Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 62(1): 67-80 (2025)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(62-1)871

Pakistan Academy of Sciences

Research Article

Uncertainty Quantification and Enhancing Panic Disorder Detection using Ensemble and Resampling Techniques

Muazzam Ali[1*], Amina Shabbir[1], Muhammad Azam[2], M.U. Hashmi[2], Umair Ahmad[1], and Affan Ahmad[2]

[1]Department of Basic Sciences, Superior University, Lahore, Pakistan
[2]Department of Computer Sciences, Superior University, Lahore, Pakistan

Abstract: Panic disorder is one of the leading mental health problems that entail serials of extreme fear that can highly hinder an individual's activities of daily life. Sometimes, intervention is very important, understanding the problem is highly crucial. Unfortunately, some methods assist with diagnostics do not provide accuracy in identification. Using machine learning approaches can resolve this problem as accuracy can be improved with data-driven models. This paper used ensemble machine learning models, Random Forest, Bagging Classifier, and Balanced Bagging Classifier to identify panic disorder. Medical datasets usually have a class imbalance problem. Therefore, we performed resampling SMOTE, ADASYN, and Tomek Links. We evaluated these models by accuracy, precision, recall, F1 score, ROC AUC, Cohen's Kappa, uncertainty measures, aleatoric uncertainty, epistemic uncertainty, and predictive entropy. In our results, The Bagging Classifier was out performance with the highest accuracy (99.97%), recall (99.66%), F1-score (99.60%), and Cohen's Kappa (99.58%), minimal uncertainty metrics (aleatoric: 0.00062, entropy: 0.002003), establishing itself as the optimal model for panic disorder diagnosis. This study proves the effectiveness of ensemble learning and resampling methods for early panic disorder diagnosis and future mental health technologies.

Keywords: Panic Attack, Machine Learning, Synthetic Samples, Decision Trees, Predictive Analytics, Mental Health.

1. INTRODUCTION

The mental health known as panic disorder is typified by frequent, unplanned panic attacks [1]. A panic disorder is characterized by a sudden increase in anxiety or discomfort that lasts for only a few minutes and is accompanied by both psychological and physical symptoms [2, 3]. A fast heartbeat, perspiration, shaking, dyspnea, chest pain, lightheadedness, nausea, and a dread of losing control or passing away are some examples of these symptoms [4]. In contrast to normal anxiety, which frequently has recognizable triggers, panic attacks in people with panic disorder might happen suddenly or for no apparent reason. This unpredictability often leads to chronic fear about future attacks and behavioral adjustments, such as avoiding circumstances or places where attacks have previously happened [5, 6]. About 2-3% of the population is affected by panic disorder each year, which typically manifests in late adolescence or early adulthood. Compared to men, women are more likely to receive a diagnosis. Although the precise causes of panic disorder are unknown, a mix of biochemical, psychological, environmental, and hereditary variables are thought to be responsible [7, 8]. An individual's life may be severely impacted by untreated panic disorder, which can lead to substance abuse, hopelessness, agoraphobia, and avoidance behaviors.

However, effective solutions are available. Cognitive-behavioral therapy (CBT), the most recommended psychological remedy, helps people recognize and change the mental processes that trigger panic attacks [9-11]. One class of drugs that is effective in alleviating symptoms is selective serotonin reuptake inhibitors (SSRIs) [12-14]. Additionally, regular exercise, stress management, and relaxation techniques are lifestyle adoptions

Received: January 2025; Revised: February 2025; Accepted: March 2025
* Corresponding Author: Muazzam Ali <muazzamali@superior.edu.pk>

that aid recovery. The treatment of panic disorder, including early diagnosis and intervention, allows affected individuals to restore order to their lives, improve their functioning, and alleviate symptoms. Advanced data science and machine learning have further enhanced treatment and diagnostic options available to individuals suffering from this disorder. The mental health paradigm concerning the research, diagnosis, and even forecasting of panic disorder is undergoing a profound change due to the impact of data science and machine learning technologies [15]. These technologies analyze and predict outcomes using sophisticated algorithms, large datasets, and modern computation techniques. By applying data from wearable, psychological evaluations, and medical records, these systems can provide an effective early diagnosis and identification of panic disorder. Through predictive analytics, customized assessments can be conducted that link stressors, genetic indicators, and panic attacks, something that traditional methods would miss [16]. With wearable and mobile applications, physiologic and behavioral data can be monitored as they happen, enabling instant capture of data [17]. Machine learning models can boost therapeutic effectiveness and lessen the trial-and-error process by predicting the therapies most likely helpful to a specific patient [18]. Machine learning also aids in the discovery of unknown phenomena within massive quantities of data, such as social networks and electronic health records, to better comprehend the etiology, course, and comorbidities of panic disorder. The amalgamation of data science and machine learning into mental health creates an opportunity for more thorough and precise identification, diagnosis, and treatment of panic disorder [19, 20]. It, in turn, would minimize societal burden and optimize patient care.

Although there have been advancements in the diagnosis of mental disorders over the years, accurately diagnosing panic disorder is still particularly challenging due to its changing nature. It seems that traditional approaches to diagnosis are not adequate. However, the use of automatic diagnostic systems based on machine learning and data science models shows potential for better accuracy in diagnosis. An example of this is the class imbalance problem in medical datasets, which is a situation where one category contains a significantly lower sample size. This frequently leads to skewed results as well as challenges in detecting the minority cases. This paper looks at these issues applying Random Forest, Bagging Classifier, Balanced Bagging Classifier ensemble learning methods, and advanced resampling techniques like SMOTE, ADASYN, and Tomek Links that are focused on improving classification accuracy. This study proposes a novel framework for panic disorder detection which removes class imbalance while including quantification of uncertainty to increase model trustworthiness. Unlike past work on anxiety disorders, this research is focused on providing robust estimations which minimizes misclassification risk, enabling dependable mental health diagnostics.

