**Department of Electrical and Computer Engineering**

**North South University**



**CSE445 Report**

**Panic Disorder Detection**

**Pranjal Saha Arnab ID # 2031531642**

**Abdullah Al Sayem ID # 2022207042**

**Abu Bakar Siddik Minhaz ID # 2031589642**

**Anusree Roy ID # 2011364042**

**Md Tahsinul Islam**  **ID # 1921232042**

**Faculty:**

**Riasat Khan**

**Assistant Professor**

**ECE Department**

**Spring, 2024**

**Individual Contribution Table**

|  |  |  |
| --- | --- | --- |
| **Section** | **Contributing Member Name** |  |
| IEEE/LaTEX formatting | Pranjal Saha Arnab |  |
| Grammarly check | Abdullah Al Sayem | Grammarly Score: 93 |
| Abstract | Abu Bakar Siddik Minhaz |  |
| Keywords | Abu Bakar Siddik Minhaz & Anusree Roy |  |
| Introduction Motivation | Abdullah Al Sayem |  |
| Paper Review 1 | Abdullah Al Sayem | [5] |
| Paper Review 2 | Pranjal Saha Arnab | [6] |
| Paper Review 3 | Abu Bakar Siddik Minhaz | [7] |
| Paper Review 4 | Anusree Roy | [8] |
| Paper Review 5 | Md Tahsinul Islam | [9] |
| Introduction Second-Last Paragraph (describe your work) | Abdullah Al Sayem |  |
| Proposed System (Dataset and Preprocessing) | Anusree Roy &  Md Tahsinul Islam |  |
| Proposed System (Model description) | Decision tree | Anusree Roy |
| SVM |
| Random forest |
| Gradient Boosting |
| Ada Boosting |
| Results and Discussion | Abu Bakar Siddik Minhaz |  |
| Figure and Table Title Formatting | Pranjal Saha Arnab |  |
| Conclusions | Abdullah Al Sayem |  |
| Equations formatting | Abdullah Al Sayem |  |
| References Formatting in IEEE format | Pranjal Saha Arnab |  |

Panic Disorder Detection Using Machine Learning

Arnab Pranjal Saha   
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
pranjal.arnab@northsouth.edu

Md Tahsinul Islam  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
tahsinul.islam08@northsouth.eduAbu Bakar Siddik Minhaz  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
abu.minhaz@northsouth.edu

Abdullah Al Sayem  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
abdullah.sayem01@northsouth.edu Anusree Roy  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
anusree.roy@northsouth.edu

*Abstract*— This project aims to identify the applicability of several machine learning algorithms in diagnosing panic disorder by employing an open-access database that includes 1,20,000 data records complemented by multiple attributes associated with panic attacks. The study evaluated five models: The algorithms used are Decision Tree, Random Forest, Gradient Boosting Classifier, Ada Boost Classifier, and Support Vector Machine (SVM) in both default form and with optimized hyperparameters. The Gradient boosting Classifier and the Support Vector Machine are the two algorithms that showed near-perfect results in Accuracy, Precision, Recall, and F1-score, showing that these two algorithms are best for detecting panic disorders. The results achieved by Decision Tree and Random Forest were also good, although recall values were a bit lower than those of the highest-performing algorithms; AdaBoost performed worse. There was not much improvement of performance based on the hyperparameter tuning for most of the models indicating that the initial hyperparameters were already best for the models. The current study encourages the use of machine learning methods, especially Gradient Boosting and SVM in clinical diagnosis of panic disorder, thus helpful in the development of efficient diagnostic tools. Therefore, future studies should confirm these outcomes with larger and more diverse sample sizes and examine the influence of machine learning techniques better.

