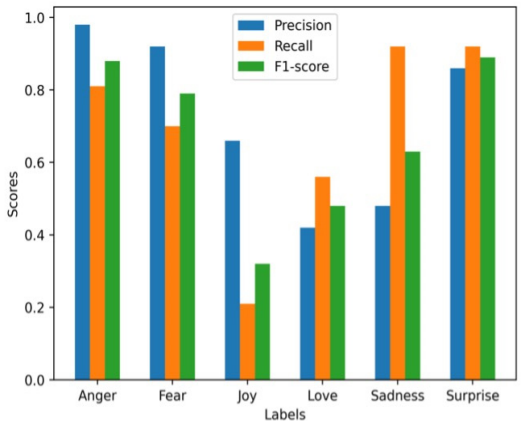

Fig. 8. Classification report for RNN on emotions sentiment dataset, showing precision, recall, and F1-score for each emotion.

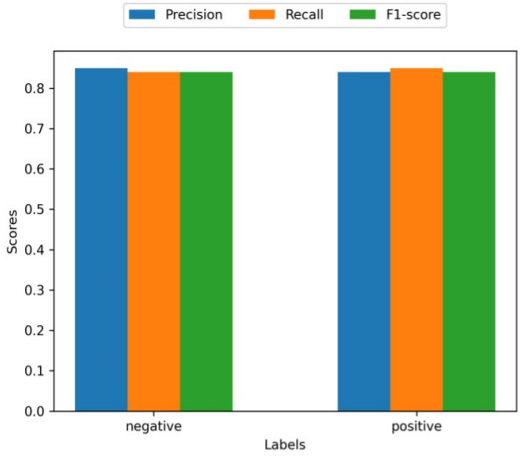

Fig. 9. Classification report for BiLSTM on tweets sentiment dataset, showing precision, recall, and F1-score.

Table 2. RNN classification report on emotions sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Anger	0.98	0.81	0.88	581
Fear	0.92	0.70	0.79	275
Joy	0.66	0.21	0.32	159
Love	0.42	0.56	0.48	66
Sadness	0.48	0.92	0.63	224
Surprise	0.86	0.92	0.89	695
Micro avg	0.79	0.79	0.79	2000
Macro avg	0.72	0.69	0.67	2000
Weighted avg	0.83	0.79	0.79	2000
Samples avg	0.79	0.79	0.79	2000

Table 3. BiLSTM classification report on tweets sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Negative	0.85	0.84	0.84	39989
Positive	0.84	0.85	0.84	40011
Macro avg	0.84	0.84	0.84	80000
Weighted avg	0.84	0.84	0.84	80000

Table 4. BiLSTM classification report on emotions sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Anger	0.97	0.96	0.97	581
Fear	0.94	0.93	0.93	275
Joy	0.74	0.92	0.82	159
Love	0.79	0.74	0.77	66
Sadness	0.88	0.90	0.89	224
Surprise	0.96	0.91	0.94	695
Macro avg	0.92	0.92	0.92	2000
Macro avg	0.88	0.90	0.89	2000
Weighted avg	0.93	0.92	0.92	2000
Sample avg	0.92	0.92	0.92	2000

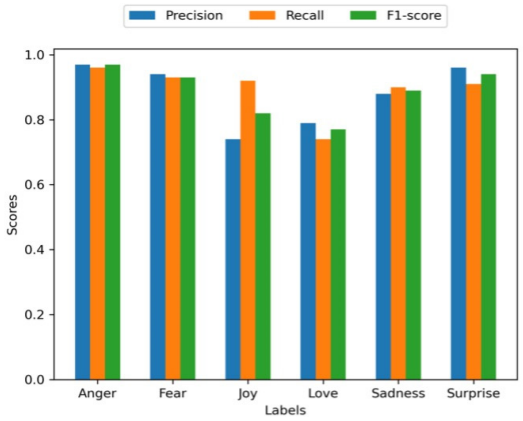

Fig. 10. Classification report for BiLSTM on emotions sentiment dataset, showing precision, recall, and F1-score for each emotion.

embeddings. We trained BiLSTM with embedding Matrixes are introduced and discussed by Yue and Li in [30]. We distributed data into categories and apply preprocessing to remove special characters or non-readable content. We don't need to clean tweets' emotion-based data and remove non-usable content. To achieve the best result, we used Adam Optimizer, which performs better, and for the loss function, we used binary cross-entropy loss function. Using these techniques, we approach our model accuracy at 84.52%, which is better. After applying BiLSTM, we used the BERT model based on the transformers [31] and achieved the best performance compared to BiLSTM.

3.2. Experimental Setup

3.2.1. Se-BERT setup

Every productivity element is connected to every input element by the BERT model, which is utilised for sentiment analysis. BERT is an established and pre-trained model based on transformers. BERT is the encoded process of transformer architecture. Pre-trained language models provide contextual data into tokens, representing the existence and representation of words from training data. BERT is considered the word from both the right and left side simultaneously. Although the idea is modest, it expands outcomes at numerous NLP tasks such as user reviews, sentiment examination, and question-and-answering systems when order is compared separately to training left and right. The input representation of the BERT can signify a sentence in a sequence of tokens [32]. The input representation of a particular token is created by adding up its relevant segments, tokens, and positional embeddings. For the implementation of classification tasks, the first word of each sequence is an exclusive classification embedding [33]. The Se-BERT model performance improves in both datasets, and the results are better than BiLSTM [28].

In sentiment analysis, accuracy measures the percentage of correctly classified positive tweets, while recall indicates the proportion of actual positives identified. The F score combines accuracy and recall, balancing precision and recall providing a comprehensive view of model performance. By analysing accuracy, recall, and F- score, we can better understand the strengths and weaknesses of a sentiment analysis model and make informed decisions about how to improve its performance. We trained three models in two sentiment analysis datasets for all the research and outcomes. We expected the models to improve the results after combining our dataset because there would be more features and emotional categories. The state-of-the-art model is BERT and uses transformers, which transform words into tokens and tokens captured by the corpus to find sentiments in a text. On behalf of both models, RNN and BiLSTM, we achieve the best result by using the BERT model, which is based on transformers and converts words into the embedding form [30].

Using the BERT model, we applied a deep learning technique to add three dense layers with one dropout proposed as a Se-BERT model. We achieved 86.77% and 97.29% accuracy in two datasets of tweets and emotional sentiments, respectively.

BERT uses positional word embeddings to generate distinct word embeddings for each word, depending on its position in the text, in contrast to the prior best models that generate one vector for each word [34]. Another factor could be using sentence-pair classification instead of the previous top models. Which relied on single sentence sorting to identify which feature was present in the text. These results are found using GPU, and we achieve our best result with 80,000 records of a dataset in 5 hours with three epochs using the Se-BERT model. The classification report is shown in Table 5 and Table 6. We show tweets' sentiment analysis results of the Se- BERT model precision, recall, and F1-score in Figure 11 and Table 7.

Table 5. Datasets accuracy on all models.

Datasets	Epochs	Models	Accuracy	Support
Tweets dataset	10	RNN	82%	581
	10	BiLSTM	84.52%	275
	6	Se-BERT	82%	159
Emotional Dataset	10	RNN	34%	66
	15	BiLSTM	92.15%	224
	20	Se-BERT	94.29%	695

Table 6. Se-BERT model classification report on emotional sentiments.

Parameters	Precision	Recall	F1-score	Support
Anger	0.92	0.94	0.93	813
Fear	0.90	0.88	0.89	712
Joy	0.96	0.95	0.96	2028
Love	0.86	0.86	0.86	492
Sadness	0.97	0.96	0.97	1739
Surprise	0.77	0.87	0.82	216
Accuracy	---	---	0.94	6000
Macro avg	0.90	0.91	0.90	6000
Weighted avg	0.94	0.94	0.94	6000

Table 7. Se-BERT classification report on tweets sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Negative	0.82	0.82	0.82	39989
Positive	0.82	0.81	0.82	40011
Accuracy	---	---	0.82	80000
Macro avg	0.82	0.82	0.82	80000
Weighted avg	0.82	0.82	0.82	80000

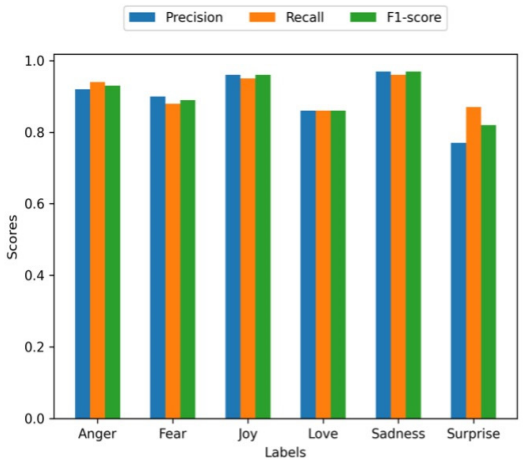

Fig. 11. The Classification report for Se-BERT on the sentiment's dataset shows precision, recall, and F1-score.

4. CONCLUSIONS

In conclusion, this study introduces Se-BERT, a novel sentiment analysis model derived from BERT. It evaluates performance against established RNN and BiLSTM models across emotional and tweet sentiment datasets. Se-BERT exhibits exceptional accuracy, scoring 97.29% for emotional sentiments and 94.84% for tweet sentiments. Sentiment analysis, a crucial aspect of natural language processing, seeks to identify and understand sentiments expressed in text. Machine learning models like RNN, BiLSTM, and BERT have propelled sentiment analysis research, and Se-BERT contributes as a specialised variant of the BERT model for enhanced sentiment analysis. The evaluation involves two datasets: tweets sentiments and emotional sentiments. Se-BERT surpasses RNN and BiLSTM, achieving a final score of 97.29% on the emotional sentiment dataset and 94.84% on the tweet sentiments dataset. These results underscore Se-BERT's effectiveness in deciphering nuanced emotional expressions in social media data. Beyond academic interest, sentiment analysis finds practical applications in social media monitoring, customer feedback analysis, and market research. With its capacity for training on extensive datasets and fine-tuning for specific tasks, Se-BERT emerges as a robust tool for these applications. Understanding user behaviour on platforms like Twitter is crucial, and sentiment analysis of social media data provides valuable insights into users' emotions, opinions, and attitudes. This research's comparative analysis of RNN, BiLSTM, and pre-trained BERT guides researchers and practitioners

in selecting appropriate models. Se-BERT not only outperforms existing models but also offers a highly accurate method for emotional sentiment analysis. Future research could explore sentiment analysis applications in different contexts and develop new models for improved accuracy and efficiency. This study significantly advances sentiment analysis methodologies and their practical.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. E. Adamopoulou, and L. Moussiades. Chatbots: History, technology, and applications. *Machine Learning with Applications* 2: 100006 (2020).
2. K. Machova, M. Szaboova, J. Paralic, and J. Micko. Detection of emotion by text analysis using machine learning. *Frontiers in Psychology* 14: 1190326 (2023).
3. E. Cambria, and B. White. Jumping NLP curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence Magazine* 9(2): 48-57 (2014).
4. J.C. Pereira-Kohatsu, L. Quijano-Sanchez, F. Liberatore, and M. Camacho-Collados. Detecting and Monitoring Hate Speech in Twitter. *Sensors* 19(21): 4654 (2019).
5. F.M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *Preprint ArXiv1*: 2305.17473 (2023).
6. K. Du, F. Xing, R. Mao, and E. Cambria. Financial Sentiment Analysis: Techniques and Applications. *ACM Computing Surveys* 56(9): 1-42 (2024).
7. U. Singh, K. Abhishek, and H.K. Azad. A Survey of Cutting-edge Multimodal Sentiment Analysis. *ACM Computing Surveys* 56(9): 1-36 (2024).
8. A. Alslaity, and R. Orji. Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour and Information Technology* 43(1): 139-164 (2024).
9. C.R. Sugimoto, S. Work, V. Lariviere, and S. Haustein. Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* 68(9): 2037-2062 (2017).
10. J.P. Bharadiya. A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. *European Journal of Technology* 7(1): 58-66 (2023).
11. Z. Jin, Y. Yang, and Y. Liu. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications* 32: 9713-9729 (2020).
12. J. Zhao, Z. Dalin, Y. Xiao, L. Che, and M. Wang. User personality prediction based on topic preference and sentiment analysis using LSTM model. *Pattern Recognition Letters* 138: 397-402 (2020).
13. M. Li, W. Li, F. Wang, X. Jia, and G. Rui. Applying BERT to analyse investor sentiment in stock market. *Neural Computing and Applications* 33: 4663-4676 (2021).
14. M. Munikar, S. Shakya, and A. Shrestha. Fine-grained Sentiment Classification Using BERT. *Preprint ArXiv1*: 1910.03474 (2019).
15. M. Singh, A.K. Jakhar, and S. Pandey. Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining* 11: 33 (2021).
16. M. Tripathi. Sentiment Analysis of Nepali COVID19 Tweets Using NB SVM and LSTM. *Journal of Artificial Intelligence* 3: 151-168 (2021).
17. T. Chen, R. Xu, Y. He, and X. Wang. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications* 72: 221-230 (2017).
18. M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali. A CNN-BiLSTM model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction* 1(3): 832-847 (2019).
19. M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U.R. Acharya. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Generation Computer Systems* 115: 279-294 (2021).
20. X. Li, L. Bing, W. Zhang, and W. Lam. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. *Preprint ArXiv*2: 1910.00883 (2019).
21. M. Hoang, O.A. Bihorac, and J. Rouces. Aspect-based sentiment analysis using bert. *Proceedings of the 22nd Nordic Conference on Computational Linguistics, Turku, Finland* (September 30 – October 02, 2019) pp. 187-196 (2019).
22. M. Li, L. Chen, J. Zhao, and Q. Li. Sentiment analysis of Chinese stock reviews based on BERT model. *Applied Intelligence* 51: 5016–5024 (2021).
23. M. Heidari, and J.H. Jones. Using BERT to Extract Topic-Independent Sentiment Features for Social Media Bot Detection. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New*

York, NY, USA, pp. 0542-0547 (2020).
24. E.F. Can, A. Ezen-Can, and F. Can. Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data. *Preprint ArXiv*: 1806.04511 (2018).
25. M.M. KazAnova. Sentiment140 dataset with 1.6 million tweets. *Kaggle*: (2017). https://www.kaggle.com/datasets/kazanova/sentiment140 (accessed 22 July 2023)
26. P. Praveen. Emotions dataset for NLP. *Kaggle*: (2020). https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp (accessed 22 July 2023)
27. S. Cornegruta, R. Bakewell, S. Withey, and G. Montana. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. *Preprint ArXiv*:1609.08409 (2016).
28. G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu. Sentiment Analysis of Comment Texts based on BiLSTM. *IEEE Access* 7: 51522–51532 (2019).
29. K. Das, M. Shehryar, F. Abid, M. Ashraf, M. Adil, S. Inam, E.U. Haq, and H. Mushtaq. Impact of using e-learning tools on Student's Psychological Health during covid-19. *VFAST Transactions on Software Engineering* 9(3): 120-127 (2021).
30. W. Yue, and L. Li. Sentiment Analysis using Word2vec-CNN-BiLSTM Classification. *Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), Paris, France* (December 14–16, 2020) pp. 1-5 (2020).
31. L. Zhao, L. Li, X. Zheng, and J. Zhang. A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts. *Preprint ArXiv*: 2001.05326 (2020).
32. Y. Song, J. Wang, Z. Liang, Z. Liu, and T. Jiang. Utilizing BERT Intermediate Layers for Aspect Based Sentiment Analysis and Natural Language Inference. *Preprint ArXiv*: 2002.04815 (2020).
33. A. Karimi, L. Rossi, and A. Prati. Adversarial Training for Aspect-Based Sentiment Analysis with BERT. *Preprint ArXiv*: 2001-11316 (2021).
34. S.M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi. Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications* 117: 139–147 (2019).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 167-179 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)861

Pakistan Academy of Sciences

Research Article

Dynamic Changes in Rainfall Necessitate Efficient Rainwater Harvesting in Different Agro-Ecologies of Pakistan for Sustainable Development

Arshad Ashraf[1*], Awais Ahmed[1], Muhammad Bilal Iqbal[1], Ahsan Mukhtar[2], Naveed Mustafa[1], Rehan Ahmad[3], and Salma Khan[1]

[1]Climate, Energy and Water Research Institute, National Agricultural Research Center, Islamabad, Pakistan
[2]Center for Agriculture and Biosciences International (CABI), Rawalpindi, Pakistan
[3]Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China

Abstract: Changes in climate together with rapid urbanization are putting immense pressure on the existing agriculture and natural resources of South Asia. The agriculture productivity and livelihoods of a large number of communities have become highly vulnerable to inadequate supplies of water, especially in the rainfed regions. In the present study, spatio-temporal changes in rainfall patterns have been analyzed in different agro-ecologies of Pakistan during 1960-2019 period for sustainable agriculture and natural resource management in the country. Major agro-ecologies identified in the country were western dry mountains over 19.1% area, western dry plateau over 14.4% area, sandy desert over 14% area and northern irrigated plain over 11.3% area of the country. An increase of 30.6% in annual rainfall was observed in the rainfed plateau zone, 14.2% in the piedmont plain and 5.9% in the western dry mountain zone during the 1960-2019 period. In contrast, the rainfall exhibited a 3.2% decrease in the western dry plateau and 6.2% in the coastal zone which is critical for the subsistence agriculture in these arid ecologies. The excess rainwater may be conserved through developing storage/farm ponds and reservoirs for subsequent use during dry periods and recharging groundwater to build resilience against drought/flood conditions. The rainwater harvesting (RWH) has the potential to sustain agricultural productivity and fulfil the growing needs of the population in the arid ecologies of the country. Regular monitoring of the water resources is essential in the context of the rapidly changing environment and growing needs of the population in this arid region in future.

Keywords: Climate Change, Drought, Flood Management, Water Conservation.

1. INTRODUCTION

Global freshwater resources are under continuous pressure due to climate change and population growth [1]. The growing water requirements for agriculture, industrial and urban developments are placing demands and constraints on the natural environment of rainfed regions where rainfall is a primary water source [2, 3]. Pakistan is facing challenge of reduced water availability for her agriculture sector owing to rapid urbanization and changing climate [4]. Groundwater is generally utilized to supplement canal water supplies in the irrigated areas and overcome water-stressed condition in the rainfed areas of the country. As a result of growth in urbanization, extension in industrial and agriculture activities [5-7], groundwater levels are declining putting high stress on the underlying aquifer in the Indus basin [8-11]. According to Qureshi [12], extensive groundwater use in the Indus basin led to an annual decline of about 1.5 m in the watertable, as a result of which aquifer started to drain faster than the natural recharge process [9]. The groundwater levels are declining owing to increase in groundwater discharge from growing number of domestic tube-wells [13, 14] and less rainwater conservation. Rainwater harvesting (RWH) system is generally adopted to mitigate risk of rainfall shortages during drought condition [15, 16]. The system

Received: April 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Arshad Ashraf <mashr22@yahoo.com>

comprises of rainwater collection, e.g., through micro to macro catchments, terracing, check dams, rooftops, for direct-use or via some storage facility or recharging of groundwater [17]. Effective RHW can sustain agricultural development, maintain natural water cycle [18] and improve livelihoods of the local communities in the region. There is a need to identify water harvesting and conservation potential to recharge groundwater and enhance agricultural productivity and economic growth in different agro-ecologies. Agro-ecological zones provide relevant information of existing status of environment important for prioritizing research in agriculture and natural resources. It is hypothesized that there exists sufficient rainwater available for storing and utilization for enhancing agricultural and economic productivity in the country. The present study is aimed to analyze spatio-temporal changes in climate during 1960-2019 period in different agro-ecologies of Pakistan for sustainable agriculture and natural resource management in the country.

1.1. Geographical Setup

Pakistan has diverse physiography and climate from south to north. The climate varies from subtropical hyper-arid and arid in the southern plains to alpine in the northern mountainous high lands. The annual precipitation is lowest in the southwest and increases gradually towards northeast (Figure 1). About 60-70% of the annual rainfall occurs during the monsoon season (July to September). Annual rainfall reaches more than 1200 mm during summer monsoon and heavy snowfall occurs during the winter season. Winter rainfall storms moving from Quetta valley lose their moisture after reaching central areas. The landform is generally mountainous having topography characterized by Indus River system and numerous intermittent and perennial nullahs/streams. In the east, alluvium deposits occur in the Indus irrigated plains of the Punjab and Sindh provinces and in the west, gravel fans form distinct piedmont zones consisting of finer sediments (sand and silt) and gravels. Major land cover consists of irrigated agriculture over 22% area, forest cover 4% and rock outcrops over 42% mainly in the northern and western mountanous region. Most of the area is used for grazing where livestock provides not only a major source of income but also of food security.

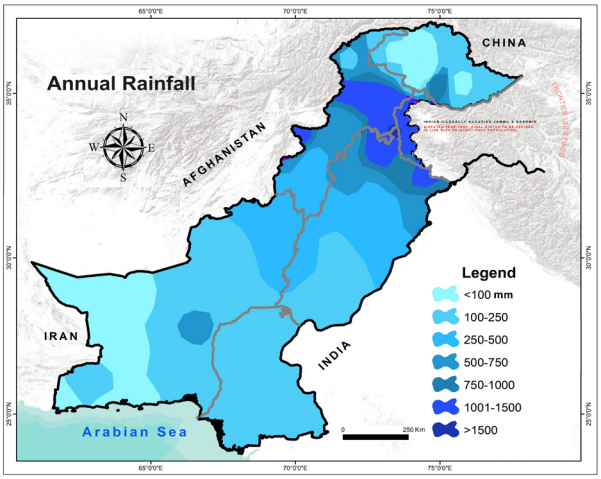

Fig. 1. Annual rainfall distribution in Pakistan.

2. MATERIALS AND METHODS

The climate data of 1960-2019 period was acquired of 59 stations of the country on monthly level from Pakistan meteorological department. The temperature and rainfall data were segregated into 1960-1989 and 1990-2019 climatic normals to perform spatio-temporal analysis of climate in different agro-ecologies of the country. Pakistan Agricultural Research Council [19] updated agro-ecological zones of the country during 2020-2023 based on the existing climate, land use, soils/landforms and irrigation conditions of the country. The ancillary data of topography, soils/landforms, hydrology, irrigation and agriculture were collected from various source departments like Survey of Pakistan, Water and Power Development Authority, and provincial soil survey, irrigation and agriculture departments. The base maps of agro-ecological zones, i.e., physiography, agro-climate, soils/landforms and land use were prepared in ArcGIS software using Universal Transverse Mercator projection system for integration and overlay analysis (Figure 2). The physiography of the country indicated 42% lowlands (< 300 m elevation), 14% plateaus (300 - 700 m), 14% Siwaliks (700 - 1200 m), 13% lower mountains (1200 - 2000 m), 6% Middle mountains (2000 - 3000 m) and 11% high mountains (> 3000 m). The reference evapotranspiration (ETo) along with rainfall data was used to develop agro-climate zones of the country. The ETo was calculated using the Penman-Monteith method (Equation 1) following Allen *et al*. [20].

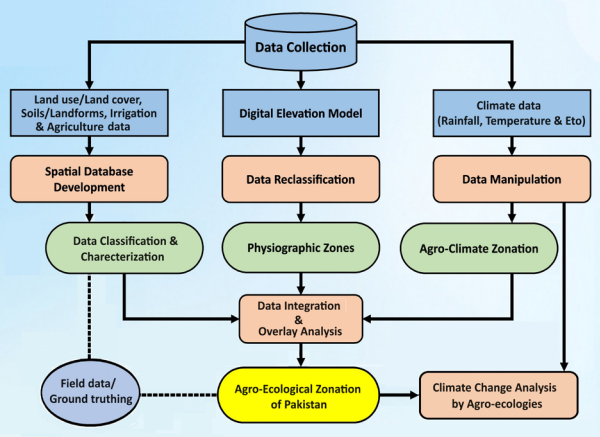

Fig. 2. Flowchart of methodology adopted in the present study.

$$ETo = \frac{0.408\Delta(R_a - G) + \gamma\frac{900}{T+273}u2\,(e_s - e_a)}{\Delta + \gamma(1+0.34u_2)} \quad (1)$$

Where, *ETo* is reference evapotranspiration (mm per day); R_a is net radiation at the crop surface (MJ/m² per day); *G* is soil heat flux density (MH/m² per day); *T* is mean daily air temperature at 2 m height (°C); U_2 is wind speed at 2 m height (m/s); e_s is saturation vapor pressure (kPa); e_a is Actual vapor pressure (kPa); Δ represents Slope of vapor pressure curve(kPa per °C) and γ represents Psychometric constant (kPa per °C). The performance of Penman-Monteith method is quite satisfactory in the diverse climatic conditions of Pakistan as compared with other methods [21].

The ETo data was also used to highlight crop water requirements in different agro-ecologies during 1960-2019 period. The rainfall, temperature and ETo data was interpolated using Inverse distance weighting (IDW) method. The zonal statistics function of ArcGIS was used to determine mean annual values of each parameter for the two normals for each agro-ecological zone of the country. The ranking of climate change as given in Table 1 was used to highlight intensity in different agro-ecologies of the country. The rank 1 denotes minimum change in the climate and rank 10 maximum change in the climate of a zone. One represntative climate station was selected in each agro-ecology based on long-term data availability for consistancy to study trend and change in annual and monthly rainfall during 1960-2019 period. The monthly rainfall data of 1960-1989 and 1990-2019 normals were illustrated graphically for comparison and shift analysis at respective climate station of the ecology. Finally benefits of adoption of RWH interventions were described for different agro-ecologies to improve agriculture productivity in the country.

Table 1. Ranking of climate change used to represent intensity in different agro-ecologies in this study.

S. No.	Intensity	Rank
1	Very low	1
2	Low	2
3	Slightly-low	3
4	Slight	4
5	Below average	5
6	Average	6
7	Moderate	7
8	Moderate-high	8
9	High	9
10	Very high	10

3. RESULTS AND DISCUSSION

3.1. Climate Data Analysis

According to 1990-2019 climatic normal, maximum mean annual rainfall was 1321 at Kakul followed by 869 mm at Jhelum, 378 mm at Faisalabad, 301 mm at D.I. Khan, 245 mm at Quetta and 226 mm at Chhor station. Other stations have < 200 mm rainfall, e.g., 196 mm at Karachi, 164 mm at Hyderabad, 137 mm at Gilgit and least at Panjgur, i.e., 97 mm. The annual rainfall indicated variation from 35 mm at Nokundi in the southwest to over 1827 mm at Murree in the northeast of the country. Over 68.4% area comprising major parts of the Balochistan, Sindh and lower Punjab province receives < 300 mm rains. The rainfall ranges between 300 - 500 mm in 19.1% area of the western and central Punjab, lower and upper Khyber Pakhtunkhwa (KPK) province, and some parts of the Gilgit-Baltistan (GB). About 6.9% area comprising of northern half of the Punjab and KPK provinces receives 500 mm to 1000 mm rains. More than 1000 mm rainfall occurs over 5.6% area comprising northern elevated parts of the Punjab and KPK provinces. The ETo was observed within range of 1500 - 2000 mm in the Punjab, KPK and 70% of the Balochistan province. The Lasbella district in Balochistan showed highest ETo, i.e., within range of 2500 - 2800 mm, while minimum ETo range i.e., 900 - 1000 mm was observed at Astore in GB. The 1000 - 1250 mm ETo range was observed in major parts of the GB and Kashmir

areas. Around 70% area of the Sindh province exhibited ETo within range of 2000 mm and 2500 mm.

