Novelty Journals

ISSN 2394-966X

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (18-23), Month: January - February 2025, Available at: www.noveltyjournals.com

Panic Disorder Detection Using Machine Learning

[1]Faezeh Norouzi, [2]Bruna Lino Modes Santos Machado

[1]Psychiatry and Behavioral Science, Isfahan University of Medical Science, Iran

[2]Instituto Luterano de Ensino Superior de Itumbiara, Brazil

DOI: https://doi.org/10.5281/zenodo.14738528

Published Date: 25-January-2025

Abstract: **Panic disorder is a debilitating mental health condition that requires early and accurate detection for effective treatment. This study employs a Gradient Boosting Classifier to detect panic disorder using a dataset of 120,000 instances, split into training, validation, and testing subsets. Preprocessing included one-hot encoding and standardization to ensure data suitability. The model achieved a training accuracy of 99.997% and a testing accuracy of 98%. A confusion matrix analysis revealed excellent performance, with 19,158 true negatives, 841 true positives, one false positive, and no false negatives, highlighting high precision and recall. These results demonstrate the potential of Gradient Boosting for reliable mental health diagnostics, supporting early intervention and improved outcomes.**

Keywords: **Diagnosis; Gradient Boosting; Machine Learning; Panic Disorder.**

I. INTRODUCTION

Mental health is fundamental to an individual's overall well-being, influencing their thoughts, emotions, and behaviors. It affects how people handle stress, build relationships, and make decisions in their daily lives. Despite its importance, mental health is often overlooked or stigmatized, leading to inadequate attention and resources devoted to addressing mental health issues on a global scale. Mental health disorders, also known as mental illnesses, are conditions that disrupt a person's cognitive, emotional, and behavioral functioning. They arise from complex genetic, biological, environmental, and psychological interactions.

Common mental disorders include anxiety, depression, bipolar disorder, schizophrenia, and post-traumatic stress disorder (PTSD). These conditions significantly impact an individual's quality of life and productivity, contributing to a substantial global health burden. Efforts to improve mental health care and reduce stigma are critical to supporting individuals living with these conditions and improving public health outcomes [1]. Panic disorder is a type of anxiety disorder characterized by recurring episodes of sudden and intense fear, known as panic attacks. These attacks often occur unexpectedly and without an obvious cause, leaving individuals in a constant state of fear and apprehension about when the next attack might occur. This persistent anxiety can lead to avoidance behaviors, where individuals steer clear of certain situations or places where they fear a panic attack might happen. The disorder is more than just a momentary feeling of fear or nervousness; it is a chronic condition that can significantly disrupt daily life. Without treatment, panic disorder can worsen over time, leading to complications such as phobias, depression, or other anxiety-related conditions [2]. Early recognition and intervention are crucial in managing panic disorders and helping individuals regain control of their lives. The primary symptoms of panic disorder are panic attacks, which are sudden and intense episodes of fear that can manifest physically and emotionally. During a panic attack, individuals may experience a racing heart, shortness of breath, chest pain, dizziness, or a sense of impending doom. These physical sensations can be so severe that people often mistake them for life-threatening conditions, such as heart attacks. In addition to the physical symptoms, panic attacks often lead to overwhelming feelings of helplessness, fear of losing control, or fear of dying. The unpredictability of these attacks contributes to heightened anxiety, as individuals may live in constant dread of another episode. Over time, this can result in avoidance behaviors, further isolating the individual and interfering with their daily functioning. Analyzing mental health data across diverse applications, such as understanding health misinformation [3] and assessing mental workload [4], is essential for advancing research and improving outcomes in the field. Advanced data analysis techniques and modern AI technologies offer powerful tools to uncover complex patterns in mental health data, enabling more accurate predictions, personalized interventions, and deeper insights into behavioral and cognitive processes.

