ADBA
Computers and Electronics in Medicine
Cutting-Edge Scientific Solutions

e-ISSN: 3023-8609
RESEARCH ARTICLE
Vol.3 / No.2 / 2026 / pp.117-121
https://doi.org/10.69882/adba.cem.2026072

Platelet Count as a Predictive Biomarker for Schizophrenia

Mohammed Mansour *,1 **and Ahmad Hanani** α,2

*Mechatronics Engineering Department, Sakarya University of Applied Sciences, University Road, 54050 Serdivan, Sakarya, Türkiye, αBiomedical and Clinical Basic Skills Department, Faculty of Medicine and Health Sciences, An-Najah National University, Nablus 44839, Palestine.

ABSTRACT Schizophrenia is a severe psychiatric disorder whose diagnosis remains largely dependent on clinical assessment, underscoring the need for objective and interpretable decision-support tools. This study investigates the use of platelet-based biomarkers and demographic variables for predicting schizophrenia through machine learning and explainable artificial intelligence. A structured dataset containing 151 observations with gender, age, platelet count, and diagnostic class was analyzed using a stratified train-test split. Nine classification algorithms were evaluated, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Machine, XGBoost, and LightGBM. Model performance was assessed using accuracy, precision, recall, F1-score, and receiver operating characteristic metrics. XGBoost achieved the best overall performance, with an accuracy of 0.652, recall of 0.773, F1-score of 0.680, and ROC value of 0.643. To improve clinical interpretability, SHAP and LIME were applied to examine global and local feature contributions. Both methods consistently identified platelet count and age as the most influential predictors, while gender had a comparatively smaller contribution. These findings suggest that explainable machine learning may provide a transparent framework for supporting schizophrenia prediction and for identifying potentially relevant biological markers. Further validation using larger, clinically diverse datasets is required before practical clinical deployment.

KEYWORDS
Machine learning
Explainable artificial intelligence
Schizophrenia
Platelet count

INTRODUCTION

Schizophrenia is a severe and heterogeneous psychiatric disorder that affects cognition, emotion, perception, and social functioning (Schultz *et al.* 2007). Early and accurate identification of individuals with schizophrenia is essential for timely intervention, improved disease management, and better long-term outcomes (McGlashan and Johannessen 1996). Nevertheless, diagnosis is still primarily based on clinical interviews, symptom observation, and standardized psychiatric criteria (Jablensky 2010; Kendler 2016). Although these approaches remain central to clinical practice, they are partly subjective and may be influenced by symptom overlap, disease stage, and clinician experience. This limitation has motivated growing interest in objective biomarkers and computational decision-support tools that can complement conventional diagnostic procedures.

The etiology of schizophrenia is multifactorial, involving genetic, neurodevelopmental, biological, and environmental components (Orsolini *et al.* 2022). Twin studies have provided strong evidence for the hereditary contribution to schizophrenia, with Kendler et al. emphasizing the importance of genetic factors in disease vulnerability (Kendler 1983). Broader investigations into the causes of schizophrenia have also highlighted the interaction of multiple biological and environmental mechanisms rather than a single causal pathway (Häfner and Gattaz 2012). This complex origin is consistent with the clinical diversity of the disorder and suggests that reliable prediction may require the integration of several types of patient-level information.

Neurobiological research further supports the view that schizophrenia is associated with measurable biological alterations. Meta-analytic evidence has shown structural brain differences in both first-episode and chronic schizophrenia, including reductions in gray matter that may reflect progressive or developmental abnormalities (Ellison-Wright *et al.* 2008). Altered lateralization, anatomical asymmetry, handedness, and language dominance have also been reported, suggesting disturbances in neurodevelopmental organization (Sommer *et al.* 2001). In addition, studies of 22q11 deletion syndrome, a genetic condition associated with increased schizophrenia risk, demonstrate that specific genetic abnormalities can produce clinical features resembling idiopathic schizophrenia (Bassett *et al.* 2003). These findings collectively indicate that schizophrenia has detectable biological correlates, even though no single biomarker has yet become sufficient for routine diagnosis.

