APPLICATION OF ARTIFICIAL INTELLIGENCE TO DEVELOP A MODEL FOR IDENTIFYING MANIA BASED ON VOICE: PRELIMINARY RESULTS

Nguyen Hoang Long[1*], Nguyen Tra Giang[2],
Hoang Dinh Tri Dung[2], Ngo Thi Tam[1], Nguyen Trong Hung[3], Nguyen Van Phi[2]

[1]*Institute of Advanced Technology, Ton Duc Thang University*
[2]*Hanoi Medical University, Thanh Hoa Branch -*
722 Quang Trung, Thanh Hoa Ward, Thanh Hoa City, Thanh Hoa Province, Vietnam
[3]*People's Police Academy*

Received: 24/04/2026
Revised: 04/05/2026; Accepted: 18/05/2026

ABSTRACT

Objective: This study aimed to develop and evaluate an artificial intelligence model based on voice features to support the identification of mania, and to build an integrated system for data collection, analysis, and decision support in mental health.

Methods: The training dataset consisted of 46 international patients labeled using the Young Mania Rating Scale, combined with a validation dataset of 4 patient samples from Vietnam. The model was trained using deep learning methods on acoustic features and evaluated using accuracy, ROC, and AUC metrics.

Results: The results demonstrated stable performance on both training and validation datasets, with clear discrimination between different states. When applied to Vietnamese data, the model correctly predicted 2 out of 4 cases, indicating limitations in adapting to regional phonetic variations. An integrated software system was developed, enabling real-time data collection, analysis, and result visualization.

Conclusion: Initial findings suggest the potential application of artificial intelligence in supporting the identification of manic symptoms; however, further data expansion and model optimization are required to improve real-world applicability.

Keywords: Artificial intelligence, mania, voice, bipolar disorder, deep learning, mental health.

1. INTRODUCTION

Bipolar affective disorder is a chronic psychiatric condition in which the manic phase plays a critical role in diagnosis, monitoring, and prognosis[1]. Clinical manifestations of mania—such as increased speech rate, prosodic alterations, heightened emotional expression, and reduced behavioral inhibition—are often clearly reflected in voice patterns. Therefore, voice can be considered a potential biomarker for the early detection of pathological states [1–3].

In clinical practice, the assessment of manic states primarily relies on psychiatric examination and clinician judgment, which introduces subjectivity and limits scalability. Artificial intelligence, particularly deep learning in acoustic signal processing, has demonstrated high accuracy in identifying mental health disorders [4]. A 2025 meta-analysis of 18 studies (3,152 participants) reported that AI achieved a sensitivity of 0.88, specificity of 0.89, and AUC of 0.94 when distinguishing mania from healthy controls, and a sensitivity of 0.84, specificity of 0.82, and AUC of 0.89 when distinguishing mania from depression. With a prevalence of 37%, the positive predictive value was approximately 74%, and the residual risk after a negative result was about 10% [4]. These findings highlight the potential of AI in supporting diagnosis, particularly when integrating voice features for early detection of manic states.

In Vietnam, the application of artificial intelligence in mental health remains limited, especially due to the lack of labeled voice datasets and the absence of integrated systems spanning data collection, processing, analysis, and decision support [5]. This gap necessitates the development of an AI model based on voice features, integrated into a comprehensive technological system that includes data acquisition, signal processing, storage, and result visualization. Such an approach not only enables early identification of manic symptoms but also establishes a foundation for real-time monitoring and clinical decision support systems, thereby improving the effectiveness of screening, follow-up, and intervention in psychiatric practice.

The objective of this study is to develop and evaluate a voice-based artificial intelligence model to support early identification of manic symptoms and to integrate it into a monitoring and decision-support system in the field of mental health.

2. SUBJECTS AND METHODS

2.1. Model Training, Validation, and Application to Vietnamese Data

2.1.1. Data Sources

The dataset comprised both international and Vietnamese data. The international dataset included 46 clinically diagnosed patients (30 males and 16 females) with

*Corresponding author
Email: nguyenhoanglong@tdtu.edu.vn **Phone:** (+84) 812098910 **Https:**//doi.org/10.52163/yhc.v67iE1.5097

a mean age of 36.5 (range: 18–53), and was used for model training and initial validation. The data were converted into Vietnamese through a pipeline involving speech recognition, translation, and speech synthesis, followed by standardization to ensure consistency. This is an open-access dataset available at: https://sites.google.com/view/avec2018/home [6].