2. MATERIALS AND METHODS

Panic disorders constitute a grave anxiety issue that poses potential risks to one's mental well-being. To deal with it, one must act quickly. Technology plays a great role in aiding early detection through data-driven methods. Classification must be on point to diagnose and treat panic disorder properly. Single- and multi-attribute generalizations performed by machine learning systems on unbalanced samples tend to overfit, resulting in poor generalization performance. Employing the Random Forest approach, this work analyzes the effectiveness of the classification of the acknowledged panic disorder datasets in the context of SMOTE handling overfitting. The results indicate that SMOTE is proven to greatly improve the Random Forest classification detection of panic disorder datasets, increasing accuracy by 15% [21]. This research proposes a technique for real-time detection of panic attacks leveraging wearable devices through the lens of human-computer interaction (HCI). The application predicts panic episodes by monitoring physiological parameters such as heart rate, heart rate variability (HRV), and electrocardiogram (ECG). These parameters are evaluated by a machine learning model based on CNN and SVM methods, where the system achieved great results revealed by eight percent accuracy, high user satisfaction, and low false positive rates. The system needs fine-tuning in further studies to help improve accuracy [22]. The work aimed to build a model capable of distinguishing panic disorder individuals from healthy subjects using machine learning. Eleven features were extracted from physiological reactions recorded during rest, stress, and recovery. The results showed key markers of

Parkinson's disease to be PT and ECG features during stress and recovery phases. The study's multilayer perceptron (MLP) achieved a maximum accuracy of 75.61% with 33 features. The study showed the capability of objective differentiation of people with Parkinson's disease by combining multimodal signals [23]. The relevance of different biometric and geographical parameters is examined for their significance in the overall process of panic detection, as described in reference [24]. A model for predicting panic attacks is based on 7 days of data gathered from wearable, passive sensors, and online questionnaires. The system integrates environmental, physiologic, and survey responses to predict panic attacks in advance. The overall accuracy is between 67.4 percent and 81.3 percent, which is driven mostly by the questions asked and the physiological attributes of the respondents. However, the focus of the study was narrowed down to patients who have panic disorder only [25]. Even though 95% of participants were able to identify a trigger, 11% of individuals suffer from panic attacks annually. A higher likelihood of panic attacks the next day was associated with lower mood, state-level mood, higher resting heart rate, and more background noise. This information could improve the specificity of one-session psycho-education programs and help direct ecological momentary treatments [26, 27].

According to cognitive-behavioral theories, perceived threat, and arousal-related bodily sensations produce a positive feedback loop that leads to panic attacks. A computational model suggests that a simulated disturbance of arousal is a good indicator of panic disorder and the likelihood of an attack. This study reviews the empirical literature on biological challenge reactions and their relationship to disorder and panic episodes. The results show a moderate association with recurring panic attacks but no association with panic disorder [28]. The study used digital phenotypes and machine learning algorithms to predict panic symptoms in patients with mood and anxiety disorders. For two years, 43 patients were monitored through wearable devices and smartphone applications. The XGBoost model also proved useful with additional features such as elevated anxiety, increased step counts, and Childhood Trauma Questionnaire results [29]. A machine learning-based technique was used to differentiate panic disorder from other forms of anxiety disorders using heart rate variability (HRV). Of the five algorithms that were employed, the L1-regularized LR showed the best accuracy. However, the study had limitations, including a limited sample size and a cross-sectional methodology. Future studies should use machine learning techniques to duplicate the diagnostic usefulness of HRV [30]. This study aimed to develop a dropout prediction model for cognitive behavioral therapy (CBT) for panic disorder using machine learning techniques. Two hundred eight patients received group cognitive behavioral therapy. The results showed excellent accuracy in predicting dropout during CBT for PD, with random forest achieving 88% accuracy and light gradient boosting machine achieving 85% accuracy. This machine-learning technology may help with clinical decision-making in typical clinical situations [31].

Panic disorder is a refractory mental condition that affects 5% of people worldwide and 10% of persons with subclinical symptoms. If untreated, it is a serious issue that can significantly harm individuals, families, and society. The prediction model of panic disorder is examined in this study using machine learning algorithms that simulate human learning processes. The results show that the artificial neural network model has the largest AUC (0.8255) and the shortest running time (less than 1 second) when compared to other models [32]. The study used machine learning to look at PD patients' brain networks. The results showed that the structural network, particularly the extended fear network, had changed. Increased connections to the insula, hippocampus, and amygdala were associated with better treatment response, while over-connectivity was associated with poor response. SVM and CPM can predict treatment outcomes based on changes in network patterns [33]. Robinaugh *et al.* [34] presented panic disorder symptoms as panic, chest pain, and tremors, afflicting 5% of the global population. Accurate diagnosis is crucial in the early stages of illness to maximize therapeutic efficacy. Among the many promising ML algorithms, artificial neural networks (ANN) appear to perform the best in predicting panic disorder, with AUC scoring 0.8255 and exceeding one second in processing speed. These discoveries fortify the argument for advancing predictive ML models for earlier discovery and preventative measures regarding panic disorder.

2.1. Methodology

This study uses a classification model with a resampling technique to develop a machine learning-based approach to detect panic disorder. Following sections demonstrate the process, which includes acquiring the dataset, data cleaning, feature engineering and transformation, exploratory data analysis, implementing techniques to resample and adjust for class imbalance, fitting the models, determining their performance, and quantifying the uncertainty.

2.1.1. Dataset acquisition and preprocessing

This study utilized a Kaggle dataset containing 100,000 entries featuring 17 attributes. Each entry is associated with an individual and captures various attributes, including a person's demographic details, family details, personal details, psychological stressors, symptoms, their severity, impact on daily activities, medical history, psychiatric history, substance use, coping strategies, social support, and lifestyle choices. The dataset also contains a binary variable capturing the existence and absence of panic disorder as the target variable. The dataset is largely imbalanced due to the large portion of not diagnosed with panic disorder, and thus, the patient group (those diagnosed with panic disorder) is significantly underrepresented in the dataset.

The dataset was uploaded and processed using Python libraries like Pandas, NumPy, and Scikit-learn. Categorical features were labeled via Label Encoding, transforming them to numerical values during the data loader's execution. This stage ensured that all features were numeric and, therefore, usable by machine learning algorithms. Examine the missing values and find no missing values. The standard scalar from Scikit-learn was used to perform feature scaling. This scalar standardizes features by removing the mean and scaling to unit variance. This process is important to ensure that models do not have their predicted outcomes skewed as a function of larger range features [35].

2.1.2. Addressing the issue of class imbalance

The study undertook measures to address class imbalance issues by leveraging multiple resampling methods such as SMOTE, ADASYN, and Tomek Links under-sampling. Through SMOTE, new instances are made in the feature space, increasing the generalization of the model by augmenting the representation of the minority class.

$$x_{new} = x_i + \gamma(x_j - x_i) \quad (1)$$

Where,
x_i is a randomly chosen minority class sample,
x_j is one of its nearest neighbors in the feature space,
γ is a random number between 0 and 1, ensuring interpolation.
ADASYN improves SMOTE by creating synthetic samples that pay more attention to the difficult cases, which allows less bias in the model's predictions [38].

$$G_i = \bar{G}.\frac{r_i}{\sum_{i=1}^{n_m} r_i} \quad (2)$$

Where,
G_i is the number of synthetic samples to generate. For instance x_i,
$\bar{G}$ is the total number of synthetic samples needed,
r_i is the ratio of majority class neighbors to total neighbors for x_i,
n_m is the number of minority class samples.
As an under-sampling technique, Tomek links focus on class separation by deleting instances of the nearest border from the majority class [39].

