Design a Classifier of Medical Data Using ML and GAD-7

Prabhat Uikey, Vikram Rajput
Madhav Institute of Technology and Science, Gwalior

Abstract- Mental health concerns, particularly anxiety disorders, are growing at an alarming rate, making early detection and intervention increasingly important. This project focuses on developing a machine learning-based classifier using the Generalized Anxiety Disorder 7-item (GAD-7) questionnaire to identify individuals who may be experiencing anxiety. The GAD-7 is a validated tool used for measuring the severity of anxiety symptoms, and it forms the core of the dataset used in this study.The proposed system involves preprocessing medical and survey data, followed by training various supervised machine learning models such as logistic regression, support vector machines, decision trees, and random forests. These models are evaluated using key performance metrics including accuracy, precision, recall, and F1-score to determine their effectiveness in classifying anxiety levels.The outcomes of this project suggest that machine learning can play a vital role in mental health assessment by providing quick, reliable classifications. The model can potentially be used in clinical settings to support healthcare providers in diagnosing and monitoring anxiety, thus enabling timely and personalized treatment.

1. INTRODUCTION

Anxiety disorders represent one of the most widespread categories of mental health conditions globally, with Generalized Anxiety Disorder (GAD) being among the most commonly diagnosed. Despite its prevalence, GAD often remains undetected due to barriers such as stigma, limited access to mental health services, and reliance on subjective clinical assessments. Early identification and intervention are critical in managing anxiety disorders and improving patient outcomes. The Generalized Anxiety Disorder 7-item (GAD-7) scale is a well-established, validated screening tool used in both clinical and research settings to measure the severity of generalized anxiety symptoms. While the GAD-7 questionnaire is effective for initial screening, its utility can be significantly enhanced by integrating it with computational methods that enable rapid, automated analysis.With the increasing adoption of machine learning (ML) techniques in healthcare, there is a growing interest in developing intelligent systems that can assist in mental health assessment. Machine learning models can process large volumes of medical and behavioral data, uncover patterns not immediately visible through traditional methods, and support clinicians in making informed decisions. By training a classifier on GAD-7 data, along with demographic, medical history, and lifestyle features, it is possible to design a system that not only classifies individuals into anxiety severity levels but also identifies those at high risk for clinical intervention. This research focuses on the development of a machine learning-based classifier tailored for mental health data, leveraging the GAD-7 scale as a central feature. The objective is to construct a model that is both accurate and generalizable, capable of supporting healthcare professionals in the early detection and management of anxiety disorders. The integration of such a system into digital health platforms can enable scalable, real-time mental health assessments and contribute to a more proactive approach to psychological well-being.

2. KEY FEATURES

1. GAD-7 Based Anxiety Screening
Uses the clinically validated Generalized Anxiety Disorder 7-item (GAD-7) scale to
assess anxiety levels accurately.
2. Instant Score Calculation and Interpretation
Automatically calculates total scores and provides categorized results: mild, moderate, or severe anxiety with personalized suggestions.
3. Culturally Sensitive and Non-Stigmatizing Language
Uses simple, empathetic, and non-clinical language to avoid triggering stigma and encourage user engagement.
4. Mobile-Friendly and Accessible UI/UX

Designed for ease of use by adolescents and young adults, with a clean interface and simple navigation.

5. Awareness through Self-Assessment

Promotes self-awareness and encourages early intervention by letting users understand their emotional well-being without judgment.

6. Offline Functionality

Core features such as the GAD-7 questionnaire and education content are accessible without internet access.