Keywords—Gradient Boosting, Ensemble Learning, Grid Search, Feature Scaling, Explainable AI

# Introduction

A panic attack is a strong, sudden act of fear or inconvenience that, as a rule, crests in some minutes. A few of the indicators include- heart palpitations, sweating, shaking, dyspnea, chest torment, sickness, vertigo, and a feeling of losing control or going insane. Despite being of an extensive kind, they can be exceptionally annoying. Panic attacks are liable to a combination of genetic predispositions and stress-related environmental precursors, like those of the major life transitions, sharply stressful events such as the death of a beloved one, and even some adversities in childhood, and often come suddenly without clear situational triggers. The consequences of such attacks are marked distress and dysfunction, leading to persistent worry about future attacks and life changes to avoid another attack [1]. Panic disorder impacts about 2.5 out of 100 people once in their lifetime, and also studies show that it is more common in women than in men [2]. Panic attacks can be detected through physical examinations like- fMRI, ECG & blood tests, along with psychological evaluations through the patient’s symptoms likestress, fears, family background, etc. [3]. It is possible to treat panic disorder with mental carefulness. Meeting a psychiatrist, taking psychotherapy sessions, and taking relevant medication would cure the process of panic disorder [4].

There are numerous research studies found in recent literature as well as systems in production that work with panic disorder and detection of panic disorder based on several approaches. For instance, Na and his team [5] applied machine learning algorithms to distinguish panic disorder from other anxiety disorders. The authors used a dataset of a medical chart record where heart rate variability (HRV) was used as input. To collect HRV input, they had to maintain the main frequency band of the HRV signal between 0 to 0.4 Hz. Among various algorithms, regularized logistic regression (LR) achieved the highest accuracy of 0.784, and its f1 score was 0.79.

Aderinwale and colleagues [6] studied to figure out a diagnosis of panic disorder using machine learning. Data was collected from 149 patients of which 40 were panic disorder patients and the rest were healthy and depression disorder patients. They used an 80-20 ratio for training and testing where the LASSO regression model was used for feature extraction. The SVM classifier was used to conclude 60% accuracy in classifying panic disorder patients along with the controls and 59% accuracy between panic disorder and Depression disorder patients.

Tsai et al. [7] developed a machine learning-based model to predict panic attacks. It uses a multifaceted dataset taken from 59 participants who were diagnosed with panic disorder over the course of a year. In the dataset, physiological data was obtained from Garmin Vivosmart 4 smartwatches, environmental data was obtained from local monitoring stations, and responses to clinical questionnaires, such as the Beck Depression Inventory and the State-Trait Anxiety Inventory, were obtained through a mobile app. Among six machine learning algorithms used in the study, the random forest model came out on top. It showed an accuracy between 67.4% and 81.3% on the test set. Through this approach, the utility of integrating wearable device data with environmental and self-reported psychological data was highlighted, as they allowed for the prediction of panic attacks up to seven days in advance. However, there were certain limitations in the study, such as a small sample size, potential recall bias from self-reported data, and possible inaccuracies in the environmental data, causing it to fail in reflecting the exact locations of panic attacks. These limitations may affect the predictive accuracy of the model.

Lazarou et al. [8] used machine-learning approaches for panic detection. They introduced biometric and spatiotemporal data containing a multimodal dataset related to panic detection and trained the dataset with various machine learning models along with a deep learning framework. Out of the different tested classifies, the Gaussian SVM classifier ranked top achieving a score of 94.5% and the LSTM approach reaching an accuracy of 94%.

Eun-Hye-Jang and associates [9] attempted to forecast panic attacks by employing a variety of machine learning techniques based on ensemble methods. They made use of a dataset that included environmental, physiological, and survey data. Out of the 67 individuals, 344 data sets were chosen in total. The authors' accuracy metric was the area under the ROC curve or AUC. The ensemble classifier that was suggested managed to get an AUC of 0.86.

The current study gathered data from a public dataset containing more than 1.2 lac records about patients with panic disorder. We employed imputation techniques for handling missing values, one-hot encoding for handling nominal categorical data, and standard scaling for feature scaling during the dataset preprocessing phase. Additionally, we handled imbalanced data using synthetic data generation techniques. Moreover, we used feature selection techniques to select a subset of effective features from the original dataset, thereby avoiding the complexity of the machine learning model. After completing data preprocessing, we applied five machine-learning models to the preprocessed dataset. We used explainable AI techniques to interpret the behavior of the machine learning models.