$$d(x_i, x_j) = \min_{x_k \in X} d(x_i, x_j) \qquad for\ all\ k \quad (3)$$

These resampling methods were evaluated using stratified k-fold cross-validation to determine their effectiveness in improving model performance. Some other techniques for dealing with class imbalance, for instance, Random Under-sampling, Random Over-sampling, Cluster-based Over-sampling, and near-miss, have weaknesses that make them unsuitable for this research. Random Under-sampling has the potential to result in the loss of important information, and Random Over-sampling is problematic in that it is prone to over fitting by copying minority class instances. In particular, Cluster-based Over-sampling may be time-consuming to process and introduce potential prejudice. At the same time, NearMiss can make wrong speculations about data distributions and may lead to high-class overlap. In contrast, SMOTE, ADASYN, and Tomek Links represent more sophisticated approaches. SMOTE creates synthetic samples that increase the generalization ability; ADASYN is committed to identifying the hard instances to lower the bias. In contrast, Tomek Links is used to eliminate the near-border instances to clarify the class boundaries, ensuring a better performance, generalization, and less over fitting of the model, thus making them very good for detecting minority classes such as panic disorder.

2.1.3. Exploratory data analysis

During the exploratory phase of the analysis, we developed correlation matrices and histograms to understand the features' relations and distributions.

2.1.3.1. Correlation matrix: The heat map style correlation matrix was useful in spotting the presence of linear relationships between the features. This was beneficial for not only identifying multi-collinearity issues, but also feature selection redundancies [35]. Figure 1 displays the correlation heatmap of the Panic Disorder dataset, elaborating on important inter-variable relationships. The most significant aggravating associations were noted both for symptoms and their severity, which was less than 0.15, and more pronounced for panic disorder diagnosis which showed a greater than 0.26 value. This shows that more severe symptoms and greater likelihood of diagnosis are strongly correlated and aggravated in individuals with severe panic disorders. Furthermore, Lifestyle Factors show a moderate positive association toward the Panic Disorder Diagnosis with a coefficient near 0.26 suggesting some degree of lifestyle influence in regard to the condition diagnosis. In general, the heatmap provides insights into the role of Severity and Symptoms of panic disorder and its clinical significance. This information is useful for other researchers if they fine-tune their efforts in managing severity and lifestyle changes.

2.1.3.2. Histogram: The analysis of solitary characteristics distribution was carried out using histograms. This method allowed for detecting potential skewed distributions that could affect the model. Based on the histograms, log transformations, and normalization techniques that improve the accuracy and reliability of the model

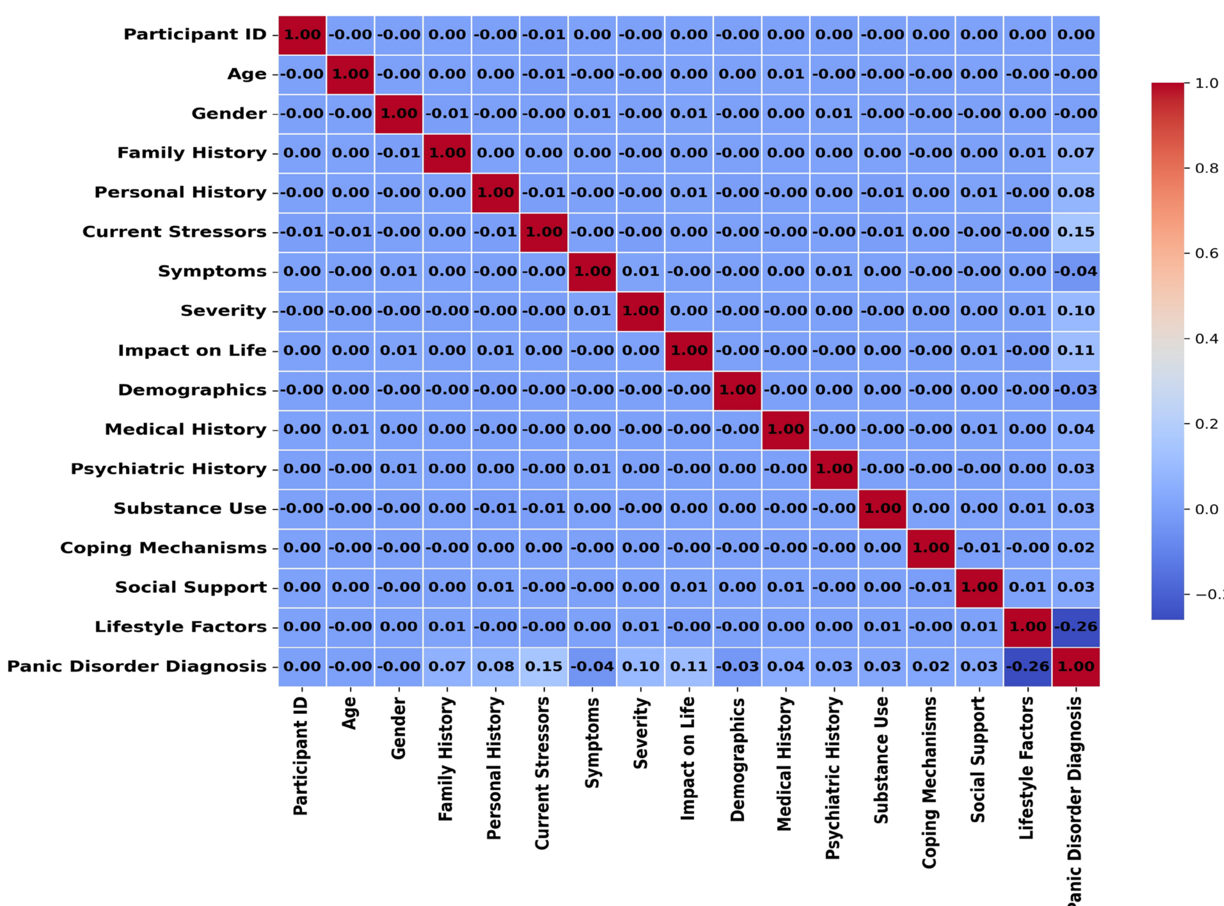

Fig. 1. Correlation heat map of variables in the panic disorder dataset.

were adopted [37]. Each histogram in Figure 2 indicates the distributions of features within the dataset's context, which gives us implications regarding the underlying patterns within the dataset. Most family history, personal history, and severity features demonstrate distinct binary or categorical features, which suggest clear categories with minimal or no intermediate cases. Features like age and gender have even distributions, which indicate a good level of diversified representation among the participants in the study. Reporting or categorizing current stressors, symptoms, and impacts on life will likely be dominated by a few prevalent patterns indicated by multimodal distribution. Medical history, psychiatric history, substance use, and coping mechanisms have features with distributions that appear to have several levels or categories, enriching the dataset with nuanced participant information. Social support and lifestyle factors showed the same level of clustering, suggesting that discrete levels of support and lifestyle habits are present. The disparity between these distributions