Artificial Intelligence (AI) and machine learning have revolutionized different sciences. These technologies excel at analyzing large volumes of complex data and identifying patterns via different applications including air pollution forecasting [5], EGG signal processing and pattern recognition [6], automated formative assessment [7], tumor reconstruction [8], analyzing the malignant glioblastoma cells [9], and cyber threat defense [10]. In particular, machine learning algorithms have demonstrated remarkable capabilities in medical diagnosis [11-15]. By learning from labeled data, these models can classify diseases, predict outcomes, and recommend personalized treatment options. The adoption of AI and machine learning in healthcare has led to improved diagnostic accuracy, faster decision-making, and more efficient resource allocation, ultimately enhancing patient outcomes [16-18]. Machine learning has shown significant promise in the field of mental health, enabling the detection and diagnosis of complex conditions such as schizophrenia, depression, and anxiety disorders [19-23]. For schizophrenia, machine learning algorithms have been used to analyze neuroimaging data, genetic information, and behavioral patterns to identify biomarkers associated with the disorder [19]. These approaches have improved the ability to diagnose schizophrenia at an earlier stage, allowing for more effective intervention. Similarly, machine learning has been applied to the detection of stress and anxiety disorders by analyzing physiological data such as heart rate variability, skin conductance, and respiratory patterns [24]. Wearable devices equipped with sensors can collect real-time data, which is then processed by machine learning algorithms to detect signs of mental distress. These innovations have the potential to transform mental health care by providing objective, data-driven assessments and enabling personalized treatment strategies. Machine learning's ability to uncover hidden patterns and relationships in data makes it a powerful tool for advancing mental health research and care. By leveraging these technologies, clinicians can gain deeper insights into mental health disorders, ultimately improving diagnostic accuracy and treatment outcomes for patients.

II. METHODS

Dataset

The dataset used for this study, titled the Panic Disorder Detection Dataset, comprises 120,000 records of individuals with labeled information about panic disorder. It is divided into two separate files to facilitate model training and evaluation. The first file, Panic_Disorder_training, contains 100,000 labeled records and is designated for training the machine learning model. The second file, Panic_Disorder_testing, includes 20,000 labeled records and is used exclusively for evaluating the model's performance on unseen data. [25-26]

The dataset contains both numerical and categorical features that capture various clinical and demographic characteristics of individuals. It also includes the target variable, which indicates the presence or absence of panic disorder. Before modeling, preprocessing steps were performed, including handling missing values, encoding categorical variables, and scaling numerical features to ensure compatibility with machine learning algorithms. The training and testing datasets were processed identically to maintain consistency.

Gradient Boosting

Machine learning algorithms work by enabling computers to learn patterns and relationships from data without being explicitly programmed. These algorithms use mathematical models to process and analyze data, allowing them to make predictions or decisions based on the learned patterns. The process begins with training, where the algorithm is fed labeled data (input features and corresponding outputs) to identify underlying patterns. During this phase, the model adjusts its internal parameters to minimize errors between its predictions and the actual outcomes. Once trained, the model is tested on new, unseen data to evaluate its ability to generalize and make accurate predictions. Depending on the task, machine learning can be supervised (trained with labeled data), unsupervised (discovering patterns in unlabeled data), or reinforcement-based (learning through trial and error). The iterative nature of these algorithms allows them to improve over

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (18-23), Month: January - February 2025, Available at: www.noveltyjournals.com

time, making them highly effective for applications in various fields, including medical diagnosis, financial forecasting, and image recognition.