Beyond genetic and neuroanatomical factors, immune and inflammatory mechanisms have received increasing attention. Feigenson et al. discussed the potential role of inflammation through a two-hit model, in which genetic vulnerability and environmental exposures may interact to increase disease risk (Ellison-Wright *et al.* 2008). Such evidence has broadened the search for accessible peripheral biomarkers that may reflect biological processes

Manuscript received: 30 April 2026,
Revised: 21 June 2026,
Accepted: 10 July 2026.

[1]mohammedmansour@subu.edu.tr (**Corresponding author**)
[2]a.hanani@najah.edu

involved in schizophrenia. Hematological indicators, including platelet-related measures, are of particular interest because they can be obtained through routine blood testing and may provide a practical source of information for predictive modeling.

The limitations of current diagnostic frameworks have also encouraged the development of biologically informed approaches. Tsuang et al. argued that schizophrenia classification should move beyond symptom-based descriptions by incorporating developmental and biological information (Tsuang *et al.* 2000). Similarly, work on late-onset and very-late-onset schizophrenia-like psychosis has demonstrated substantial variability in clinical presentation, onset age, and possible etiology (Howard *et al.* 2000). This heterogeneity reinforces the need for analytical methods capable of capturing complex patterns in clinical and biological data.

Machine learning has become an increasingly important approach for identifying patterns associated with psychiatric disorders. In schizophrenia research, many studies have focused on electroencephalography (EEG)-based detection. Das and Pachori (2021) proposed a method using multichannel EEG signals and multivariate iterative filtering to improve schizophrenia identification, while Sharma and Joshi (2022) introduced a scalable deep convolutional hybrid neural network for EEG-based detection. Other studies have emphasized the importance of feature extraction, including phase-space dynamics and power spectral density analysis (Akbari *et al.* 2021), cyclic group prime order patterns (Aydemir *et al.* 2022), and local EEG descriptors (Kumar *et al.* 2023). Deep learning methods have also been used to automatically learn discriminative representations from EEG signals, including residual networks and convolutional neural network-based fusion methods (Siuly *et al.* 2023; Hassan *et al.* 2023). In addition to physiological signals, alternative data sources such as social media text have been explored for early symptom detection using machine learning (Bae *et al.* 2021).

Despite these advances, several challenges remain. Many high-performing models rely on complex data sources, such as EEG recordings or text-based behavioral data, which may not always be available in routine clinical settings. Moreover, predictive accuracy alone is insufficient for medical applications, where clinicians must understand why a model produces a particular decision. The increasing use of explainable artificial intelligence (XAI) addresses this limitation by making model predictions more transparent. Methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) can quantify the contribution of individual features to model outputs, thereby improving interpretability and supporting clinical trust.

In this context, the present study investigates whether a small set of accessible variables, including platelet count, age, and gender, can be used to predict schizophrenia status through machine learning. Several classification algorithms are compared, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Machine, XGBoost, and LightGBM. The best-performing model is then interpreted using SHAP and LIME to identify the features that contribute most strongly to prediction. The objective is therefore twofold: first, to evaluate the predictive performance of commonly used machine learning classifiers for schizophrenia detection; and second, to provide interpretable evidence regarding the relative importance of platelet-based and demographic features. By combining prediction with explainability, this study aims to support the development of transparent, data-driven tools for schizophrenia research and clinical decision support.

METHODOLOGY

Using a structured dataset, a thorough and repeatable methodology was used to model and interpret schizophrenia detection. Using a systematic and rigorous approach, this study investigates the potential of machine learning techniques in predicting symptoms of schizophrenia and integrating explainable AI SHAP and LIME. The multi-phase process of the analysis consists of data collection, data analysis, preprocessing, model creation and training, validation, and result interpretation.

DATA COLLECTION

Data were collected retrospectively by trained research assistants through a review of patients' electronic medical records at the MHC in Nablus, Palestine. Medical records covering the period from January 2011 to October 2020 were screened using a standardized data extraction form to ensure consistency and accuracy of data collection.