3.2. Climate Analysis by Agro-Ecology

Agro-ecological zones of Pakistan were developed indicating regions of homogenous climatic, landforms/soils, physiography and land use characteristics in the country [19], as shown in Figure 3. Major zones consist of western dry mountains (WDM) about 19.1%, western dry plateau (WDP) 14.4%, sandy desert (SD) 14% and northern irrigated plain (NIP) 11.3% (Table 2). The WDM zone consists of western parts of the KPK and adjoining Balochistan province where climate is predominantly arid to semi-arid subtropical to temperate and warm Mediterranean. Forest, livestock and horticulture plantation are the major land use here. The WDP zone comprises of major part of the Balochistan province in the southwest where climate is predominantly warm desertic hyper-arid to semi-arid subtropical. Major landuse here is rangeland, livestock, fruits and minor crops. At places, groundwater resource in form of karezes and perennial streams exist to sustain scanty vegetation of drought resistant trees and some agriculture land in patches. The SD zone comprises of Thal and Cholistan deserts in the Punjab, Thar in the Sindh and Kharan desert in the Balochistan province. Climate here is desertic hyper-arid to semi-arid subtropical. Rangeland, livestock and minor crops/fruits are the major landuse in this ecology. Northern dry mountains (NDM) over 9.2% consist of massive snow and glacier resource which contributes meltwater flows to gigantic Indus River system downstream. Mono cropping is practiced here within 2000-3000 m elevation range and double cropping below 2000 m range [22]. Temperate mountains (TM) over 5.9% area, lies within 300 - 6300 m elevation range mainly in the Khyber Pakhtunkhwa and Punjab provinces. Annual rainfall here reaches more than 1200 mm during the summer and heavy snowfall occurs during the winter season. In the northern irrigated and southern irrigated plains, surface water availability sustains year-round cropping intensities [10]. Climate in the former zone is predominantly arid to semi-arid subtropical, whereas in the later zone, climate is predominantly warm desertic hyper-arid to arid subtropical. The rainfed plateau (RP) zone (also called 'Pothwar plateau') stretches over 3% area within 300 - 900 m elevation range mainly in the northwestern part of the Punjab province. Agriculture and livestock are the main landuse here. The piedmont plains (PP) over 7.4% area (Table 2) consist of nearly leveled to undulating piedmont region along Kohistan mountains in the Sindh and Balochistan, Suleiman mountains in the Punjab and KP, besides sub-Himalayan mountains in the Punjab province. During flood season (July-September), the hill-torrent (locally '*Rod Kohi' or 'Sailaba'*) flows in the piedmont plain is diverted to irrigate the agriculture land [23-25]. The coastal zone (CZ) stretches over 9.9% area along Arabian Sea in the south of Sindh and Balochistan provinces (Figure 3). This zone is rich in mangrove forests, fruits, fish and livestock resource. The agro-ecological zones support in selection of suitable crop grown areas and climate smart interventions in the country.

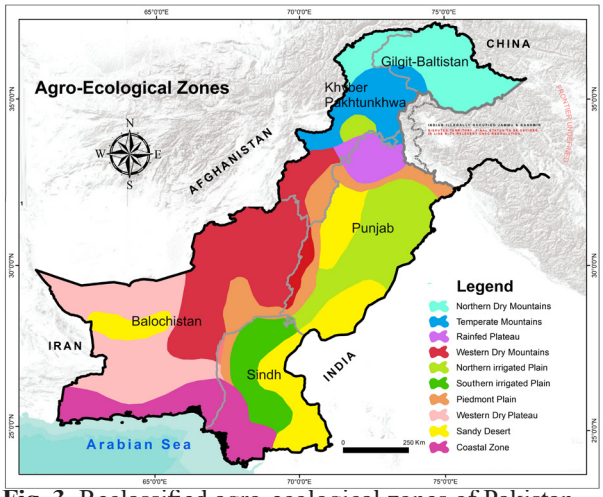

Fig. 3. Reclassified agro-ecological zones of Pakistan.

Climate change analysis was performed at agro-ecological level during the 1960-2019 period (Figure 4). Annual rainfall indicated an increase of about 30.8%, mean temperature 2.6% and ETo 4.2% in the NDM ecology during 1960-1989 and 1990-2019 climatic normals (Table 3). In the TM ecology, annual rainfall indicated an increase of about 3.9%) and ETo 8.3%, while mean temperature exhibited a decrease 5.6% during the two normals. In the RP ecology, annual rainfall indicated an increase from 668 mm to 872 mm (about 30.6%), while mean temperature exhibited a decrease from 21.6°C to 20.7°C (–4.0%) and ETo from 1496 mm to 1493 mm (–0.2%). The WDM ecology indicated an increase in annual rainfall from 271 mm to 287 mm, in mean temperature from 20.5°C to 21.6°C and in ETo from 1661 mm to 1662 mm during the

Table 2. Physical characteristics of major agro-ecological zones of Pakistan.

S. No.	Agro-ecology zone	Area (%)	Mean elevation (m)	Selected station
1	Northern Dry Mountains (NDM)	9.2	4212	Gilgit
2	Temperate Mountains (TM)	5.9	2116	Kakul
3	Rainfed Plateau (RP)	3.0	471	Jhelum
4	Western Dry Mountains (WDM)	19.1	1398	Quetta
5	Northern Irrigated Plain (NIP)	11.3	175	Faisalabad
6	Southern Irrigated Plain (SIP)	5.8	42	Hyderabad
7	Piedmont Plain (PP)	7.4	176	D.I. Khan
8	Western Dry Plateau (WDP)	14.4	902	Panjgur
9	Sandy Desert (SD)	14	179	Chhor
10	Coastal Zone (CZ)	9.9	160	Karachi

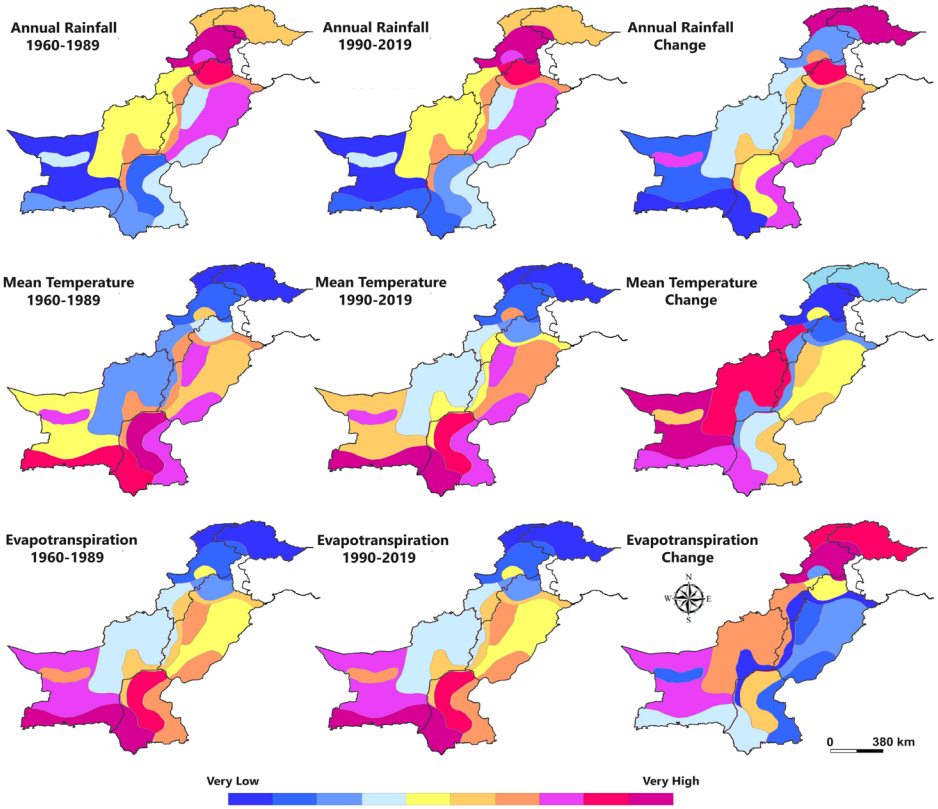

Fig. 4. Changes in climate observed in different agro-ecologies of the country during 1960-2019 period.

two climatic normals. The increase in rainfall in the RP and WDM ecologies would be beneficial for the agriculture, livestock and grazing land. Annual rainfall showed an increase of about 17.2% and mean temperature 0.4%, while ETo exhibited a decrease of 2.3% in the NIP ecology (Table 3).

Whereas, in the SIP ecology, annual rainfall indicated an increase of about 11.2%, while mean temperature exhibited a decrease of about 0.6% during the two normals. In the PP ecology, annual rainfall increased from 330 mm to 377 mm, while mean temperature decreased from 24.6 °C to 24.2 °C and ETo from 1784 mm to 1677 mm during the 1960-2019 period. In contrast, annual rainfall indicated a decrease from 105 mm to 102 mm, while mean temperature exhibited an increase from 22.4 °C to 24.2 °C and ETo from 1939 mm to 1966 mm in the WDP ecology. The SD ecology indicated an increase of about 17.6% in annual rainfall, 1.5% in mean temperature and a decrease of about 2.7% in ETo during the 1960-2019 period. Annual rainfall indicated a decrease from 162 mm to 152 mm (about –6.2%) and ETo from 2160 mm to 2121

Table 3. Climate change analysis by agro-ecology during 1960-2019 period.

Sr. No.	Zones	Rainfall (mm)			Temperature (°C)			Evapotranspiration (mm)		
		1960-1989	1990-2019	Change %	1960-1989	1990-2019	Change %	1960-1989	1990-2019	Change %
1	NDM	274	358	30.8	13.9	14.3	2.6	1077	1122	4.2
2	TM	880	914	3.9	18.3	17.3	–5.6	1182	1279	8.3
3	RP	668	872	30.6	21.6	20.7	–4.0	1496	1493	–0.2
4	WDM	271	287	5.9	20.5	21.6	5.3	1661	1662	0.1
5	NIP	353	414	17.2	24.4	24.5	0.4	1715	1675	–2.3
6	SIP	136	152	11.2	26.7	26.5	–0.6	1975	1974	0.0
7	PP	330	377	14.2	24.6	24.2	–1.7	1784	1677	–6.0
8	WDP	105	102	–3.2	22.4	24.2	8.3	1939	1966	1.4
9	SD	177	208	17.6	25.0	25.4	1.5	1900	1849	–2.7
10	CZ	162	152	–6.2	25.9	26.6	2.7	2160	2121	–1.8

mm (–1.8%), while mean temperature exhibited an increase from 25.9 °C to 26.6 °C (–2.7%) in the CZ ecology (Figure 4). Agriculture is generally vulnerable to changing climate in these areas because of receiving insufficient rains for growing crops and other vegetation [26]. Overall annual rainfall indicated an increase from 277 mm to 310 mm (about 12%) and mean temperature from 22.3 °C to 22.8 °C (2.3%), while ETo exhibited a decrease from 1730 mm to 1721 mm (–0.6%) in the country during the 1960-2019 period. Such changes in climate have resulted in increased floods frequency during the last several decades in the country [27-30].

Annual rainfall trends were observed positive at Gilgit, Jhelum, Faisalabad, D.I. Khan and Chhor, while they were negative at other stations except at Quetta where the trend was found stable (Figure 5). The monsoon rainfall (July-September) exhibited a similar pattern as that of the annual rainfall, however its contribution is much lower in the total rainfall of Gilgit, Quetta and Panjgur due to westerly's effect. The rainfall indicated predominant negative change at Kakul, Quetta, Hyderabad and Panjgur stations, while the change was positive at Gilgit, Faisalabad, D.I. Khan and Chhor stations (Figure 6). The initial positive tendency in rainfall values followed by negative values at Kakul, Jhelum and Quetta stations is likely because of decreasing monsoon intensity in these areas. Mix positive and negative deviations in rainfall were observed at Faisalabad representing NIP ecology and Hyderabad station representing SIP ecology.

3.3. Monthly Rainfall Data Analysis

The monthly rainfall at Gilgit in NDM ecology indicated positive trends for all months (Pearson correlation R values ranging within 0.01 - 0.42) except for April, May, July and October for which they were either negative or stable (Table 4). According to Gadiwala and Burke [31], the prevalence of liquid precipitation exhibits growing temperature warming in this region. The Kakul station in TM ecology exhibited negative trends for all months except for June, September, October and November for which they were positive. On the other hand, Jhelum station in RP ecology indicated prevailing rainfall pattern in most of the months (R values ranging within 0.01 - 0.18) except during May, July, November and December. Similarly, the rainfall indicated prevalence during seven months at Quetta in WDM ecology (R values ranging within 0.03 - 0.24) and seven months at Karachi in CZ ecology (R values ranging within 0.02 - 0.26). In contrast, the rainfall exhibited declining behavior in seven months at Hyderabad in SIP ecology (R values ranging from –0.16 to –0.02) and seven months at Panjgur station in WDP ecology (R values ranging from –0.34 to –0.04). There was mix pattern of monthly rainfall observed at Faisalabad station in NIP ecology and Chhor station in SD ecology. The monthly rainfall at D.I. Khan in PP ecology indicated prevalence during most of the months (R values ranging within 0.08-0.24) except during May and December (Table 4).

The monthly rainfall at Gilgit station indicated variable changes, e.g., positive for months of

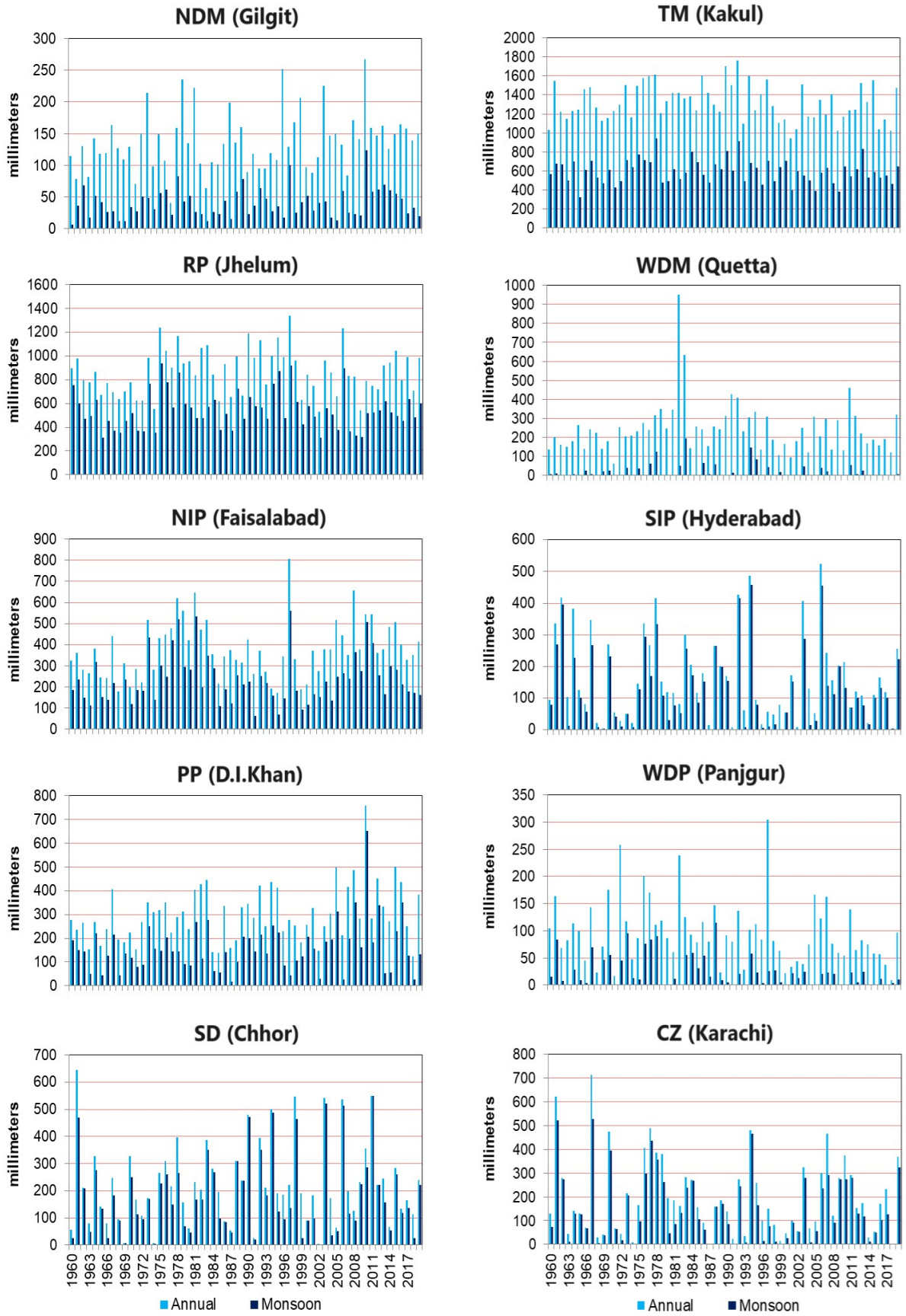

Fig. 5. Annual and monsoonal rainfall variability and trending at different climate stations of the agro-ecological zones.

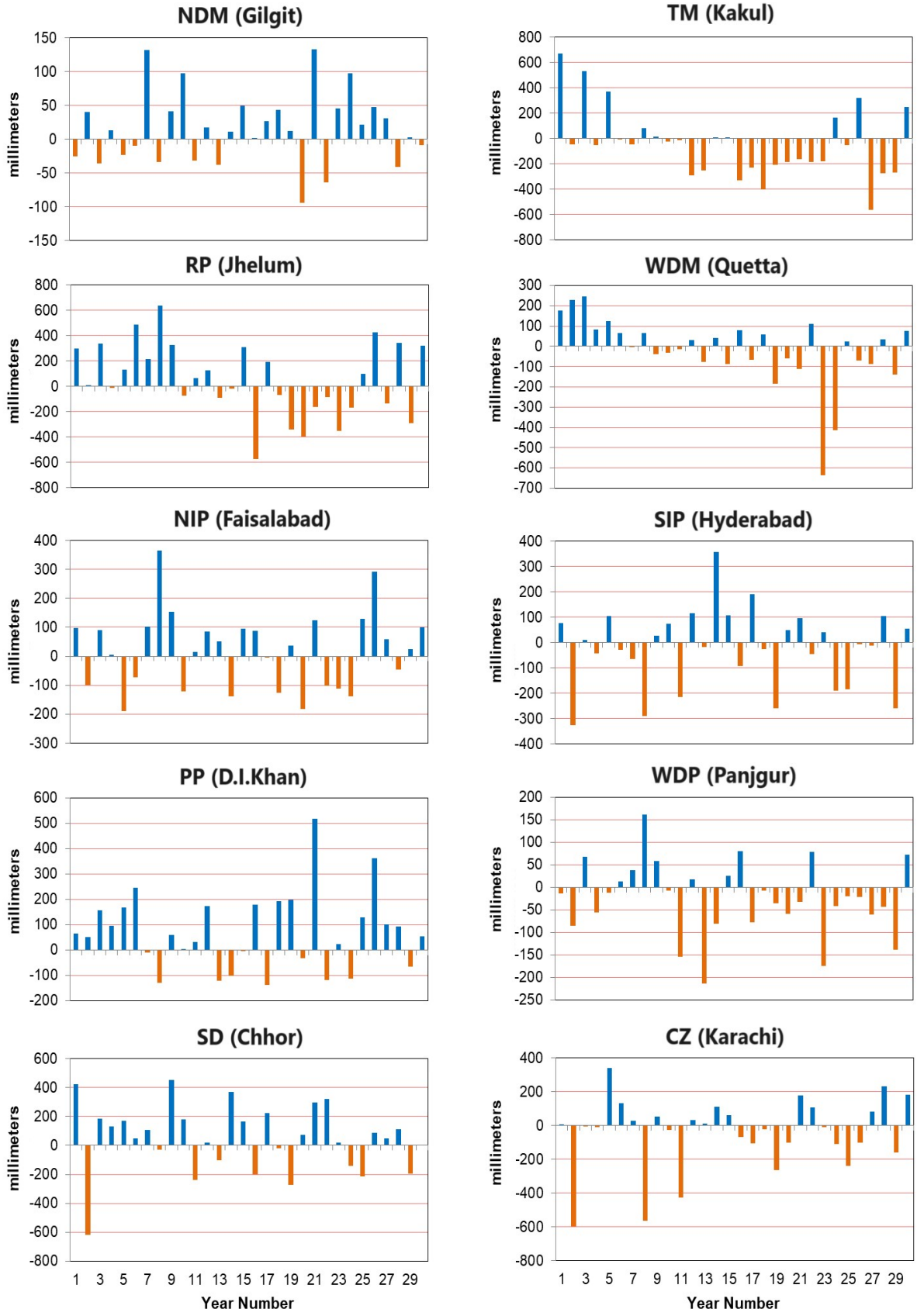

Fig. 6. Change in annual rainfall (positive change in blue and negative change in orange bars) at various climate stations of the agro-ecological zones.

February, June, September and November, and negative for May, July and October during the 1960-2019 period (Table 4). The changes in rainfall were less pronounced at Kakul and Jhelum stations, except a slight decline in peak rainfall at the former station. The peak rainfall also exhibited decline at Faisalabad, Panjgur and Karachi stations. Similarly, the monthly rainfall exhibited negative change for January, March, July and August at Quetta station during 1960-2019 period. The situation appears critical for the arid areas of the country where local communities are dependent on subsistence agriculture. In contrast, the monthly rainfall at D.I. Khan exhibited positive change which may be favorable for rainfed agriculture but risky in terms of flood occurrence. The peak rainfall at Hyderabad and Chhor stations indicated onward shifts, i.e., from July to August during 1960-2019 period (Figure 7).

3.4. Benefits of RWH in Different Ecologies

Rainwater harvesting (RWH) is a mean to reduce water demand in non-consumptive use and the negative impacts of changing rainfall pattern can be significantly reduced through proper rainwater management in the arid and semi-arid regions [32, 33]. Improved rainwater harvesting and conservation can meet future water requirements of the urban and rural areas on sustainable basis [34]. RWH system offers potential for raising high value crops, fruits and vegetables in the RP, PP, WDP and SD zones to sustain agriculture and provide resilience against high risks of climate induced hazards like droughts and floods. RWH should be aimed to collect maximum rainwater through storage, e.g. in tanks, farm ponds and mini-dams for multipurpose use [35]. At the household level, rooftop RWH system can be conveniently stored in tanks for domestic use, growing vegetables and recharging the aquifer. RWH is declared a feasible technology for the hilly areas by the World Bank and under national water policy of Pakistan [36]. The technique is also recommended under the sustainable development goals [37] to meet growing water needs of the rural and urban population. The RWH from the rooftop can be integrated with water conservation system for recharging groundwater as well as for agriculture use [38, 39].

The RWH can be used for recharging groundwater resources in the NIP, RP and WDP ecologies where depletion of groundwater is creating high risk for agriculture development [11, 40]. The culturable wasteland can be developed through utilizing hill-torrent water management techniques like establishing distribution/diversion and conservation structures. In Mithawan piedmont ecology of Punjab, the hill-torrent water generated from rainfall in several catchments had been harnessed in storage reservoirs and channelized for irrigating cultivable land. Three such earthen

Table 4. Pearson R values of rainfall trends at different climate stations during 1960-2019 period.

Period/ Station	Gilgit	Kakul	Jhelum	Quetta	Faisalabad	Hyderabad	D.I. K	Panjgur	Chhor	Karachi
Jan	0.06	–0.04	0.05	–0.07	–0.05	0.08	0.12	–0.18	0.08	0.11
Feb	0.16	–0.01	0.11	0.09	0.14	0.04	0.15	–0.04	0.08	–0.14
Mar	0.01	–0.08	0.01	0.03	0.00	–0.13	0.10	–0.05	–0.15	–0.24
Apr	0.00	–0.04	0.11	–0.06	0.10	–0.02	0.08	–0.05	–0.08	–0.08
May	–0.01	–0.18	-0.02	0.05	0.05	–0.08	-0.15	0.14	0.02	0.09
Jun	0.42	0.14	0.18	0.15	0.34	–0.16	0.30	0.08	–0.03	0.26
Jul	–0.03	–0.01	–0.07	–0.12	–0.02	–0.16	0.10	–0.34	–0.12	–0.26
Aug	0.10	–0.18	0.03	–0.04	–0.06	0.06	0.16	–0.10	0.14	0.04
Sep	0.25	0.01	0.04	0.12	0.33	–0.10	0.24	0.03	0.19	0.02
Oct	–0.04	0.01	0.17	0.04	0.23	0.10	0.19	0.05	0.10	0.09
Nov	0.32	0.06	–0.07	0.24	–0.05	–0.08	0.16	0.15	–0.10	–0.20
Dec	0.08	–0.11	–0.10	–0.08	–0.15	0.05	–0.16	–0.07	0.00	0.03
Annual	0.26	–0.13	0.08	0.00	0.22	–0.11	0.33	–0.22	0.09	–0.12
Summer	0.19	–0.10	0.05	–0.04	0.23	–0.12	0.26	–0.22	0.09	–0.10
Winter	0.18	–0.09	0.07	0.03	0.05	0.03	0.18	–0.12	0.00	–0.15

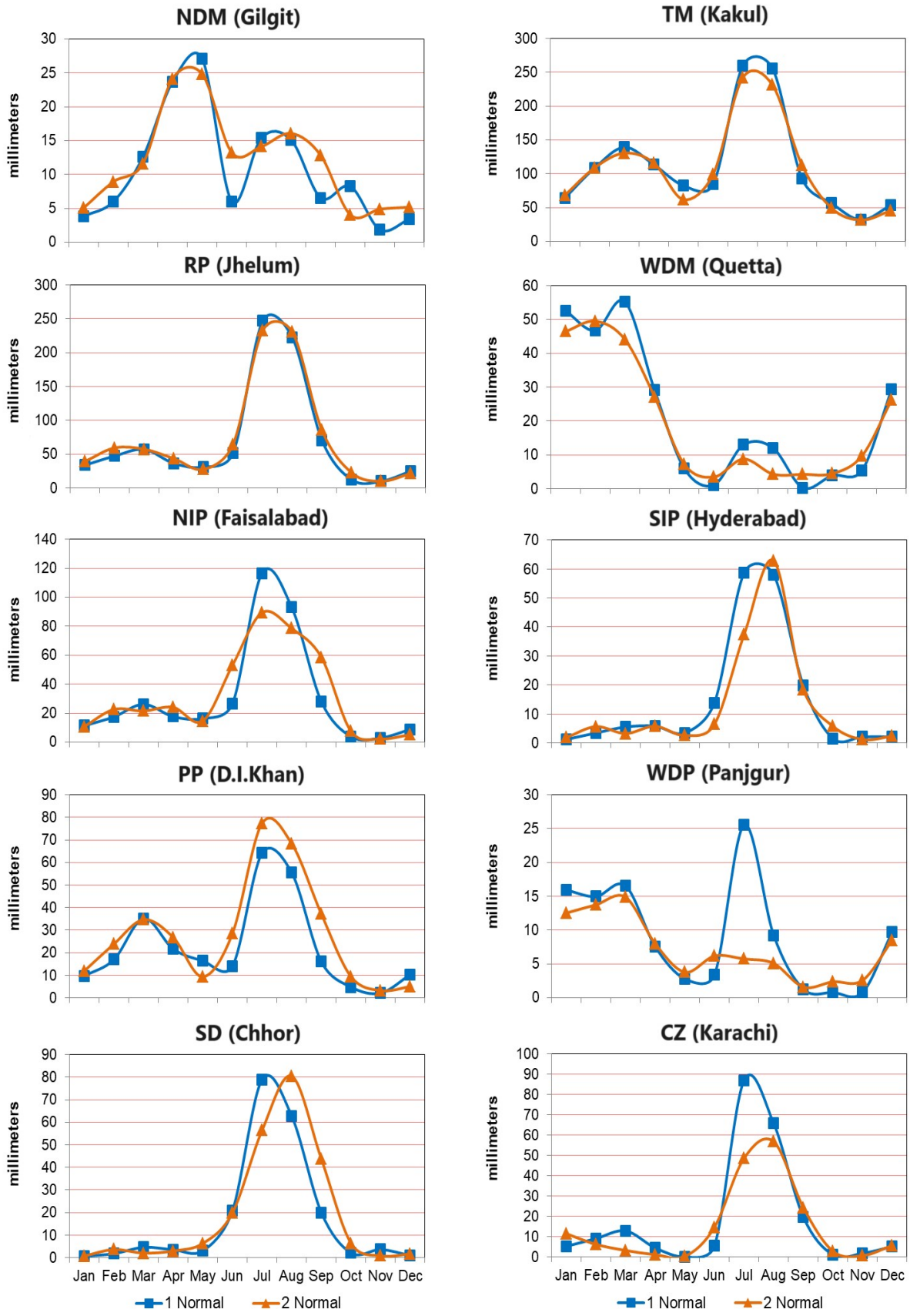

Fig. 7. Comparison between monthly rainfall of 1960-1989 and 1990-2019 climatic normals at selected stations of the ecologies.

reservoirs constructed by Pakistan Agricultural Research Council during 1995-2002 are still supplying water for irrigation use and ensure sustained water supplies during lean period of water availability besides recharging the underlying aquifer. Using the harvested water, the local communities are now growing cotton, wheat and sorghum crops coupled with farmland trees which was not possible before in that harsh mountainous terrain. At places, the stored water is used for growing high value crops through adopting high efficiency irrigation techniques. This integrated water management approach can be adopted to raise fruit orchards and woody trees to meet food, fuel and timber needs of the communities in the arid region.