The present study briefly describes the proposed machine learning approach to detecting panic disorder in four sections. We used Section II to explain the proposed system to detect panic disorder, which included tables, figures, and flowcharts of the working sequence. Section III presents the results in tables or figures, showcasing hyperparameter values and performance metrics with and without default values and comparing the proposed system. Finally, Section IV concludes the paper and discusses future improvements to this work.

# Proposed System

The proposed system uses machine learning techniques for predicting panic disorder from the dataset's features.

## Dataset

The dataset contains 1,20,000 records and 16 features, which we collected from Kaggle. The training split contains 100000, and the test split contains 20000 rows. The features are- participant ID, age, gender, family history, personal history, current stressors, symptoms, severity, impact on Life, demographics, medical history, psychiatric history, substance use for example, use of substances like alcohol or drugs, coping mechanisms, social support, lifestyle factors like influences of diet, sleep quality and exercise and panic disorder diagnosis being the label. Fig. 1 demonstrates the box plot and histogram of the age feature of the dataset. Fig. 2 illustrates the distribution of the gender features of the dataset which shows that the dataset contains both genders in equal proportions.

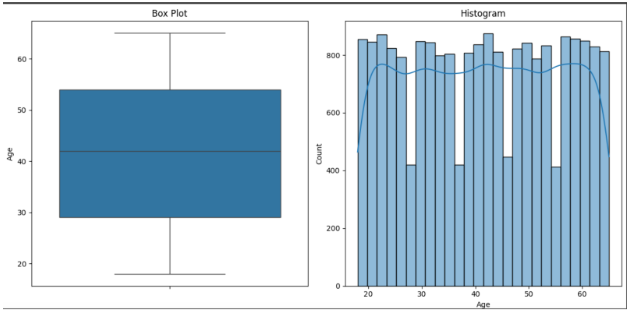


Fig. 1. Box plot and histogram of the age feature of the dataset.

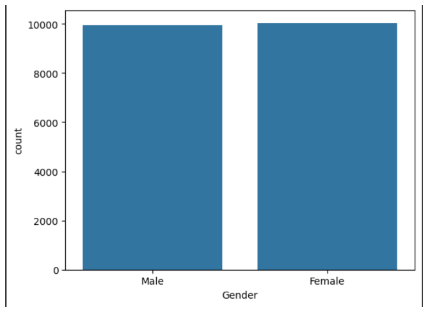


Fig. 2. Distribution of Gender features of the dataset

## Dataset Preprocessing

The dataset went through preprocessing. We implied various techniques like imputation, feature scaling, and label encoding. Then, we discarded the participant ID feature of the dataset from both train and test data. We also performed data imputation by replacing missing values with the value- unknown. For standardizing the feature values of the dataset and converting them to numeric representation, we used one-hot-encoding on all the features except the ‘age’ feature, which is already a numerical feature. Then, the age feature was transformed into a range from 0 to 1 with feature scaling as the magnitudes of this feature were greater compared to other features. Also, we used seaborn library to find a correlation matrix among the correlated features of the training dataset.

We used standardization as a feature scaling technique which mainly relied on the central tendencies and variance of the data. The equation we used to do feature scaling is,

(1)

Here, = mean of the data, = Standard deviation

## Machine Learning Models

### Random-Forest: Several weak learners are created and combined to form a strong learner whose accuracy is much higher compared to a single decision tree model. It also ensures that the model is not overfitting and that the chance of the model having a high accuracy just on the training data is reduced. Such a type of model is called an ensemble model, and the random forest is an example of it as it comprises many decision trees and yields a good learner in the training phase. The accuracy also increases in the process.

### Decision tree: A tree-type structure algorithm containing a root node where the decision tree starts, below that some other nodes represent the other attributes, and the class label is represented by the leaf node. This type of structure creates a hierarchy and the model is known as a decision tree. This model is suitable for classification as well as regression problems.