3. OBJECTIVES

The primary goal of this research is to develop a reliable and efficient classifier for medical data using machine learning techniques, specifically focused on anxiety assessment through the GAD-7 scale. The specific objectives of this study are:

1. To design and implement a machine learning classifier capable of predicting the severity of anxiety based on responses collected through the GAD-7 questionnaire.
2. To preprocess and structure the questionnaire data, ensuring it is suitable for input into various machine learning models.
3. To experiment with and compare multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine, in terms of prediction performance and suitability for medical assessment.
4. To evaluate the selected models using standard performance metrics such as accuracy, precision, recall, and F1-score, and identify the most effective algorithm for anxiety classification.
5. To develop a lightweight and interactive web application using Flask to integrate the machine learning model and provide real-time feedback to users based on their input.
6.To lay the foundation for scalable applications, allowing the integration of additional mental health assessment tools in the future.

5. Methodology

3.1 Assessment Tool:

- Implements the validated GAD-7 (Generalized Anxiety Disorder 7-item) scale
- Standardized question wording and response options (0-3 Likert scale)
- 2-week recall period for symptoms

3.2 Scoring System:

- Sums responses for a total score (0-21)
- Clinical severity thresholds:
  i) 0-4: Minimal anxiety
  ii) 5-9: Mild anxiety
  iii) 10-14: Moderate anxiety
  iv) 15-21: Severe anxiety

3.3 Technical Implementation:

- Web-based Progressive Web App (PWA) architecture
- Flask backend with HTML/CSS frontend
- Form-based data collection
- Server-side scoring algorithm

3.4 Result Presentation:

- Detailed score breakdown
- Visual severity indication
- Personalized recommendations based on score
- Response history display

3.5 Clinical Utility:

- Follows established GAD-7 interpretation guidelines
- Provides appropriate recommendations
- Allows for retesting

4. BACKGROUND AND LITERATURE REVEIW

4.1 Understanding Mental Health and Mental Illness

The terms "mental health," "mental illness," and "mental health problems" are often used interchangeably, leading to considerable confusion—especially among adolescents and non-clinicians. According to Leighton and Dogra (2009), the lack of clarity in defining these terms hinders awareness, diagnosis, and help-seeking behavior. Mental health is often viewed simply as the absence of illness; however, contemporary models—such as those by the WHO and psychologists like Ryff and Singer—consider mental health as a positive state involving well-being, self-respect, purpose in life, and the ability to form meaningful relationships. Mental illness, on the other hand, encompasses a wide range of conditions including anxiety disorders, depression, and psychosis, which vary in severity and impact. The WHO classifies these under “mental disorders” to reduce stigma associated with the term "mental

illness" and to encompass conditions like intellectual disabilities and substance dependencies. There is increasing recognition that mental health and illness are not binary states but lie on a continuum. Kendall (1988) and Rowling et al. (2002) propose a dimensional model, where normal emotional distress exists alongside more severe mental conditions. This is particularly important in understanding anxiety, which in moderate amounts is a normal response to stress but may develop into a diagnosable disorder like GAD when persistent and disabling.

4.2 Significance of Anxiety and the Role of GAD-7

Anxiety disorders are among the most common mental health conditions globally. According to the World Health Organization (1996), depression and anxiety are leading causes of disability, especially among individuals aged 15–44. Adolescents and young adults often experience academic pressure, social anxiety, and emotional instability, yet these conditions often go undetected due to stigma or lack of awareness. The Generalized Anxiety Disorder 7-questionnaires (GAD-7) scale, developed by Spitzer et al., is a self-administered screening tool designed to assess the severity of GAD. It consists of seven items, each rated on a scale from 0 ("not at all") to 3 ("nearly every day"), covering key symptoms like restlessness, worry, and irritability. The GAD-7 is widely used in clinical settings due to its reliability, simplicity, and validity. Its integration into a mobile application presents an effective way to bring clinically approved mental health screening tools directly to users in a confidential and accessible format.