Gradient Boosting is an ensemble machine-learning method that builds a strong predictive model by combining the outputs of multiple weak learners, typically decision trees. The algorithm optimizes a specified loss function in a stage-wise manner, gradually improving the model by minimizing the error in each iteration. The training process of a decision tree can be described as follows: A decision tree works by iteratively splitting a dataset into smaller and smaller subsets based on feature values until it reaches a decision or prediction at the leaf nodes. The process begins by evaluating all available features and selecting the one that provides the best split, based on specific criteria such as Gini Impurity, Information Gain, or Variance Reduction. This splitting criterion measures how well a feature separates the data into groups that are homogeneous in terms of the target variable. Once the best feature is selected, the dataset is divided into branches, with each branch representing a possible value or range of values for that feature. The algorithm then repeats this process at each branch, recursively selecting the best feature to further divide the data. This continues until a stopping condition is met, such as reaching a maximum tree depth, having too few samples in a node, or achieving complete purity in the subsets. At the end of this process, the leaf nodes of the tree provide the final prediction, either as a class label for classification tasks or as a numerical value for regression tasks. The hierarchical structure of decision trees makes them intuitive, as they mimic a series of decision-making steps, with each step narrowing down the possible outcomes.

As a result, the overall training process of a gradient-boosting classifier can be summarized as follows: The model starts with an initial prediction and iteratively adds decision trees, each trained to correct the residual errors of the previous trees. This step-by-step optimization reduces overfitting and enhances the model's accuracy on both training and testing datasets. The Gradient Boosting Classifier works by building an ensemble of decision trees in a sequential manner, where each new tree is trained to correct the errors made by the previous ones. The process begins by initializing the model with a simple prediction, such as the average class probabilities. At each iteration, the algorithm calculates the residual errors—differences between the predicted class probabilities and the actual class labels. A new decision tree is then trained to predict these residuals, focusing on the areas where the previous trees performed poorly. The predictions from this tree are scaled by a learning rate, a hyperparameter that controls the contribution of each tree to the overall model, ensuring gradual improvement. These scaled predictions are added to the ensemble, and the model is updated to minimize the loss function, typically cross-entropy for classification tasks. This iterative process continues, with each tree incrementally reducing the overall error by learning from the gradients of the loss function. By combining the predictions from all trees, the Gradient Boosting Classifier creates a powerful and accurate model capable of handling complex, non-linear relationships in the data.

Preprocessing

In machine learning, data preprocessing is a critical step to ensure the model is trained effectively and can generalize well to unseen data. A common approach involves splitting the dataset into three subsets: training, validation, and testing. This process ensures that the model is robust and prevents overfitting, which occurs when the model performs well on training data but poorly on new, unseen data. The training dataset is the largest subset and serves as the foundation for the model to learn patterns and relationships in the data. The machine learning algorithm is exposed to this dataset during the training process, where it iteratively adjusts its parameters to minimize the loss function. For our dataset, we allocated 70,000 instances (70% of the training data) for this purpose. The validation dataset is used during the training process to evaluate the model's performance and tune its hyperparameters. This subset is critical for preventing overfitting by providing feedback on how well the model generalizes to unseen data. By monitoring validation performance, we can identify when the model starts to overfit the training data and adjust accordingly. In our case, we reserved 30,000 instances (30% of the training data) for validation. The test dataset is a completely separate subset used for the final evaluation of the model. It is crucial to ensure that the test data remains unseen during the entire model-building process to avoid data leakage, which could result in an overly optimistic performance estimate. Our test dataset consists of 20,000 instances, and it provides an unbiased estimate of how the model is likely to perform in real-world scenarios.

In our preprocessing pipeline, we performed one-hot encoding for categorical features to transform them into a numerical format suitable for machine learning algorithms. Additionally, we standardized numerical features to ensure they have a mean of zero and a standard deviation of one. For example, the "Age" feature was scaled using the StandardScaler. The transformation ensures that all numerical features are on the same scale, which is especially important for gradient-based

Novelty Journals

ISSN 2394-966X

International Journal of Novel Research in Life Sciences
Vol. 12, Issue 1, pp: (18-23), Month: January - February 2025, Available at: www.noveltyjournals.com

algorithms like Gradient Boosting. The preprocessed data allows our model to be trained in well-structured and normalized inputs, enhancing its ability to learn patterns effectively and improving its predictive accuracy.