### Gradient Boosting: An ensemble model that sequentially combines multiple decision trees where each tree aims to correct the errors of the previous ones. This model improves accuracy and reduces bias by optimizing the gradient descent and reducing the loss function. It can handle complex datasets and offer strong performance, but the hyperparameters should be tuned properly to reduce overfitting.

### Ada Boost: It is called ‘Adaptive Boosting’ which is an ensemble learning technique that combines multiple weak classifiers like decision trees sequentially and creates a strong classifier by focusing on the misclassified instances of the previous ones. It is effective in reducing bias and variance. However, in the case of outliers and noisy data, it is sensitive.

### SVM: Used for both the classification and regression tasks, also known as support vector machine, SVM is a supervised machine learning model. It tries to maximize the margin between different classes in the feature space by finding the optimal hyperplane.

In the present work, accomplishing the objectives depends on the following structured workflow demonstrated in Fig.3. Finally, we were able to perform and accomplish all the general steps which range from the splitting and preprocessing of the dataset to the training and evaluation of the models in the pattern as shown in the workflow diagram.

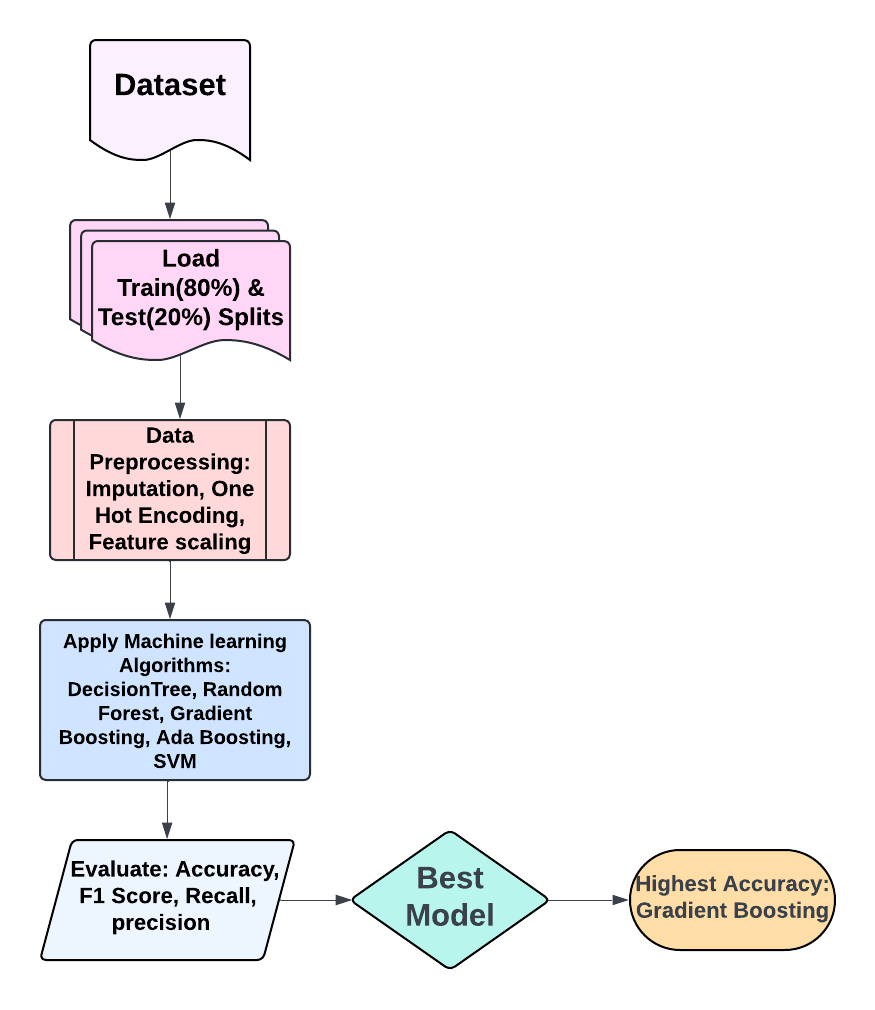


Fig. 3. Workflow of the proposed Panic Disorder Detection system.