4. CONCLUSIONS

In the present study, spatio-temporal changes in climate were studied in different agro-ecologies of Pakistan during 1960-2019 period for sustainable agriculture and natural resource management. Major agro-ecological zones identified were western dry mountains over 19.1% area, western dry plateau over 14.4% area, sandy desert over 14% area and northern irrigated plain over 11.3% area of the country. Significant changes in rainfall pattern have been observed in the country during the last several decades likely due to global and regional changes in climate and land use conditions. Annual rainfall indicated 30.6% increase in the rainfed plateau, 14.2% in the piedmont plains, 5.9% in the western dry mountains and about 3.9% in the temperate mountain ecology. In contrast, the rainfall exhibited 3.2% decrease in the western dry plateau and 6.2% in the coastal zone during the 1960-2019 period, which is critical for subsistence agriculture in these arid ecologies. The rainwater may be conserved through developing storage/farm ponds and reservoirs for subsequent use during dry periods and recharging groundwater to build resilience against drought/flood conditions. The rainwater harvesting (RWH) system can be adopted to fulfill growing water requirements of the communities and ensure food security in different ecologies. There is a need to raise awareness and build capacity of the local communities and institutions in efficient water use and advance water harvesting, and conservation techniques for sustainable development in the country. Regular monitoring of the water resource is essential in context of rapidly changing environment and growing needs of the population in the climate vulnerable arid region of the country in future.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. R. Missaoui, B. Abdelkarim, K. Ncibi, M. Gentilucci, S. Brahmi, Y. Ayadi, and Y. Hamed. Mapping groundwater recharge potential zones in arid region using remote sensing and GIS perspective, Central Tunisia. *Euro-Mediterranean Journal for Environmental Integration* 8: 557-571 (2023).
2. Z. Li, F. Boyle, and A. Reynolds. Rainwater harvesting and greywater treatment systems for domestic application in Ireland. *Desalination* 260: 1–8 (2010).
3. A. Markhi, N. Laftouhi, Y. Grusson, and A. Soulaimani. Assessment of potential soil erosion and sediment yield in the semi-arid N′fs basin (High Atlas, Morocco) using the SWAT model. *Acta Geophysica* 67: 263-272 (2019).
4. World Bank. World development indicators 2017. *The World Bank, Washington DC* (2017). *https://databank.worldbank.org/source/world-development-indicators*.
5. I. Ahmad, S.A.S. Hussain, and M.S. Zahid. Why the Green Revolution Was Short Run Phenomena in the Development Process of Pakistan: A Lesson for Future. *Journal of Rural Development and Administration* XXXV(1-4): 89-104 (2004).
6. A.S. Qureshi, M.A. Gill, and A. Sarwar. Sustainable groundwater management in Pakistan: Challenges and opportunities. *Irrigation and Drainage* 59(2): 107-116 (2010).
7. H. Heidari, M. Arabi, T. Warziniack, and S. Sharvelle. Effects of urban development patterns on municipal water shortage. *Frontiers in Water* 3: 694817 (2021).
8. A. Ashraf, and Z. Ahmad. Regional groundwater flow modeling of Upper Chaj Doab of Indus Basin, Pakistan using Finite element Model (Feflow) and Geoinformatics. *Geophysical Journal International* 173: 17-24 (2008).
9. A.N. Laghari, D. Vanham, and W. Rauch. The Indus basin in the framework of current and future water resource management. *Hydrology and Earth System Sciences* 16: 1063-1083 (2012).

10. A.S. Qureshi. Improving food security and livelihood resilience through groundwater management in Pakistan. *Global Advanced Research Journal of Agricultural Science* 4(10): 687-710 (2015).
11. M. Jabeen, Z. Ahmad, and A. Ashraf. Predicting behaviour of the Indus basin aquifer susceptible to degraded environment in the Punjab province, Pakistan. *Modeling Earth Systems and Environment* 6: 1633-1644 (2020).
12. A.S. Qureshi. Water management in the Indus basin in Pakistan: Challenges and opportunities. *Mountain Research and Development* 31(3): 252-260 (2011).
13. M.T. Sohail, Y. Mahfooz, K. Azam, Y. Yen, L. Genfu, and S. Fahad. Impacts of urbanization and land cover dynamics on underground water in Islamabad, Pakistan. *Desalination and Water Treatment* 159: 402-411 (2019).
14. U.H. Faraz, U.A. Naeem, H.F. Gabriel, N.M. Khan, I. Ahmad, H.U. Rehman, and M.A. Zafar. Impact of urbanization on groundwater levels in Rawalpindi City, Pakistan. *Pure and Applied Geophysics* 178: 491-500 (2021).
15. B. Helmreich, and H. Horn. Opportunities in rainwater harvesting. *Desalination* 248: 118–124 (2009).
16. T. Judeh, and I. Shahrour. Rainwater harvesting to address current and forecasted domestic water scarcity: Application to arid and semi-arid areas. *Water* 13: 3583 (2021).
17. T.M. Boers. Rainwater harvesting in arid and semi-arid zones. International Institute for Land Reclamation and Improvement, *Wageningen University, Netherland,* Publication 55 (1994).
18. K. Kim, and C. Yoo. Hydrological modeling and evaluation of rainwater harvesting facilities: case study on several rainwater harvesting facilities in Korea. *Journal of Hydrologic Engineering* 14: 545-561 (2009).
19. PARC. Agro-ecological zonation of Pakistan: Technical Report. *Pakistan Agricultural Research Council and Ministry of National Food Security and Research, Islamabad* (2024).
20. R.G. Allen, L. Pereira, D. Raes, and M. Smith. Crop Evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56*. Food and Agriculture Organization, Rome, Italy* (1998).
21. G. Rasul, and A. Mahmood. Performance Evaluation of Different Methods for Estimation of Evapotranspiration in Pakistan's Climate. *Pakistan Journal of Meteorology* 5(10): 25-36 (2009).
22. A.Q. Suleri, S. Munir, and S.Q. Shah. Impact of Trade Liberalisation on Lives and Livelihood of Mountain Communities in the Northern Areas of Pakistan. Research Report. *Sustainable Development Policy Institute, Islamabad, Pakistan* (2002).
23. M.A. Kamran, and T. Shamshad. Impacts of hill torrents' management on socio-economic conditions of arid land farmers: A case study of Tehsil DG Khan. *Asian Journal of Humanity, Art and Literature* 1(3): 155-162 (2014).
24. S. Farooq, R. Kausar, A. Ashraf, and B. Haider. Exploring surface runoff potential and water harvesting sites in Dera Ismail Khan Rod-kohi area using GIS and Remote sensing. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences* 57(2): 1-10 (2020).
25. T.F. Ahmed, S.U.S. Shah, A.A. Sheikh, H.N. Hashmi, M.A. Khan, and M. Afzal. Flood water management from hill torrents of Pakistan for agriculture livelihood improvement. *Pakistan Journal of Agricultural Research* 34(3): 438-445 (2021).
26. S. Adnan, K. Ullah, S. Gao, A.H. Khosa, and Z. Wang. Shifting of Agro-climatic zones, their drought vulnerability, and precipitation and temperature trends in Pakistan. *International Journal of Climatology* 37: 529-543 (2017).
27. M.T. Arslan. Unusual rainfall shift during monsoon period of 2010 in Pakistan: Flash flooding in Northern Pakistan and riverine flooding in Southern Pakistan. *African Journal of Environmental Science and Technology* 7(9): 822-890 (2013).
28. P. Ganguli, and M.J. Reddy. Evaluation of trends and multivariate frequency analysis of droughts in three meteorological subdivisions of western India. *International Journal of Climatology* 34: 911-928 (2014).
29. MOCC. Framework for implementation of climate change policy. *Ministry of Climate Change, Islamabad, Pakistan* (2014).
30. M.A. Miyan. Droughts in Asian least developed countries: vulnerability and sustainability. *Weather and Climate Extremes* 7: 8-23 (2015).
31. M.S. Gadiwala, and F. Burke. Climate change and precipitation in Pakistan-A meteorological prospect. *International Journal of Economic and Environmental Geology* 4(2): 10-15 (2013).
32. M. Bakir, and Z. Xingnan. GIS and remote sensing applications for rainwater harvesting in the Syrian Desert (Al-Badia). *Twelfth International Water Technology Conference*, *IWTC12 2008 Alexandria, Egypt* (2008).
33. M. Shahid, K.U. Rahman, K.S. Balkhair, and A.

Nabi. Impact assessment of land use and climate changes on the variation of runoff in Margalla Hills watersheds, Pakistan. *Arabian Journal of Geosciences* 13: 1-14 (2020).

34. R. Shabbir, and S.S. Ahmad. Water resource vulnerability assessment in Rawalpindi and Islamabad, Pakistan using Analytic Hierarchy Process (AHP). *Journal of King Saud University-Science* 28: 293-299 (2016).
35. F. van Steenbergen. Groundwater and surface water in the mega-irrigation systems of Pakistan: The Case for conjunctive management. *Water Knowledge Note, World Bank Group* (2019).
36. MOWR. National Water Policy. *Ministry of Water Resources, Government of Pakistan* (2018).
37. C.C. Amos, A. Rahman, J.M. Gathenya, E. Friedler, F. Karim, and A. Renzaho. Roof-harvested rainwater use in household agriculture: Contributions to the sustainable development goals. *Water* 12: 332 (2020).
38. K. Kalhor, and N. Emaminejad. Sustainable development in cities: Studying the relationship between groundwater level and urbanization using remote sensing data. *Groundwater for Sustainable Development* 9: 100243 (2019).
39. D. Zhang, X. Ding, J. Liu, and C. Mei. Review on mechanism and technical measures of urban rainwater harvesting. *IOP Conference Series: Earth and Environmental Science* 983: 012106 (2022).
40. F. van Steenbergen, A.B. Kaisarani, N.U. Khan, and M.S. Gohar. A case of groundwater depletion in Balochistan, Pakistan: Enter into the void. *Journal of Hydrology: Regional Studies* 4: 36-47 (2015).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 181-191 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)832

Pakistan Academy of Sciences

Research Article

Design and Development of Fractional Order Convolutional Neural Network Based Fractional Order Nonlinear Reactor Power Simulator for CANDU-PHWR

Arshad Habib Malik[1*], Feroza Arshad[2], and Aftab Ahmad Memon[3]

[1]Faculty of Engineering, Information and Technology,
Sindh Institute of Management and Technology, Karachi, Pakistan

[2]Department of Information System Division, Karachi Nuclear Power Generating Station, Pakistan Atomic Energy Commission, Karachi, Pakistan

[3]Faculty of Electrical, Electronic and Computer Engineering,
Mehran University of Engineering and Technology, Jamshoro, Sindh, Pakistan

Abstract: A highly complex nonlinear Reactor Regulating System (RRS) of Canadian Deuterium Uranium Pressurized Heavy Water Reactor (CANDU-PHWR) based Nuclear Power Plant (NPP) simulated in the present research. The internal design of RRS is secured and vendor controlled which is embedded in AC-132 Programmable Logic Controller (PLC). Therefore, the problem of the identification of the RRS controller model is addressed. A data-driven Fractional Order Nonlinear MIMO Hammerstein Model (FO-NC-MIMO-HM) of NPP is identified using an Adaptive Immune Algorithm (AIA) based on a Global Search Strategy (GSS) and Auxiliary Model Recursive Least Square Method (AMRLSM). Parameters of FO-MIMO-HM are identified using Innovative Real-Time Plant Operational Data (IRTPOD). The original PLC-based controller is replaced with a new Fractional Order Convolutional Neural Network (FO-CNN) based Fractional Order Nonlinear Controller (FO-NC). Therefore, a visual Simulator is developed for detailed modeling, control, simulation, and analysis of the proposed design scheme for RRS in Visual Basic (VB) Software. The performance of the proposed design scheme is tested and validated for different modes of RRS against benchmark data obtained from Plant Data Recorder (PDR) and found in close agreement well within the design bounds.

Keywords: Fractional Order, Convolutional Neural Network, Nonlinear Control, MIMO System, Visual Basic, Simulator, Reactor Regulating System, CANDU-PHWR.

1. INTRODUCTION

The reactor regulating system is a discrete logic base multivariable reactor power controller. This controller uses various inputs from primary and secondary side of CNADU-PHWR NPP. An auxiliary least squares identification method for Hammerstein model is discussed by Feng *et al.* [1]. ARX based LS algorithm is developed for multivariate output system by Qinyao *et al.* [2]. The concept of adaptive immune genetic algorithm is introduced for multivariable system using global optimization by Dai *et al.* [3]. Later an adaptive immune inspired multi-objective algorithm is designed using multiple DE strategies by Qiuzhen *et al.* [4]. An attempt is made using subspace identification for FO Hammerstein systems by Zeng *et al.* [5]. An iterative method is devised for online FO and parameter identification of system by Oliver *et al.* [6]. A parameter identification method for FO Hammerstein systems is formulated by Zhao *et al.* [7]. A FO Hammerstein state space model is developed for nonlinear dynamic system using input-output measurement data by Rahmani *et al.* [8]. A system identification method is suggested for hybrid fractional order Hammerstein-Wiener model in continuous time domain by Allafi *et al.* [9]. A recursive identification approach

Received: July 2023; Revised: April 2024; Accepted: June 2024
* Corresponding Author: Arshad Habib Malik <engr.dr.arshad.habib.malik@gmail.com>

is chosen for FO Hammerstein model based on Adaptive Differential Evolution with the Local Search strategy by Jin *et al.* [10]. Later, a real time semantic segmentation network is introduced using the concept of Proportional Integral Derivative (PID) controller algorithm by Xu *et al.* [11]. A computer vision control based Convolutional Neural Network- Proportional Integral Derivative (CNN-PID) scheme is devised for mobile robot by Farkh *et al.* [12]. Research is further explored in which a PID controller approach is adopted for stochastic optimization of DNN based systems by An *et al.* [13]. Further research is addressed for CNN back propagation neural network optimized by a fractional order gradient method by Taresh *et al.* [14]. An investigation is performed for the realization and comparative analysis of fractional order controllers using different discretization techniques by Calderon [15]. A convolutional neural network based on FO order momentum term is trained and tuned by Kan *et al.* [16]. A fractional order integral sliding mode controller is designed for Pressurized Water Reactor (PWR) type NPP by Surjagade *et al.* [17].

In this research work, a new fractional order convolutional neural work based fractional order nonlinear sliding mode reactor power controller is designed for a nonlinear MIMO model of CANDU-PHWR NPP. The suggested methodology is the new algorithm which is synthesized for the first time in Visual Basic for the NPP with special emphasis on CANDU-PHWR type nuclear power with state-of-the-art Graphical User Interface (GUI) for modeling and controller design. The proposed scheme provides optimal solution with fast convergence and stability. The FO gradient method is chosen to enhance the dynamic significance of neural parameters for CNN. The fractional order nonlinear sliding mode controller is used for fast convergence and stability and is much better than conventional sliding mode controller because it offers minimum or smaller steady state error and the control law is more flexible due to the variable order of the fractional term with a great advantage of scalability of the fractional order to the sliding mode surface. The proposed design is a novel design with a coupled FO convolutional neural network and FO-SMC for a fractional order multivariable nonlinear model of CANDU-PHWR reactor regulating system optimized by AIA based on a GSS and AMRLSM.

2. MATERIALS AND METHODS

2.1. Reactor Regulating System

The reactor regulating system (RRS) uses three type reactivity mechanisms. One is fine reactivity insertion mechanism called moderator level variation method while two are coarse reactivity insertion mechanism called absorber rod method and booster rod method. Both coarse reactivity insertion method operates in a specific moderator level band. These all methods in turn are responsible for reactor power level change from 0% to 100%.

The reactor regulating system is designed on five control modes such as:

1. SD mode
2. ML mode
3. LL mode
4. NP mode
5. SP mode

These modes work in parallel as group of overrides. Thus, a mode controlling low power is always backed up by the modes ready to control high power levels. The automatic mode reductions provided in the event of loss of permissive for the activated mode. In addition, three overrides are present in all modes.

1. The high rate of log power override also called the rate log N override.
2. The Neutron power override also called the high linear N override.
3. The high moderator level override.

The reactor regulating system is made by CEGLEC ACEC Company of Belgium. This is a 186-machine consisting of bus driver, memory expander, Ethernet (LN1, LN2, and LN3) each of 256 Kbps. Its operating system is RNM-86 based on COGITO configuration.

2.2. CANDU-PHWR Modeling

The reactor regulating system is a multivariable CANDU power control system. It is the most important, critical and most complex control system of CANDU-PHWR type NPP. It is coupled with many other important systems of nuclear power plant. Various parameters/symbols and variables used hereafter in the process

of RRS modeling are defined in Annexure I.

α_i are variable FO of differential $D_t^{\alpha_i}$ and integral $I_t^{\alpha_i}$ operators used for the model identification of the reactor regulating system. In system identification, the first step is to use AIA based on a GSS called AIAGS for initialization. Once the coefficients are calculated then second step is to use these model coefficients to identify the actual or true ones of FO-NC-MIMO-HM using AMRLSM. The AIAGS algorithm which is an intelligent optimization algorithm consists of stimulation function, mutation function and simulated annealing function [4].

$$f(x_i) = min\, f(y_k(t), \alpha, m, M, \Omega, \delta, \beta, \gamma) \quad (1)$$

Subject to the linear constraints:

$$g(y(t)^k) \geq 0$$
$$y(t)^k \leq y(t)^{k,max}$$

Where x_i, is i^{th} individual of population and $y(t)$ is a variable vector, $f(y(t)$ is the objective function or cost function, $y(t)^{k,\max}$ is the upper bound of variable $y(t)^k$ and $g(y(t)^k)$ is the inequality constraint of output variable in an intelligent optimization problem.

Fractional order HM is a nonlinear fractional order model consisted of cascaded fractional order static nonlinear model and fractional order dynamic linear model. The fractional order dynamic linear model can be defined as:

$$y_k(t) = G_{k,1}u_1^{NL} + G_{k,2}u_2^{NL} + \ldots + G_{k,p}u_p^{NL} \quad (2)$$

The static nonlinear model can be defined as [8]:

$$u_p^{NL} = a_{p,1}f_{p,1}(u_p(t)) + a_{p,2}f_{p,2}(u_p(t)) + \ldots\ldots + a_{p,p}f_{p,p}(u_p(t)) \quad (3)$$

The fractional order Hammerstein model can be modeled as:

$$y_k(t) = G_{k,1}^{'}u_1 + G_{k,2}^{'}u_2 + \ldots + G_{k,p}^{'}u_p \quad (4)$$

The dynamic linear model with Gaussian White Noise can be defined as:

$$y_k^{GWN}(t) = y_k(t) + \varepsilon(t) \quad (5)$$

Now, with Gaussian White Noise, the auxiliary model is computed as:

$$y_k^{AM}(t) = G_{k,1}^{AM}u_1^{AMNL} + G_{k,2}^{AM}u_2^{AMNL} + \ldots + G_{k,p}^{AM}u_p^{AMNL} \quad (6)$$

The static nonlinear model can be modified as:

$$u_p^{AMNL} = a_{p,1}^{AM}f_{p,1}^{AM}(u_p^{AM}(t)) + a_{p,2}^{AM}f_{p,2}^{AM}(u_p^{AM}(t)) + \ldots\ldots + a_{p,p}^{AM}f_{p,p}^{AM}(u_p^{AM}(t)) \quad (7)$$

The fractional order Auxiliary Hammerstein model can be modeled as:

$$y_k(t) = G_{k,1}^{'AM}u_1^{AM} + G_{k,2}^{'AM}u_2^{AM} + \ldots + G_{k,p}^{'AM}u_p^{AM} \quad (8)$$

Now, by using the Recursive Least Square Method, the value of model parameter vector θ_k and the value of auxiliary model parameter vector θ_k^{AM} are estimated using input-output data and defining a cost function in terms of θ. Then, using the accurate values of coefficients, the value of fractional order of auxiliary model α is estimated by minimizing another cost function defined in terms of α [2]. The accuracy of the estimated model is computed using Mean Square Error (MSE).

2.3. Fractional Order Convolutional Neural Network Modeling

The fractional order deep convolutional neural network (FO-CNN) is a four-layer network. One is input layer, two convolution layers and one output layer. The convolution layer is an important part of FO-CNN and its main function is to extract dynamic modes of transient behavior. The two convolution layers have been chosen depending on the complexity of problem, size of dataset and specific architecture being used. However, there is no fixed number of layers that is optimal for all tasks. The output of FO-CNN is given as [14]:

$$\varepsilon = f_0(\varepsilon_2) = w_0\varepsilon_2 + b_0 \quad (9)$$

The weights and biases are optimized using fractional order gradient method [16].

2.4. Fractional Order Nonlinear Controller Modeling

Now, FO-SMC is modeled as fractional order

nonlinear controller. The novel fractional order siding surface can be designed for fractional order nonlinear system as:

$$S = h_0 D_t^{1-\alpha} e + h_0 D_t^{-\alpha} H e^{\frac{c}{d}} \quad (10)$$

Where,

$$H = diag(h_1, h_2, \ldots\ldots, h_n) \quad (11)$$

2.5. Coupling Between FO-CNN and FO-NC

Now, the weights and biases described in equation (9) are updated as:

$$w_{NEW} = w - \frac{1}{\eta_w} h_0 D_0^t w \quad (12)$$

$$b_{NEW} = b - \frac{1}{\eta_b} h_0 D_0^t b \quad (13)$$

Now, the weights and biases described in equation (9) are calculated as:

$$h_0 D_t^{\alpha} w_0^{(i,q)} = S_i \varepsilon_2^q \quad (14)$$

$$b_0^i = h_0 I_t^{\alpha} S_i \quad (15)$$

The modeling error vector given as:

$$\Delta = [\Delta_1 \Delta_2 \Delta_3 \ldots\ldots \Delta_n]^T \quad (16)$$

Where Δ_n is the modeling error in each output variable affected by noise which is implemented using the concept of Gaussian White Noise. In the subsequent derivation, the dealing of modeling error in terms model uncertainties are described in detail.

The output of FO-CNN without any Gaussian White Noise is the ideal behavior which is given as:

$$\varepsilon^{Ideal} = w_0^{Ideal} \varepsilon_2 + b_0^{Ideal} \quad (17)$$

The uncertainty is addressed in terms of noise in the modeling which is responsible for uncertain dynamics in plant operation and control. Therefore, the lumped noisy parameters can be modeled as:

$$\varepsilon^{Lumped} = w_0^{Ideal} \varepsilon_2 + b_0^{Ideal} + \Delta \quad (18)$$

The estimated noise error function is given as:

$$\hat{\varepsilon} = \varepsilon^{Lumped} - \varepsilon$$

$$\hat{\varepsilon} = (w_0^{Ideal} - w_0)\varepsilon_2 + (b_0^{Ideal} - b_0) + \Delta \quad (19)$$

The error weights and biases matrices are given as:

$$\hat{w}_0 = w_0^{Ideal} - w_0 \quad (20)$$

$$\hat{b}_0 = b_0^{Ideal} - b_0 \quad (21)$$

Now, equation (19) becomes:

$$\hat{\varepsilon} = \hat{w}_0 \varepsilon_2 + \hat{b}_0 + \Delta \quad (22)$$

The Lyapunov function can be defined as:

$$V = \frac{1}{2} S^T S + \frac{1}{2} \eta_w \sum_{i,q} (\hat{w}_0^{(i,q)})^2 + \frac{1}{2} \eta_b \hat{b}_0^{\ T} \hat{b}_0 \quad (23)$$

The control law can be deduced from the following inequality as:

$$h_0 D_t^{\alpha} V \le S^T \left(\dot{e} + H e^{\frac{c}{d}} \right) + \eta_w \sum_{i,q} (\hat{w}_0^{(i,q)}) h_0 D_t^{\alpha} (\hat{w}_0^{(i,q)}) + \eta_b \hat{b}_0^T h_0 D_t^{\alpha} \hat{b}_0 \quad (24)$$

The error dynamics of FO-NC is given as:

$$e = y_k^{AM} - y_k^{Reference}, \dot{e} = \dot{y}_k^{AM} - \dot{y}_k^{Reference} \quad (25)$$

On modifying equation (24) based on equation (22) and hence on simplifying using equation (25) with an assumption that Caputo fractional derivative for a constant is zero, the desired FO-CNN based FO-NC control law is given as:

$$u_{CNN}(t) = \left(G_{k,p}^{AM} \right)^{-1} \left[\left(\dot{y}_k^{Reference} - H e^{\frac{c}{d}} \right) - \hat{\varepsilon} + \left(\eta S + K_{SMC} \frac{S}{\|S\|} \right) \right] \quad (26)$$

2.6. Controller Simulator Development

The controller simulator is developed in Visual Basic (VB) 6.0. VB supports well for event driven programming. In GUI various active X controls are designed dedicated to specific functions. For single and multiple graphs, tee chart active X control is used. As power Regulating system run on different

sample time intervals, so timer control is used to show real time simulation of the transients scenarios. But for long time transients, fast simulation options are available with multipliers of 2, 5, 10, and 25.

3. RESULTS AND DISCUSSION

The entire synthesis of reactor regulating system is carried out in closed loop in the subsequent sections.

3.1. Design Analysis of Proposed Configuration and Simulator Development

The proposed designed scheme of FO-CNN-FO-NC based reactor regulating system is shown in Figure 1. The proposed framework is composed of CANDU-PHWR type nuclear power plant, structure of existing reactor regulating system, proposed modeling and controller synthesis scheme in testing and evaluation phases. The testing loop incorporates the estimated model and proposed controller while the validation loop incorporates the actual plant and the proposed controller. There are five set points for five different operational and control modes of nuclear power plant for RRS. Each control mode represents a sub-controller in RRS design. It takes several inputs from the plant and generates various control and output signals which are defined in detail in section 2.2. As reactor power increases from 0% to any desired power level or maximum up to 100%, the modes are configured sequentially, and signals are generated accordingly as the interlocks are met.

The proposed controller design scheme is shown in Figure 2 for neutron power mode as example for better understanding of problem formulation. Error and error rate signals are specially design to cater the requirements of proposed control design scheme.