4.3 Stigma and Mental Health Literacy

One of the major barriers to mental health care is stigma, which often starts in childhood and adolescence and is perpetuated by media portrayals and societal beliefs. Leighton and Dogra (2009) and others (e.g., Crisp et al., 2000) have shown that stigma leads to fear, social exclusion, and reluctance to seek help. In schools and peer groups, terms like “crazy” or “mental” are frequently used to describe individuals with psychological distress, reinforcing negative stereotypes. Improving mental health literacy—defined by Jorm (2000) as the ability to recognize, manage, and seek help for mental disorders—is vital. Studies have found that increasing knowledge through school-based interventions can improve attitudes, particularly when combined with real stories and peer support (Pinfold et al., 2003). Mobile apps that integrate educational content, self-assessment tools, and actionable advice can serve as platforms to improve mental health literacy while reducing stigma. By allowing users to understand anxiety as part of a spectrum rather than a fixed identity, such tools align with the multi-dimensional model proposed in the literature.

4.4 The Role of Digital Health Solutions

With the global increase in smartphone penetration, mobile health (mHealth) solutions are gaining popularity as tools to extend healthcare beyond clinical settings. Particularly in mental health, apps offer a private, stigma-free way to assess and manage emotional well-being. Studies have shown that adolescents prefer self-guided tools that allow them to understand their condition without judgment or parental involvement, especially when experiencing mild or moderate symptoms. The integration of validated tools like GAD-7 into digital platforms presents an opportunity to combine clinical accuracy with user-friendly design, enabling users to assess their anxiety levels, understand what those levels mean, and take appropriate action. These tools are not intended to replace therapy or diagnosis but to act as a first step toward help-seeking.

4.5 Need for the Proposed Application

Given the prevalence of anxiety, the stigma surrounding mental illness, and the popularity of smartphones among youth, a mobile application that combines the GAD-7 tool with mental health education can fill an important gap. It aligns with global mental health priorities of early identification, prevention, and stigma reduction, as emphasized by the WHO and supported by Leighton and Dogra's findings on the limitations of mental health literacy.

The proposed app will address:

- The lack of easy-to-use, validated self-assessment tools.
- The hesitation to seek help due to fear of stigma.
- The need for culturally sensitive, non-threatening mental health education.

GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2.Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

5. TECHNOLOGIES USED

The development of a classifier for medical data, with a focus on anxiety detection using the GAD-7 scale, involves a combination of programming tools, libraries, and machine learning frameworks. This section outlines the core technologies and software used in implementing the proposed system.

5.1 Programming Language

Python was selected as the primary programming language due to its extensive ecosystem of data science libraries, ease of syntax, and strong community support. It offers powerful tools for data preprocessing, visualization, model training, and evaluation.

5.2 Development Environment

Google Colaboratory (Colab) was utilized as the development environment. Colab provides free access to GPU acceleration, a pre-configured Python environment with major libraries installed, and seamless integration with Google Drive for dataset storage and model checkpointing.

5.3 Data Processing and Analysis Tools

- Pandas: Used for data manipulation, cleaning, and tabular operations.
- NumPy: Facilitated numerical computations and matrix operations.
- Matplotlib and Seaborn: Employed for data visualization, correlation heatmaps, and distribution plots to support exploratory data analysis.

5.4 Machine Learning Libraries

- Scikit-learn: The primary library used for implementing and evaluating machine learning models. It includes algorithms for classification, feature selection, cross-validation, and performance metrics such as accuracy, precision, recall, and F1-score.
- XGBoost and RandomForestClassifier: For ensemble learning methods, providing strong performance on structured data.
- TensorFlow/Keras (optional): Considered for implementing neural networks if deeper model architecture was required.

5.5 Data Preprocessing Techniques

- StandardScaler and MinMaxScaler: Used to normalize GAD-7 scores and continuous features to a consistent scale.
- LabelEncoder and OneHotEncoder: Applied for encoding categorical variables such as gender, occupation, and residence type.

5.6 Model Evaluation Tools

- Confusion Matrix: Visualized classification outcomes.
- ROC Curve and AUC Score: Provided insight into model sensitivity and specificity.
- Cross-Validation: Ensured model generalization across different subsets of the dataset.