# Results and Discussion

This study evaluated several machine learning models for detecting panic disorders. These included Decision Tree, Random Forest, Gradient Boosting Classifier, AdaBoost, and Support Vector Machine (SVM), with the top performers being Gradient Boosting Classifier and SVM. They boasted near-perfect metrics with accuracy and precision of 1 and recall and an F1 score of 0.99 or greater. Then comes Decision Tree (accuracy is around 0.997) and Random Forest(accuracy is around 0.993) but with slightly lower recall values. AdaBoost had much lower precision (0.86) and recall (0.62) even though it had a high accuracy. Hyperparameter optimization had a negligible impact on the models. Therefore, we can conclude that their initial configurations are robust. In conclusion, the most effective model for this project would be the Gradient Boosting Classifier.

.

TABLE I. HYPERPARAMETER VALUES’ RANGES FOR VARIOUS ML MODELS

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameter Value Range | Optimized value |
| SVM | C: [0.1, 1, 10, 100],  gamma: [1, 0.1, 0.01, 0.001],  kernel: ['rbf', 'poly', 'sigmoid'] | **C: 0.1,**  **gamma: 1**  **kernel: poly** |
| Random Forest | n\_estimators: [100, 200, 300]  max\_depth: [None, 10, 20, 30]  criterion: [gini, entropy] | criterion: entropy **max\_depth:** None n\_estimators: 200 |
| Decision Tree | criterion:[gini, entropy],  max\_depth: [None, 10, 20, 30] | **criterion: entropy**  **max\_depth: 10** |
| **Gradient Boosting** | **criterion: [friedman\_mse]**  **loss: [deviance],**  **learning\_rate: [0.01, 0.1]**  **n\_estimators: [10, 50]** | **criterion: friedman\_mse**  **loss: deviance**  **learning\_rate: 0.01**  **n\_estimators: 50** |
| AdaBoost | **n\_estimators: [50, 100, 200]**  **learning\_rate: [0.01, 0.1, 1],**  **base\_estimator\_\_max\_depth: [1, 2, 3]** | **n\_estimators: 50**  **learning\_rate: 0.01**  **base\_estimator\_\_max\_depth: 1** |

Table I illustrates the hyperparameter values’ ranges and the corresponding optimized hyperparameters for all the ML models.

TABLE II. PERFORMANCE METRICS OF VARIOUS ML MODELS WITH DEFAULT HYPERPARAMETERS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy/MAE | Precision/MSE | Recall/RMSE | F1-score/R2 coefficient |
| SVM | **0.99915** | **1** | **0.99** | **0.99** |
| Random Forest | **0.99337** | **0.98** | **0.94** | **0.96** |
| **Gradient Boosting** | **0.99985** | **1** | **1** | **1** |
| Decision Tree | **0.99716** | **0.99** | **0.98** | **0.98** |
| Ada Boost | **0.96382** | **0.86** | **0.62** | **0.67** |

Performance metrics of various ML models with default hyperparameters have been illustrated in Table II.

TABLE III. PERFORMANCE METRICS OF VARIOUS ML MODELS WITH OPTIMIZED HYPERPARAMETERS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy/MAE | Precision/MSE | Recall/RMSE | F1-score/R2 coefficient |
| SVM | **0.99915** | **1** | **0.99** | **0.99** |
| Random Forest | **0.99337** | **0.98** | **0.94** | **0.96** |
| **Gradient Boosting** | **0.99985** | **1** | **1** | **1** |
| Decision Tree | **0.99716** | **0.99** | **0.98** | **0.98** |
| Ada Boost | **0.96382** | **0.86** | **0.62** | **0.67** |

Performance metrics of various ML models with optimized hyperparameters have been illustrated in Table III.