The GUI for controller and simulation configuration is shown in Figure 3. Different GUI features such as mode selection, analysis selection, mode comparison, transient analysis, transient initialization schemes, online parameter adjustments and simulator administrative controls are provided in Figure 3. The transfer functions are solved using discrete factional order Tustin Approximation [15] as shown in Figure 4. Recursive solver uses previous value of input, output variables and previous value of output variable with sample time.

There are three different type of reactivity insertion mechanism. Booster rods method is one of the coarse power transient mechanisms as shown in Figure 5. More the booster rods inserted in the reactor core more will be power generated as introduces positive reactivity.

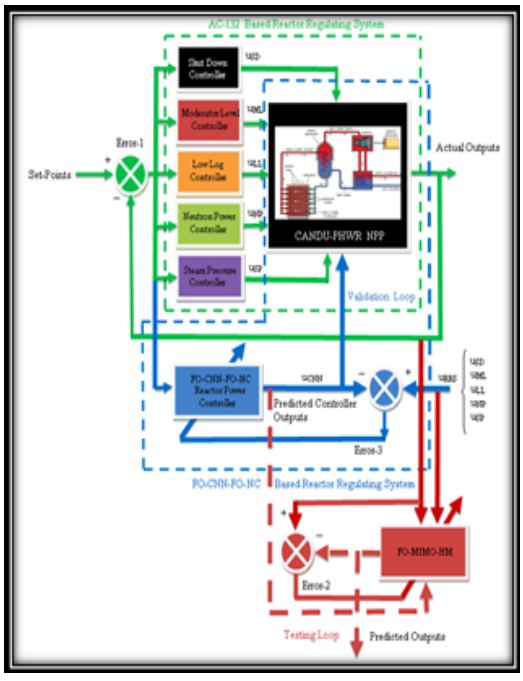

Fig. 1. Proposed configuration of FO-CNN-FO-NC based reactor regulating system.

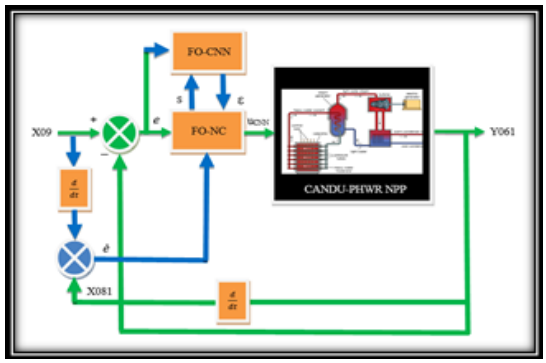

Fig. 2. Design configuration of FO-CNN-FO-NC in neutron power mode.

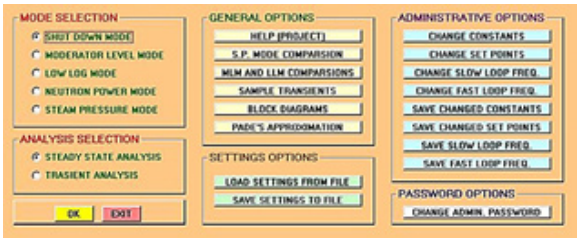

Fig. 3. GUI for controller and simulation selections.

Absorber rods method is another coarse power transient mechanism as shown in Figure 6. More the absorber rods inserted in the reactor core lesser will be power generated as introduces negative reactivity.

The GUI for the selection of transient is shown in Figure 7. Three transient mechanisms are provided for the insertion or withdrawal of reactivity depending on the type and direction of motion.

The GUI for the selection and design of standard and patent transients is shown in Figure 8. Reactivity transient selection menu is one of the unique features of the simulator. Once the type reactivity mechanism is chosen, the detailed menu will pop up for parametric adjustment and transient design.

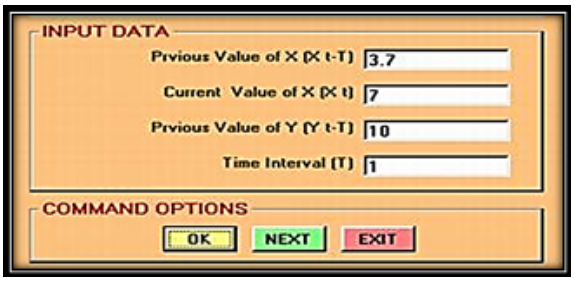

Fig. 4. GUI for discrete solvers.

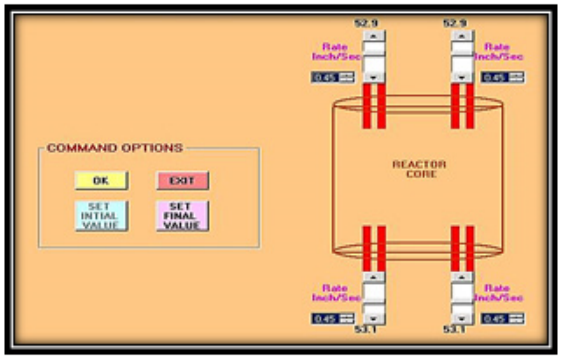

Fig. 5. GUI for booster rod transients.

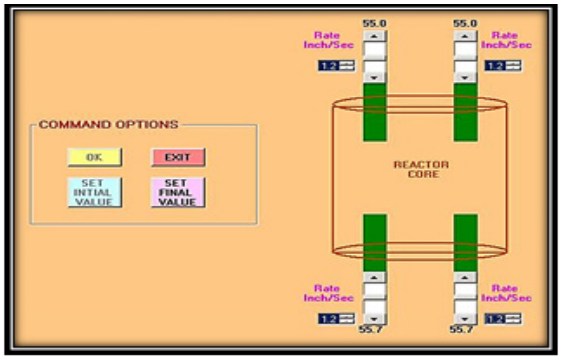

Fig. 6. GUI for absorber rod transients.

3.2. Evaluation of Proposed Controller Design for Moderator Level Controller

The proposed closed loop system is initialized using functions programmed using equation (1). The initialization of moderator level controller in reactor regulating system is shown in Figure 9.

The dynamics of existing and proposed RRS in the simulation framework is designated as Plant Data Recorder (PDR) and Simulator respectively and hence act like legends for entire simulator development and display system. Initialization is performed at 35 inches of moderator level and various parameters of interest are observed against it. New controller is initialized in moderator level mode as shown in Figure 9.

Now, the transient simulation of moderator level controller is shown in Figure 10. A special plant operational transient is configured in moderator level mode and various parameters of interest are visualized. There is a high level of fluctuations on parameters X04 and X05 but final control demand signal Y011 and X02 are tracking excellently. This proves the successful realization of proposed

Fig. 7. GUI for reactivity transient selection of moderator level mode in low power operation.

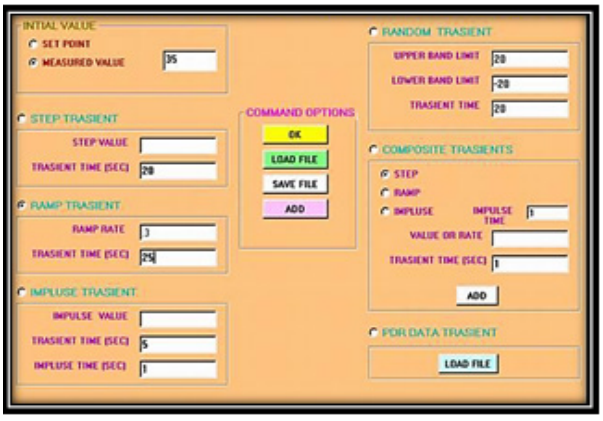

Fig. 8. GUI for selection and design of transients for moderator level controller.

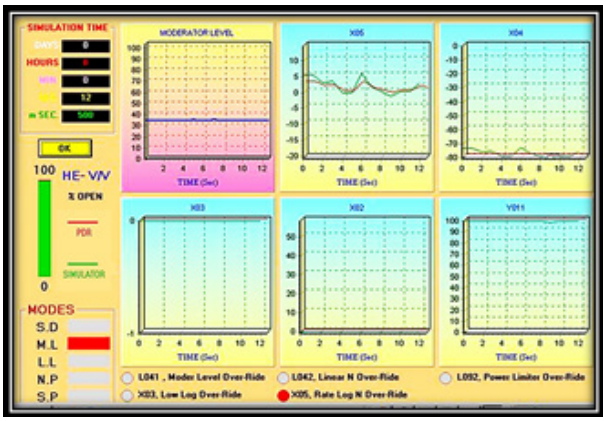

Fig. 9. Initialization of moderator level controllers.

moderator level controller and optimal design performance of parameters.

3.3. Evaluation of Proposed Controller Design for Neutron Power Controller

Now, the dialed reactor power is dropped from 70% to 30% via step change as shown in Figure 11. In this simulation scenario, a transient is configured in neutron power mode and various parameters of interest are visualized. In neutron power mode, under step down power transient, the dynamics of X08, Y061, Y083, Y112 and X12 with proposed scheme follows the benchmark transient data. This dynamic behavior proves the excellent trackability of proposed controller. This optimal performance is achieved due to a transient followed by step change in dialed reactor power. In this transient, the requirements or constraint imposed on the controller is that power ramp down for X08, Y061 and X12 must not exceed -0.25 % RP/sec. This proves the successful realization of proposed neutron power controller and design parameters are optimal in performance in high power operation of the plant.

3.4. Evaluation of Proposed Controller Design for Steam Pressure Controller

The front panel design for steam pressure mode is shown in Figure 12. The front panel design for steam pressure mode is meant for steam pressure, steam flow, primary system temperatures and primary system flows on representative channels.

A special plant operational transient is configured in steam pressure mode as shown in Figure 13 and various parameters of interest are observed. There is a very smooth tracking and very close performance of X08, Y061, X07, Y112 and X12 is achieved. This proves the successful realization of proposed steam pressure controller and optimal design performance of parameters.

The zoomed simulation comparison of X04 and X05 are shown in Figures 14 and 15 respectively.

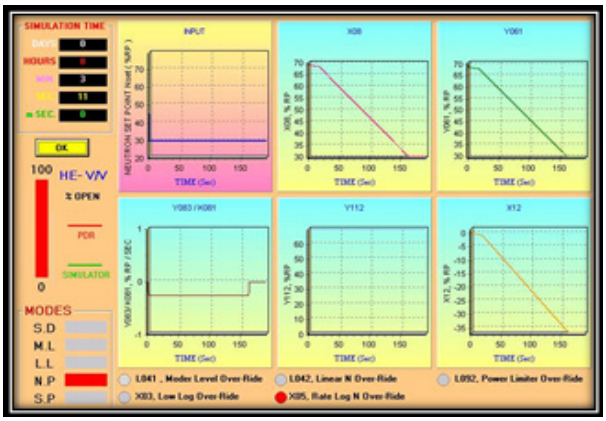

Fig. 11. Simulation of multiple parameters for neutron power controller.

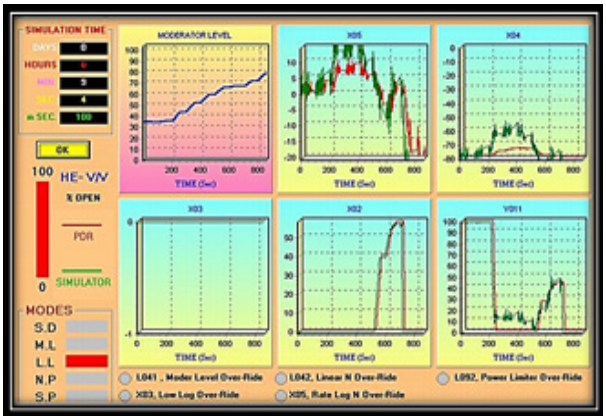

Fig. 10. Transient simulation of multiple parameters for moderator level controller.

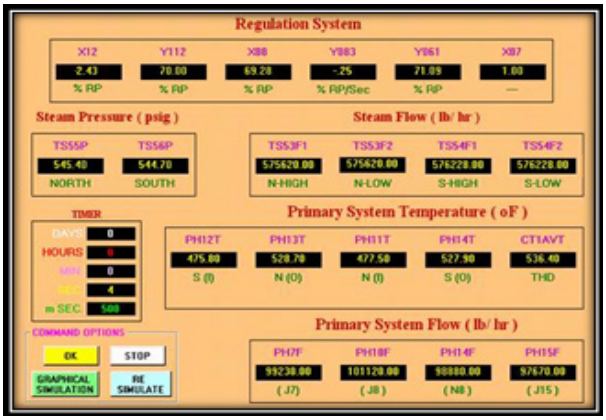

Fig. 12. Front panel design for steam pressure mode.

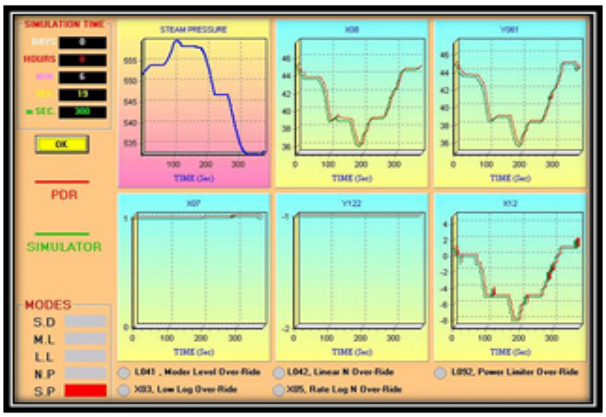

Fig. 13. Transient simulation of multiple parameters for steam pressure controller.

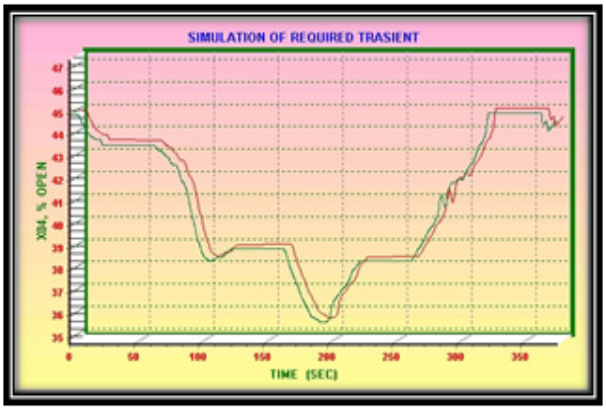

Fig. 14. Simulation of X04 in steam pressure mode.

The comparison is assessed based on the degree of relative errors in distribution of performance and found 0.43% and 0.42% respectively. The optimized design parameters of model and controllers' framework are obtained using proposed design modeled in equations (1) through (25). The optimized parameters are tabulated in Table 1.

4. CONCLUSIONS

The reactor regulating system of CANDU-PHWR is composite complex power controller of nuclear power generating station. A data driven nonlinear MIMO model of NPP is developed with desired parameters of interest. Model parameters are optimized using adaptive algorithms. Constrained fractional order CNN based FO-SMC is synthesized for RRS. Controller parameters are optimized for five modes of plant operation. Synthesized controller is evaluated and compared with permissives, interlocks, compensators and conventional controllers oriented RRS. A simulator is developed with visual environment

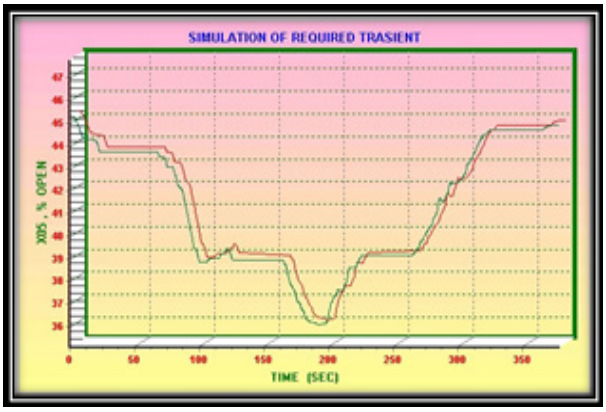

Fig. 15. Simulation of X05 in steam pressure mode.

Table 1. Optimized parameters of model and controllers.

Parameters	Values	Parameters	Values
Ω	1	A_2	20
δ	0.11	C	3
γ	0.21	D	7
β	0.64	A	0.333
MSE	0.012	η_w	14
P	2	η_b	16
A_1	10	K_{SMC}	7.7

for parametric analysis and simulation. Simulator is developed with advanced features of zooming, steady state analysis, transient analysis, customized transient design, fast simulation, multiple plots and online controller parameters adjustment. Therefore, the proposed control design scheme exhibits much robust performance against the model uncertainties and online controller parameters adjustments. The effectiveness and robustness of the proposed manifolds are proved by different configured modes of control to verify their efficacy. This simulator is a step towards other plant systems, controllers and new configuration schemes for advanced research and development in nuclear industry.

5. ACKNOWLEDGEMENTS

The support of Sindh Institute of Management and Technology, PAEC and Mehran University of Engineering and Technology is gratefully acknowledged.

6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

1. D. Feng, S. Yang, and C. Tongwen. Auxiliary least squares identification methods for Hammerstein output error systems. *Systems and Control Letters* 56: 373-380 (2007).
2. L. Qinyao, and D. Feng. Auxiliary model based recursive generalized least squares algorithm for multivariate output error autoregressive systems using the data filtering. *Circuits Systems and Signal Processing* 38: 590-610 (2019).
3. Y. Dai, Y. Li, L. Wei, J. Wang, and D. Deling. Adaptive immune genetic algorithm for global optimization to multivariable function. *Journal of Systems Engineering and Electronics* 18(03): 655-660 (2007).
4. L. Qiuzhen, M. Yueping, C. Jianyong, Z. Qingling, C. A. Carlo, W. Ka-Chun, and C. Fei. An adaptive immune inspired multi-objective algorithm with multiple differential evolution strategies. *Information Sciences* 431: 46-64 (2018).
5. L. Zeng, Z. Zhuting, L. Shu, P. Cheng, and W. Yong. Subspace identification for fractional order Hammerstein systems based on instrumental variables. *International Journal of Control, Automation and Systems* 10(5): 947-953 (2012).
6. S. Oliver, K. Philipp, and H. Soren. Iterative method for online fractional order and parameter identification. *IEEE Conference on Decision and Control, Republic of Korea* (14-18 December 2020) pp. 5159-5166 (2020).
7. Y. Zhao, Y. Li, and Y.Q. Chen. Complete parameter identification of fractional order Hammerstein systems. *ICFDA'14 International Conference on Fractional Differentiation and Its Application, Catania, Italy* (23-25 June 2014) pp. 1-6 (2014).
8. M.R. Rahmani, and M. Farrokhi. Fractional order Hammerstein state space modeling of nonlinear dynamic systems from input-output measurements. *ISA Transactions* 96: 177-184 (2020).
9. W. Allafi, I. Zajic, K. Uddin, and K.J. Burnham. Parameter estimation of the fractional order Hammerstein-Wiener model using simplified refined instrumental variable fractional order continuous time. *IET Control Theory and Applications* 11(15): 2591-2598 (2017).
10. Q. Jin, Y. Ye, W. Cai, and Z. Wang. Recursive identification for fractional order Hammerstein model based on ADELS. *Mathematical Problems in Engineering* 2021: 6629820 (2021).
11. J.Xu, Z. Xiong, and S.P. Bhattacharyya. Real time semantic segmentation network inspired by PID controllers. *Computer Vision and Pattern Recognition, Cornell University, New York* (4th June 2022) pp. 1-11 (2022).
12. R. Farkh, M.T. Quasim, K.A. Jaloud, S. Alhuwaimel, and S.T. Siqqiqui. Computer vision-control-based CNN-PID for mobile robot. *Computers, Materials and Continua* 68(1): 1065-1079 (2021).
13. W. An, H. Wang, Q. Sun, J. Xu, Q. Dai, and L. Zhang. A PID controller approach for stochastic optimization of deep networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA* (18-23 June 2018) pp. 1-6 (2018).
14. M.M. Taresh, N. Zhu, T.A. Ali, and W. Guo. Using a novel fractional order gradient method for CNN back propagation. *Computer Vision and Pattern Recognition, Cornell University, New York* (1st May 2022) pp. 1-10 (2022).
15. C.A. Calderon. Realization and comparative analysis of fractional order controllers for different discretization methods. *IEEE International Conference on Automaton, Curicó, Chile* (24-28 October 2022) pp. 5159-5166 (2022).
16. T. Kan, Z. Gao, C. Yang and J. Jian. Convolutional neural networks based on fractional order momentum for parameter training. *Neurocomputing* 449: 85-99 (2021).
17. P.V. Surjagade, J. Deng, V. Vajpaeyee, V.M. Becerra, S.R. Shimjith, and A.J. Arul. Fractional order integral sliding mode control for PWR nuclear power plant. *European Control Conference, London, United Kingdom* (12-15 July 2022) pp. 987-992 (2022).

Annexure I. Symbols/parameters of RRS.

Parameters	Definitions
X02	Moderator Level Controller Output
X03	Low Log Controller Output
X04	Dynamic Power Mismatch Compensator Output
X05	Transient Compensator Output corresponding to Reactivity Effects
X07	Compensator Power Correction Factor corresponding to Moderator Level and Power Mismatch of Thermal and Actual Reactor Power
X08	Turbine Power
NSETD /X09	Demanded Reactor Power
X081	Rate of Change of Reactor Power
X12	Reactor Power corresponding to Steam Pressure
Y011	Equilibrium Helium Valve Position
Y061	Actual Reactor Power
Y122	Compensated Rate corresponding to Steam Pressure
u_{SD}	Equilibrium Helium Valve Position for Shutdown Mode
u_{ML}	Equilibrium Helium Valve Position for Moderator Level Mode
u_{LL}	Equilibrium Helium Valve Position for Low Log Mode
u_{NP}	Equilibrium Helium Valve Position for Neutron Power Mode
u_{SP}	Equilibrium Helium Valve Position for Steam Pressure Mode
u_{RRS}	Control Signal of Reactor Regulating System
u_{CNN}	Control Signal of Fractional Order Convolutional Neural Network based Fractional Order Nonlinear Controller
S	Sliding Surface
E	Gaussian White Noise Compensation Vector
ε_2	Vector of Feature Map of Convolutional Kernel Layer 2
w_0	Weights of Output Layer
b_0	Biases of Output Layer
α_i	Variable Fraction Orders
m	Current Iterations
δ	Total Iterations
δ	Design Parameter for Current Population
δ	Parameter Estimation Error
β	Adaptive Operator
γ	Adaptive Variable for Mutation
P	Number of Inputs of FO-MIMO-HM
L	Number of Outputs of Static Nonlinear Model
K	Output Variables of FO-MIMO-HM

u_L^N	*L*-th Output of Nonlinear Model
u_L	*L*-th Output of Linear Model
$f_{L,n}$	Coefficients
$f_{L,n}$	Basic Functions
$G_{k,p}$	Fractional Order Transfer Function of Dynamic Linear Model
$G'_{k,p}$	Fractional Order Transfer Function of Hammerstein model
$G'^M_{k,p}$	Fractional Order Transfer Function of Auxiliary Hammerstein model
η	Positive Constant
η_w	Learning Rate associated with Weights
K_{SMC}	Learning Rate associated with Biases
K_{SMC}	Sliding Mode Controller Switching Gain
h_0	Design Positive Constant
c	First Positive Odd Number
d	Second Positive Odd Number
p	Number of Convolution Layers = 1, 2
q	Number of Feature Maps = 1,2, …, A_p
j	Number of Convolution Kernels in a Layer = 1, 2, …., n
A_p	Amount of Kernels in a Layer

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 193-201 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)868

Pakistan Academy of Sciences

Research Article

An Effective Paradigm for Selecting Channels in 6G Wireless Networks with Improved Quality of Service

Humaira Afzal[1], Kainat Sajid[1], Muzaffar Hameed[1], Muhammad Rafiq Mufti[2*], and Humera Batool Gill[3]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Pakistan
[3]Institute of Computer Science and Information Technology, The Women University Multan, Pakistan

Abstract: Faster transmission of information and improvements to existing spectrum access procedures will define 6G wireless networks. The requirements of communication systems to serve the expanding demands for more services are critical because spectrum resources are quite limited. Due to the vast number of equipment, there is scarcely any spectrum left to offer future technologies. It provides access based on methods that use licensed spectrum to resolve spectrum crowding issues in 6G wireless networks and satisfy the future demands for wireless communications. This results in dynamic access, where licensed access must provide equitable allocation mechanisms that raise quality of service (QoS) while also posing minimum user interference. In this study, a channel selection algorithm for optimizing frequency, power, and penetration rate in a 6G communication system utilizing a technique that allocates frequency dynamically in a multi-user scenario is presented and evaluated. The proposed approach enables secondary users (SUs) to exchange relevant information prior to channel allocation so they can choose the channel with the highest probability of being available. A hybrid approach based on technique for order of preference by similarity to ideal solution (TOPSIS) and analytical hierarchy process (AHP) is used to perform evaluation and validation with the metrics of frequency usage, SU power consumption, and transmitted signal -penetration rate. Simulated outcomes show that the proposed productive channel allocation technique outperforms the traditional channel allocation strategies in terms of minimal frequency utilization, minimal power consumption, and minimal information loss.

Keywords: 6G Wireless Networks, Channel Selection, Dynamic Spectrum Allocation, QoS Optimization.

1. INTRODUCTION

There is no sector that can avoid the utilization of cellular connection and Internet of Things (IoT) as the usage of wireless equipment has reached its peak in recent decades. Anyone having a smartphone or a smart sensor is aware of how crucial wireless connectivity is to modern life. The concept that these equipment and devices are linked to the computer system changes the overall insight of internet as shown in Figure 1. These are capable of communicating with one another without the aid of humans, after all. In the era of developing telecommunications industry trends, one of the important elements in the adoption of the IoT standard is wireless sensor network. Due to the rise in cellular communications and the popularity of the IoT, mobile carriers consistently focus on enhancing the quality of service (QoS). In contrast to earlier cellular generations, 5th generation long term evaluation (5G-LTE) utilizes a significantly higher proportion of the radio frequency (RF) band that enables the 5G system to transfer data much faster with significantly lower latency [1]. The current capabilities of 4G/5G wireless networks allow high-speed transmission. However, none of them are fully equipped to manage the massive volume of data produced by mobile and IoT devices yet.

Received: April 2024; Revised: May 2023; Accepted: June 2024
* Corresponding Author: Muhammad Rafiq Mufti <rafiq_mufti@ciitvehari.edu.pk>

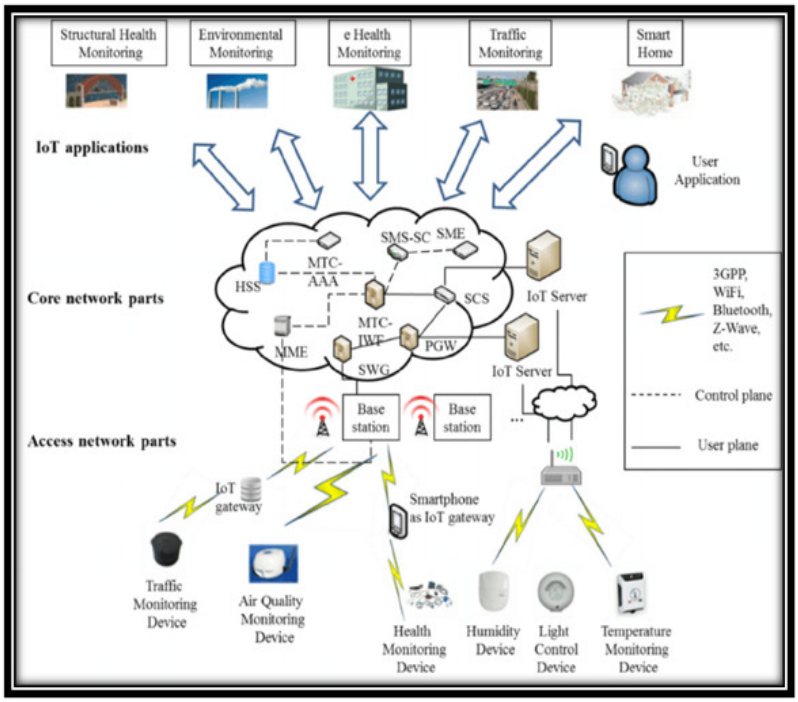

Fig. 1. IoT Architecture.