III. RESULTS

The Gradient Boosting Classifier demonstrated exceptional performance, achieving a near-perfect training accuracy of 0.999. This high accuracy indicates that the model effectively learned the patterns and relationships in the training data. To evaluate its performance on the test dataset, a confusion matrix was used, providing a detailed breakdown of the model's predictions. A confusion matrix is a tool that summarizes the performance of a classification model by showing the number of correct and incorrect predictions for each class. It consists of four components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives and true negatives represent correctly classified instances, while false positives and false negatives are the misclassified ones. For this model, the confusion matrix revealed that 28,709 instances were correctly classified as class 0 (absence of panic disorder), and 1,290 instances were accurately identified as class 1 (presence of panic disorder). Remarkably, there was only one false positive, where the model incorrectly predicted the presence of panic disorder, and there were no false negatives, meaning the model successfully identified all actual cases of panic disorder. This performance highlights the classifier's high precision, with almost no false alarms, and its excellent recall, ensuring that no cases were missed. The combination of these metrics underscores the reliability and robustness of the Gradient Boosting Classifier, making it a powerful tool for detecting panic disorder. The results emphasize its potential to support early diagnosis and intervention in mental health diagnostics, particularly in clinical applications where accuracy is critical.

The testing results for the Panic Disorder Detection model reveal an impressive performance, achieving an accuracy of 98%. This high accuracy demonstrates the model's ability to generalize well to unseen data. To further evaluate its performance, the confusion matrix provides a detailed breakdown of the predictions. A confusion matrix is a tool used to assess the performance of classification models by comparing actual and predicted values. It consists of four components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True positives and true negatives represent correctly classified instances, while false positives and false negatives indicate misclassifications. In this case, the confusion matrix shows that the model correctly identified 19,158 instances as class 0 (absence of panic disorder) and 841 instances as class 1 (presence of panic disorder). It made only 1 false positive, where it incorrectly predicted the presence of panic disorder, and there were 0 false negatives, meaning it successfully detected all actual cases of panic disorder without missing any. This indicates the model's high precision in minimizing false alarms and its excellent recall in identifying all positive cases. These results highlight the robustness and reliability of the Gradient Boosting Classifier for detecting panic disorder. The high accuracy, coupled with the nearly perfect precision and recall, underscores its potential as a valuable tool in clinical diagnostics. By accurately distinguishing between individuals with and without panic disorder, the model demonstrates significant promise for supporting early diagnosis and treatment in mental health care.