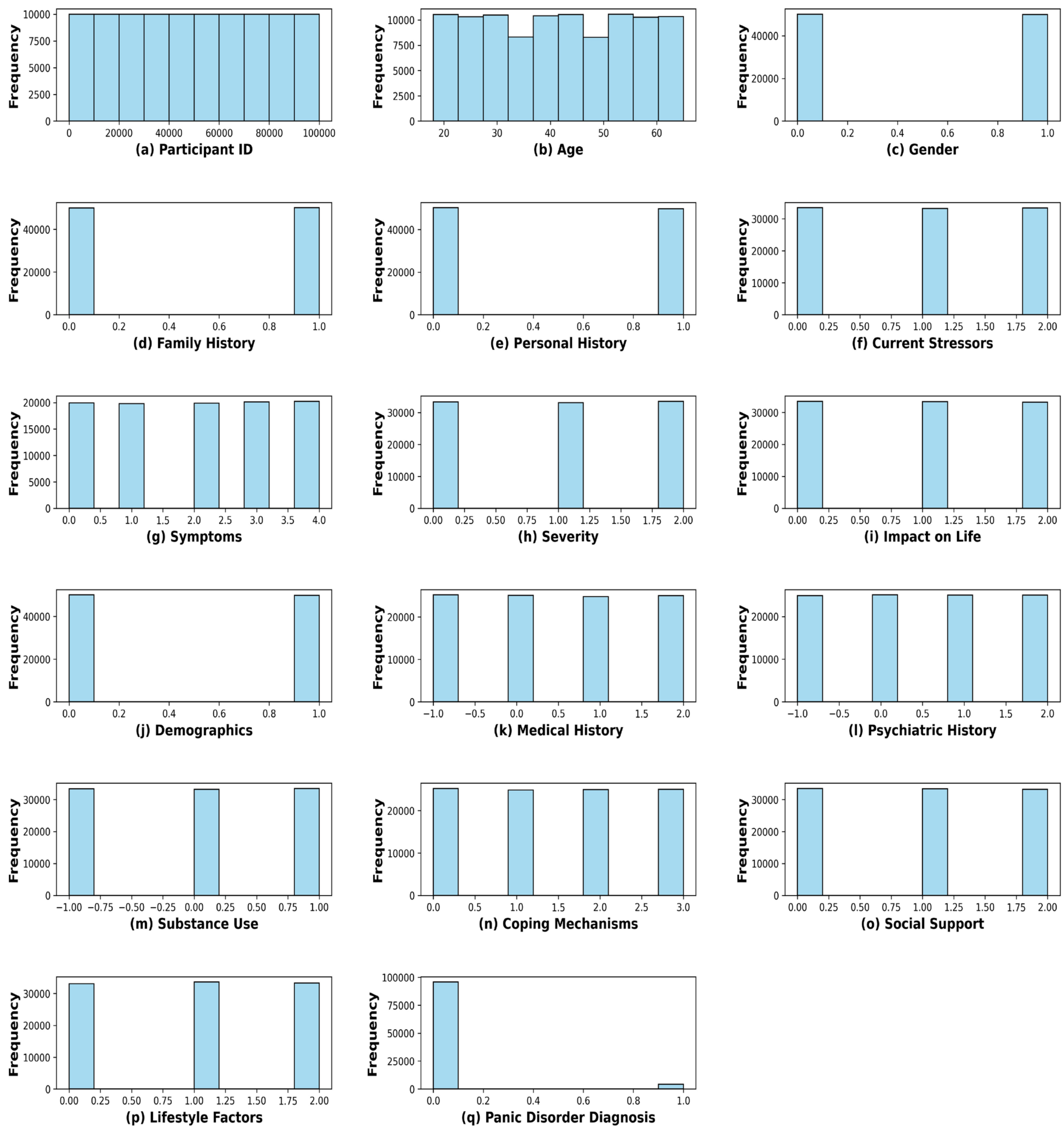

Fig. 2. Distribution of variables in the panic disorder dataset.

indicates that the dataset will easily be subjected to classification techniques since it relies heavily on categorical or binary features.

2.1.4. Model training and evaluation

The classifiers used for this work included Random Forest, and Balanced Bagging Classifier. Random forest is an ensemble of classifiers with decision trees that works well with imbalanced data sets as class weight can be assigned. The Balanced Bagging Classifier is a class from an imbalanced learn library that uses bootstrap aggregation to balance the distribution of classes among samples [40-42]. Each model was developed using stratified k-fold cross-validation and five splits for the most accurate performance estimation. Accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic Curve (ROC AUC) were determined for each model-resample configuration. Accuracy, while a good measure of the overall correctness of model predictions, is difficult to interpret with imbalanced datasets. So precision assesses the value of positively predicted positives (PPV) while recall determines the rate of true positives captured by the model. The F1 score provides a balance between precision and recall and, thus, is a more appropriate metric with class imbalance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$precision = \frac{TP}{\mathrm{TP}+FP} \tag{5}$$

$$Recall = \frac{PT}{TP+FN} \tag{6}$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false positive.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \tag{7}$$

$$\mathcal{K} = \frac{P_0 - P_e}{1-P_e} \tag{8}$$

2.1.5. Estimation of uncertainty

This research focused on estimating aleatoric and epistemic uncertainties, as well as predictive entropy, to increase model accuracy. As stated in the previous section, aleatoric uncertainty stems from underlying noise in data sources and was measured as the variance of predicted probabilities. Higher aleatoric uncertainty indicates class distribution overlaps, and therefore, individual predictions become ambiguous due to noise. The variance of the predictive distribution is used to quantify aleatoric uncertainty:

$$Aleatoric\ Uncertainty = E_{p(y|x,\theta)[(y-E|y)^2]} \tag{9}$$

It can be approximated using the variance of Softmax probabilities:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} p(y_i|x,\theta)(1-p(y_i|x,\theta)) \tag{10}$$

$p(y_i|x,\theta)$ is predictive probability of class y.
N is the number of Monte Carlo dropout samples
θ represents the model parameter.
Due to the lack of training data, epistemic uncertainty denotes model uncertainty, which was estimated through the variance of probabilistic predictions from ensemble models.

$$Epistemic\ Uncertainty = Var\left[E_{p(\theta|D)}p(y_i|x,\theta)\right] \tag{11}$$

Monte Carlo dropout-based estimation is:

$$\sigma^2 = \frac{1}{T}\sum_{i=1}^{T} p(y_i|x,\theta)^2 - \left(\frac{1}{T}\sum_{i=1}^{T} p(y_i|x,\theta_t)^2\left(1-p(y_i|x,\theta_t)\right)\right)^2 \tag{12}$$

T is the number of stochastic forward passes (MC Dropout), $p(y_i|x,\theta_t)$ is the probability of class y given x and model parameters at $t-th$ forward pass.