To address spectrum crowding issues and satisfying the rising needs for wireless technology, licensed frequency access is essential. To achieve dynamic access, the registered users adopt appropriate allocation techniques to improve QoS with the least amount of interference to other users. These characteristics will make cognitive radio (CR) an essential component of upcoming 6G wireless connections. Joseph Mitola [2] was the first who introduced the concept of cognitive radio network (CRN). Later, the CR has emerged as a strategy for resolving shortage difficulties by employing dynamic spectrum allocation. It might be characterized as listening, learning, establishing judgment, and reacting to current network environments. Figure 2 illustrates the CR process. Utilizing all CR Network's features is still challenging due to the limited accuracy of user decisions [3].

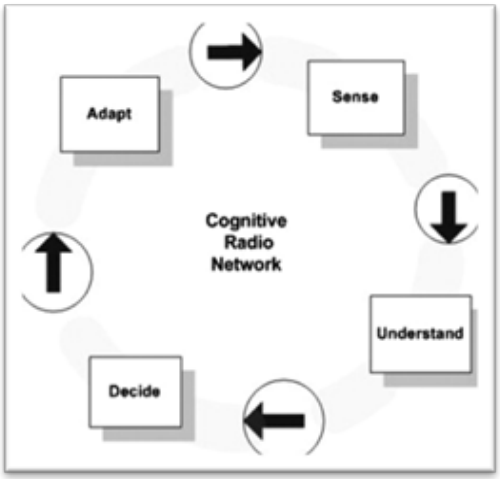

Fig. 2. CR process.

Among the most challenging parts of implementing next-generation systems is radio communication across numerous consumers. Users of CR must be capable of recognizing changes in the surrounding environment in order to collect measurements and expand their partial understanding of the current condition of the system as described in Figure 3. The acquired information can assist to explain the system's indefinite condition and will improve the decision efficiency leading to improve the effectiveness of the system. Each secondary user (SU) must consider channel quality in order to make channel access decisions. Each SU experiences poor performance due to collision among SUs if additional customers access the same frequency band, which is recognized as a negative network externality.

1.1. IoT Enabled CR

The majority of recent research has been towards connection, computation, and communication. The IoT is not considered to be able to handle potentials

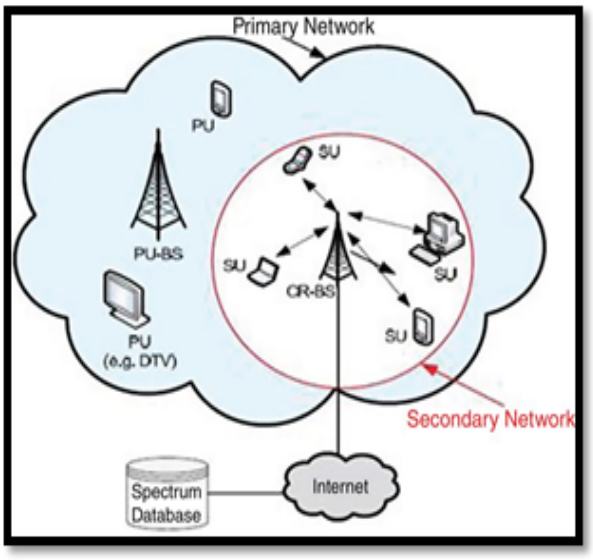

Fig. 3. Radio communication environment.

of growing problems in the absence of complete cognitive capabilities [1]. Therefore, including CR networks into the IoT is essential and the following factors will make it necessary:

- Future sensor communication will rely heavily on wireless technologies such as Bluetooth and Zigbee, etc., and due to its restricted range, the CRN-based IoT framework is to be emerged.
- The scenario will get worse for the enormous number of IoT objects since it would not be able to provide enough bandwidth to all of the devices. Additionally, the price of obtaining bandwidth will be high. This encourages novel solutions and CRNs could be the one.
- Secondary or unlicensed consumers can access the licensed spectrum when not in use by the licensed user or primary users (PUs).
- As more IoT devices are added, interference issues will arise due to their movement from one location to another. The CRN-based IoT devices have the ability to opportunistically seek for interference-free channels to improve transmission.
- In future, there will be expected billions of IoT devices that require constant spectrum operations. Conventional systems will not be able to deal this scenario and hence it would be necessary to change ordinary IoT devices into cognitively connected intelligent devices in order to take advantage of spectrum overcrowding problems.

The contribution of this work is as follows:

i. To increase the effectiveness of radio resource usage by enhancing the decision-making procedure, the dynamic channel selection technique for 6G wireless network is developed.
ii. The proposed hybrid approach efficiently manages the allocation of channels utilized by user equipment.
iii. Next, using the decision-making technique, the best channel is to be chosen based on higher probability of availability.

The overloading problem in 3rd generation partnership project LTE (3GPP-LTE) influencing both machine-to-machine (M2M) and human-to-human (H2H) communications was investigated by Beshley *et al.* [1]. A better priority algorithm design that enables interaction between H2H and M2M traffic and an improved radio resource allocation method was created for LTE and other wireless communications. This approach was based on adaptive channel bandwidth selection taking QoS requirements and priority traffic consolidation into account. An analytical hierarchy process (AHP) based on network interface was introduced by Kim and Kim [4] for interface selection and the suggested approach was expected to have a minimum delay, a high packet reception rate, a long network lifetime, and strong stability. An improved method for channels selection was presented by Thakur *et al.* [5], where access was made using an overlapping mechanism to boost the throughput of CR users. The complexities like random, enhanced, and customized allocation techniques were provided and the results showed the significant performance. A joint channel selection method for mobile CRNs based on link-oriented channel availability and channel-quality was reported by Rahman *et al.* [6]. The link-based probability of channel availability was computed using the distances between PU and CR users in a specific time interval. A useful channel allocation method that utilizes both link-oriented channel availability probability and link-oriented channel quality measurements was also developed to increase the throughput of the network.

The efficient free-channel selection method to enhance the QoS of cooperative CRNs was suggested by Jayakumar and Janakiraman [7]. To determine unused spectrum bands, a combined method employing multi-criteria decision-making

approaches including efficient-free AHP called EFAHP and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was also developed by considering several parameters. Moreover, the performance evaluation was tested using calculation time and switching rate with numerous channel sizes, network services, and PU traffic rates. The suggested approach was claimed to outperform the traditional approaches.

An effective technique for channel assignment was developed by Li *et al.* [8] by considering the optimization of system throughput using channel idle probability, channel gain and spectrum access probability of SU. This approach with genetic algorithm could improve system throughput, transmission latency and channel loss for CR users. The probabilistic-based channel allocation strategy was suggested by Salameh *et al.* [9] in order to remove the need for a jamming detection stage or redundant network resources. A constrained method considering the characteristics of the CRN environment and the effects of jamming assaults was incorporated to determine the signal authenticity ratio. The results showed that there was a great enhancement in throughput as compared to conventional channel allocation methods. A novel method for dynamically choosing a realistic free spectrum band was discussed by Khan and Zeeshan [10]. A versatile adaptive modulation method called orthogonal frequency division multiplexing was then applied evaluating various parameters such as retransmission probability, achieved throughput and latency.

An idea of a decentralized system offering a lot of implementation freedom and very little transmission overload was developed by Wang *et al.* [11]. Each pair of SUs competes against the other SUs in a non-cooperative game to maximize their transmission correlation matrix. It is noted that if energy efficiency increases, then multi-user congestion decreases and the effects of erroneous spectrum status information are mitigated by the designed methodology. The work on channel adaptability and dynamic rate in a CRN was reported by Qureshi and Tekin [12] that handles various kinds of applications with changing dynamic spectrum accessibility and rate limitations. It is termed as a Bayesian learning problem by the experts. A volatile constrained Thompson sampling was also proposed which is a cutting-edge learning method that takes each channel-rate pair as a two-dimensional activity. It effectively optimizes action exploration and exploitation while minimizing regrets by taking advantage of the transmission success probabilities monotonicity. In comparison to other methods, it is revealed that volatile constrained Thompson sampling offers substantial improvement in throughput. The solution to the non-convex and dynamic power optimization problem was proposed by Rahimi *et al.* [13] as a hierarchical deep learning-based resource allocation framework. The offered learning automata determines the most effective remedy for the issue by employing a two-time scale-based sophisticated learning automaton. The suggested architecture might provide an improved performance under transmission power limits. Almasri *et al.* [14] independently assessed each channel opportunity without coordination or prior knowledge of the available channels. The priority-based access strategy could be useful both for fixed and dynamic users. However, it is lack of reliance on prior information or user collaboration in contrast to other collaborative methods.

The optimal resource allocation problem is one of the challenges in wireless networks. The traditional mechanisms aim to attain sub-million latency and can support several users but might experience overcrowding in highly populated areas. To address this issue with spectrum allocation in 5G and beyond, several approaches have been investigated to handle resource allocation problem. However, a little effort has been made to address the issue of spectrum overloading in 6G networks. Therefore, 6G innovation needs to be developed for the best spectrum usage. To increase the effectiveness of radio resource usage in 6G wireless networks, this work aims to design a hybrid channel selection mechanism for 6G wireless networks that allocates channels dynamically with a higher probability of availability, low latency, and highly responsive services.

2. PROPOSED RADIO RESOURCE ALLOCATION ALGORITHM TO ENSURE QoS

To satisfy the transmission demands, the base station (BS) first gathers information from the available resources and then examines it as shown in Figure 4.

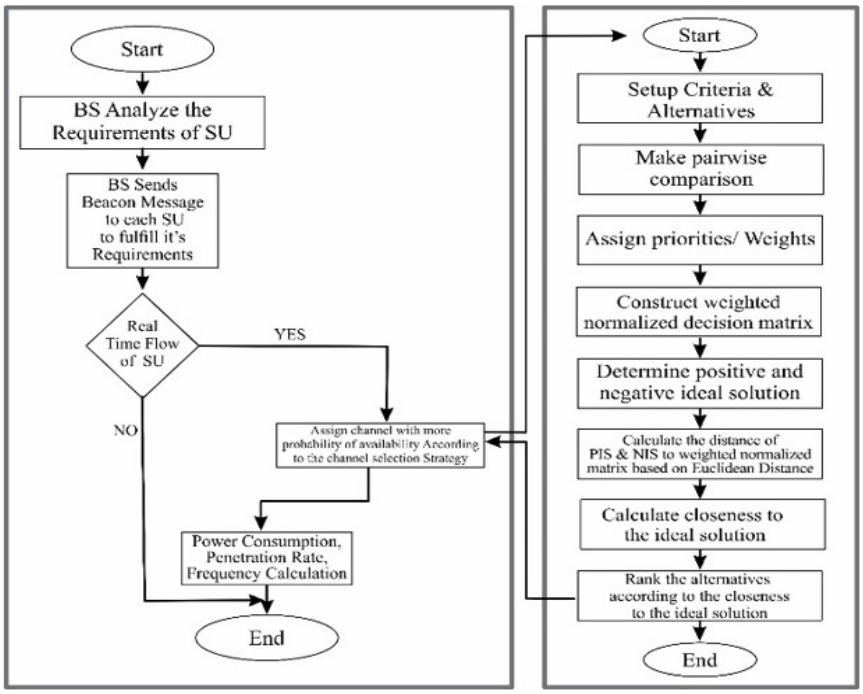

Fig. 4. Block diagram of the proposed strategy.

After analyzing the user's requirements, the BS allocates the smallest available bandwidth to the user. Each SU receives a beacon message from BS in order to fulfill their requirements. This procedure helps to determine the frequency-time constraints and the channel state. To assess whether it is real-time flow or not, the BS monitors the traffic streams. In the case of real-time flow, the BS assigns the channel with the higher chance of availability. On the other hand, if communication is not real-time, the devices are either sorted according to channel selection procedure or will be excluded by the BS. The process of choosing the resource order for mobile users is to be started. In this stage, a buffer is created and cleaned using the resource allocation mechanism and weighting coefficients. The CR users will successfully make use of a channel once find it. After this, frequency, power consumption and penetration rate are computed.

2.1. Hybrid Channel Selection Method based on Multi-Criteria Decision-Making (MCDM)

In this section, a combined channel allocation method is proposed. The weights are assigned using MCDMs to prioritize a number of evaluation parameters. The alternatives are ranked utilizing TOPSIS algorithm, while empty or channels with a greater possibility of being available are weighted with the AHP approach. Its input parameters include penetration rate, power, and frequency. The TOPSIS method uses the normalized vector in order to rank the accessible channels. The best channel will be selected depending upon its availability and SU will have access for data transmission. The final component of MCDM Method is the outcome, which is equivalent to a ranking matrix.

To calculate weights in the proposed hybrid approach, the following expression is used.

$$A = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots b_{nn} \end{pmatrix} \quad (1)$$

where, $b_{ii} = 1$, $b_{ji} = \frac{1}{b_{ij}}$

The decision matrix is normalized as:

$$N_{ij} = \frac{b_{ij}}{\sum_{i=1}^{n} b_{ij}} \quad (2)$$

The weighted decision matrix is calculated using expression given in equation (3). At this stage, the weights sum is also computed. If the sum = 1, the weights are given appropriately, otherwise, the

weights are reassigned.

$$Z_i = \frac{\sum_{j=1}^{n} N_{ij}}{n} \quad \text{and} \quad \sum_{j=1}^{n} Z_i = 1 \tag{3}$$

The all positive ideal and all negative ideal solutions are expressed as:

$$\begin{cases} I^+ = [R_1{}^+, \ldots\ldots, R_m{}^+] \\ I^- = [R_1{}^-, \ldots\ldots, R_m{}^-] \end{cases} \tag{4}$$

Desirable or positive ideal measures are calculated as:

$$\begin{cases} R_i{}^+ = \max\{w_{ij}\} \\ R_i{}^- = \min\{w_{ij}\} \end{cases} \tag{5}$$

Undesirable or negative ideal measures are calculated as:

$$\begin{cases} R_i{}^+ = \min\{w_{ij}\} \\ R_i{}^- = \max\{w_{ij}\} \end{cases} \tag{6}$$

The degree of similarity to the ideal solution is computed as:

$$S_j{}^+ = \sqrt{\sum_{j=1}^{m} \left(R_i{}^+ - w_{ij}\right)^2} \tag{7}$$

$$S_j{}^- = \sqrt{\sum_{j=1}^{m} \left(w_{ij} - R_i{}^-\right)^2} \tag{8}$$

where, $j = 1, 2, 3, \ldots, n$

The exact closeness to the optimal solution can be determined as:

$$C_j{}^* = \frac{S_j{}^-}{S_j{}^+ + S_j{}^-}, j = 1, 2, 3, \ldots, n \tag{9}$$

2.2. Power Consumption

Using control messaging energy (E_m) at the start of each frame, every SU sends its status information to the gateway in t_f time. The pole i used the required amount of power during a control messaging operation is computed as:

$$E_m = P_{max} t_f \tag{10}$$

Assume that the SU transmits at power P_{max} during the control message interval. After receiving the control message, the gateway shifts SU to allocated channels by utilizing the channel shifting power, E_{sp}. When channel j is assigned, the total energy spent by SU is calculated as:

$$E_{sp} = P_u T_u \tag{11}$$

Where, P_u represents the energy spent by the SU while switching channels. The amount of power consumed during transmission is calculated as:

$$E_t = P_{max} |G_{i,j}| \tau_t \tag{12}$$

The energy usage caused by circuit power C_p is as follows:

$$E_u = P_c \tau_t \tag{13}$$

The total power consumed (E_{tot}) by the SU during data transmission can be calculated by adding equations (10), (11), (12), and (13) as follows:

$$E_{tot} = \left(P_{max} t_f\right) + \left(P_u T_u\right) + \left(P_{max} |G_{i,j}| \tau_t\right) + \left(P_c \tau_t\right) \tag{14}$$

2.3. Penetration Rate

The penetration rate is expressed as the sum of information loss on all the channels which can be computed as:

$$P_r = \sum_{i=1}^{4} L(C_i) \tag{15}$$

where, $L(C_1)$, $L(C_2)$, $L(C_3)$, and $L(C_4)$ are the information losses in the channels C_1, C_2, C_3 and C_4, respectively.

3. RESULTS AND DISCUSSION

3.1. Average Frequency

The comparison between AHP, TOPSIS, and proposed hybrid approach is depicted in Figure 5. When compared to TOPSIS and the proposed hybrid approach, the AHP decision-making model is recognized to have 25% and 38% higher frequency utilization, respectively, which results in lowest performance. Additionally, it has been

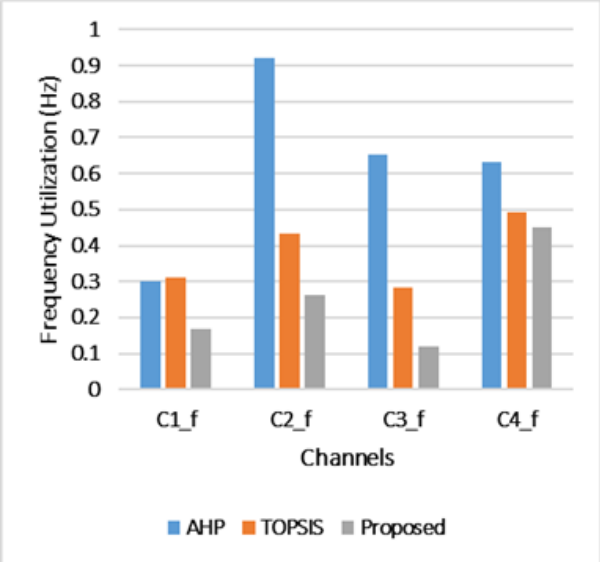

Fig.5. Frequency comparison.

determined that the TOPSIS algorithm performs better than the AHP method but worse when used 13% more frequency than the suggested technique. This indicates that the suggested hybrid technique, which dynamically assigns channels, uses the spectral band's least average to satisfy user equipment.

3.2. Power Comparison

The developed technique is more energy-efficient than AHP and TOPSIS techniques. In case of transmission on the allocated channel, AHP and TOPSIS may cause repeated spectrum handoffs resulting in more power consumption. In comparison to the proposed strategy, the AHP and TOPSIS both use 22% and 8% more power, respectively. The presented method attains a considerable improvement in power consumption and can transfer faster data with less power consumption, because of the dynamic channel allocation mechanism as illustrated in Figure 6.

3.3. Penetration Rate Comparison

The percentage of frame loss, which is determined by dividing the number of lost frames by the overall number of transmitted frames, is used to indicate the penetration rate [15]. The proposed approach offers better transmission with a small percentage of data lost. The penetration rate of the proposed technique is 26% smaller than AHP procedure and 15% lesser than TOPSIS as shown in Figure 7.

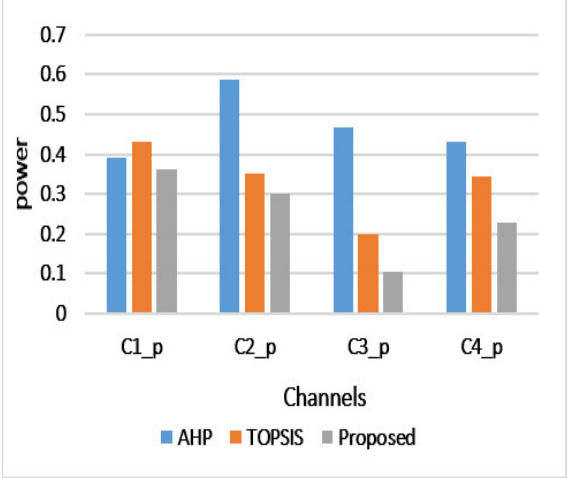

Fig. 6. Power consumption.

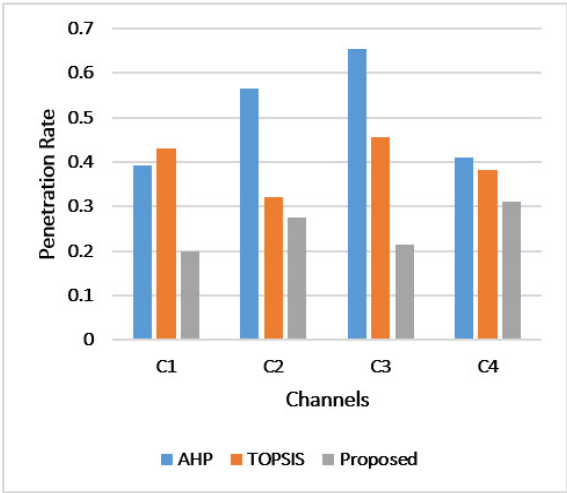

Fig. 7. Penetration rate comparison.

3.4. Discussion

3.4.1. *Parameters selection*

Based on an extensive review of existing 6G network standards and real-world implementation scenarios, the simulation parameters were chosen. By integrating our parameters with these requirements, we make sure that the model's results are appropriate and applicable to real-world network problems. This gives us significant insights about the viability and effectiveness of the suggested solutions.

3.4.2. *Determining real-time data*

We have enabled CR which can sense and adapt traffic according to the network condition. The

CRs examine the entire spectrum frequently in order to identify and assess the type of incoming traffic. It determines whether traffic is real-time or not by evaluating characteristics including data rate, latency requirements, transmission patterns, and the kind of application that is producing the traffic. Services like transferring files or email may be classified as non-real-time, but VoIP and live streaming which have strict latency requirements can be characterized as real-time.

3.4.3. *Scalability of the proposed framework*

By employing a dynamic allocation method that is adaptable to different network sizes and user densities, the proposed model is made to be scalable. Based on real-time network conditions, this approach continuously updates channel availability and optimizes resource allocation, as demonstrated by our experimental results.

4. CONCLUSIONS

This work proposed a channel selection technique that select channel with the probability of their availability. The channel which has more probability of availability or which is free for maximum time duration is selected according to the presented approach. The proposed method allocates channel bandwidth dynamically according to QoS requirements resulting in efficient frequency utilization, less power consumption and reduced information loss in the process of SU data transmission. The feasibilities and use cases of this hybrid channel selection method along with potential implications in real life includes smart cities, IoT, and autonomous vehicles. The proposed method considers only real-time traffic. Future research will concentrate on expanding the proposed strategy to manage different kinds of traffic that go beyond real-time, like bursty and delay-tolerant traffic. We intend to explore the integration of artificial intelligence methodologies to increase the flexibility and capacity for decision-making on channel allocation and selection to improve the methodology. Indeed, our methodology's key concepts can be applied to other cutting-edge technologies like mmWave and Li-Fi.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. H. Beshley, M. Beshley, M. Medvetskyi, and J. Pyrih. QoS-Aware Optimal Radio Resource Allocation Method for Machine-Type Communications in 5G LTE and beyond Cellular Networks. *Wireless Communications and Mobile Computing* 2021: Article Id: 9966366 (2021).
2. J. Mitola. Cognitive radio. Doctoral dissertation, *KTH Royal Institute of Technology, Stockholm, Sweden* (2000).
3. Y.B. Reddy, and H. Smith. Congestion game model for efficient utilization of spectrum. *Conference on Defense Transformation and Net-Centric Systems, (28th April 2010), Orlando, Florida, United States* (2010).
4. B. Kim, and S. Kim. An AHP-Based Interface and Channel Selection for Multi-channel MAC Protocol in IoT Ecosystem. *Wireless Personal Communications* 93(1): 97–118 (2017).
5. P. Thakur, A. Kumar, S. Pandit, G. Singh, and S. N. Satashia, Performance analysis of cognitive radio networks using channel-prediction-probabilities and improved frame structure. *Digital Communications Networks* 4(4): 287–295 (2018).
6. M.A. Rahman, A.T. Asyhari, M.Z.A. Bhuiyan, Q.M. Salih, and K.Z.B. Zamli. L-CAQ: Joint link-oriented channel-availability and channel-quality based channel selection for mobile cognitive radio networks. *Journal of Network and Computer Applications* 113: 26–35 (2018).
7. L. Jayakumar, and S. Janakiraman. A novel need based free channel selection scheme for cooperative CRN using EFAHP-TOPSIS. *Journal of King Saud University-Computer and Information Sciences* 34(4):1326-1342 (2022).
8. Y. Li, S. Li, S. Zhang, and Q. Zhang. Optimal channel selection strategy based on maximizing throughput in Cognitive Radio Network. *5th IEEE International Conference on Information Technology, Networking, Electronic and Automation Control, (15th – 17th October 2021), Xi'an China* (2021).
9. H.A.B. Salameh, S. Almajali, M. Ayyash, and H. Elgala. Spectrum assignment in cognitive radio networks for internet-of-things delay-sensitive applications under jamming attacks. *IEEE Internet of Things Journal* 5(3): 1904–1913 (2018).
10. M.W. Khan, and M. Zeeshan. QoS-based dynamic channel selection algorithm for cognitive radio based smart grid communication network. *Ad Hoc Networks* 87: 61–75 (2019).

11. N. Wang, S. Han, Y. Lu, J. Zhu, and W. Xu. Distributed Energy Efficiency Optimization for Multi-User Cognitive Radio Networks over MIMO Interference Channels: A Non-Cooperative Game Approach. *IEEE Access* 8: 26701–26714 (2020).
12. M.A. Qureshi, and C. Tekin. Rate and channel adaptation in cognitive radio networks under time-varying constraints. *IEEE Communications Letters* 24(12): 2979–2983 (2020).
13. A.M. Rahimi, A. Ziaeddini, and S. Gonglee. A novel approach to efficient resource allocation in load-balanced cellular networks using hierarchical DRL. *Journal of Ambient Intelligence and Humanized Computing* 13(5): 2887–2901 (2022).
14. M. Almasri, A. Assoum, A. Mansour, C. Osswald, C. Moy, and D.L. Jeune. All-powerful learning algorithm for the priority access in cognitive network. *27th European Conference on Signal Processing, (2nd – 6th September 2019), Coruña, Spain* (2019).
15. D.R. Bhadra, C.A. Joshi, P.R. Soni, N.P. Vyas, and R.H. Jhaveri. Packet loss probability in wireless networks: A survey. *International Conference on Communications and Signal Processing (ICCSP), (2nd – 4th April 2015), Melmaruvathur, India* (2015).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 203-216 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)794

Pakistan Academy of Sciences

Research Article

Abstract Art as an Inspiration to Create Textile Patterns through Computer Aided Designing

Zunaira Jamil* and Mehreen Ijaz

Department of Home Economics, Lahore College for Women University, Lahore, Pakistan

Abstract: Art breaks away from the traditional representation of everyday objects and familiar subjects. The viewer is not distracted by meaningful images, so the mind is stirred into felling the energy and spirit of a design. Art does not reflect any form of realism, in fact, it breaks the rules. This study aimed at creating unique patterns for textiles through computer-aided design by taking abstract art as a source of inspiration. It was completed in three phases. First phase comprised of a selection of various inspirational themes from Abstract Art. In the second phase, an experimental approach was taken by developing various patterns through a range of colours and layouts to discover the best and most suitable for the end products. The created patterns were then applied to each category of apparel as 2D and 3D models through Adobe Illustrator and Adobe Photoshop. In the third phase, an online survey was conducted to investigate the market demand of consumers towards developed designs. The results showed that 60% respondents strongly appreciated the uniqueness of design, 56% respondents liked the colour scheme, and 62% of respondents preferred to have this design in their wardrobe. 57% respondents were strongly agreed that the principles of design were well balanced. It was concluded that Marbling Abstract art becoming more popular pattern for apparel due to its bright colour shades, different lines and triangle shapes which attract the participants. The study will be beneficial for the textile designers to opt for multiple categories from abstract art for developing patterns in apparel.