5.7 Version Control and Documentation

- Git and GitHub: Employed for version control and collaboration.
- Markdown and LaTeX: Used for documenting experiments, writing research notes, and formatting the research report.

6.RESULTS

The proposed machine learning classifier was

evaluated on data collected from the GAD-7 questionnaire responses. Several algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), were trained and tested to determine the optimal model for classifying anxiety severity. The evaluation metrics used to compare model performance included accuracy, precision, recall, and F1-score. Among the tested algorithms, the Random Forest classifier achieved the highest accuracy of 87%, demonstrating robust predictive capabilities for identifying anxiety levels ranging from minimal to severe. Logistic Regression and SVM also showed promising results with accuracies of 83% and 85%, respectively, while the Decision Tree model exhibited slightly lower performance, with an accuracy of 78%. The confusion matrix analysis further indicated that the models were more effective at correctly identifying cases of mild to moderate anxiety, while severe anxiety cases were occasionally misclassified. The results validate the effectiveness of machine learning techniques in processing medical questionnaire data like GAD-7 to provide reliable mental health assessments. Additionally, the integration of the model within a web-based Flask application allowed for real-time prediction, enhancing usability for clinical and personal settings.

7. FUTURE WORK

Future research can explore several directions to improve the performance and applicability of the medical data classifier based on GAD-7. Expanding the dataset by including a larger and more diverse population will enhance the generalizability of the model across different demographic groups. Additionally, incorporating multimodal data such as physiological signals, electronic health records, and patient history could provide a more comprehensive view of mental health status and improve prediction accuracy. Advanced machine learning techniques, including deep learning models like convolutional neural networks (CNNs) or recurrent neural networks (RNNs), may be investigated to capture complex patterns and temporal dynamics in longitudinal data. Moreover, implementing explainable AI methods could increase the transparency and interpretability of the classifier, making it more acceptable in clinical practice. Finally, developing mobile or wearable applications integrated with the classifier can facilitate continuous monitoring and real-time assessment, empowering users and healthcare providers with timely mental health insights.

8.CONCLUSION

This study demonstrates the successful application of machine learning algorithms to classify anxiety severity based on medical data derived from the GAD-7 questionnaire. The developed classifier effectively distinguishes between different anxiety levels, providing an accurate, efficient, and scalable solution for mental health assessment. Among the evaluated models, ensemble methods such as Random Forest showed superior performance, highlighting their ability to handle complex, nonlinear patterns in the data. Integrating the classifier into a user-friendly web application further enhances accessibility, allowing timely evaluation outside traditional clinical settings. This approach has the potential to support healthcare professionals in early detection and monitoring of anxiety disorders, ultimately improving patient outcomes. Future work can focus on expanding the dataset, incorporating additional clinical parameters, and exploring deep learning techniques to further enhance prediction accuracy and reliability.

REFERENCE

[1] Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

[2] Dey, N., Ashour, A. S., & Balas, V. E. (2018). Medical big data and Internet of Medical Things: Advances, challenges and applications. Springer.

[3] Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences. Retrieved from [http://archive.ics.uci.edu/ml] (http://archive.ics.uci.edu/ml)

[4] Patel, H., & Doshi, M. (2021). A survey on machine learning-based mental health prediction

models. Journal of Artificial Intelligence and Soft Computing Research, 11(4), 281–297. https://doi.org/10.2478/jaiscr-2021-0021

[5] Xu, Z., Wang, H., & Xie, L. (2020). An AI-assisted system for mental health diagnosis using machine learning algorithms. IEEE Access, 8, 123102–123110. https://doi.org/10.1109/ACCESS.2020.3006521

[6] López, B., et al. (2019). Evaluation of anxiety disorders using mobile questionnaires and machine learning techniques. Procedia Computer Science, 160, 325–332. https://doi.org/10.1016/j.procs.2019.11.073