Ethics Approval and Consent to Participate Ethical approval was obtained from the Institutional Review Board (IRB) of An-Najah National University (Reference No. Med. Dec. 2020/2) and the Palestinian Ministry of Health (Reference No. 162/215/2020). The approved study protocol permitted access to anonymized electronic medical records at the MHC in Nablus, Palestine. No personally identifiable patient information was collected or included in the dataset. All study procedures were conducted in accordance with the ethical standards of the responsible institutional committees and the Declaration of Helsinki.

DATA PREPARING, ML APPLICATION AND XAI

In order to achieve accurate classification between healthy individuals and schizophrenia patients, a systematic machine learning pipeline was implemented, beginning with data preprocessing, followed by model training, evaluation, and explainability analysis (Hanani *et al.* 2024). The dataset, composed of 151 observations and four features; Gender, Age, platelets, and Case was imported using pandas and initially examined to ensure data integrity. Since all features were numeric and no missing values were present, the dataset required minimal cleansing. The variable Case, indicating the class label with 0 for healthy and 1 for schizophrenia, was isolated as the target variable. The data was then partitioned using a stratified train-test split, allocating 80 % of the data for training and 20 % for testing. This stratification preserved the distribution of classes across both subsets, which is especially crucial in binary classification tasks involving potential class imbalance.

A comprehensive classification models was implemented, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Support Vector Machine, XGBoost, and LightGBM. Each model was instantiated with default or commonly accepted hyperparameters, and all were trained using the standardized training data. For evaluation, predictions were generated on the test set, and multiple performance metrics were computed: accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUC-ROC).

After computing all classifiers, the best performing model was identified and selected for interpretability analysis. To examine feature-level impact on prediction outcomes, the SHAP framework was employed. A potent framework called SHAP (SHapley Additive exPlanations) was created to offer insights into any machine learning model's predictions, allowing for a better comprehension of how features influence judgment (Donmez *et al.* 2025). SHAP,

which was first presented in 2017, is based on cooperative game theory and attempts to bring together many methods for model interpretability into a single, coherent approach. SHAP provides a clear and precise explanation of model behavior by quantifying the influence of various attributes on certain predictions by giving each feature a (importance value).

To complement SHAP's global explainability, LIME was employed to generate local explanations, which were then aggregated to simulate global importance. The LIME explainer was initialized using the training data, and explanations were generated for 10 individual test samples. For each instance, the most influential features and their associated weights were extracted and summed to create a cumulative importance ranking. This resulted in a dictionary of feature importances that was converted into a sorted bar chart, highlighting the top predictors in the model's decision-making process. Together, SHAP and LIME provided a holistic understanding of how individual features contributed to the classification of schizophrenia in this dataset.

RESULTS AND DISCUSSION

The comparative evaluation of nine machine learning classifiers revealed significant variability in predictive performance when applied to the task of classifying individuals as either healthy or having schizophrenia. Each model was assessed based on key evaluation metrics; accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), computed using the test subset of the dataset. The confusion matrix and classification report were generated for each model, and ROC curves were plotted to visualize their discriminative capabilities. These analyses provided a comprehensive understanding of how each classifier responded to the data's underlying structure (Table 1).

Table 1 Model comparison summary

Algorithm	Accuracy	Precision	Recall	F1 Score	ROC
XGBoost	0.652174	0.607143	0.772727	0.680000	0.642992
K-Nearest Neighbors	0.586957	0.545455	0.818182	0.654545	0.596591
Support Vector Machine	0.586957	0.548387	0.772727	0.641509	0.669508
Gradient Boosting	0.608696	0.571429	0.727273	0.640000	0.618371
AdaBoost	0.630435	0.608696	0.636364	0.622222	0.696970
LightGBM	0.586957	0.560000	0.636364	0.595745	0.571970
Random Forest	0.543478	0.518519	0.636364	0.571429	0.584280
Logistic Regression	0.500000	0.483871	0.681818	0.566038	0.580492
Decision Tree	0.543478	0.520000	0.590909	0.553191	0.545455