In addition, predictive entropy for model outputs was estimated using Shannon entropy [43]. These measures of uncertainty were used to evaluate the confidence of the model predictions and the additional data that needs to be collected to make the model more accurate.
It is computed using the Shannon entropy:

$$H[y_i|x] = -\sum_{i=1}^{C} p(y_i|x)\log p(y_i|x) \tag{13}$$

C is the number of classes and $p(y_i|x)$ is the predicted probability for class i.

2.1.6. Random forest

Random Forest is an ensemble learning approach that combines several weak classifiers to handle complicated problems. It comprises several decision trees that make predictions based on the majority vote of forecasts. Regression and classification issues are addressed by the supervised machine learning method, which uses decision trees. Decision trees display predictions from a sequence of feature-based splits using a tree structure resembling a flowchart. The root, decision, and leaf nodes are their three constituent parts. The root node is where population division begins; the nodes remaining after splitting a root node are known as the decision node, and the node that cannot be further divided is the leaf node. The *Gini Index* is used to identify which feature will be the root node. Mathematically, it can be written as:

$$Gini\ Index = 1 - \sum_{i=1}^{n}(p_i)^2$$
$$= 1 - \ [(p_+)\ ^2 + (p_-)\ ^2] \quad (14)$$

Where p_+ the probability of positive is class and p_- is the probability of negative class. Unlike other decision tree models, the Random Forest model is an ensemble learning algorithm that increases the model's accuracy and stability by training multiple decision trees in parallel. Each of the trees is trained using a different subset of the data, and the final result is determined by aggregating the outcomes of the trees. The class_weight = 'balanced' parameters are also of great use for datasets with imbalance problems, as they modify the weight of each class based on its prevalence. Other parameters of importance include n_estimators, which controls the number of decision trees in the forest with a default value of 100, max_depth, which limits how deep each tree can get and also starts as None, which allows trees to grow until all nodes are pure, and min_samples_split which determines the minimum sample count needed to split a node which is also set at two by default. Additionally, min_samples_leaf refers to the minimum number of samples to be required to form a leaf node, which starts at 1. max_features controls the number of features available for consideration of levels splits, and bootstrap = True signifies that without reserve sampling is used, meaning each of the trees is trained on a different sample set of the data. Using the random_state parameter guarantees that results will be the same across runs by setting a random seed to create trees (see Table 1).

2.1.7. Balanced bagging classifier

The Balanced Bagging Classifier is another ensemble method, which, like Random Forest, resolves the problem of imbalanced data by automatically combining imbalance handling classes with multiple decision tree learners. Balanced Bagging differs from Random Forest in that it also balances the class distribution for each bootstrapped sample by oversampling the minority class to the size of the majority class; this is done for all bootstrapped samples. This particular model helps manage highly imbalanced datasets. Key parameters are base_estimator, which by default implements a DecisionTreeClassifier as the base model, and n_estimators, with a default value of 10, which sets the count of base models in the ensemble. max_samples controls the proportion of the dataset for each model, and max_features defines the proportion of features to be used when fitting each base model. The former is set to 'None,' thus enabling all data to be used, whereas the latter is overridden to guarantee some degree of sparsity defined by the ratio s. The bootstrap parameter indicates whether or not bootstrapped sampling will be used, while the sampling_strategy='auto' parameter guarantees that the minority class will be scaled to the dominant class. Like the Random Forest, random_state sets a seed for reproducibility (see Table 1).

3. RESULTS AND DISCUSSION

The results showcase the performance of different machine learning models Random Forest, Bagging Classifier, and Balanced Bagging Classifier each evaluated with various resampling techniques, including SMOTE, ADASYN, and Tomek Links. A comparative analysis of key performance metrics such as accuracy, precision, recall, F1-score, ROC AUC, Cohen's Kappa, and uncertainty measures (aleatoric uncertainty, epistemic uncertainty, and predictive entropy) provides valuable insights into the models' effectiveness in detecting panic disorder.

3.1. Performance Evaluation of Random Forest

Table 2 shows that the accuracy performed by the baseline Random Forest model was 0.99805, giving perfect precision with a value of 1.0. However, the model's recall was comparatively lower at 0.955682. Due to this disparity, the model is quite sensitive as it demonstrates missing a notable portion of true positive cases while identifying positive cases with high confidence. The SMOTE also automated the recall improvement and F1 score to 0.977363 and

Table 1. Hyperparameters used for the model.

Model	Hyperparameter	Value
Random Forest	class_weight	'balanced' (handles imbalanced classes)
	n_estimators	Default (100)
	max_depth	Default (None, trees expand until pure leaves)
	min_samples_split	Default (2)
	min_samples_leaf	Default (1)
	max_features	Default ('auto,' square root of the number of features)
	Bootstrap	Default (True, bootstrap sampling used)
	random_state	Default (None, random seed set by the OS)
Balanced Bagging Classifier	base_estimator	Default (DecisionTreeClassifier)
	n_estimators	Default (10)
	max_samples	Default (1.0, 100% of data used for each estimator)
	max_features	Default (1.0, 100% of features used for each estimator)
	Bootstrap	Default (True, bootstrap sampling used)
	sampling_strategy	Default ('auto,' balances data by making minority class size equal to majority class)
	random_state	Default (None, random seed set by the OS)

Table 2. Evaluation metrics results.

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Cohen's Kappa
Random Forest	0.99805	1	0.95568	0.97733	0.99999	0.97632
Random Forest with SMOTE	0.99903	1	0.97736	0.98854	0.99999	0.988043
Random Forest with ADASYN	0.99915	0.999762	0.98039	0.98997	0.99999	0.989527
Random Forest with TomekLinks	0.99839	1	0.96242	0.98083	0.99999	0.979999
Bagging Classifier	0.99965	0.99546	0.99659	0.99602	0.99991	0.995842
Balanced Bagging with SMOTE	0.99931	0.9923	0.99159	0.99194	0.99996	0.991585
Balanced Bagging with ADASYN	0.99942	0.99232	0.99416	0.99323	0.99997	0.992934
Balanced Bagging with TomekLinks	0.99738	0.942666	0.99976	0.97035	0.99996	0.968983

0.988549, respectively. These changes indicate much better recall, especially in the lower class instances. This enhancement comes from SMOTE's ability to balance the dataset automatically, and as the model learns, it can define issues more cohesively. Like that, recall for Random Forest with ADASYN rose to 0.980397 while achieving a higher F1 score of 0.989971 compared to SMOTE. ADASYN performs better than SMOTE when learning the model's generalization because it is based on harder-to-classify instances requiring new synthetic samples. The opposite was true for the Random Forest with Tomek Links. When the comparison was made, the recall dropped to 0.962427, lower than previously achieved. However, the precision remained the same at 1.0. Although, the Tomek Links methodology suffers sensitivity as it is an under-sampling approach. In Table 3, uncertainty metrics revealed additional insights. Compared to models that utilized SMOTE and ADASYN, the baseline Random Forest model had relatively greater values in its aleatoric uncertainty (0.010129) and predictive entropy (0.036515). These values did drop somewhat with the use of SMOTE and ADASYN. These drops in values do demonstrate a greater degree of confidence in the predictions due to better balanced training data. Nonetheless, epistemic uncertainty fluctuated much less, suggesting consistent model performance irrespective of the resampling technique (Figure 4).