Keywords: Abstract Art, Computer Aided Design, Patterns, Adobe Photoshop, Adobe Illustrator.

1. INTRODUCTION

Textile is one of the largest worldwide industries that play a strong role in the development and growth of any country. It has a long history and in present day can be observed as a prominent sector in apparel, fashion, home accessories, upholstering furniture, and other commercial and industrial products [1]. Textile products are accepted as the first trade-able goods. They give us identity and offer warmth and safety to the wearer. They may reveal one's social standing and can be used for decoration and ornamentation purposes. Colour and appearance of textile products are considered as two most important aspects to be considered by the end consumer. Therefore, a textile designer's understanding of colour and aesthetics is a crucial factor towards the commercial success of any product [2]. Computer Aided Design (CAD) is one of the digitalized procedures currently used by textile industries. It helps to plan, design, sketch and present the product before it is produced. Various types and versions of this software are offered by the marketers to be used by textile and fashion designers [3]. Today, it is used in businesses all around the world. It becomes most widely used tool in the textile sector. It offers opportunity to create new design processes, patterns, motifs and adds aesthetic attributes to these design elements. The designer must go through a learning process to be able to handle this medium to communicate ideas or build designs. Although, this technology may be simple to adapt, yet challenging to stimulate new thoughts and perceptions [4]. In a study by Jing [5], it was found that using visual design thinking can help to convey the design concepts and can be a helpful way to create a more vivid picture of a design. A computer-aided design technology offers the chance for mass customization in clothing and textile sector to satisfy the current market needs [6]. Abstract art is also known as a non-representational and non-objective art used for painting, textiles and

Received: November 2022; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Zunaira Jamil <zunairajamil47@gmail.com>

sculpture. It does not define any single form and can be represented in various ways such as curvilinear, colour based, light based, geometric abstraction, emotional, minimalist, and institutional art [7]. Abstract art is an interesting tool that can be taken as source of inspiration by many designers all over the world. Abstract patterns lack a recognizable shape and hence unrelated to any other reference. It uses shapes, colours, forms, and gestural markings to create desired impact rather than attempting to accurately portray a visual reality. Modern abstract art may be seen in a variety of media. It can be used entirely abstractly or in conjunction with the realistic art. Abstract indicates an element of impressionism and greater freedom found in most of the engineered designs. It is characterized by its simplicity and can be enhanced through the use of colour and light. Patterns in objects have influenced the visual qualities of materials used in technologies that fascinate the eye [8]. An Abstract art is an incredibly versatile and abstract images are psychologically powerful for the viewer. Design has long been considered as an intentional problem-solving action and has become the focus of events prepared for professional activities almost in every sector. The three main elements of clothing design include colour, style and fabric. The designer can attract the consumers with a combination of various colour schemes, textures, materials, and unique designs [9]. Creativity not only caters an individual's own thoughts but also influenced by the surrounding environment, cultural influences, and the availability of resources. It assists in identifying the gaps observed in the existing knowledge and providing the better solutions. New tools and equipment have been developed especially an internet and computers have revolutionized the design process. These help to make the manufacturing and finishing procedures more efficient and effective [10]. Creativity is often viewed as imagination of new ideas involved in the creation of various designs in the field of art and textiles. In the clothing design process, colour is the most direct expression of individuality. Therefore, designers must carry out creative procedures and reasonable colour layouts to enhance the visual impact of clothing design. The creative process begins with the exploration of customer's needs and demands, which can turn into feasible ideas for offering various products or services. The study by Musa [8] discussed the ways of developing the creative process. Although designers choose a wonderful source of inspiration, it is useless if they cannot translate it into reality or interpret it appropriately. However, it is not enough to find the right source of inspiration to be creative, or less often to invent something completely new. Therefore. to improve the creativity of designers, they must look for unique ways to observe various mediums so as to generate more ideas.

The consumers of 21st century become aware of latest trends with the advent of technology. They want innovation in designing process to maintain their individuality. The textile and fashion designers face many challenges to keep satisfied their ultimate consumers. Therefore, designers must combine their creative innovation with a thorough understanding of the properties and applications of fabrics to meet the growing needs of their consumers [9]. For many years, textile and apparel sector focused only on answering the orders received from the market and fulfilling the trends opposed by others. Though textiles have many wider scientific implications from automotive industry to medical applications, yet have long been regarded as 'beautiful and aesthetic'. Every day, we use products that are often decorated with interesting patterns. When people are looking to purchase a product, aesthetics is an important factor to consider. Decorating the surface of products plays an important role in attracting and persuading the consumer to purchase. Not all these designs are inspired by exotic cultures or created uniquely and originally. There are many sources of inspiration that can trigger a designer to select theme for different types of textile products [9]. Computer aided designs help in creating aesthetically appealing patterns using an abstract art as theme for female apparel. Very little research has been conducted on abstract art and its form and their implementation in garment designing. It helps to highlight the efficiency of digital designing compared with the traditional methods. It can lead to explore the art forms that presents better adoption of theme in dress designing. It encourages the designers to work with different styles and patterns in a unique way. This study also gained the feedback from the potential users about their preferences to aid in the marketing of these products.

2. MATERIALS AND METHODS

The theme of Abstract art was taken as a source of inspiration in apparel designing. Abstract motifs comprised of various colours, sizes, patterns and

shapes were made. These motifs are flexible and easy to change their shape and pattern in a unique way. It involves a combination of small and large motifs with soft and dark shades in various directions and at various angles.

This study was conducted in three phases. The first phase comprised of selection of various inspirational themes from Abstract Art. The themes were searched out from the internet, well-known and trustworthy literary sources, books and journals. The shortlisted themes for pattern development used in apparel were line art, colour block, marbling, drip and splash, and curvilinear. The characteristics of each theme that can well suited to be used for apparel designing were purposively selected in this phase of the study. The purposive sampling method is useful as it helps to select an unbiased sample thus increasing the creditability of the obtained findings. Most research publications use purposeful sampling techniques because they are ubiquitous across all research paradigms and aid in locating high-quality samples free from bias, hence enhancing the dependability and credibility of the results [11].

During the second phase of the study, an experimental approach was adopted by developing various patterns through a range of colours and layouts to discover the best and most suitable for the end product. Then the selected patterns were applied on apparel (shirt front, shirt back, sleeves, dupatta) as 2D and 3D models. The patterns were designed with the help of computer aided software such as Adobe Illustrator and Photoshop for 2D and 3D modes of application. After getting inspiration from an abstract art, the designing process was initiated. The sketches of the designs were made by opening a new document. The patterns were made and colours added to them. Shadows and highlights were given to achieve a realistic approach. Some of the filters were also used to soften the harsh edges. The researcher started making sketches on computer and playing with the layouts and colours until a satisfactory balance was found. Later on, samples were displayed on product to understand the possible outcomes. Since there were a large number of design samples only five were displayed. During the design process the layouts, directions, colour, space, and other elements and principles of design had been experimented to find the best results. The study was limited to the pattern making for female apparels only. The specifications are depicted in Table 1.

Third phase of the study was based on an online survey to investigate the feedback of female customers. A questionnaire was designed with 3D representation to identify the responses of targeted group towards apparel design. It was based on structured statements about patterns, colours, creativity, and representation. The questionnaire was reviewed by the experts and after several reviews, improvements and suggestions, final version was prepared and distributed among the targeted population. A comprehensive analysis based on the responses of respondents was made against each category. An online questionnaire was made in Google Form and then shared through Facebook and email with the respondents. Respondents made

Table 1. Design specifications.

Specification of design	Design 1	Design 2	Design 3	Design 4	Design 5
Theme	Line Art	Curvilinear Art	Marbling Art	Block Colour Art	Drip and Splash Art
Colour scheme	Analogous	Triad	Tetrad	Triad	Split-complementary
Target market	Female adults	Teenagers	Teenagers	Female adults	Teenagers
Season	Midsummer/spring collection	Midsummer/spring collection	Midsummer/spring collection	Midsummer/spring collection	Midsummer/spring collection
Suggested fabrics	Shirt-Cotton	Shirt-Cotton	Shirt-Cotton	Shirt-Cotton	Shirt-Cotton
	Trousers-Cambric	Trousers-Cambric	Trousers-Cambric	Trousers-Cambric	Trousers-Cambric
	Dupatta / Stole-Chiffon	Dupatta / Stole-Chiffon	Dupatta / Stole-Chiffon	Dupatta / Stole-Chiffon	Dupatta / Stole-Chiffon

responses against each statement using a likert scale such as strongly agree, agree, neutral, disagree and strongly disagree.

2.1. Description of Design No. 1

Figure 1 shows the description of design no. 1; the design depicts 'line art' in abstract theme, the colours were selected according to the theme. The shades include cyan and mustard, with combination of three neutral colours black, brown, and off-white. The design was based on eastern concept for female adults.

The front of shirt was designed with continuous vertical lines in the background to give a softer and graceful visual impact with off-white and black colour. The border was designed with a unique, creative, and dramatic zigzag and diagonal lines to attract the customer's attention. The border of the shirt was made with a combination of lines following variants triangles. Triangles were broken down with very fine lines and colour tones. The neckline was designed from top till the start of border at the edge. A geometric pattern was repeated throughout the border of neckline. A very fine motif was repeated along with this border. A vertical solid line with base colour was added in the centre of neckline that can be embellished by using decorative buttons. The front was emphasized with simple and continuous lines to give the impression of restfulness and spirituality to cater the demand of young adults (Figure 1(a)). The back of the shirt was designed with the help of verticals lines in the background. It was very similar with the pattern made for the front with same border design. The design was created in such a way to gives a well-blended appearance to the outfit. The design was perfectly in harmony with the shirt front. Neckline area had been omitted at the back side. Moreover, border length was also reduced to give more prominence to the front area (Figure 1(b)).

Sleeves were designed with simple and creative verticals lines used at the front and back of the shirt. The border was designed with straight horizontal lines with cyan, mustard and brown colours. The sleeves were designed with mixed vertical and horizontal lines to gives stability to the design. No pattern was added in the sleeves to avoid monotonous look (Figure 1(c)). Dupattas and shawls have become more of adorable fashion attire with the choice of the variable designs. This is seen as an increased trend among the women's fashion. The center of Dupatta was covered with variations of lines in different tones to give the effect of optical illusions. The borders were designed in a similar manner to the design of the sleeves. A marbling effect was created in the center with the neutral colours. A larger view of border was tried with the broken triangles (Figure 1(d)). Figure 2 shows the draped dress of Design No. 1.

2.2. Description of Design No. 2

Figure 3 explains the description of design no. 2; the design depicts ''curvilinear Art' in Abstract Theme, the colours were chosen according to the

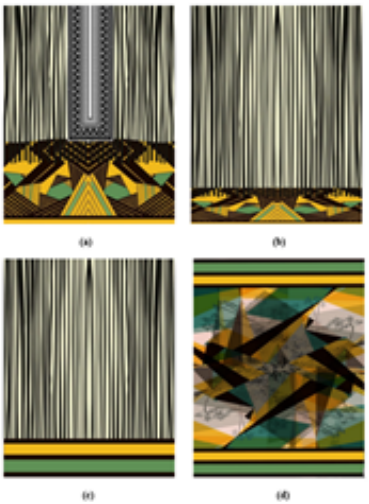

Fig. 1. Design No 1 (a) Front of shirt, (b) Back of shirt, (c) Sleeves of shirt (d) Dupatta.

Fig. 2. Draped Dress of Design No 1.

theme. The shades include red, rust, yellow and bottle green accompanied by three neutral colours such as white, black and grey. The design was based on eastern concept for teenage girls.

The pattern was made in the form of a semi-circle that covers the front of shirt with the above-mentioned colour scheme. The impression of abstract curved lines was present at the background and the circle in darker tones covers the foreground of the pattern. The center was emphasized with red tone to cater the demands of young teenage girls. Very fine curved and straight lines in horizontal and vertical direction were added in patches to create harmony with the overall look of the design (Figure 3(a)).

The back of the shirt was designed with the help of curved abstract lines. It was very similar to the pattern made for the front of shirt in its background. The colours become brighter for the back to create harmony with the front pattern. The lines were drawn in such a way that gives a well-blended appearance to the outfit. A circle had been omitted at the back to avoid the monotonous look (Figure 3(b)).

Sleeves were designed with neutral colours of black and grey to break the monotony of front and back of shirts. The sleeves were well-balanced with the same trendy curved lines throughout whereas, three solid lines with warm colours are added at the edge to harmonize it with the rest of the pieces (Figure 3(c)).

Dupatta is one of the important elements in eastern clothing. It was designed in such a way that looks attractive in combination with the shirt. It was an amalgamation of warm colours and neutral tones. The center was covered with variations of circle in different tones and the border was designed in a similar manner to the edges of the sleeves. By adding three solid lines at the edge of the Dupatta neutral colour was highlighted throughout (Figure 1(d)). Figure 4 depicts the draped dress of Design No. 2.

2.3. Description of Design No. 3

Figure 5 shows the description of design no. 3; the design represents ‘’Marbling’’ in abstract art theme, the colours were selected according to the theme. The colour shades included red, rust, magenta, navy blue, orange and mustard accompanied by three neutral colours black, white, and brown. The design was based on eastern concept for a teenage girl.

The pattern in the center of shirt for the front area was made in the form of diagonal lines to gives the effect of motion. Rust colour was added in the background and foreground covered with diagonal lines and zigzag patterns. The side of the shirt front was emphasized with vertical lines filled with impression of marbling abstract art in darker tones. The border of shirt was designed with horizontal lines and triangles to give a well-balanced appearance to the overall design. The base of the border was highlighted with solid rust and

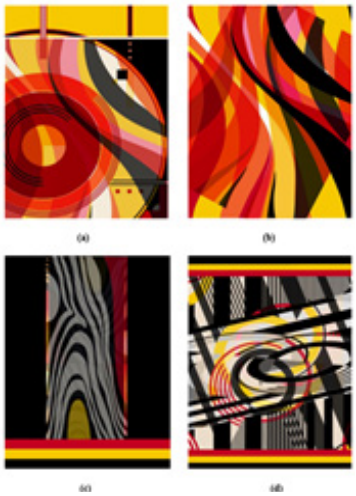

Fig. 3. Design No 2 (a) Front of shirt, (b) Back of shirt, (c) Sleeves of shirt (d) Dupatta.

Fig. 4. Draped Dress for Design No. 2.

indigo coloured line in combination with repeated triangles. The upper half border was designed with colour blocks in combination with solid blue and rust colour lines to create balance (Figure 5(a)).

The back of the shirt was designed in the form of vertical lines. The pattern was made in the form of marbling and zig-zag vertical lines with the above-mentioned colour scheme. The half pattern made for shirt was similar to the design made for sleeves. The zig-zag pattern presents in the back added energy to the design. The colours became brighter for the back to create harmony with the front of the shirt. Random effect of lines was prominent at the half side of back. White colour was added to create attractive look. Sleeves were designed with similar pattern in the back of the shirt to create monotony (Figure 5(b)).

The sleeves were designed with same pattern present in the back with a solid line of marbling abstract art. The colours were chosen according to the theme to give a balanced look to overall design. Sleeves were made attractive and in contrast with the pattern of shirt (Figure 5(c)).

Dupattas and stoles have become more of adorable fashion attire with the choice of the variable designs. It is observed as an increased trend among the women's fashion. The impression of marbling abstract art was added in the background and the border in vertical and zig-zag lines covered foreground of the pattern. The center of dupatta was emphasized with brighter tones to cater the demand of teenage girls. The border of dupatta was designed in a similar manner to the edge of the shirt front. The zigzag patterns showed movement in the design (Figure 5(d)). Figure 6 shows the draped dress of Design No. 3.

2.4. Description of Design No. 4

Figure 7 shows the description of design no. 4; the design depicts ''Block colour" in Abstract Theme, the colours were chosen according to the theme. The shades include tea-pink, dark pink, turquoise, off-white and blue with two bright colours orange and Magenta. The design was based on eastern concept for young female adults.

The front was made in the form of "block colour" that covers the front of shirt with the above-mentioned colour scheme. The impression of off-kilter images of statues, objects, and people cover the half of the front while rest of the half part presents the effect of horizontal straight and broken lines. The design was basically used to achieve a distorted version of reality by manipulating realist imagery. The image of statue in the half of front was added to give an artistic and unique look with combination to the creative horizontal lines in the side of front to give the harmony to the overall look of design. Very fine dots were added to these lines (Figure 7(a)).

The back of the shirt was designed by using pastel colours separated by thick and thin lines making squares of different colours, most

Fig. 5. Design No 3 (a) Front of shirt, (b) Back of shirt, (c) Sleeves of shirt (d) Dupatta.

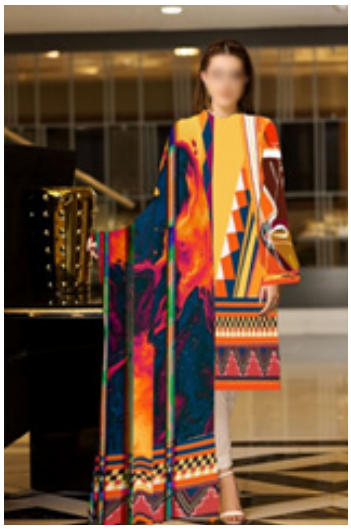

Fig. 6. Draped Dress of Design No 3.

commonly, purple, pink, turquoise, and off-white. The design showed the psychological and dramatic potential of block colour to give artistic look. The pattern in the back was designed to surprise and delight the viewer and catches her attention. Back of the shirt was opposing to the shirt of the front. By adding a solid black triangle at one end and a pale pink triangle at the other, it gives the sense of being a geometric image. Off-white was used to break tints and shades of the back (Figure 7(b)).

Sleeves were designed with pastel colours to give harmony with the overall design. The half horizontal portion of sleeves was emphasized with irrational and unconventional lines. Simple blocks of colour and stunning backdrop to grid-base layouts were added in the border to give a well-balanced look. All colours of the selected theme were added at the edge of the sleeves (Figure 7(c)).

The dupatta is seen as a representation of culture and tradition among us. The impression of dupatta was soft and smooth with different styles and patterns at border. The center was designed with stunning backdrop to grid-based layouts relatively paired with border of sleeve. The design of dupatta was emphasized with pastel colours to give harmony with the overall appearance of the outfit. Block theme become prominent at the edges with a combination of squares and triangles. All related colours were added (Figure 7(d)). Figure 8 shows the draped dress of Design No. 4.

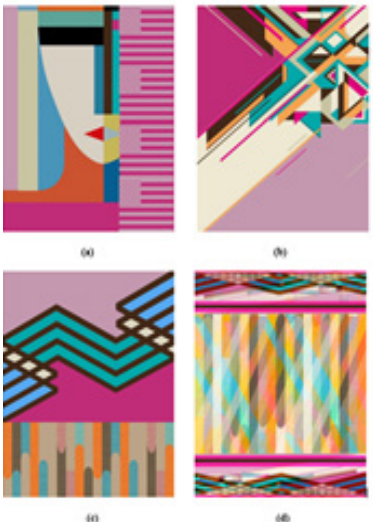

Fig. 7. Design No 4 (a) Front of shirt, (b) Back of shirt, (c) Sleeves of shirt (d) Dupatta.

2.5. Description of Design No. 5

Figure 9 shows the description of design no. 5; the design depicts ‘’Drip and Splash’’ in combination with ‘’line art’’ in abstract theme, the colours were chosen according to the theme. The shades include blue, mustard and pink with two neutral colours black and white. The design was based on eastern concept for teenagers.

The pattern was made in the form of creative lines to give the effect of optical illusion. Line is the most basic element of design. Line can give direction or a feeling of movement to a design. The impression of abstract lines was present in the overall design in the pastel colours. The design was emphasized with colourful dramatic lines to surprise and delight the viewer.

The front of the shirt was designed with variation of lines to create dominant illusions. The overlapping of solid curve lines and zigzag lines add interest in the design. The blending of these pastel lines gives balance to the outfit and very calming to eyes. The interesting flow of lines in front of the shirt attracts the attention by giving it focal point. However, the flow of these lines gives prominence to the front of shirt (Figure 9(a)).

The back of the shirt was designed with simple but creative vertical lines. The pattern was designed with continuous vertical lines to give rigidity and stability. The design was different from the pattern

Fig. 8. Draped Dress for Design No. 4.

made for the front of shirt to create asymmetrical balance. The back of the shirt was emphasized with simple and continuous lines to give the impression of restfulness and calmness to cater the demand of teenagers. Very fine lines were added at the back by adding dots in between them by following a zigzag motion throughout (Figure 9(b)).

The sleeves were designed in such a way that looks attractive in combination with the back of shirt. The impression of drip and splash abstract art was present in the background and foreground is emphasized with abstract curved lines. These curve lines were used in repeated pattern to create rhythm in the design. The sleeves were designed with colour combination to create harmony with the back pattern (Figure 9 (c)).

Dupatta is a symbol of femininity, and it enhances the beauty of the attire. It has been worn in different occasions, such as wedding ceremonies, festivals etc., and it has become a part of all cultural and social life of the women. It was designed with abstract curved lines to give soft and smooth effect. The overall design was covered with variations of lines in different tones to give the effect of movement. The curved lines in random direction were added to break monotony of sleeves and back of shirt. Abstract lines were added in a random way to harmonize with the back and front of shirt. Overlapping of lines add more beauty to the overall look of an outfit (Figure 9(d)). Figure 10 shows the draped dress of Design No. 5.

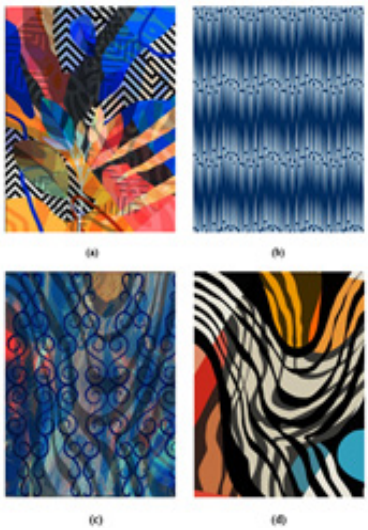

Fig. 9. Design No 5 (a) Front of shirt, (b) Back of shirt, (c) Sleeves of shirt (d) Dupatta.

3. RESULTS AND DISCUSSIONS

Demographic profile of the respondents was given in the Table 2. Table 3. depicts that in Design No 1 most of the respondents were strongly agreed with the given statements whereas, very few disagreed and almost none of them strongly disagreed. The goal of motifs and patterns, according to the researcher of another study, is to ascertain how they relate to people's emotions. Patterns have a vital role in human civilization because they affect how individuals view and interpret the world. Customers' emotions might be pleasantly or adversely impacted by colours. According to Mete [12], emotions are linked with various hues. It has been observed that various colours are commonly associated with different emotions. For instance, off white in this dress connotes mental rest and peace to meet the demands of female adults, while the hue "Green" is calming and linked with nature and typically delivers relaxing qualities. According to Pride and Ferrell [13], the perceptions and attitudes of the peers and social group to which one belongs demonstrate how much they value a person based on their appearance.it is also considered as a powerful tool to communicate with others. Textile and fashion designers present their products which are aesthetically appealing using various motifs, patterns, textures, and elements of design. The contemporary styles use natural elements to present the creativity, skill and art for innovative approaches in this field. Designers can produce goods that are not only aesthetically pleasing but also beneficial to the users by utilizing floral motifs, elements, and patterns into textile design [14].

Fig. 10. Draped Dress for Design No. 5.

Table 2. Demographic profile of respondents.

Demographic variables	Options	%age of responses
Age (N = 100)	18-24	62
	25-29	14
	30-35	13
	35-46	11
Monthly Income	Less than Rs.30, 000	8
	Rs.30, 000 – Rs. 60,000	55
	Rs. 60,000 – Rs.90,000	20
	Above 90,000	17
Marital Status	Unmarried	88
	Married	12

Table 3. Responses (%) of respondents against each designed pattern.

	Design No-1					Design No-2					Design No-3					Design No-4					Design No-5				
	SA	A	N	D	SD	SA	A	N	D	SD	SA	A	N	D	SD	SA	A	N	D	SD	SA	A	N	D	SD
Pattern is unique and different	38	39	20	3	0	46	34	17	2	1	60	26	11	3	0	44	35	10	9	2	56	31	12	1	0
Colours are looking attractive	36	45	17	2	0	50	34	13	2	1	56	32	9	2	1	50	30	12	7	1	55	31	13	1	0
Theme is beautifully translated into design	30	47	20	3	0	42	36	18	3	1	61	26	10	2	1	42	35	15	7	1	53	33	12	2	0
I like the style details	27	47	21	4	1	38	35	21	5	1	57	28	12	3	0	45	31	16	8	0	47	36	14	2	1
Current fashion trends are depicted	31	35	29	5	0	41	30	21	6	2	57	25	15	1	2	48	27	19	6	0	50	35	12	3	0
It can be a good addition to my wardrobe	36	35	22	6	1	45	34	13	6	2	62	22	11	3	2	46	30	14	8	2	52	31	13	3	1
Design is well balanced	28	44	23	4	1	39	43	14	3	1	56	28	14	1	1	40	39	13	6	2	53	31	13	3	0
Elements of design are aesthetically utilized	25	45	26	4	0	37	42	19	1	1	56	27	14	2	1	39	39	15	7	0	44	38	15	2	1
Principles of design are well managed	30	47	21	2	0	38	41	20	0	1	57	28	13	1	1	42	34	18	5	1	54	28	16	2	0
I am willing to purchase these products from the market	27	43	23	3	4	43	27	22	5	3	61	23	11	2	3	43	31	14	10	2	56	27	12	3	2

*SA = Strongly Agree, A = Agree, N = Neutral, D = Disagree, SD = Strongly Disagree.

It has been observed that most of the respondents were strongly agreed that the current fashion trends were depicted in the given dress. There are several ideological connotations to clothes and fashion. Fashion trends display a visual aesthetic and method of dressing that expresses the personality of the consumer in a particular environment. These fashion trends serve the social objective of expressing the distinctiveness of the customer via lifestyle and attitude decisions [15]. According to 28% of the survey participants, design is nicely balanced. The most important design element in the visual arts, according to many authors on western art, is balance because it unifies the structural elements of a pictorial presentation into a single, cohesive framework that helps determine the purpose of each element within a composition. Respondents appeared to agree that the design components were used in an attractive way. Wilson [1] elaborated on this concept from the perspective of a textile designer, stating that design is the visual arrangement of design elements to achieve effects using space, line, shape, form, colour, value, and texture. Textile designers and artists handle these design elements with an attention on balance, movement, repetition, emphasis, contrast, and unity. Some of the respondents firmly believed that design concepts were handled properly. Studies show that people's purchasing decisions are largely impacted by the colour and appearance of the items. It is unexpected that cost and usefulness come afterwards. Because of this, a textile designer's knowledge of colour and aesthetics is essential to the success of a product from a sales perspective. Colour variation in clothing depicts information regarding wearer's personality, thinking, psychology, and even to some extent the level of education. It gives individuality which most of the fashion leaders want to maintain in every aspect of their life. It gives the self-confidence and colour also leads to strengthen communication with others. Colour choice is also affected by the season, occasion and surrounding environment [16].