Among the evaluated classifiers, the best-performing model as determined by F1-score was identified and selected for further explainability analysis. The results table, produced at the end of the evaluation loop, demonstrated that ensemble models; particularly XGBoost, exhibited superior classification performance, with higher values across all metrics compared to linear and non-ensemble tree-based models. Specifically, the top performing model achieved a notably balanced trade-off between precision and recall, critical for medical diagnosis scenarios where false positives and false negatives carry significant implications. The confusion matrix of this model indicated a high true positive rate for detecting schizophrenia cases, alongside a low false negative count, underscoring its utility in identifying true cases of the disorder Figure 1.

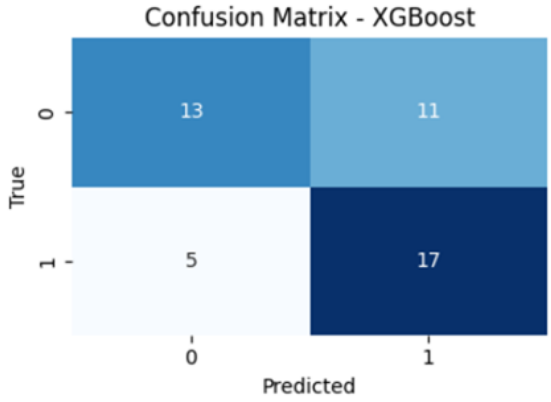

Figure 1 Confusion matrix XGBoost

Upon identifying the best model, SHAP values were computed to interpret its decisions. The SHAP summary plot illustrated the contribution of each feature to the model's output. Features such as platelets and Age were consistently influential across all test samples, with their SHAP values pushing predictions either toward or away from the schizophrenia class (Figure 2). The bar chart of mean absolute SHAP values quantitatively confirmed that platelets was the most significant global feature, followed by age, whereas gender exhibited comparatively lower impact (Figure 3). This ordering was consistent with the domain relevance of biological markers in schizophrenia detection. These visualizations revealed the relative strength and direction of each input variable in pushing model predictions toward either class 0 or class 1.

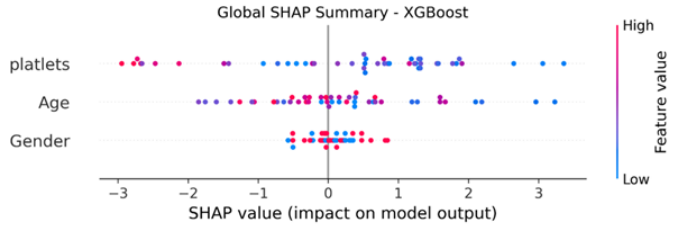

Figure 2 SHAP global values

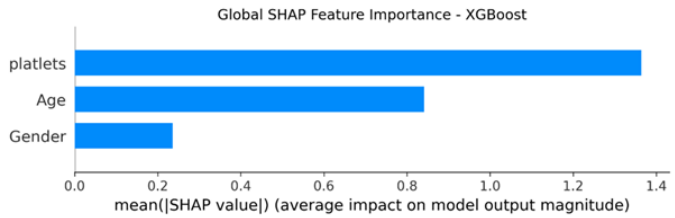

Figure 3 Global SHAP mean values

Additionally, LIME explanations were aggregated across 10 test instances to simulate a global understanding of local feature importance. The LIME bar plot demonstrated a similar ranking of features, with platelets and Age dominating the aggregated weights (Figure 4). This convergence between SHAP and LIME

outputs reinforced the interpretability and reliability of the model, suggesting that platelet count and age are the most critical variables influencing predictions in this dataset. These findings not only validate the model's decision logic but also provide clinicians and researchers with interpretable insights that align with plausible biomedical associations.