3.2. Performance of Bagging Classifier and Balanced Bagging Classifier

The Bagging Classifier demonstrated exceptional performance, achieving an accuracy of 0.99965 alongside a remarkable recall of 0.996591 and an F1 score of 0.996025. The model performed optimally due to the ability of the ensemble to cut down on variance and issues arising from overfitting. Surprisingly, aleatoric uncertainty was much lower (0.00062), and predictive entropy was also very low (0.002003), suggesting that the model was highly confident in its predictions and, therefore, the model was quite reliable. The balanced ensemble approach already solves the problem of disproportionate class distribution and, therefore, does not require resampling. Once again, applying SMOTE and ADASYN techniques proved beneficial for the Balanced Bagging Classifier. Balanced Bagging with SMOTE achieved a 0.99931 accuracy score, 0.991599 recall, and F1-score of 0.991945. ADASYN achieved slightly better results than SMOTE, with a recall of 0.994166 and an F1 score of 0.993237 (Figure 3). It proves that the targeted sampling techniques employed by ADASYN accurately overcame restrictions posed by the minority class (see Table 2). It is also worth noting that both models retained low levels of uncertainty, with aleatoric uncertainty being approximately 0.00068 and predictive entropy sitting at 0.0021. It can be taken as a sign of the robustness of the models. As for the Balanced Bagging Classifier trained with Tomek Links, the results were mixed. Although the model achieved a remarkable recall of 0.999767, the precision dropped to 0.942666, resulting in a lower F1-score of 0.970352, as shown in Table 3. It suggests that while the model could correctly identify almost all positive cases, it also had too many false positives. The increase in aleatoric uncertainty (0.004502) and predictive entropy (0.013946) indicates that the model had

Table 3. Uncertainty results of the data set with different classifiers.

Model	Aleatoric Uncertainty	Epistemic Uncertainty	Predictive Entropy
Random Forest	0.010129	0.030084	0.036515
Random Forest with SMOTE	0.009757	0.032701	0.036485
Random Forest with ADASYN	0.009711	0.032793	0.03621
Random Forest with TomekLinks	0.010639	0.028976	0.038413
Bagging Classifier	0.00062	0.04143	0.002003
Balanced Bagging with SMOTE	0.00068	0.040308	0.002127
Balanced Bagging with ADASYN	0.000749	0.040301	0.002311
Balanced Bagging with TomekLinks	0.004502	0.042529	0.013946

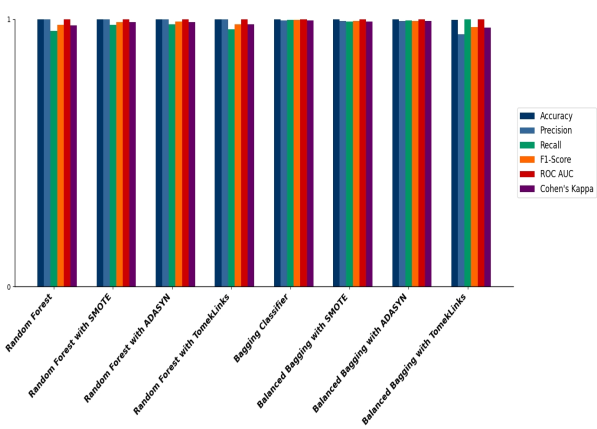

Fig. 3. Comparative analysis of evaluation matrices.

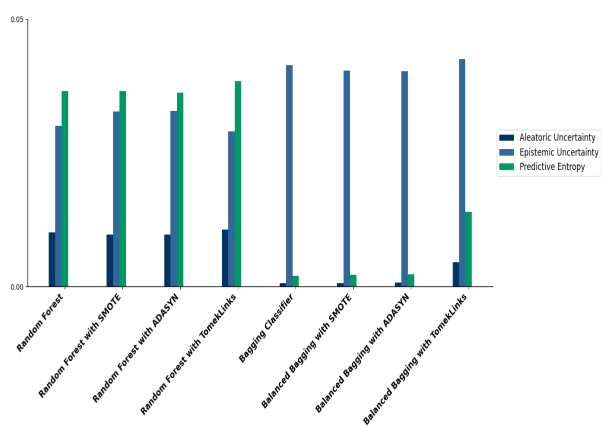

Fig. 4. Comparative analysis of uncertainties.

less confidence due to the more radical removal of borderline samples done by Tomek Links (see Figure 4).

3.3. Comparative Insights

While comparing to [24], it can be noticed that our models, particularly Random Forest with ADASYN (99.915% Accuracy), have shown marked and significant improvement in accuracy over the Gaussian SVM (94.5%). The application of resampling methods, particularly ADASYN, has greatly improved classification performance, surpassing raw features and HRMAD results [24]. These findings resonated in [21], where SMOTE was reported to increase the accuracy of Random Forest from 82% to 97%, demonstrating the positive effects of resampling on class imbalance and overall model performance. In regards to precision and recall, our models outperform previously published results. For example, Random Forest with ADASYN also attained one of the highest recorded precision scores of 0.999762 and recall scores of 0.98039, outperforming the precision (0.632) and recall (0.667) scores in [30] for Random Forest. The results from [30] indicate that integrating ADASYN and other ensemble changes is much more effective in decreasing false positives and true negatives than models. Additionally, our F1-Score results with Random Forest with ADASYN (0.98997) and Bagging Classifier (0.99602) far exceed [30] SVM (0.733) and Logistic Regression (0.790), meaning that our models achieve a preferable balance between recall and precision. It indicates that ensemble methods such as Bagging are more effective at optimizing the balance between false positive and negative rates than traditional methods. Our models achieve almost perfect values (0.99999) for ROC AUC, far exceeding the AUC of 0.8255 recorded for ANN in [34]. This sharp deviation reinforces the claim that my models, particularly those with resampling techniques, remarkably outperformed previous models in distinguishing panic disorder from other conditions. Moreover, the remarkably high Cohen's Kappa score (0.989527) accompanies my results, demonstrating divergence from those values against the consensus that he was predicting outcome targets, bolstered by the previously cited Kappa of 0.298 from Random Forest in [30], highlighting that his assertion was closer to the reality.