In the Design No. 2 Apparel should be designed considering three major aspects: structure, function and decoration. It should have an efficient structural design that is suitable for the needs of the clients and current trend. Functionally, it should allow the activity that the user does while wearing it, and aesthetically, it should be suitable for both the garment and the wearer. These three traits may be noticed in a garment when the elements and principles of visual design have been explained as the essential elements or components from which visual design is generated [17]. According to the data, 46% of respondents strongly agreed that the pattern is distinct and different. Design provides people with a platform to express both their originality and their adaptability to the way of living [18]. The kids who were encountered claimed that they dressed in unique ways to signify both their loyalty to and segregation from various groups. The majority of respondents (50%) strongly felt that colours appear appealing. They frequently play such a significant role because they are connected to the traditions or beliefs of the customer. It should be noted that the effects that are seen are not caused by the colours themselves but rather by the symbolic meanings that different civilizations have given to colours [19].

It was believed that style compelled extreme orientation character, encapsulated methods meant to externalize and control women, bound them to guarded and untrue forms of introduction, and strengthened their social relationships with narcissism and technicality. Women's activists who have been influenced by postmodernism recently have adopted a more optimistic stance, realizing that concerns of style and social structure in relation to the body and appearance are inevitable [6]. Current fashion trends are depicted in the given design no.2. Clothes are never a frivolity; they are always an expression of the fundamental social and economic pressures of the time [20]. Fashion is one of the most evident and well-known mediums for society and individual expression. Although it is a communication tool, it does not represent a revolution in and of itself. The advent of street style is another example of how clothing is an essential means of expressing oneself [11]. They may want to add it to their wardrobe. It was determined that by creating a strong mental connection between customers and the products, emotional design would be able to enhance a product's usability [21]. Based on James' peripheral theory, which stated that emotion might impact design or vice versa, Choi [22] explored how diverse commodities may provoke distinct emotions [23]. This demonstrated that designers may evaluate and manage the emotional impact of their products by comprehending the relationship between design components and sentiments.

Most of the respondents liked Design No.3 as a creative theme-based apparel design. According to recent experimental aesthetic studies, humans have very strong preferences for specific forms and relate specific emotional states to abstract geometry [24]. For example, structural patterns may provide viewers a wealth of semantic information that extends beyond form choice. This has an impact on the new design, emotion, and semantics study, which contends that our relationships with the items in our life are incredibly complex and that they have the potential to affect us almost instinctively. 60% of those surveyed firmly agreed that the pattern is distinct and distinctive. In a static material example, textile designs are often meant to stay a stable articulation. This implies, for instance, that the shading structure is intended to give a similar articulation in the determined lifetime of the example [25].

As previously noted, colours have always had an impact on how humans feel physically or psychologically. The psychological effects of colour on people have an indirect impact on social standards, emotional responses, and personal conduct [26]. Aghdaie and Honari [27] explored the style, as a simplest way of knowledge process that may influence the perceptions of the customers within the procedures. Thus, it's vital for designers to know that style (with each visual and verbal strategy) will encourage feeling (the perceptions of the consumers). 57% of the respondents were strongly agreed that current fashion trends are depicted in design 3. 56% of the respondents were strongly agreed and only 1% strongly disagreed that elements of design are aesthetically utilized. The elements of design are the pieces, the components, the building blocks of design. Elements are like the ingredients in a recipe [28], the parts of a machine [29], or the notes in music. The design principles serve as guides for combining elements to provide clear information [10]. In an experience market, design acted as a semiotic instrument, as described by Evans and Thomas [30]. Because of emotion, the user's own expectations and experiences of a product that was objectively delivered in terms of its function were transformed from an objective expectation into a subjectively disposed expectation.

Patterns are geometrically multifarious varying in complexity and structure for the Design No. 4. The "adaptive function" of emotions, which may be impacted by encounters with form, has been characterized as an evaluation of this occurrence. 44% of the respondents were strongly agreed that pattern is unique and different in the given dress. Past studies have reported that those consumers with a high need for uniqueness tend to have a high interest in new products and brands; and they constantly look out for self-differentiating and non-out-of-date goods [31]. Most respondents overwhelmingly concurred that they liked the dress's stylistic features. Consumption develops as a gratuitous and creative activity as individuals reinterpret and reorganize the objects, they have acquired in accordance with a certain style which they are continuously engaged in acquiring [32]. 46% of the respondents were strongly agreed while very few were disagreed that the given dress can be a good addition to their wardrobe.

With the advancements in the garment industry, computer systems, designing software's and artificial intelligence procedures laid a foundation to produce garments with ease and creativity. CAD mapping presents 3D effect to attract the new customers and retain the existing ones. The use of technology in designing process significantly limits the involvement of human resource yet increased the productivity and efficiency in less time [33]. Consumers typically develop an interest in a certain sort of clothing when they are shopping for clothes based on their sensory observation of the sight, texture, and feel of the clothing goods. A sense of how well or poorly made, designed, and constructed the clothing would result from the sensory inspection [34]. Most of the respondents were strongly agreed while only 1% was not willing to purchase these products from the market. Among consumers of different age groups, it has been found that young generation consumers play an important role in the consumption of fashion products [35]. By properly utilizing the elements of design, designers produce unique clothing that adheres to established fashion trends [36, 37].

More than half of the respondents strongly agreed that the pattern depicted in Design No.5 is unique and different. Consumers who want to be different from others almost always use clothing products to display their unique selves and social image [38, 39]. Most of them agreed that the colours are looking attractive, and the theme is aesthetically

translated into an end-product. Individualistic gender-based purchasing behaviour is seen in the purchase of fashion clothing. According to the various studies colour, brand, fashion, and material, clothing tastes might vary from person to person. For instance, a lot of individuals communicate their emotions via colour and contrast. In addition, important aspects including product features, styles, comfort, and uniqueness are influencing consumers' decisions to purchase garments [40]. It is essential to comprehend customer behaviour accurately to fulfill in a novel and imaginative way. Every design book includes design principles. This is because no work of art utilizes only one element or idea [41, 42]. It involves "relating and organizing components or elements graphically to get results." The elements and rules of design are also referred to as the language of art and design. 56% of the respondents strongly agreed that they were willing to purchase these products from the market.

It was observed that most of the respondents with the age group (18-24) preferred the designs with an abstract theme. One of the possible reasons is the use of bright colours that attract the teenagers group. They have an average income of Rs. 30,000 to 60,000 and were unmarried. They were happy and satisfied with the presented options as these patterns give an elaborated and rich look within the limited budget. The patterns are also bold that corelates with their wish to have a unique identity and stand out from the friends and peers. They have the high level of energy so that they would love to wear bright and bold colours. Wearing such colours gives them a feeling of acceptance among their group. This age group particularly attract to the various forms of abstract art due to a combination of creativity and contemporary appeal. Abstract art provides a platform for personal expression and freedom.

Table 4. shows the statistical analysis of responses through the calculation of percentages of highest and lowest responses. It has been observed that most respondents preferred all the designs, out of which Design 3 was liked by most of them for each of the given statements. More than 60 % customers preferred this design for its uniqueness followed by Design 5, Design 4, Design 2 and Design 1.

4. CONCLUSIONS

The computer-aided design software was used to create textile patterns. Five different abstract art sub-themes were selected to produce these designs for female consumers. The prepared designs were shown to the respondents to get their feedback.

Table 4. Statistical analysis of responses (%) against each designed pattern.

	Design No-1		**Design No-2**		**Design No-3**		**Design No-4**		**Design No-5**	
Statements	HV	LV	HV	LV	HV	LV	HV	LV	HV	LV
Pattern is unique and different	38	0	46	1	60	0	44	2	56	0
Colours are looking attractive	36	0	50	1	56	1	50	1	55	0
Theme is beautifully translated into design	30	0	42	1	61	1	42	1	53	0
I like the style details	27	1	38	1	57	0	45	0	47	1
Current fashion trends are depicted	31	0	41	2	57	2	48	0	50	0
It can be a good addition to my wardrobe	36	1	45	2	62	2	46	2	52	1
Design is well balanced	28	1	39	1	56	1	40	2	53	0
Elements of design are aesthetically utilized	25	0	37	1	56	1	39	0	44	1
Principles of design are well managed	30	0	38	1	57	1	42	1	54	0
I am willing to purchase these products from the market	27	4	43	3	61	3	43	2	56	2

*HV = Highest value of responses, LV = Lowest value of respondents.

It was interesting to note that most participants chose marbling abstract art as their preferred pattern because of its vivid colour tones, unique lines, and triangle forms that are appropriate for feminine apparel. In conclusion, by having a broad understanding of the market group, it may be possible to concentrate on understanding the project's commercial component, conducting more research, and marketing in the future. Future research might be done to develop upholstery patterns inspired by various abstract art components.

5. CONFLICT OF INTERST

The authors declare no conflict of interest.

6. REFERENCES

1. J. Wilson (Ed.). Handbook of textile design: principles, processes and practice. *Woodhead Publishing*, *UK*, pp. 10-18 (2001).
2. C. Gale. The Robot Seamstress. *Advance Research in Textile Engineering* 1: 1003 (2016).
3. E.I. Ugwu, M.N. Ezeaku, B.I. Attah, U.M. Emeghebo, and E.C. Eze. Application of Computer Aided Design (CAD) and Flat Techniques in Teaching Pattern Drafting by Clothing Lecturers in Universities in South East, Nigeria. *International Journal of Home Economics, Hospitality and Allied Research* 2: 29-43 (2023).
4. N. Baba. Using Innovative Surface Pattern to Express Cultural Identity in Textile Printing Education. Masters Thesis. *Ains Shams University, Cairo, Egypt* (2018).
5. F. Jing. The Analysis of Computer Aided Design and Software Application in Textile industry. *Journal of Physics: Conference Series* 1952: 042068 (2021).
6. M. Tabraz. Importance of Fashion CAD (Computer-Aided Design) Study for Garment Industry in Bangladesh. *International Journal of Scientific& Technology Research* 6(10): 26-28 (2017).
7. G.L. McLaughlin. Abstract Art: *Independently published, UK*, pp. 25-33 (2023).
8. Z.B. Musa. The Impact of Motifs and Patterns on Surface Design in Fashion and Textile Design on People's Emotions. Masters Thesis. *Limkokwing University of Creative Technology, Malaysia* (2019).
9. K. Niinimaki, G. Peters, H. Dahlbo, P. Perry, T. Rissanen, and A. Gwilt. The Environmental Price of Fast Fashion. *Nature Reviews Earth & Environment* 1(4): 189-200 (2020).
10. S. Liggett, R. Earnshaw, and J. Townsley. Creativity in Art, Design and Technology. *Springer Nature, USA,* pp. 50-76 (2023).
11. F. Nyimbili, and L. Nyimbili. Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies. *British Journal of Multidisciplinary and Advanced Studies* 5(1): 90-99 (2024).
12. F. Mete. The Creative Role of Sources of Inspiration in Clothing Design. *International Journal of Clothing Science and Technology* 18(1): 78-93 (2006).
13. W. Pride, and O.C. Ferrell (Eds.). Marketing: Concepts and Strategies. *Houghton Mifflin, Boston, MA, USA* (2003).
14. V. Stoykoya, J. Illieva, and Z.D. Zlativ. Application of Floral Elements in Textile Patterns for Contemporary Fashion. *Applied Researches in Technics, Technologies and Education* 11(2): 61-73 (2023).
15. M. Barnard (Ed.). Fashion as Communication. *Routledge, London, UK* (2013).
16. G. Hui, A. Aris, and R.D. Rusli. Research on the Application of Color Psychology in Fashion Design. *Advances in Educational Technology and Psychology* 7(18): 1-5 (2023).
17. N. Venkatasamy. Fashion Trends and their Impact on the Society. *International Conference on Textiles, Apparels and Fashion,* (2-9 February 2015) *Tamilnadu India* (2015).
18. D. Anitha (Ed.). Fashion and Apparel Designing. *State Council of Educational Research and Training, Kerala, India* (2005).
19. C.J. Thompson. and D.L. Haytko. Speaking of Fashion: Consumers' Uses of Fashion Discourses and the Appropriation of Countervailing Cultural Meanings. *Journal of Consumer Research* 24(1): 15-42 (1997).
20. J.C. Chebat, and M. Morrin. Colours and Cultures: Exploring the Effects of Mall Decor on Consumer Perceptions. *Journal of Business Research* 60(3): 189-196 (2007).
21. E. Moers (Ed.). The Dandy Brummell to Beerbohm. *Viking Press, New York, USA* (1968).
22. S. Choi. Emotional Universal Design Beyond Usability of Products. *Proceedings of the 5th International Conference on Design & Emotion, Gothenburg, Sweden* (2006).
23. P. Desmet. Designing Emotions. *The Design Journal* 6(2): 60-62 (2002).
24. R. Mauro, K. Sato, and J. Tucker. The Role of Appraisal in Human Emotions: A Cross-Cultural Study. *Journal of Personality and Social Psychology*

62(2): 301-317 (1992).
25. M. Bertamini. L. Palumbo. T.N. Gheorghes, and M. Galatsidas. Do observers like curvature or do they dislike angularity? *British Journal of Psychology* 107(1): 154-178 (2016).
26. L. Worbin. Designing Dynamic Textile Patterns. PhD Thesis. *Chalmers University of Technology, Gothenburg, Sweden* (2010).
27. S.F. Aghdaie, and R. Honari. Investigating the Psychological Impact of Colors on Process of Consumer Shopping Behavior. *International Review of Management and Business Research* 3(2): 1244-1253 (2014).
28. L.G. Tassinary, and J.T. Cacioppo. Unobservable Facial Actions and Emotion. *Psychological Science* 3(1): 28-33 (1992).
29. P. Faimon, and J. Weigand (Eds.). The Nature of Design. *HOW Design Books, Cincinnati, Ohio, USA* (2004).
30. P. Evans, and M. Thomas (Eds.). Exploring the Elements of Design. *Thomson Delmar Learning, New York, USA* (2012).
31. A.G. Ho, and K.W. Siu. Emotion Design. Emotional Design, Emotionalize Design: A Review on their Relationships from a New Perspective. *The Design Journal* 15: 9-32 (2012).
32. K.T. Tian, W. Bearden, and G. Hunter. Consumers Need for Uniqueness: Scale Development and Validation. *Journal of Consumer Research* 28(1): 50-66 (2001).
33. P. Jin, J. Fan, R. Zheng, Q. Chen, L. Liu, R. Jiang, and H. Zhang. Design and Research of Automatic Garment-pattern-Generation System based on Parameterized Design. *Sustainability* 15(2): 1268-1285 (2023).
34. R. Sassatelli (Ed.). Consumer Culture: History, Theory and Politics. *Sage Publications, California, USA* (2007).
35. R. Tsiotsou. The Role of Perceived Product Quality and Overall Satisfaction on Purchase Intentions. *International Journal of Consumer Studies* 30(2): 207-217 (2006).
36. P.K. Tee, B. Gharleghi, and B.Y.F. Chan. Malaysian Young Consumer Preferences in Choosing International Fashion Brand. *Journal of Human and Social Science Research* 1(1): 31-38 (2013).
37. E. Renfrew, and C. Renfrew (Eds.). Basics Fashion Design: Developing a Collection. *AVA Publishing SA, West Sussex, UK* (2009).
38. J. Watson. The Components of Design. A Survey of the Design Elements and Principles and the Development of Design Components. PhD Thesis. *University of Central Oklahoma*, *USA* (2003).
39. A. Kumar, H.J. Lee, and Y.K. Kim. Indian Consumers' Purchase Intention toward a United States versus Local Brand. *Journal of Business Research* 62(5): 521-527 (2009).
40. P. Vikkraman, and N. Sumathi. Purchase Behaviour in Indian Apparel Market: An Analysis. *Zenith International Journal of Business Economics and Management Research* 2(2): 1-12 (2012).
41. M. Pereira, S.G. Azeved. J. Ferreira, R.A. Migul, and V. Pedroso. The Influence of Personal Factors on Consumer Buying Behaviour in Fashion. *International Journal of Management Cases* 12(2): 509-518 (2010).
42. S. Pammi. Creativity in Fashion. *Journal of Textile and Apparel, Technological and Management* 2(4): 1-16 (2002).

Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 217-225 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)847

Pakistan Academy of Sciences

Research Article

A Dual-Channel MAC Protocol with Fibonacci Backoff for Enhanced Efficiency in UAV-Based Sensor Networks

Owais Khan, Muhammad Ismail, and Imad Ali*

Department of Computer Software and Technology, University of Swat, Khyber Pakhtunkhwa, Pakistan

Abstract: Unmanned aerial vehicles (UAVs) are highly effective in collecting data from challenging environments equipped with Wireless Sensor Networks (WSNs), overcoming retrieval challenges. However, using a single-channel Medium Access Control (MAC) protocol for synchronization can lead to potential data collisions among multiple sensors sharing the same medium and result in high power consumption. In this article, we propose a dual-channel MAC protocol specifically designed for UAV-based data collection from WSNs. The protocol includes features such as varying transmission power levels for UAVs, dedicated channels for control and data packets, and a Fibonacci Backoff strategy. The UAV optimizes power usage by initially using low-power transmission and gradually increasing it. The dual-channel communication allows for separate channels for wakeup signals and data transmission, enhancing efficiency. Additionally, the sleep and wakeup mechanism conserves sensor node battery power during inactivity. We developed a discrete event simulator to evaluate the proposed protocol's performance. Our simulation results show that the average for each node count, the proposed protocol with the Fibonacci Backoff strategy improves network throughput by 20.68%, reduces delay by 22.32%, and decreases power consumption by 21.84% compared to the conventional Exponential Backoff method.

Keywords: Wireless Sensor Networks, Unmanned Air Vehicles, Dual-Channel MAC Protocol, Fibonacci Backoff Strategy.

1. INTRODUCTION

In the era of wireless connectivity, the wireless sensor network (WSN) stands as a technological marvel, interlinking sensor nodes with a central sink/base station node [1]. These sensor nodes diligently collect data from their surroundings, orchestrating a symphony of information transmitted to the sink. WSNs are utilized in a variety of fields, including military operations, geographic location tracking, transportation monitoring, healthcare applications, environmental and maritime surveillance, construction project management, cost efficiency initiatives, and agricultural operations. [2]. Presently, a significant emphasis is placed on establishing network connectivity among deployed sensors. This enables the transmission of collected data to a central location for subsequent processing and implementation of necessary actions [3]. Significant strides have been made in data collection from sensors in WSNs, particularly for scenarios where sensors are deployed in challenging environments lacking adequate network infrastructure for data transmission [4]. The provision of permanent network infrastructure proves cost-prohibitive in such scenarios, prompting the exploration of alternative mechanisms [5]. Unmanned aerial vehicles (UAVs) have emerged as a recent solution for data collection in these challenging environments [6]. UAVs, controlled remotely or autonomously, have garnered attention for their ability to navigate areas such as forests, deserts, mountains, glaciers, seas, battlefields, and borders, where conventional infrastructure is lacking [7]. Khan *et al*. [8] explored the use of UAVs and Intelligent Reflecting Surfaces to enhance Public Safety Communication networks in disaster scenarios where cellular base stations fail.

Numerous studies have delved into the realm of Medium Access Control (MAC) protocols for data collection, with a specific focus on enhancing

Received: March 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Imad Ali <imadali@uswat.edu.pk>

the capabilities of UAVs in WSNs. Notable contributions include genetic algorithm-based protocols ensuring optimal UAV path selection for power conservation [9]. Distributed Aerial Data Collection Algorithm has been proposed as a coordinated approach to streamline data collection processes involving multiple UAVs [10]. A cross-layer protocol is introduced to boost efficiency and reliability in UAV-sensor node communication [11]. Pan *et al.* [12] dynamically adjusted UAV speed to maximize data collection efficiency. Goudarzi *et al.* [13] addressed the shortest path problem using the Traveling Salesman's approach and introduced an adaptation system for autonomous flight re-planning, effectively reducing UAV power consumption and optimizing network throughput. These diverse strategies collectively contribute to an overarching improvement in the efficiency of UAV-assisted data collection from WSNs.

Existing MAC protocols for UAV-based data collection in WSNs face several significant challenges. They rely on a single channel for synchronization and data collection. This can result in high power consumption and potential data collisions when multiple sensors attempt to transmit data simultaneously on the same channel. Also, the standard Binary Exponential Backoff (or simply, Exponential Backoff) algorithm used in protocols such as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is not effective in high-density sensor networks, leading to increased collisions and retransmissions [14]. Moreover, high power consumption is a critical issue in WSNs, especially in remote or challenging environments. Existing protocols do not sufficiently address energy conservation while maintaining efficient data transmission. This situation demands a novel approach to enhance efficiency, reduce power consumption, and address the collision problem in WSNs.

In response to the significant challenges posed by existing MAC protocols, this article proposes a dual-channel MAC protocol leveraging the Fibonacci series to address these issues. We propose a dual-channel MAC protocol that uses a dedicated control channel for wakeup signals and synchronization, and a separate primary channel for data transmission. This separation reduces collisions and lowers power consumption. This innovative protocol incorporates two distinct channels for the data collection procedure: a designated control channel for sensor wakeup and synchronization, and a primary channel for transmitting the accumulated data to the UAV. Communication initiation involves transmitting a wakeup signal on the control channel, followed by the actual data transmission on the main channel. Before data transmission, individual nodes undergo a backoff process, during which a random backoff time is selected within the contention window (CW) range, calculated using the Fibonacci series [15].

Our protocol incorporates a Fibonacci Backoff strategy, which provides a more adaptive and effective backoff mechanism. Using Fibonacci intervals for the backoff time reduces collision probability and improves network throughput and efficiency. This approach aims to enhance network performance by avoiding collisions in UAV-based sensor networks and increasing throughput through the implementation of a Fibonacci-based backoff strategy. The proposed protocol optimizes power usage by dynamically adjusting transmission power. The UAV starts with low-power transmission and increases power only when necessary. Additionally, the dual-channel approach and sleep/wakeup mechanism for sensor nodes reduce power consumption by allowing nodes to remain inactive when not transmitting data.

To address the problem of potential data collisions and high power consumption in UAV-based WSNs using a single-channel MAC protocol, this article presents the following contributions:

i. We introduce a dual-channel MAC protocol featuring a Fibonacci Backoff strategy that leverages Fibonacci distribution for collision avoidance in wireless networks deployed on UAVs. The use of a dedicated control channel for wakeup signals and synchronization, and a separate primary channel for data transmission, reduces collisions and lowers power consumption.

ii. Our protocol optimizes power usage by dynamically adjusting transmission power. The UAV starts with low-power transmission and increases power only when necessary. Additionally, the dual-channel approach and sleep/wakeup mechanism for sensor nodes reduce power consumption by allowing nodes to remain inactive when not transmitting data.

iii. We develop an event simulator to assess the protocol's performance, considering factors such as throughput, delay, and power consumption. Our simulation results show that the average for each node count, the proposed protocol with the Fibonacci Backoff strategy improves network throughput by 20.68%, reduces delay by 22.32%, and decreases power consumption by 21.84% compared to the conventional Exponential Backoff method.

2. PROPOSED PROTOCOL

In this section, we present the proposed dual-channel protocol and its pseudocode.

2.1. Dual Channel MAC Protocol With Fibonacci Backoff Strategy

The proposed protocol is specifically designed for situations where a network of sensors is deployed in a field to gather data related to environmental factors such as fire, movement, temperature, and pollution levels. Following the data collection, the acquired information is then transmitted to a central base station for thorough analysis. In regions with limited or no existing network infrastructure, retrieving data from sensor nodes becomes a formidable challenge. To tackle this issue, the proposed solution involves periodically deploying a UAV to the field. The UAV is tasked with collecting data from the sensors and returning, with the frequency of these UAV visits dictated by the specific requirements of the data collection process.

In the context of this protocol, the CSMA [16] strategy plays a pivotal role. CSMA minimizes collision risks by requiring stations to sense the medium before transmitting, significantly reducing collision probabilities. While Carrier Sense Multiple Access with Collision Detection (CSMA/CD) [16] addresses collisions post-occurrence, the wireless network-oriented CSMA/CA [17] employs interframe space, contention window, and acknowledgments to avoid collisions. It introduces a delay, Interframe Space (IFS), before transmission to prevent collisions with distant nodes. Despite prioritizing stations with shorter IFS, the risk of collisions persists, necessitating acknowledgments and timeout timers to ensure successful frame reception. The proposed protocol incorporates dual-channel communication, utilizing both a low-power control channel for synchronization and a high-power main channel for data transmission. During idle periods, sensor nodes conserve power by entering sleep mode and deactivating the main channel. The always-active control channel receives control signals and wakeup signals, prompting sensor nodes to resume full functionality. This dual-channel approach effectively reduces power consumption, particularly beneficial for low-powered devices like sensor nodes. With modern low-power transmission antennas capable of sustaining sensor nodes for extended periods, this strategy significantly prolongs the lifespan of WSNs.

The dual-channel MAC protocol facilitates UAVs in collecting data from sensor nodes in a field, where each sensor is equipped with two channels. One channel is dedicated to UAVs for data collection, while the other serves as a low-power wake-up channel, ensuring prolonged battery life. In their default state, sensors operate in sleep mode, preserving power by deactivating all resources except the wakeup antenna. Upon the UAV's arrival in the field, it transmits a wakeup radio signal to all sensors using its dedicated wakeup channel. Upon receiving this signal, sensors activate all resources, including their primary transmission antenna. Communication between the UAV and sensor nodes on the main channel utilizes the CSMA/CA protocol.

The proposed protocol introduces a novel approach by employing the Fibonacci series to determine the CW in the CSMA/CA protocol on the main channel. Traditionally, CSMA/CA relies on the Exponential Backoff mechanism, where the contention window size doubles on each unsuccessful attempt, leading to increased network latency and performance degradation as the number of nodes rises. In contrast, the Fibonacci-based backoff strategy offers advantages, particularly in small-sized networks. The CW size in the Fibonacci series is calculated using equation (1):

$$\mathrm{CW}_k = \mathrm{CW}_{k-1} + \mathrm{CW}_{k-2} \tag{1}$$

where k denotes the number of attempts a node makes to sense the medium before determining its idleness. Initially, CW_{k-1} and CW_{k-2} are set to 1, resulting in an initial CW of 2.

The Fibonacci series leads to a gradual start with smaller contention windows, providing transmitting nodes more opportunities to sense the medium frequently and a higher probability of early data transmission attempts. As the process continues, the CW grows by following the Fibonacci sequence (3, 5, 8, ….), while ensuring

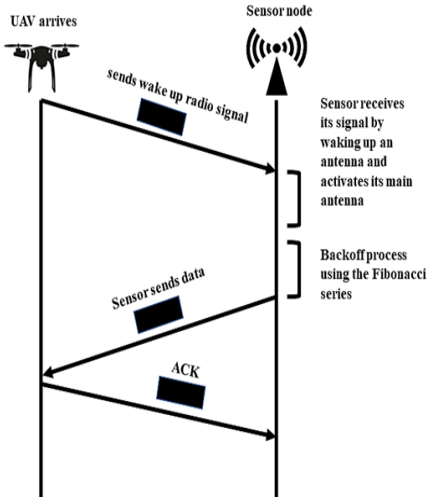

Fig. 1. Data flow diagram illustrating the interaction between UAV and sensors and sensors.

$maxCWk \leq CWmax$ where k represents the number of attempts. This approach optimizes network performance, particularly in a smaller network context. In the illustrated data flow diagram (Figure 1), the interaction between the UAV and a sensor node is depicted. Upon arriving at the location where sensor nodes are deployed, the UAV initiates the process by broadcasting a wake-up radio signal to all nodes. The nodes, equipped with wakeup antennas, receive this signal and activate their main antennas for subsequent data transmission. Following the activation, each node engages in a backoff process on the main channel, CSMA/CA, and utilizes Fibonacci distribution. Subsequently, the node transmits the collected data and awaits acknowledgment from the UAV to complete the communication cycle.