The Vietnamese dataset consisted of 4 manic patient samples (2 males and 2 females), aged 15 to 73, including Northern and Central dialects. These data were collected from conversational recordings, with only the patients' speech segments used to evaluate the model's adaptability.

2.1.2. Labeling, Training, and Validation Process

Data were labeled using the Young Mania Rating Scale and categorized into two states: remission and mania [7]. After preprocessing, the data were transformed into spectral features (Short-Time Fourier Transform, Mel spectrogram, log scale, standardized to 224×224). The dataset was split into training (60%), validation (20%), and testing (20%) sets. The model was trained using the Adam optimizer with a Categorical Cross-Entropy loss function and evaluated on an independent dataset.

2.1.3. Application to Vietnamese Data

The trained model was applied to real-world Vietnamese data. Input data consisted solely of patient speech and were processed using the same pipeline as in the training phase. Prediction results were used to assess the model's adaptability to Vietnamese language data.

2.1.4. Data Processing and Analysis

Data underwent preprocessing and extraction of acoustic features, including MFCC and Mel spectrogram, before being input into the deep learning model for classification. Model performance was evaluated using accuracy, Receiver Operating Characteristic (ROC) curves, and Area Under the Curve (AUC). Predictions were compared with clinical assessments for validation and were subsequently used for model retraining to improve performance.

2.2. Development of a System for Supporting the Identification of Manic Symptoms

2.2.1. System Architecture

The system was designed using a layered architecture consisting of data acquisition, transmission, storage, processing, and visualization components. Audio data were collected via Internet of Things (IoT) devices integrated with Edge AI modules to perform preprocessing and preliminary analysis.

The MinIO object storage system was used to store audio files, while Apache Kafka ensured stable data transmission. PostgreSQL was used to store analysis results and metadata.

2.2.2. Operational Workflow

The system operates through a sequential pipeline: audio data collection, edge-level preprocessing, data storage, metadata transmission via Kafka, server-side processing, and storage of results in the database. The web interface queries the storage system to display information, statistics, and visualized results.

2.2.3. Model Deployment

The trained artificial intelligence model was integrated into the system for real-time audio data analysis. Input data were processed following the same steps as in the training phase before being fed into the model for state prediction. Results were stored in the database and displayed via the web interface. They were also compared with expert evaluations and used in the retraining process to enhance system performance and adaptability.

3. RESULTS

3.1. Model Performance

3.1.1. Results on Training and Validation Data

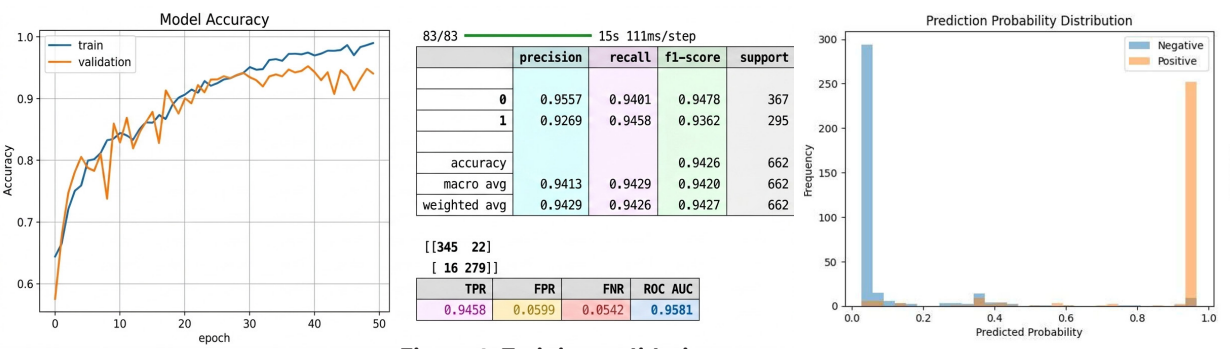

Figure 1. Training–validation accuracy and predicted probability distribution of the AI model for mania detection

Figure 1 shows that model accuracy increased rapidly with the number of epochs and stabilized after approximately 30–40 epochs. The training set achieved nearly 0.98 accuracy, while the validation set reached approximately 0.93–0.95, indicating good generalization without evidence of overfitting. An epoch is defined as one complete training cycle in which the entire training dataset is passed through the model once.

