Iran J Public Health, Vol. 55, No.1, Jan 2026, pp.177-179

Letter to the Editor

Application of Explainable AI for Enhanced Depression Prediction among Community Adults in South Korea

*Haewon Byeon

Worker's Care & Digital Health Lab, Department of Future Technology, Korea University of Technology and Education (KOREA TECH), Cheonan 31253, South Korea

*Correspondence: Email: bhwpuma@naver.com

(Received 12 Aug 2025; accepted 21 Aug 2025)

Dear Editor-in-Chief

Depression is a pervasive mental health challenge globally, with significant implications for public health systems around the world (1). This is particularly true in South Korea, where the burden of depression and associated suicide rates are among the highest in the OECD (2). The societal and economic impacts of depression are profound, affecting not only individuals but also families and communities. Traditional diagnostic methods, though valuable, often fall short due to their reliance on subjective assessments, which can lead to inconsistent outcomes and limited scalability (3). These methods typically involve clinical interviews and self-reported questionnaires (3), which are inherently subjective and can be influenced by the patient's current mental state or willingness to disclose personal information. As a result, there is an urgent need for more objective and scalable solutions to accurately diagnose and predict depression in diverse populations.

This study aimed to bridge this gap by integrating explainable artificial intelligence (AI) to enhance the prediction of depression in community adults. This approach leverages advanced AI techniques to analyze complex datasets, providing more reliable and consistent diagnostic results. This research utilized data from the Korea National Health and Nutrition Examination Survey (KNHANES), which encompasses a comprehensive dataset of over 10,000 individuals. This rich dataset includes a wide range of variables, from demographic information to detailed health metrics, allowing for a multifaceted analysis of factors contributing to depression. I employed a combination of machine learning (ML) and deep learning (DL) models, including the Feature Tokenizer Transformer (FT-Transformer), to analyze this data. The FT-Transformer, with its ability to handle tabular data effectively, was particularly instrumental in achieving high accuracy and interpretability. Its design allows for the integration of both categorical and numerical data, making it ideal for the diverse dataset provided by KNHANES.

A key innovation of our study is the use of Tabular Variational Autoencoder (TVAE) to generate synthetic data. This approach addresses the challenges of data imbalance and privacy concerns inherent in working with sensitive health data. Data imbalance, a common issue in health data, can lead to biased models that do not generalize well across different populations (4). By generating synthetic data, TVAE helps create a balanced

DOI: https://doi.org/10.18502/ijph.v55i1.20985

dataset that reflects the true diversity of the population. Moreover, the synthetic data approach preserves patient confidentiality by obfuscating individual identifiers, thus mitigating privacy concerns and complying with data protection regulations. This not only enhances model training but also results in improved generalizability across different populations, making the model applicable to a broader range of clinical settings.

I also incorporated SHapley Additive exPlanations (SHAP) to provide insights into model predictions, offering both global and local interpretability. This dual approach allows healthcare professionals to understand the significance of different features in predicting depression, thereby increasing trust in AI-driven diagnostic tools. SHAP values provide a clear indication of how each feature contributes to a model's prediction, which is critical for clinical acceptance and integration. This transparency helps build confidence among healthcare providers and patients, fostering greater adoption of AI technologies in clinical practice.

The integration of synthetic data and advanced AI models significantly enhances the prediction of depression, with the FT-Transformer model demonstrating superior performance compared to traditional methods (Fig. 1). This research has the potential to transform depression screening processes, making them more efficient and accessible, and ultimately improving patient outcomes. By providing a more accurate and scalable solution, our approach can facilitate early intervention and treatment, reducing the long-term impact of depression on individuals and healthcare systems.

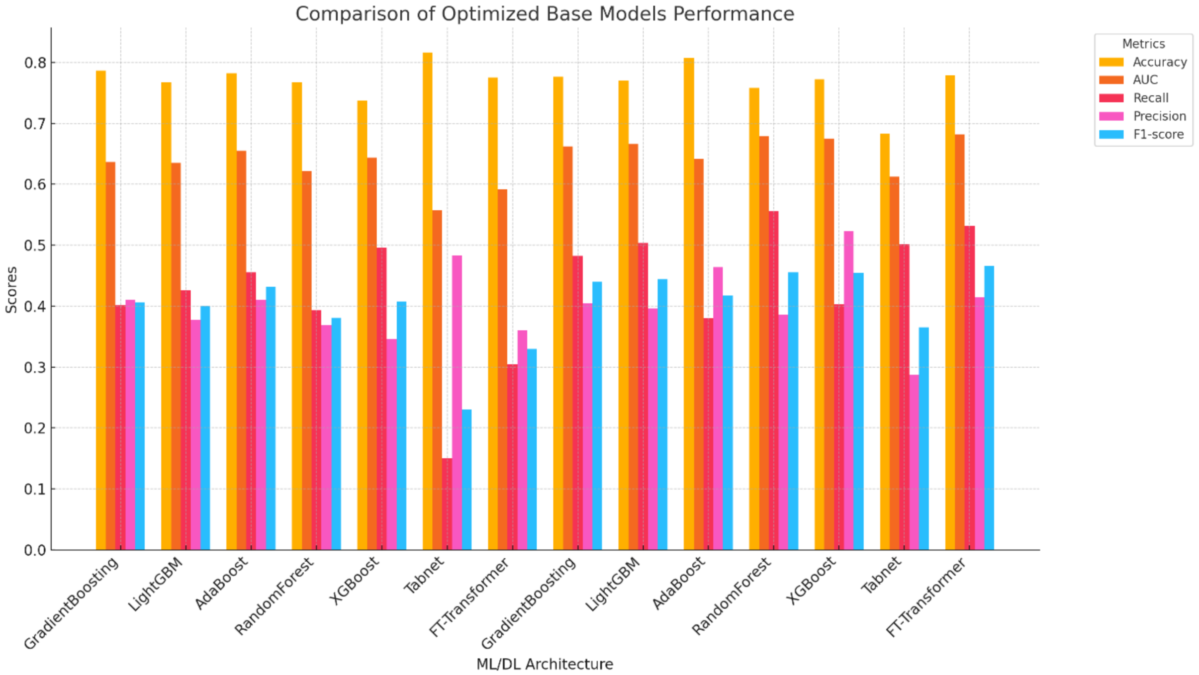

Fig. 1: Comparison of optimized base model performance

The implications of this study extend beyond depression prediction, offering a framework for applying explainable AI to other health conditions where privacy and data imbalance are concerns.

Conflict of interest

The authors declare no conflict of interests.

References

1. Moreno-Agostino D, Wu YT, Daskalopoulou C, Hasan MT, Huisman M, Prina M (2021). Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. *J Affect Disord*, 281:235-243.
2. Jang H, Lee W, Kim YO, Kim H (2022). Suicide rate and social environment characteristics in South Korea: the roles of socioeconomic, demographic, urbanicity, general health behaviors, and other environmental factors on suicide rate. *BMC Public Health*, 22(1):410.
3. Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J (2023). Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digital Medicine*, 6(1):84.
4. Eom G, Byeon H (2023). Searching for optimal oversampling to process imbalanced data: Generative adversarial networks and synthetic minority over-sampling technique. *Mathematics,*11(16):3605.