2.2. Proposed Protocol Pseudocode

The pseudocode outlined in Table 1 describes a recursive strategy for the proposed protocol, where nodes repeatedly attempt data transmission. Key elements guiding the transmission process include

Table 1. Procedural Pseducode depicting the operation of the proposed dual-channel protocol.

Step	Pseudocode
1	**Input:** X, prevBackoff, prevPrevBackoff, CWmin, CWmax, backoffCounter, maxTries
2	**Output:** True if data transmission successful, False otherwise
3	Initialization: boPrev = 0, boPrevPrev = 0, CWmin = 0, CWmax = 12, backoffCounter = 0, maxTry = 5, P = initialPower, ΔP = powerIncrement
4	while (P < Pmax) and (N < Nmix) do
5	Send signal S with power P
6	P = P + ΔP
7	data ← prepareData()
8	procedure DualChannelTransmission(data, boPrev, boPrevPrev, CWmin, CWmax, backoffCounter, maxTry)
9	if backoffCounter == 0 then CW ← CWmin + CWmax, boPrev ← CWmax else CW ← boPrev + boPrevPrev
10	backoffTime ← random (0, CW)
11	while backoffTime > 0 do wait (), backoffTime ← backoffTime - 1
12	if CCA () == idle then sendPackets(data), return True
13	boPrevPrev ← boPrev, boPrev ← CW, backoffCounter ← backoffCounter + 1
14	if backoffCounter > maxTry then return False else DualChannelTransmission(data, boPrev, boPrevPrev, CWmin, CWmax, backoffCounter, maxTry)
15	if DualChannelTransmission(data, boPrev, boPrevPrev, CWmin, CWmax, backoffCounter, maxTry) == True then sleep() else stopTransmission()

parameters such as X, previous backoff values, CWmin, CWmax, the backoff counter, and the maximum number of attempts.The desired output is a binary indication: True if data transmission succeeds, and False if it fails.

In the initial phase, the UAV engages in a repetitive process of sending signals with increasing power until the power, P, is less than Pmax and the number of responses, N, is less than Nmix (Lines 4-6). Following this, input variables are initialized: backoff parameters boPrev and boPrevPrev are set to 0, CW size limits CWmin and CWmax are established, and counters for backoff attempts (backoffCounter) and a threshold for maximum attempts (maxTry) are initialized, and data is prepared for transmission (Line 7).

The dual-channel transmission procedure is then defined and invoked with the initialized input variables (Line 8). During the first backoff attempt, if the backoff counter is 0, the CW is set to the sum of CWmin and CWmax, and boPrev is assigned the value of CWmax (Line 9). For subsequent backoff attempts, CW is calculated as the sum of boPrev and boPrevPrev (Line 9), and a random backoff time is chosen (Line 10).

The procedure then waits during this backoff time by decrementing it in a loop. Following the wait, a Clear Channel Assessment (CCA) is performed (Line 11). If the channel is idle, packets are sent one by one, and the procedure returns True, indicating successful transmission (Line 12). If the channel is busy, boPrevPrev is updated to the value of boPrev, boPrev is updated to the current CW value, and the backoff counter is incremented. If the backoff counter exceeds maxTry, the transmission is terminated (Line 13). Otherwise, the procedure is called again with updated values (Line 14).

After the dual-channel transmission procedure, the algorithm checks if the returned value is True (Line 15). If True, the node goes to sleep again. If False, the node quits transmission for an indefinite time. In cases where the channel is not idle, the node re-enters the backoff process. In such instances, the CW is calculated using the Fibonacci series, relying on CW values stored in the previous backoff processes. The backoff timer, a countdown timer, expires upon reaching zero, leading to a resumption of the channel assessment. The sensor node exits the backoff loop only when an idle channel is detected during channel assessment, allowing data transmission to the UAV to commence.

The proposed protocol, using a Fibonacci backoff mechanism, has a computational complexity of O(CWmax). This is more efficient compared to the exponential backoff algorithm which has a complexity of $O(2^k)$, where k is the number of attempts. The Fibonacci backoff mechanism provides a more predictable and controlled backoff time, leading to better performance and lower collision probability.

3. MATERIALS AND METHODS

3.1. Simulation Setting

We design an event-driven simulator in Python to evaluate the performance of the proposed protocol. The simulation outcomes are geared towards analyzing network throughput, delay, and power consumption. A comparative analysis is conducted against the standard CSMA/CA protocol. This Python simulator, crafted for wireless communication, introduces a modified CSMA/CA protocol, substituting the Exponential Backoff strategy with a more sophisticated Fibonacci Backoff strategy.

The simulation specifically focuses on communication within the MAC layer, disregarding any complexities associated with other layers of the OSI model. Throughout the simulation, the channel bandwidth and packet size remain constant. Various transmission overheads are overlooked, and the calculation of packet transmission time spans from the initiation of Clear Channel Assessment (CCA) by a node to the moment an acknowledgment is received for the respective packet. The hardware used for the simulation includes an Intel Core i7-9700K CPU @ 3.60GHz, 32GB of DDR4 RAM, and a 1TB SSD, running on Ubuntu 20.04 LTS. The software environment consists of Python 3.8 as the programming language, SimPy 4.0.1 for simulation, and NumPy 1.19.2 and Matplotlib 3.3.2 for data analysis and visualization.

Notably, the simulation entails multiple runs, progressively increasing the number of network nodes. Each simulation involves an increment of 10 nodes, with a total of 50 simulation runs, each

having a distinct node count. The packet size remains consistent at 60 bits, and the network initiates with two nodes. Subsequently, in each simulation run, one node is added to the network. Our protocol optimizes power use with a sleep and wake-up strategy, activating nodes only when data is ready and the UAV is available. It utilizes two channels, i.e., a low-power receptor channel and a main channel for data transmission—accounting for power consumption in both. For simplicity, we assume constant data rates and fixed transmission ranges, focusing on comparing power efficiency across backoff mechanisms.

3.2. Evaluation Metrics

In this article, we employ several key metrics to assess the performance of the proposed protocol:

Throughput: Throughput, expressed in bits per second (bit/s), quantifies the rate at which data bits are successfully delivered over a communication channel. The formula used for calculating throughput is:

$$Throughput = No.\ of\ Bits\ Transferred / Total\ Time$$

This formula quantifies how efficiently bits can be sent across the network in a given unit of time.

Delay: In our simulation, delay refers to the time taken for a packet to travel from the source to the destination and receive confirmation through an acknowledgment. Although the distance is not a factor in our simulation, the delay is dissected into various components, including processing delay, queuing delay, transmission delay, and any additional delays. The total network delay is computed using the following formula:

$$Network\ Delay = Transmission\ Delay + Propagation\ Delay + Processing\ Delay$$

This formula considers the entire cycle, from the initiation of the packet send the process to the receipt of the acknowledgment.

Power Consumption: Power consumption refers to the amount of energy utilized by sensor nodes during their operation. It is a critical performance metric as it directly influences the lifespan and efficiency of the network.

These metrics collectively provide a comprehensive evaluation of the proposed protocol's effectiveness, considering factors such as data delivery rate, network delay, and power efficiency.

4. RESULTS AND DISCUSSIONS

Fibonacci Backoff strategy provides higher throughput**.** Figure 2 presents a comparative analysis between networks utilizing Fibonacci and Exponential Backoff strategies. The red solid line represents the throughput of the network employing the Fibonacci Backoff algorithm, while the green dotted line represents a network utilizing the traditional Exponential Backoff strategy. The initial increase in throughput with the addition of more nodes is observed, as the higher number of nodes leads to increased frame transmission within the network. This leads to a more efficient delivery of bits across the network. However, around a network size of 20, the trend reverses due to the network's limited capacity caused by backoff and collision issues. Beyond a network size of three, a noticeable disparity in throughput becomes evident between the Fibonacci and exponential networks.

Consistently, the Fibonacci network demonstrates superior throughput compared to the exponential network. Throughput increases for both the Fibonacci Backoff and Exponential Backoff strategies as nodes increase from 2 to 17 due to lower congestion. The throughput peaks around 17 nodes, after which it gradually declines, with the Fibonacci Backoff strategy showing better performance due to superior collision avoidance. There is a slight dip around 20 nodes, likely from temporary congestion, and a significant decline as the node count nears 47, where the Fibonacci Backoff strategy still manages

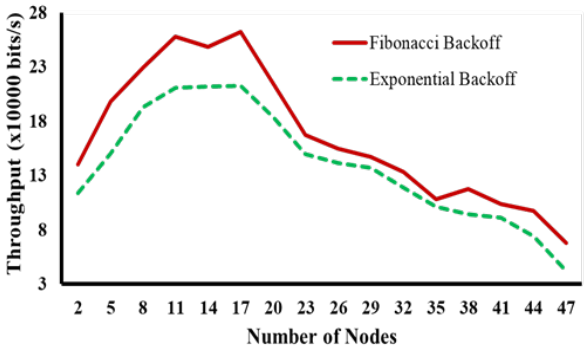

Fig. 2. Comparison of throughput vs. number of nodes using Fibonacci and Exponential Backoffs.

collisions more effectively than the Exponential Backoff strategy. The average for each node count, the proposed protocol with the Fibonacci Backoff strategy improves network throughput by 20.68% compared to the conventional Exponential Backoff method.

The Fibonacci Backoff strategy has lower network delay. Figure 3 presents the delay in seconds for networks employing the Fibonacci Backoff mechanism and the conventional Exponential Backoff strategy. The findings illustrate the enhanced performance of Fibonacci concerning delay. With a network size of up to 40 nodes, the delay within the Fibonacci-based network consistently maintains a lower profile compared to the exponential-based network. However, beyond 40 nodes, Fibonacci-based networks experience higher delays than their exponential-based counterparts. Occasional spikes in the graph denote variations in network conditions, yet the overarching pattern indicates that as the number of nodes increases, so does the delay. This upsurge in nodes results in an increased number of frames for transmission, leading to heightened traffic. Given the limited capacity of a transmission channel, increased traffic inevitably extends the time it takes for a frame to travel from one point in the network to another. Consequently, it can be inferred that Fibonacci networks exhibit commendable performance concerning delay. These results show that the average for each node count, the proposed protocol with the Fibonacci Backoff reduces the delay by 22.32%, compared to the conventional Exponential Backoff method.

Figure 4 illustrates a comprehensive comparison of the power consumption of nodes in two distinct types of networks: one employing the Fibonacci Backoff mechanism and the other utilizing the Exponential Backoff strategy. Power measurements are expressed in milli-watts. Our protocol employs a sleep and wake-up strategy to conserve power. In the absence of data to transmit, sensor nodes enter sleep mode, turning off all resources. They activate when the UAV is available and data is ready for transmission. This approach incorporates the utilization of two channels: a low-power unidirectional receptor channel, consistently accessible, and a main channel designated for data transmission to the UAV. The computation of power consumption is carried out using the pertinent formula. It is crucial to acknowledge the intricate interplay between power consumption and network throughput. Typically, higher throughput results in increased power consumption, establishing a proportional connection. Conversely, lower throughput is associated with reduced power consumption. These results show that the average for each node count, the proposed protocol with the Fibonacci Backoff strategy decreases power consumption by 21.84% compared to the conventional Exponential Backoff method.

Fibonacci Backoff strategy provides a higher throughput-to-power ratio. Nevertheless, the proposed protocol demonstrates lower power consumption compared to the baseline for the same throughput. To illustrate this, we present the throughput-to-power ratio plotted against the number of nodes in Figure 5. The Efficiency Ratio, calculated as throughput divided by power consumption, serves as a metric for the effectiveness of each protocol in balancing data transfer capacity with energy usage. Analyzing Figure 5, a higher Efficiency ratio for the proposed protocol indicates superior efficiency, emphasizing

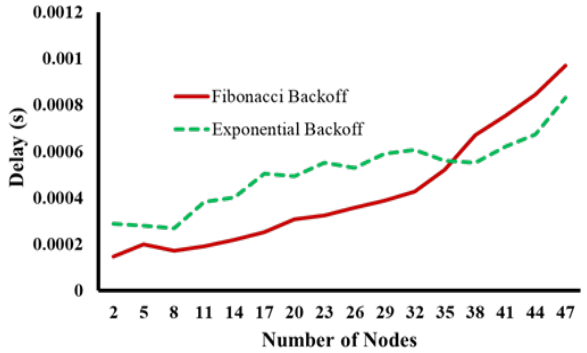

Fig. 3. Comparison of delay vs. number of nodes using Fibonacci and Exponential Backoffs.

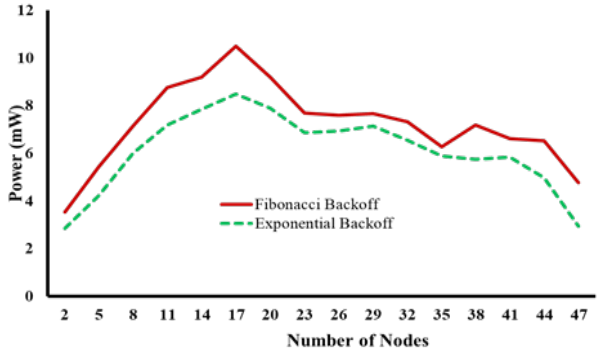

Fig. 4. Comparison of power consumption vs. number of nodes using Fibonacci and Exponential Backoffs.

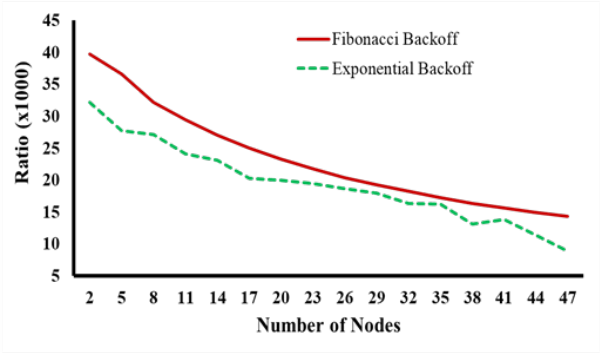

Fig. 5. Comparison of throughput-to-power ratio vs. number of nodes using Fibonacci and Exponential Backoffs.

its ability to achieve greater throughput relative to power consumption.

The empirical observations from our study affirm the superior performance of the dual-channel approach combined with the Fibonacci Backoff strategy compared to the conventional method. Moreover, this innovative approach not only enhances network efficiency but also ensures an extended operational lifespan for both UAVs.

5. CONCLUSIONS

Our proposed protocol enhances data collection from sensor nodes in infrastructure-limited areas using UAVs. Key contributions include the implementation of a dual-channel communication strategy and the use of the Fibonacci series for contention window size calculation, which collectively improve network performance and efficiency. The UAV's dynamic power adjustment and the sensor nodes' sleep mode contribute to significant power conservation. Simulation results show that the protocol achieves nearly 100% efficiency with about 20 nodes and extends network lifespan through its dual-channel approach. This provides a robust solution for UAV-assisted data collection from WSNs. Future research directions include exploring dual-channel communication in UAV swarms, developing UAV power conservation strategies, optimizing data collection routes, implementing geometric backoff for collision avoidance, and investigating cluster-based methods for energy-efficient multi-stage data collection.

6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

1. S. Aiswariya, V.J. Rani, and S. Suseela. Challenges, technologies and components of wireless sensor networks. *International Journal of Engineering Research and Technology* 6: 1-5 (2018).
2. D. Puccinelli and M. Haenggi. Wireless sensor networks: applications and challenges of ubiquitous sensing. *IEEE Circuits and Systems Magazine* 5: 19-31 (2005).
3. D. Hemanand, C. Senthilkumar, O.S. Saleh, B. Muthuraj, A. Anand, and V. Velmurugan. Analysis of power optimization and enhanced routing protocols for wireless sensor networks. *Measurement: Sensors* 25: 100610 (2023).
4. R.F. Miranda, C.H. Barriquello, V.A. Reguera, G.W. Denardin, D.H. Thomas, F. Loose, and L.S. Amaral. A Review of Cognitive Hybrid Radio Frequency/ Visible Light Communication Systems for Wireless Sensor Networks. *Sensors* 23(!8): 7815 (2023).
5. D. Mohammed, M. Omar, and V. Nguyen. Wireless sensor network security: Approaches to detecting and avoiding wormhole attacks. *Journal of Research in Business, Economics and Management* 10: 1860-1864 (2018).
6. S.A.H. Mohsan, N.Q.H. Othman, Y. Li, M.H. Alsharif, and M.A. Khan. Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends. *Intelligent Service Robotics* 16: 109-137 (2023).
7. Y. Liu, H.N. Dai, Q. Wang, M.K. Shukla, and M. Imran. Unmanned aerial vehicle for Internet of everything: Opportunities and challenges. *Computer Communications* 155: 66-83 (2020).
8. N. Khan, A. Ahmad, S. Ali, T. Jan, and I. Ullah. IRS and UAV-relay assisted public safety networks for video transmission: optimizing UAVs deployment and resource allocation. *Telecommunication Systems* 86(3): 433-449 (2024).
9. R.A. Nazib and S. Moh. Energy-efficient and fast data collection in UAV-aided wireless sensor networks for hilly terrains. *IEEE Access* 9: 23168-23190 (2021).
10. B. Olivieri and M. Endler. DADCA: An efficient distributed algorithm for aerial data collection from wireless sensor networks by UAVs. *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems,* pp. 129-136 (2017).
11. Y. Qin, D. Boyle, and E. Yeatman. Efficient and reliable aerial communication with wireless sensors. *IEEE Internet of Things Journal* 6: 9000-9011 (2019).

12. Q. Pan, X. Wen, Z. Lu, L. Li, and W. Jing. Dynamic speed control of unmanned aerial vehicles for data collection under the Internet of Things. *Sensors* 18: 3951 (2018).
13. S. Goudarzi, N. Kama, M.H. Anisi, S. Zeadally, and S. Mumtaz. Data collection using unmanned aerial vehicles for Internet of Things platforms. *Computers & Electrical Engineering* 75: 1-15 (2019).
14. B.J. Kwak, N.O. Song, and L.E. Miller. Performance analysis of exponential backoff. *IEEE/ACM Transactions on Networking* 13: 343-355 (2005).
15. A.J. Gopinath and B. Nithya. Mathematical and simulation analysis of contention resolution mechanism for IEEE 802.11 ah networks. *Computer Communications* 124: 87-100 (2018).
16. F.A. Tobagi and V.B. Hunt. Performance analysis of carrier sense multiple access with collision detection. *Computer Networks* 4: 245-259 (1980).
17. J. Peng and L. Cheng. Revisiting carrier sense multiple access with collision avoidance (CSMA/CA). *IEEE Annual Conference on Information Sciences and Systems,* pp. 1236-1241 (2006).

Instructions for Authors

Manuscript Format

The manuscript may contain Abstract, Keywords, INTRODUCTION, MATERIALS AND METHODS, RESULTS, DISCUSSION (or RESULTS AND DISCUSSION), CONCLUSIONS, ACKNOWLEDGEMENTS, CONFLICT OF
INTEREST and REFERENCES, *and any other information that the author(s) may consider necessary*.

Abstract (font size 10; max 250 words)**:** Must be self-explanatory**,** stating the rationale, objective(s), methodology, main results, and conclusions of the study. Abbreviations, if used, must be defined on the first mention in the Abstract as well as in the main text. Abstract of review articles may have a variable format.

Keywords (font size 10)**:** Three to eight keywords**,** depicting the article.

INTRODUCTION: Provide a clear and concise statement of the problem, citing relevant recent literature, and objectives of the investigation.

MATERIALS AND METHODS: Provide an adequate account of the procedures or experimental details, including statistical tests (if any), concisely but sufficient enough to replicate the study.

RESULTS: Be clear and concise with the help of appropriate Tables, Figures, and other illustrations. Data should not be repeated in Tables and Figures, but must be supported with statistics.

DISCUSSION: Provide interpretation of the RESULTS in the light of previous relevant studies, citing published references.

ACKNOWLEDGEMENTS: (font size 10)**:** In a brief statement, acknowledge the financial support and other assistance.

CONFLICT OF INTEREST: State if there is any conflict of interest.

REFERENCES (font size 10)**:** Cite references in the text **by number only** in **square brackets**, e.g. "Brown *et al.* [2] reported ..." or "... as previously described [3, 6–8]", and list them in the REFERENCES section, in the order of citation in the text, Tables and Figures (not alphabetically). Only published (and accepted for publication) journal articles, books, and book chapters qualify for REFERENCES.

Declaration: Provide a declaration that: (i) the results are original; (ii) the same material is neither published nor under consideration elsewhere; (iii) approval of all authors have been obtained; and (iv) in case the article is accepted for publication, its copyright will be assigned to *Pakistan Academy of Sciences*. Authors must obtain permission to reproduce, where needed, copyrighted material from other sources and ensure that no copyrights are infringed upon.

Manuscript Formatting

Manuscripts must be submitted in Microsoft Word (Latest Version .doc or .docx format); **pdf** files are not acceptable. Figures can be submitted separately in TIFF, GIF, JPEG, EPS, or PPT. Manuscripts, in *Times New Roman*, 1.15 spaced (but use single-space for Tables, long headings, and long captions of tables & figures).The Manuscript sections must be numbered, i.e., **1. INTRODUCTION, 2. MATERIALS AND METHODS,** and so on… (a) **Title** of the article **(**Capitalize initial letter of each main word; font-size 16; **bold**), max 160 characters (no abbreviations or acronyms), depicting article's contents; (b) Author' first name, middle initial, and last name (font size 12, **bold**), and professional affiliation (i.e., each author's Department, Institution, Mailing address and Email and Contact number; but no position titles) (font size 12); (c) Indicate the corresponding author with *; (d) **Short running title**, max 50 characters (font size 10).

Headings and Subheadings (font size 11): All flush left

LEVEL-1: **ALL CAPITAL LETTERS; Bold**

Level-2: **Capitalize Each First Letter (**Except prepositions**); Bold**

Level-3: Capitalize the first letter only (Sentence case)***; Bold, Italic***

Level-4: *Run-in head*; *Italics*, *in the normal paragraph position. Capitalize the first letter only and end in a colon (i.e.,* ***:)***

List of REFERENCES must be prepared as under:

a. Journal Articles *(Name of journals must be stated in full)*

1. I. Golding., J. Paulsson., S.M. Zawilski, and E.C. Cox. Real time kinetics of gene activity in individual bacteria. *Cell* 123: 1025–1036 (2005).
2. W. Bialek, and S. Setayeshgar. Cooperative sensitivity and noise in biochemical signaling. *Physical Review Letters* 100: 258–263 (2008).
3. R. K. Robert, and C.R.L.Thompson. Forming patterns in development without morphogen gradients: differentiation and sorting. *Cold Spring Harbor Perspectives in Biology* 1(6) (2009).
4. D. Fravel. Commercialization and implementation of biocontrol. *Annual Reviews of Phytopathology* 43: 337359 (2005).

b. Books

5. W. R. Luellen. Fine-Tuning Your Writing. *Wise Owl Publishing Company, Madison, WI, USA* (2001).
6. U. Alon, and D.N. Wegner (Ed.). An Introduction to Systems Biology: Design Principles of Biological Circuits. *Chapman & Hall/CRC, Boca Raton, FL, USA* (2006).

c. Book Chapters

7. M.S. Sarnthein, and J.D. Stanford. Basal sauropodomorpha: historical and recent phylogenetic developments. In: The Northern North Atlantic: A Changing Environment. P.R. Schafer, & W. Schluter (Ed.), *Springer, Berlin, Germany*, pp. 365–410 (2000).
8. J.E. Smolen, and L.A. Boxer. Functions of Europhiles. In: Hematology, 4th ed. W.J. Williams., E. Butler and M.A. Litchman (Ed.), *McGraw Hill, New York, USA*, pp. 103–101 (1991).

d. Reports

9. M. D. Sobsey & F. K. Pfaender. Evaluation of the H2S method for Detection of Fecal Contamination of Drinking Water, Report WHO/SDE/WSH/02.08, *Water Sanitation and Health Programme, WHO, Geneva, Switzerland* (2002).

e. Online References

These should specify the full URL for reference and give the date on which it was consulted. Please check again to confirm that the work you are citing is still accessible:

10. L. Branston. SENSPOL: Sensors for Monitoring Water Pollution from Contaminated Land, Landfills and Sediment (2000). http://www.cranfield.ac.uk/biotech/senspol/ (accessed 22 July 2005)

f. Conference Proceedings

11. M. Almasri, A. Assoum, A. Mansour, C. Osswald, C. Moy, and D.L. Jeune. All-powerful learning algorithm for the priority access in cognitive network. *27th European Conference on Signal Processing, (2nd – 6th September 2019), Coruña, Spain* (2019).

Tables and Figures

Insert all tables as editable text, not as images. Number tables consecutively following their appearance in the text, Figures should appear in numerical order, be described in the body of the text, and be positioned close to where they are first cited. Each figure should have a caption that describes the illustration, and that can be understood independently of the main text (Caption Table 1. and Fig 1. font size 10; Bold; Captions should be in sentence case; left-aligned). All Figures should have sufficiently high resolution (minimum 1000 pixels width/height, or a resolution of 300 dpi or higher) to enhance the readability. Figures may be printed in two sizes: column width of 8.0 cm or page width of 16.5 cm; number them as **Fig. 1**, **Fig. 2**, … in the order of citation in the text. Parts in a figure can be identified by A, B, C, D, ... and cited as Figure 2A, Figure 2B, Figure 2C. Captions to Figures must be concise but self-explanatory. Laser printed line drawings are acceptable. Do not use lettering smaller than 9 points or unnecessarily large. Photographs must be of high quality. A scale bar should be provided on all photomicrographs.

Tables: with concise but self-explanatory headings must be numbered according to the order of citation (like **Table 1.**, **Table 2.**). *Do not* abbreviate the word "Table" to "Tab.". Round off data to the nearest three significant digits. Provide essential explanatory footnotes, with superscript letters or symbols keyed to the data. Do not use vertical or horizontal lines, except for separating column heads from the data and at the end of the Table.

Figures: Figures may be printed in two sizes: column width of 8.0 cm or page width of 16.5 cm; number them as **Fig. 1**, **Fig. 2**, … in the order of citation in the text. Captions to Figures must be concise but self-explanatory. Laser-printed line drawings are acceptable. Do not use lettering smaller than 9 points or unnecessarily large. Photographs must be of high quality. A scale bar should be provided on all photomicrographs.

Note: The template of the manuscript is available at http://www.paspk.org/proceedings/; http://ppaspk.org/

Reviewers: Authors may suggest four relevant reviewers (reviewers should not be from the author's institutions), two National and two International (with their Designation, Affiliation, Contact no. and Institutional E-mail addresses).

SUBMISSION CHECKLIST

The following list will be useful during the final checking of an article before sending it to the journal for review.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

- E-mail address (Correct and valid)
- Full address of Institute/organization
- Keywords
- All figure captions
- All tables (including title, description, footnotes)

Further considerations

- Manuscript has been 'spell-checked' and 'grammar checked'
- References are in the correct format for this journal
- All references mentioned in the Reference list are cited in the text, and vice versa
- Permission has been obtained for the use of copyrighted material from other sources (including the Internet)

In case of any difficulty while submitting your manuscript, please get in touch with:

Editor

Pakistan Academy of Sciences
3-Constitution Avenue
G-5/2, Islamabad, Pakistan
Email: editor@paspk.org

Tel: +92-51-920 7140
Websites: http://www.paspk.org/proceedings/; http://ppaspk.org/

PROCEEDINGS OF THE
PAKISTAN ACADEMY OF SCIENCES: PART A
Physical and Computational Sciences

CONTENTS

Volume 61, No. 2, June 2024 **Page**

Review Articles

Research Articles

Instructions for Authors

PAKISTAN ACADEMY OF SCIENCES, ISLAMABAD, PAKISTAN

HEC Recognized; Scopus Indexed

Websites: http://www.paspk.org/proceedings/; http://ppaspk.org