At the same time, the distribution of predicted probabilities shows that negative samples are concentrated near 0 and positive samples near 1, with only a small number of cases in the intermediate range. This demonstrates strong discriminative ability and high prediction confidence of the model.

3.1.2. Validation Results on Vietnamese Subjects

Table 1. Model validation results on Vietnamese data

Characteristics	BN01	BN02	BN03	BN04
Gender	Male	Male	Female	Female
Age	15–19	73	15–25	34–40
Clinical status	Mania	Mania	Mania	Mania
Region	Northern	Northern	Central	Central
Data collection method	Phone	Direct	Direct	Direct
Model prediction	Mania	Mania	Normal	Normal
Confidence (%)	89.43	99.52	91.62	92.82
Assessment	Correct	Correct	Incorrect	Incorrect

Table 1 summarizes the model validation results on 4 Vietnamese patient samples based on system analysis. The model correctly predicted 2 out of 4 cases (BN01, BN02) with high confidence. The remaining two cases (BN03, BN04) were misclassified despite confidence levels above 90%, indicating limitations in adapting to regional dialect variations.

3.2. System for Supporting the Identification of Manic Symptoms

The system for supporting the identification of manic symptoms was developed as an integrated platform combining an artificial intelligence model with management software, aimed at monitoring, analyzing, and supporting decision-making in mental health.

The system features an intuitive interface that allows users to track psychological status, retrieve data, and receive early warnings of abnormal manifestations, while ensuring data security and privacy. (Figure 2 and Figure 3)