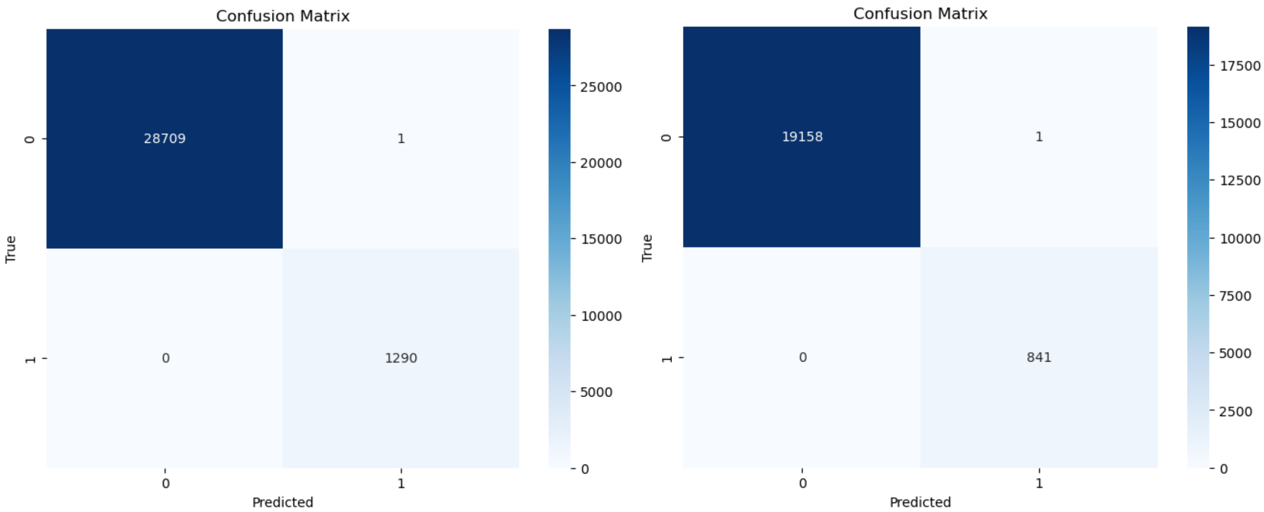

Fig. 1. Confusion matrix for (left) training and (right) testing data

IV. CONCLUSION

In this study, we developed and evaluated a Gradient Boosting Classifier to detect panic disorder, leveraging a large dataset of 120,000 instances. The model achieved exceptional performance, with a training accuracy of 99.997% and a testing accuracy of 98%, demonstrating its ability to generalize effectively to unseen data. Preprocessing steps, including one-hot encoding and standardization, ensured that the dataset was optimized for machine learning. The confusion matrix analysis revealed strong precision and recall, with minimal misclassifications, highlighting the model's reliability in distinguishing between individuals with and without panic disorder. These findings underscored the potential of machine learning, particularly Gradient Boosting, in mental health diagnostics, offering a promising tool for early detection and intervention.

Despite these successes, challenges remained for future research. The dataset used in this study was structured and labeled, which may not reflect real-world scenarios where missing or noisy data are common. Additionally, while the model performed well in binary classification, extending it to detect comorbid mental health conditions could enhance its clinical utility. Ethical considerations, such as data privacy and bias in predictions, must also be addressed to ensure fair and responsible deployment. Future work could explore the integration of multimodal data, including physiological signals and imaging, to improve diagnostic accuracy further. Addressing these challenges would enhance the applicability and robustness of machine learning models in mental health care.

REFERENCES

[1] Cosci F, Fava GA. Staging of mental disorders: systematic review. Psychotherapy and psychosomatics. 2012 Nov 6;82(1):20-34.

[2] Carleton RN, Duranceau S, Freeston MH, Boelen PA, McCabe RE, Antony MM. “But it might be a heart attack”: Intolerance of uncertainty and panic disorder symptoms. Journal of anxiety disorders. 2014 Jun 1;28(5):463-70.

[3] Kamali D, Romain J, Liu H, Peng W, Meng J, Kordjamshidi P. Using Persuasive Writing Strategies to Explain and Detect Health Misinformation. arXiv preprint arXiv:2211.05985. 2022 Nov 11.

[4] Entezarizarch E, Zakerian SA, Madreseh E, Abbasinia M, Abdi H. Comparative analysis of mental workload and performance between young and elderly drivers: Implications for road safety and age-related driving challenges. Work. 2024 Mar 30(Preprint):1-2.

[5] Karami M, Hamzehei S, Rastegari F, Akbarzadeh O. Exploring the Relationship Between Air Pollution and CNS Disease Mortality in Italy: A Forecasting Study with ARIMA and XGBoost. In2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE) 2023 Jul 24 (pp. 46-52). IEEE.

[6] Rafiei M, Pour MR, Akbari H. Comparing linear, nonlinear and time frequency-based features in the EEMD domain for depression detection application using EEG signals. Signal Processing and Renewable Energy. 2024 Dec 30;8(4):1-1.

[7] Karizaki MS, Gnesdilow D, Puntambekar S, Passonneau RJ. How Well Can You Articulate that Idea? Insights from Automated Formative Assessment. InInternational Conference on Artificial Intelligence in Education 2024 Jul 2 (pp. 225-233). Cham: Springer Nature Switzerland.

[8] Najafi H, Savoji K, Mirzaeibonehkhater M, Moravvej SV, Alizadehsani R, Pedrammehr S. A Novel Method for 3D Lung Tumor Reconstruction Using Generative Models. Diagnostics. 2024 Nov 20;14(22):2604..