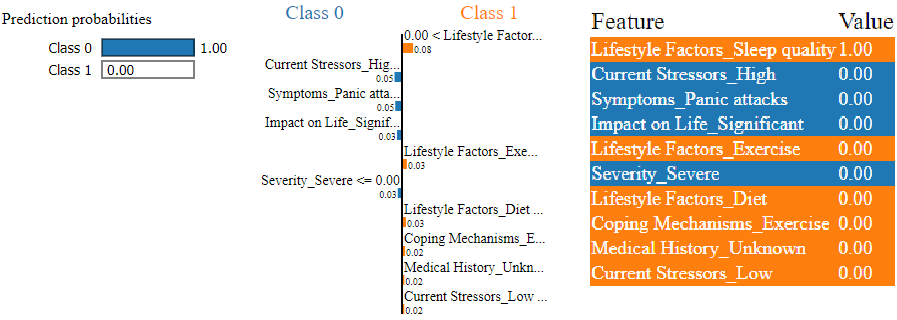


Fig.4: Machine learning model prediction interpretation by LIME explainable AI library.

Fig. 4 shows the machine learning model prediction interpretation by LIME explainable AI library. It basically shows the reasoning behind our model's decision-making for the predictions it made.

Table IV. COMPARISON OF THE PROPOSED SYSTEM WITH EXISTING WORKS

|  |  |  |  |
| --- | --- | --- | --- |
| Ref. | Model | Accuracy | F1-score |
| [2] | LSTM | 94% | N/A |
| [3] | Logistic Regression | 78.4% | 79% |
| This work | **Gradient Boosting** | **99%** | 100% |

Table IV illustrates comparison of the panic disorder detection system with other existing works on the Panic Disorder Detection Dataset.

# Conclusions

In this work, we aimed to detect panic disorder which is a classification-type problem using several features related to panic attacks. To do the job, we built a proposed system that contains data preprocessing, feature selection, model building, and explainable AI techniques. We have employed 5 different machine learning models. Regarding machine learning, the best classification results came from a gradient boosting model, reaching an accuracy of 0.9995. Also on Gaussian SVM classifiers, almost similar results have been shown after hyperparameter optimization whose accuracy was 0.9991. After all, we applied explainable AI techniques to interpret the machine learning model. This proposed system further demonstrated that a machine-learning approach can be the basis for an efficient panic disorder detection system. Future work involves an approach where deep learning algorithms can be applied to detect panic disorder. Also, a hybrid dataset can be used where a combination of public and private data will be present in the entire dataset.

##### References

1. D. Bonevski and A. Naumovska, ‘Panic Attacks and Panic Disorder’, Psychopathology - An International and Interdisciplinary Perspective. IntechOpen, Jan. 22, 2020..
2. M. G. Craske and M. B. Stein, "Anxiety," The Lancet (London, England), vol. 388, no. 10063, pp. 3048-3059, 2016.
3. C. H. Tsai, P. C. Chen, D. S. Liu, Y. Y. Kuo, T. T. Hsieh, D. L. Chiang, F. Lai, C. T. Wu, "Panic Attack Prediction Using Wearable Devices and Machine Learning: Development and Cohort Study," JMIR Med Inform, vol. 10, pp. e33063, 2022.
4. National Institute of Mental Health, “Panic disorder: When Fear Overwhelms,” *U.S. Department of Health and Human Services,* NIH Publication No. 22-MH-8077. [Online]. Available: https://www.nimh.nih.gov/health/publications/panic-disorder-when-fear-overwhelms. [Accessed: May. 23, 2024]
5. 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, vol. 278, pp. 1–4, 2021.
6. A. Aderinwale, G. B. Tolossa, Ah Y. Kim, E. H. Jang, Y. Lee, H. J. Jeon, H. Kim, H. Y. Yu, J. Jeong, “Two-channel EEG based diagnosis of panic disorder and major depressive disorder using machine learning and non-linear dynamical methods,” Psychiatry Research: Neuroimaging, vol. 332, 2023.
7. 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, vol. 10, 2022.
8. I. Lazarou, A. L. Kesidis, G. Hloupis, and A. Tsatsaris, “Panic Detection Using Machine Learning and Real-Time Biometric and Spatiotemporal Data,” International Journal of Geo-Information, vol. 11, pp. 552, 2022.
9. 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, vol. 45 2023, pp. 105–118.