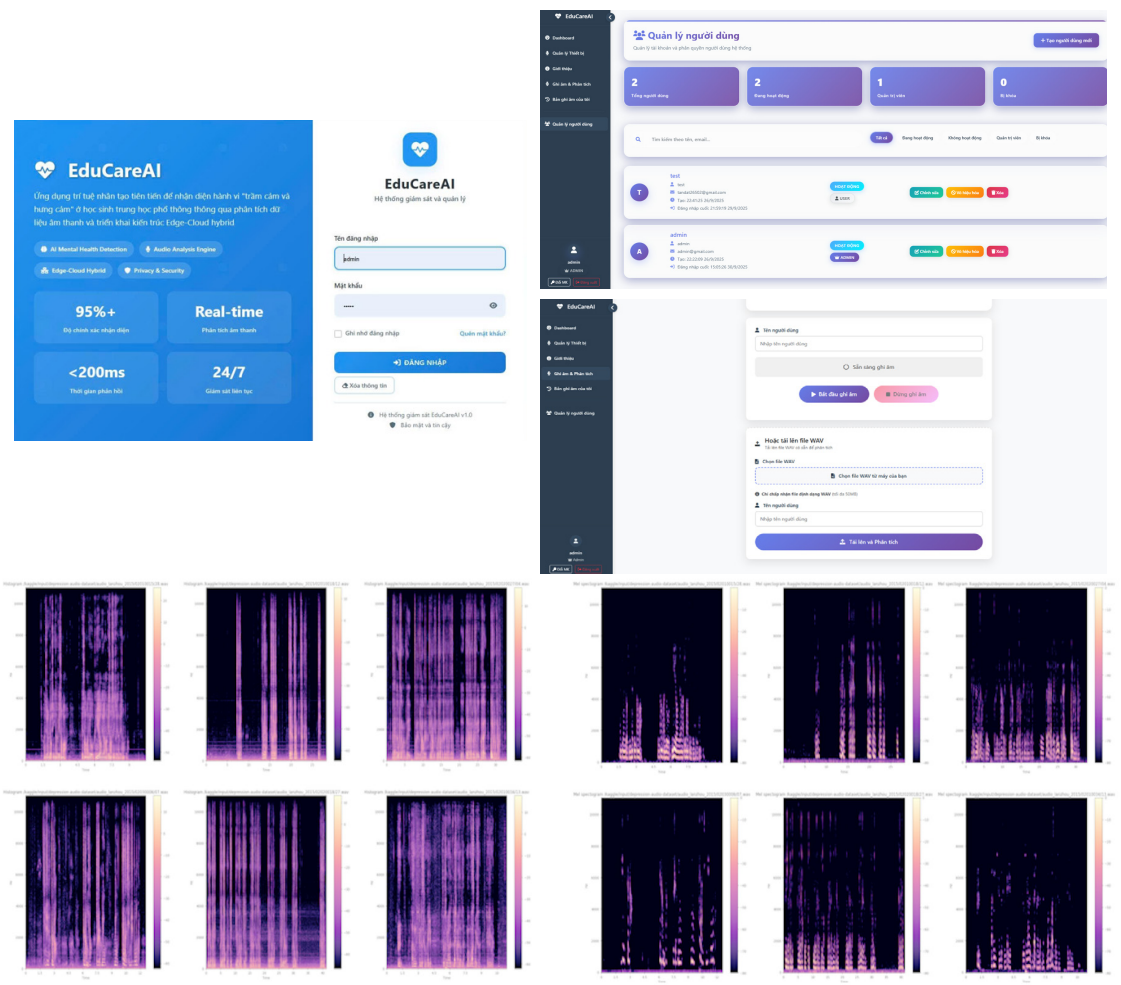

Figure 3. Visualization of extracted features after analysis

The system was developed to integrate the artificial intelligence model with a software platform, supporting monitoring, analysis, and decision-making in the field of mental health.

+ User management: Enables registration, login, and user information management; tracks assessment history, number of evaluations, and longitudinal psychological state changes; supports retrieval and playback of audio data, ensuring security via JWT and PostgreSQL.

+ Overview dashboard: Displays the number of analyses, distribution of "normal"/"mania" states, and temporal trends; integrates visual charts with filtering and detailed interaction; updated in real time via WebSocket.

+ Data analysis: Integrates the AI model to process audio signals, extract features, and classify psychological states; provides detailed outputs for expert monitoring and supports identification and optimization of underperforming modules.

+ Real-time reporting: Provides daily, monthly, and yearly reports; supports PDF export; aggregates monitoring lists and state progression to assist experts in making timely intervention decisions.

The system has been deployed in practice on a web-based platform at: https://www.ducbao-edu-careai.com/. Source code and installation guidelines are available at:

+ CodeAI:https://drive.google.com/drive/folders/1PTn-Vnh2xiYu76HKgtRnFKmfktujyjQ0V

+ CodeWeb:https://drive.google.com/drive/folders/1BO6uIUvb3EK5bqu8cpKr2QYTGZLraVx9

+ CodeAgent:https://drive.google.com/drive/folders/1r5yBxuFDD8I1UhevKWREOBG6UfCp8Lo1

4. DISCUSSION

This study developed and evaluated an artificial intelligence model based on voice features to identify manic states, and simultaneously implemented an integrated system for data acquisition, analysis, and decision support in mental health. Initial findings demonstrate the feasibility of this approach at both the model and real-world deployment levels, aligning with current trends in AI applications in psychiatry.

On the training and validation datasets, the model achieved stable performance, reflecting strong learning and generalization capacity. These findings are consistent with recent reviews reporting that machine learning models for detecting bipolar disorder using voice data achieve accuracy ranging from 70.9% to 96.9%, with AUC values for mania detection reaching up to 0.89 [8]. The high accuracy and clear probability separation between classes indicate that acoustic features such as rhythm, intensity, and prosody provide strong discriminative signals, consistent with prior evidence on the utility of speech analysis in mood state assessment.

When applied to real-world Vietnamese data, model performance was inconsistent across cases, particularly in the presence of regional dialect variations. This finding aligns with previous studies indicating that model performance may decline in naturalistic and heterogeneous datasets, as well as in complex classification tasks such as distinguishing bipolar disorder from depression [8, 9]. In contrast, more clearly defined tasks, such as identifying mania or distinguishing it from healthy controls, tend to yield higher performance, with AUC values around 0.89 or accuracy exceeding 80% in some studies [10, 11]. These results suggest that the current model is highly dependent on data homogeneity and is sensitive to linguistic variability. Therefore, future work should focus on developing region-specific or dialect-adaptive models, as well as expanding the training dataset to include larger and more diverse speech samples. In addition, incorporating geographic or regional metadata as input features may improve model robustness and enhance performance across different linguistic contexts.