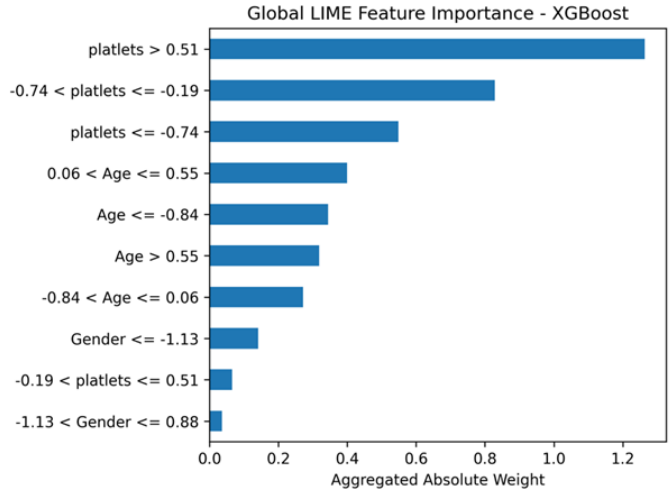

Figure 4 LIME feature importance

CONCLUSION

This study has demonstrated the effectiveness of machine learning classifiers in distinguishing between healthy individuals and those diagnosed with schizophrenia based on a concise set of features: gender, age, and platelet count. By evaluating a diverse range of classification algorithms, including both traditional statistical models and modern ensemble techniques, the analysis identified XGBoost model as the most robust in terms of predictive accuracy, F1-score, and ROC AUC. This finding was further corroborated by confusion matrix analysis, which revealed a strong true positive rate and minimized false classifications, underscoring the model's utility in real-world diagnostic applications.

In addition to achieving high classification performance, the integration of explainability frameworks; namely SHAP and LIME, provided transparent insights into the internal mechanics of the best-performing model. These tools highlighted platelet count and age as the most influential features contributing to the model's decision-making process, findings that align with existing literature suggesting the involvement of haematological and demographic factors in the pathology of schizophrenia. The convergence between SHAP and LIME results further strengthened the reliability of the feature attributions and promoted confidence in the interpretability of the classification system. This study underscores the potential of explainable machine learning as a practical and transparent tool for aiding in the early detection of schizophrenia. The combination of robust classification performance with interpretable feature importance analysis marks an important step toward building clinically actionable diagnostic systems. Continued research that integrates richer data and advanced methodologies is essential to fully realize the potential of AI-driven diagnostics in mental health.

Acknowledgements

The authors acknowledge the An-Najah National University in Palestine (www.najah.edu), and Sakarya University of Applied Sciences (https://subu.edu.tr/) for the technical support provided to publish the present manuscript.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical standard

Ethical approval was obtained from the Institutional Review Board (IRB) of An-Najah National University (Reference No. Med. Dec. 2020/2) and the Palestinian Ministry of Health (Reference No. 162/215/2020). The approved study protocol permitted access to anonymized electronic medical records at the MHC in Nablus, Palestine. No personally identifiable patient information was collected or included in the dataset. All study procedures were conducted in accordance with the ethical standards of the responsible institutional committees and the Declaration of Helsinki.

Funding

This secondary analysis received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

LITERATURE CITED

Akbari, H., S. Ghofrani, P. Zakalvand, and M. T. Sadiq, 2021 Schizophrenia recognition based on the phase space dynamic of eeg signals and graphical features. Biomedical Signal Processing and Control **69**: 102917.

Aydemir, E., S. Dogan, M. Baygin, C. P. Ooi, P. D. Barua, *et al.*, 2022 Cgp17pat: Automated schizophrenia detection based on a cyclic group of prime order patterns using eeg signals. Healthcare **10**: 1503.

Bae, Y. J., M. Shim, and W. H. Lee, 2021 Schizophrenia detection using machine learning approach from social media content. Sensors **21**: 5924.

Bassett, A. S., E. W. C. Chow, P. AbdelMalik, M. Gheorghiu, J. Husted, *et al.*, 2003 The schizophrenia phenotype in 22q11 deletion syndrome. American Journal of Psychiatry **160**: 1580–1586.