The uncertainty metrics set out in Table 3 are indicators of considerable disparities in a model's performance. It has been noted that using a Random Forest with TomekLinks (a preprocessing method for cleaning up the data and reducing the imbalanced training set) has increased the aleatoric uncertainty, which is the noise of the prediction and the class overlap, indicating that the model is facing greater difficulty in distinguishing between the classes. This higher uncertainty resulted in unreliable classification. However, epistemic uncertainty, which exhibited model uncertainty due to either a limit in the data or insufficient learning, did not exhibit consistent fluctuations instead of moving up and down among models, which would have suggested that the model, even with the resampling technique used, is still having a good performance. It is also notable that the predictive entropy, which is a concept of what the prediction's uncertainty is about, is quite much lower in the case of the Bagging Classifier models as a result of their solid confidence and stability in classification (see Figure 3). It gives a cue that methods like Bagging, which uses ensembles, are much more capable of giving accurate predictions while at the same time reducing the uncertainty compared to other approaches.

In summary, ensemble models, specifically the Balanced Bagging Classifier with ADASYN, recorded the best balance between performance metrics and uncertainty measures. This research also illustrates the importance of resampling methods to improve model sensitivity and reliability, which requires a specific framework depending on the nature of the data set.

4. CONCLUSIONS

This research shows the advanced applications of ensemble machine learning techniques using Random Forest, Bagging Classifier, and Balanced Bagging Classifier that diagnose panic disorder utilizing SMOTE, ADASYN, and Tomek Links for class imbalance resampling. Notably, the Balanced Bagging Classifier with ADASYN outperformed the competition by achieving greater accuracy, recall, precision, and F1 scores combined with low uncertainty metrics, which signify strong and dependable predictions. The addition of uncertainty quantification reduced the chances of misclassification while increasing rationale within the model, thereby making the model more useful from a clinical standpoint. Despite promising results, further analysis utilizing diverse datasets alongside real-time data from wearable devices can augment performance and adaptability. This work illustrates the significant impacts of using machine learning in diagnostics within the mental health field, thereby providing a benchmark for developing efficient and tailored treatments for panic disorders and psychiatry in general.

5. ACKNOWLEDGMENT

The authors are greatly indebted to the Superior University, Lahore, for providing all the necessary material and funding for this research. We thank the colleagues of the Department of Basic Sciences and the Department of Computer Sciences for their valuable comments and guidance.

6. ETHICAL STATEMENT

All research involving medical data was done ethically using best practices, which was done by the protocols set up. The Institutional Review Board reviewed the protocol for this study at the Superior University of Lahore, Pakistan. In this study, all materials were handled confidentially and remained anonymous.

7. CONFLICT OF INTEREST

The authors have no conflict of interest regarding this article.

8. REFERENCES

1. A.E. Meuret, J.L. Kroll, and T. Ritz. Panic disorder comorbidity with medical conditions and treatment implications. *Annual Review of Clinical Psychology* 13(1): 209-240 (2017).
2. L. Spytska. Exploration of personality characteristics of individuals with panic attacks: A personality profile study. *Journal of Infant, Child, and Adolescent Psychotherapy* 84(1): 84-94 (2025).
3. M.G. Abrignani, N. Renda, V. Abrignani, A. Raffa, S. Novo, and R. Lo Baido. Panic disorder, anxiety, and cardiovascular diseases. *Clinical Neuropsychiatry* 11(5): 130-144 (2014).
4. A.M. Krypotos, G. Mertens, D. Matziarli, I. Klugkist, and I. M. Engelhard. Fear generalization in individuals with subclinical symptoms of panic disorder. *Behaviour Research and Therapy* 1(184): 104649 (2025).
5. P. Porcelli. Fear, anxiety, and health-related consequences after the COVID-19 epidemic. *Clinical Neuropsychiatry* 17(2): 103-111(2020).
6. J.F. Brosschot, B. Verkuil, and J.F. Thayer. The default response to uncertainty and the importance of perceived safety in anxiety and stress: An evolution-theoretical perspective. *Journal of Anxiety Disorders* 41(1): 22-34 (2016).
7. M.G. Craske, M.B. Stein, T.C. Eley, M.R. Milad, A. Holmes, R.M. Rapee, and H.U. Wittchen. Anxiety disorders. *Nature Reviews. Disease Primers* 3: 17024 (2017).
8. B. Bandelow, D. Baldwin, M. Abelli, C. Altamura, B. Dell'Osso, K. Domschke, N.A. Fineberg, E. Grünblatt, M. Jarema, E. Maron, and D. Nutt. Biological markers for anxiety disorders, OCD, and PTSD–a consensus statement. Part I: Neuroimaging and genetics. *The World Journal of Biological Psychiatry* 17(5): 321-365 (2016).
9. S.G. Hofmann (Ed.). *An Introduction to Modern CBT: Psychological Solutions to Mental Health Problems. John Wiley & Sons* (2011).
10. N.R. Rayburn and M.W. Otto. Cognitive-behavioral therapy for panic disorder: A review of treatment elements, strategies, and outcomes. *CNS Spectrums* 8(5): 356-362 (2003).
11. S.G. Farri, L. Derby, and M.M. Kibbey. Getting comfortable with physical discomfort: A scoping review of interoceptive exposure in physical and mental health conditions. *Psychological Bulletin* 151(2): 131-191 (2025).
12. G. Chouinard and V.A. Chouinard. A new classification of selective serotonin reuptake inhibitor withdrawal. *Psychotherapy and Psychosomatics* 84(2): 63-71 (2015).
13. S.S. Clevenger, D. Malhotra, J. Dang, B. Vanle, and W.W. Ishak. The role of selective serotonin reuptake inhibitors in preventing relapse of major depressive disorder. *Therapeutic Advances in Psychopharmacology* 8(1): 49-58 (2018).
14. D. Lochmann and T. Richardson. Selective serotonin reuptake inhibitors. In: Antidepressants. Handbook of Experimental Pharmacology. M. Macaluso and S. Preskorn (Eds.). Volum 250. *Springer Nature Switzerland* pp. 135-144 (2018).
15. U. Madububambachu, A. Ukpebor, and U. Ihezue. Machine learning techniques to predict mental health diagnoses: A systematic literature review. *Clinical Practice and Epidemiology in Mental Health* 20: e17450179315688 (2024).
16. F. Akhtar, M.B.B. Heyat, A. Sultana, S. Parveen,

H. Muhammad Zeeshan, S.F. Merlin, B. Shen, D. Pomary, J. Ping Li, and M. Sawan. Medical intelligence for anxiety research: Insights from genetics, hormones, implant science, and smart devices with future strategies. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 14(6): e1552 (2024).