The support system developed in this study helps bridge the gap between research models and real-world applications. Integrating AI into a software platform enables a closed-loop workflow from data collection to processing, analysis, and real-time reporting. This approach aligns with trends in mobile health (mHealth) for mood disorder monitoring, where studies using voice data from mobile devices have demonstrated the ability to track disease progression with AUC values around 0.70–0.76 in longitudinal models [12]. This suggests the system's potential for continuous monitoring and clinical support.

Based on these findings, expanding the training dataset in terms of size and phonetic diversity is essential to improve model generalization. Furthermore, previous studies indicate that personalized models outperform group-based models (correlation 0.78 vs. 0.44) [8], suggesting that individualized adaptive modeling may be a promising direction. At the same time, integration with clinical assessment remains necessary to ensure accuracy and safety in real-world applications.

This study has several limitations, including a small validation sample size, reliance on transformed training data, and limited representation of Vietnamese linguistic characteristics. These limitations are consistent with prior studies, which often face issues such as small sample sizes, single-center data, and reduced performance under real-world conditions. Therefore, future studies should adopt larger sample sizes and multicenter designs.

5. CONCLUSION

This study initially developed and evaluated an artificial intelligence model based on voice features for identifying manic states, while also implementing an integrated support system for data collection and analysis. The results demonstrate that the model is capable of distinguishing clinical states and shows potential for real-world application. However, performance on Vietnamese data remains limited, indicating the need to expand and diversify the training dataset, as well as to further optimize the model before large-scale deployment in clinical practice.

REFERENCES

[1] Nguyen Thi Hoai Thuong, Nguyen Van Tuan, Le Thi Thu Ha, and Pham Xuan Thang, Agitation in bipolar disorder during the manic phase: a cross-sectional study at the National Institute of Mental Health. Vietnam Medical Journal, 2022. 517(1).

[2] Nong Duc Dung, Tran Nguyen Ngoc, and Nguyen Van Tuan, Clinical characteristics of the manic phase in inpatients with bipolar I disorder. Journal of Medical Research, 2023. 170(9): p. 162–172.

[3] Le Thi Thu Ha, Study on clinical characteristics and treatment status of depression in patients with bipolar disorder at the Institute of Mental Health, V. Nguy-

en Kim and B. Tran Huu, Editors. 2018, Hanoi Medical University.

[4] Pan, Y., et al., Machine learning for the diagnosis accuracy of bipolar disorder: a systematic review and meta-analysis. Frontiers in Psychiatry, 2025. 15.

[5] Le, P.T.V., et al., Cross-Cultural Identification of Acoustic Voice Features for Depression: A Cross-Sectional Study of Vietnamese and Japanese Datasets. Bioengineering (Basel), 2025. 13(1).

[6] Ringeval, F., et al. AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition. In Proceedings of the Audio/Visual Emotion Challenge and Workshop (AVEC'18). 2018. Seoul, Republic of Korea: Association for Computing Machinery (ACM).

[7] Young, R.C., J.T. Biggs, V.E. Ziegler, and D.A. Meyer, A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry, 1978. 133: p. 429–435.

[8] Briganti, G. and J. Lechien, Voice Quality as Digital Biomarker in Bipolar Disorder: A Systematic Review. Journal of Voice, 2025.

[9] Pan, W., et al., Exploring the ability of vocal biomarkers in distinguishing depression from bipolar disorder, schizophrenia, and healthy controls. Frontiers in Psychiatry, 2023. 14.

[10] Faurholt-Jepsen, M., et al., Voice analysis as an objective state marker in bipolar disorder. Translational Psychiatry, 2016. 6.

[11] Ji, J., et al., Depressive and mania mood state detection through voice as a biomarker using machine learning. Frontiers in Neurology, 2024. 15.

[12] Min, S., et al., Automated Speech Analysis for Screening and Monitoring Bipolar Depression: Machine Learning Model Development and Interpretation Study. JMIR Medical Informatics, 2025. 13.