[9] Vahedi MM, Shahini A, Mottahedi M, Garousi S, Shariat Razavi SA, Pouyamanesh G, Afshari AR, Ferns GA, Bahrami A. Harmaline exerts potentially anti-cancer effects on U-87 human malignant glioblastoma cells in vitro. Molecular Biology Reports. 2023 May;50(5):4357-66.

[10] Ramezani A. Fusion models for cyber threat defense: integrating clustering with kmeans, random forests, and SVM against windows malware. In2024 IEEE World AI IoT Congress (AIIoT) 2024 May 29 (pp. 465-470). IEEE.

[11] Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, D'Amico NC, Sardanelli F. AI applications to medical images: From machine learning to deep learning. Physica medica. 2021 Mar 1;83:9-24.

[12] Abbasi H, Afrazeh F, Ghasemi Y, Ghasemi F. A Shallow Review of Artificial Intelligence Applications in Brain Disease: Stroke, Alzheimer's, and Aneurysm. International Journal of Applied Data Science in Engineering and Health. 2024 Oct 5;1(2):32-43.

[13] Ghasemi F, Minoo S. Dental Imaging Analysis with Artificial Intelligence. Available at SSRN 5008235. 2024 Oct 21.

[14] Chen H, Kim S, Hardie JM, Thirumalaraju P, Gharpure S, Rostamian S, Udayakumar S, Lei Q, Cho G, Kanakasabapathy MK, Shafiee H. Deep learning-assisted sensitive detection of fentanyl using a bubbling-microchip. Lab on a Chip. 2022;22(23):4531-40.

[15] Afrazeh F, Shomalzadeh M. Revolutionizing Arthritis Care with Artificial Intelligence: A Comprehensive Review of Diagnostic, Prognostic, and Treatment Innovations. International Journal of Applied Data Science in Engineering and Health. 2024 Sep 10;1(2):7-17.

[16] Orouskhani M, Zhu C, Rostamian S, Zadeh FS, Shafiei M, Orouskhani Y. Alzheimer's disease detection from structural MRI using conditional deep triplet network. Neuroscience Informatics. 2022 Dec 1;2(4):100066.

[17] Abbasi H. Transfer Learning and Advanced CNN Models for Detecting Brain Tumors using MRI. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):92-103.

[18] Sharifi S, Donyadadi A. Detection and Diagnosis of Congenital Heart Disease from Chest X-Rays with Deep Learning Models. International Journal of Applied Data Science in Engineering and Health. 2025 Jan 2;1(1):1-9.

[19] Norouzi F, Machado BL, Nematzadeh S. Schizophrenia Diagnosis and Prediction with Machine Learning Models. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):113-22.

[20] Nova K. Machine learning approaches for automated mental disorder classification based on social media textual data. Contemporary Issues in Behavioral and Social Sciences. 2023 Apr 2;7(1):70-83.

[21] Pintelas EG, Kotsilieris T, Livieris IE, Pintelas P. A review of machine learning prediction methods for anxiety disorders. InProceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion 2018 Jun 20 (pp. 8-15).

[22] Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychological medicine. 2019 Jul;49(9):1426-48.

[23] Chen S, Chen G, Li Y, Yue Y, Zhu Z, Li L, Jiang W, Shen Z, Wang T, Hou Z, Xu Z. Predicting the diagnosis of various mental disorders in a mixed cohort using blood-based multi-protein model: a machine learning approach. European Archives of Psychiatry and Clinical Neuroscience. 2023 Sep;273(6):1267-77.

[24] Mentis AF, Lee D, Roussos P. Applications of artificial intelligence− machine learning for detection of stress: a critical overview. Molecular Psychiatry. 2024 Jun;29(6):1882-94.

[25] https://www.kaggle.com/datasets/muhammadshahidazeem/panic-disorder-detection-dataset

[26] https://www.kaggle.com/code/vinciusparede/panic-disorder-detection-gradient-boosting