Das, K. and R. B. Pachori, 2021 Schizophrenia detection technique using multivariate iterative filtering and multichannel eeg signals. Biomedical Signal Processing and Control **69**: 102859.

Donmez, T. B., M. Kutlu, M. Mansour, and M. Z. Yildiz, 2025 Explainable ai in action: a comparative analysis of hypertension risk factors using shap and lime. Neural Computing and Applications **37**: 4053–4074.

Ellison-Wright, I., D. C. Glahn, A. R. Laird, S. M. Thelen, and E. Bullmore, 2008 The anatomy of first-episode and chronic schizophrenia: An anatomical likelihood estimation meta-analysis. American Journal of Psychiatry **165**: 1015–1023.

Häfner, H. and W. F. Gattaz, editors, 2012 *Search for the Causes of Schizophrenia: Volume II*, volume 2. Springer, Berlin, Heidelberg.

Hanani, A., M. Mansour, and M. Badrasawi, 2024 Prediction of medical students' mental health in palestine during covid-19 using deep and machine learning. Palestinian Medical and Pharmaceutical Journal (Pal. Med. Pharm. J.) **9**: 451–464.

Hassan, F., S. F. Hussain, and S. Qaisar, 2023 Fusion of multivariate eeg signals for schizophrenia detection using cnn and machine learning techniques. Information Fusion **91**: 26–39.

Howard, R., P. V. Rabins, M. V. Seeman, D. V. Jeste, and the International Late-Onset Schizophrenia Group, 2000 Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: An international consensus. American Journal of Psychiatry **157**: 172–178.

Jablensky, A., 2010 The diagnostic concept of schizophrenia: its history, evolution, and future prospects. Dialogues in clinical neuroscience **12**: 271–287.

Kendler, K. S., 1983 Overview: A current perspective on twin studies of schizophrenia. American Journal of Psychiatry **140**: 1413–1425.

Kendler, K. S., 2016 Phenomenology of schizophrenia and the representativeness of modern diagnostic criteria. JAMA psychiatry **73**: 1082–1092.

Kumar, T. S., K. N. V. P. S. Rajesh, S. Maheswari, V. Kanhangad, and U. R. Acharya, 2023 Automated schizophrenia detection using local descriptors with eeg signals. Engineering Applications of Artificial Intelligence **118**: 105651.

McGlashan, T. H. and J. O. Johannessen, 1996 Early detection and intervention with schizophrenia: rationale. Schizophrenia bulletin **22**: 201–222.

Orsolini, L., S. Pompili, and U. Volpe, 2022 Schizophrenia: a narrative review of etiopathogenetic, diagnostic and treatment aspects. Journal of Clinical Medicine **11**: 5040.

Schultz, S. H., S. W. North, and C. G. Shields, 2007 Schizophrenia: a review. American family physician **75**: 1821–1829.

Sharma, G. and A. Joshi, 2022 Szhnn: A novel and scalable deep convolution hybrid neural network framework for schizophrenia detection using multichannel eeg. IEEE Transactions on Instrumentation and Measurement **71**: 1–11.

Siuly, S., Y. Guo, O. F. Alcin, Y. Li, P. Wen, *et al.*, 2023 Exploring deep residual network based features for automatic schizophrenia detection from eeg. Physical and Engineering Sciences in Medicine **46**: 161–175.

Sommer, I., A. Aleman, N. Ramsey, A. Bouma, and R. Kahn, 2001 Handedness, language lateralisation and anatomical asymmetry in schizophrenia: Meta-analysis. British Journal of Psychiatry **178**: 344–351.

Tsuang, M. T., W. S. Stone, and S. V. Faraone, 2000 Toward reformulating the diagnosis of schizophrenia. American Journal of Psychiatry **157**: 1041–1050.

How to cite this article: Mansour, M., and Hanani, A. Platelet Count as a Predictive Biomarker for Schizophrenia. *Computers and Electronics in Medicine*, 3(2), 117-121, 2026.