17. G. Prieto-Avalos, N.A. Cruz-Ramos, G. Alor-Hernandez, J.L. Sánchez-Cervantes, L. Rodriguez-Mazahua, and L.R. Guarneros-Nolasco. Wearable devices for physical monitoring of heart: A review. *Biosensors* 12(5): 292 (2022).
18. E.A. Poweleit, A.A. Vinks, and T. Mizuno. Artificial intelligence and machine learning approaches to facilitate therapeutic drug management and model-informed precision dosing. *Therapeutic Drug Monitoring* 45(2): 143-150 (2023).
19. N.K. Iyortsuun, S.H. Kim, M. Jhon, H.J. Yang, and S. Pant. A review of machine learning and deep learning approaches on mental health diagnosis. *Healthcare* 11(3): 285 (2023).
20. K. Nova. Machine learning approaches for automated mental disorder classification based on social media textual data. *Contemporary Issues in Behavioral and Social Sciences* 7(1): 70-83 (2023).
21. D. Nurmalasari, H.R. Yuliantoro, and D.H. Qudsi. Improving panic disorder classification using SMOTE and random forest. *Journal of Applied Informatics and Computing* 8(2): 272-279 (2024).
22. P. Sahebi. Enhancing user experience for real-time panic attack detection with wearable technology: A human-computer interaction approach with machine learning integration. *International Journal of Advanced Human-Computer Interaction* 2(2): 55-66 (2024).
23. E.H. Jang, K.W. Choi, A.Y. Kim, H.Y. Yu, H.J. Jeon, and S. Byun. Automated detection of panic disorder based on multimodal physiological signals using machine learning. *ETRI Journal* 45(1): 105-118 (2023).
24. I. Lazarou, A.L. Kesidis, G. Hloupis, and A. Tsatsaris. Panic detection using machine learning and real-time biometric and spatiotemporal data. *ISPRS International Journal of Geo-Information* 11(11): 552 (2022).
25. C.H. Tsai, P.C. Chen, D.S. Liu, Y.Y. Kuo, T.T. Hsieh, D.L. Chiang, F. Lai, and C.T. Wu. Panic attack prediction using wearable devices and machine learning. Development and cohort study. *JMIR Medical Informatics* 10(2): e33063 (2022).
26. E.W. McGinnis, B. Loftness, S. Lunna, I. Berman, S. Bagdon, G. Lewis, M. Arnold, C.M. Danforth, P.S. Dodds, M. Price, and W.E. Copeland. Expecting the unexpected: Predicting panic attacks from mood, Twitter, and Apple Watch data. *IEEE Open Journal of Engineering in Medicine and Biology* 5(1): 14-20 (2024).
27. I.H. Bell, S.L. Rossell, J. Farhall, M. Hayward, M.H. Lim, S.F. Fielding-Smith and N. Thomas. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): Feasibility, acceptability and preliminary clinical outcomes. *Schizophrenia Research* 216: 479-487 (2020).
28. D.J. Robinaugh, M.J. Ward, E.R. Toner, M.L. Brown, O.M. Losiewicz, E. Bui, and S.P. Orr. Assessing vulnerability to panic: A systematic review of psychological and physiological responses to biological challenges as prospective predictors of panic attacks and panic disorder. *General Psychiatry* 32(6): e100140 (2019).
29. S. Jang, T.H. Sun, S. Shin, H.J. Lee, Y.B. Shin, J.W. Yeom, Y.R. Park, and C.H. Cho. A digital phenotyping dataset for impending panic symptoms: A prospective longitudinal study. *Scientific Data* 11(1): 1264 (2024).
30. K.S. Na, S.E. Cho, and S.J. Cho. Machine learning-based discrimination of panic disorder from other anxiety disorders. *Journal of Affective Disorders* 278: 1-4 (2021).
31. S. Ogawa. Predicting dropout from cognitive behavioral therapy for panic disorder using machine learning algorithms. *Journal of Clinical Medicine Research* 16(5): 251 (2024).
32. C.H. Tsai, P.C. Chen, D.S. Liu, Y.Y Kuo, T.T Hsieh, D.L Chiang, F. Lai and C.T Wu. Panic attack prediction using wearable devices and machine learning: development and cohort study. *JMIR Medical Informatics* 10(2): e33063 (2022).
33. C. Pae, H.J. Kim, M. Bang, C.I. Park, and S.H. Lee. Predicting treatment outcomes in patients with panic disorder: Cross-sectional and two-year longitudinal structural connectome analysis using machine learning methods. *Journal of Anxiety Disorders* 106(2): 102895 (2024).
34. D.J. Robinaugh, J. Haslbeck, L.J. Waldorp, J.J. Kossakowski, E.I. Fried, A.J. Millner, R.J. McNally, O. Ryan, J. de Ron, H.L. van der Maas and E.H. van Nes. Advancing the network theory of mental disorders: A computational model of panic disorder. *Psychological Review* 131(6): 1482 (2024).
35. R. Rajendran and A. Karthi. Heart disease prediction using entropy-based feature engineering and

ensembling of machine learning classifiers. *Expert Systems with Applications* 20(7): 117882 (2022).

36. X. Peng, W. Wu, Y. Zheng, J. Sun, T. Hu, and P. Wang. Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. *Scientific Reports* 10(1): 10451 (2020).
37. O.L. Usman, R.C. Muniyandi, K. Omar, and M. Mohamad. Gaussian smoothing and modified histogram normalization methods to improve neural-biomarker interpretations for dyslexia classification mechanism. *PLoS One* 16(2): e0245579 (2021).
38. F. Gurcan and A. Soylu. Synthetic boosted resampling using deep generative adversarial networks: A novel approach to improve cancer prediction from imbalanced datasets. *Cancers* 16(23): 4046 (2024).
39. S. Nouas, L. Oukid, and F. Boumahdi. Syngo: Synthetic genetic oversampling technique for textual data. *Social Network Analysis and Mining* 15(1): 9 (2025).
40. S.S. Dhaliwal, A.A. Nahid, and R. Abbas. Effective intrusion detection system using XGBoost. *Information* 9(7): 149 (2018).
41. A.S. More and D.P. Rana. Review of random forest classification techniques to resolve data imbalance. *Proceedings of the 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* pp. 72-78 (2017).
42. S. Hido, H. Kashima, and Y. Takahashi. Roughly balanced Bagging for imbalanced data. *Statistical Analysis and Data Mining. The ASA Data Science Journal* 2(5-6): 412-426 (2009).
43. 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).