NeuroInk: Transparent AI for Mental Health Assessment Using Handwriting Dynamics

Dhairya Korgaonkar[1], Payal Patil[2], Samina Siddiquie[3]

[1]*Dhairya Korgaonkar, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic*
[2]*Payal Patil, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic*
[3]*Samina Siddiquie, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic*

---***---

Abstract - The detection of schizophrenia (SZ) and bipolar disorder traditionally relies on expensive imaging techniques such as MRI, which can limit accessibility and patient compliance. To overcome these challenges, this study proposes a cost-effective handwriting-based analysis approach that leverages motor abnormalities commonly associated with these disorders. The DataRepository SAV dataset from Figshare is used, containing handwriting-derived features for classification.

Data preprocessing includes robust feature selection using Recursive Feature Elimination (RFE) and handling class imbalance through cost-sensitive learning with class weighting. Several machine learning models are implemented, including XGBoost, Logistic Regression, Linear Discriminant Analysis (LDA), Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM with linear and RBF kernels). An ensemble Voting Classifier combining XGBoost and Logistic Regression is also developed to improve predictive performance.

Model evaluation is conducted using accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and confusion matrix. The ensemble model achieves the highest test accuracy. For interpretability, SHAP-based Explainable AI (XAI) techniques are applied to identify key contributing features. Finally, the optimized model is deployed through a Flask-based web application for real-time and transparent patient classification.

Key Words: Handwriting analysis, Schizophrenia detection, bipolar disorder classification, Machine learning, Explainable Artificial Intelligence, Cost-sensitive learning.

1.INTRODUCTION

Schizophrenia (SZ) and bipolar disorder are chronic psychiatric conditions characterized by impaired perception of reality, disorganized thinking, emotional dysregulation, and abnormal social behavior [1]. The diagnosis of these disorders traditionally relies on structured clinical interviews and standardized guidelines such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [2]. Clinical assessment tools, including the Positive and Negative Syndrome Scale (PANSS), are widely used to evaluate symptom severity and disease progression in patients with schizophrenia [3]. However, these conventional diagnostic procedures are inherently subjective. Patients may underreport or inconsistently describe symptoms, while clinicians often face challenges in distinguishing schizophrenia from other psychiatric or affective disorders with overlapping features. Similarly, the manual diagnosis of bipolar disorder is time-consuming, highly dependent on practitioner expertise, and prone to diagnostic variability. These limitations highlight the need for more objective, reliable, and data-driven diagnostic techniques.

With rapid advancements in artificial intelligence, machine learning (ML) has emerged as a powerful tool for disease detection and classification across healthcare domains. Automated ML systems have demonstrated promising results in diagnosing conditions such as Parkinson's disease [4], heart failure [5], breast cancer [6], and stroke [7], offering improved accuracy and scalability. In psychiatric research, Winterburn et al. applied logistic regression, linear discriminant analysis, and support vector machines to MRI datasets, achieving up to 73.5% accuracy in distinguishing schizophrenia patients from healthy controls [8]. Similarly, Messinger et al. reported 92.9% accuracy using ensemble models such as random forests [9]. Although these studies validate the feasibility of ML in psychiatric diagnostics, neuroimaging-based approaches remain costly and difficult to scale in clinical practice.

High operational expenses, limited equipment availability, and strict patient compliance requirements restrict the widespread use of neuroimaging techniques. Patients with schizophrenia or bipolar disorder may experience agitation or restlessness, affecting imaging quality. Consequently, there is growing interest in affordable, noninvasive, and easily deployable alternatives. Handwriting analysis has emerged as a promising behavioral biomarker. Research by Gawda demonstrated significant differences in handwriting density and spatial characteristics among schizophrenia patients, bipolar patients, and healthy individuals [10].

Motor abnormalities and cognitive impairments associated with these disorders are reflected in handwriting dynamics, including variations in velocity, pressure, rhythm, and spatial organization. These measurable motor signatures provide valuable insights into underlying neurological dysfunction.

In response, the present study proposes a handwriting-based machine learning framework for accurate and scalable detection of schizophrenia and bipolar disorder. By integrating robust feature selection, cost-sensitive learning, ensemble classification, and explainable artificial intelligence, the system aims to enhance diagnostic accuracy while ensuring transparency and clinical interpretability, thereby supporting objective decision-making and early intervention.

2 Working Methodology

The proposed system is designed for the handwriting-based detection of schizophrenia (SZ) and bipolar disorder using machine learning techniques. The framework utilizes the DataRepository SAV dataset obtained from Figshare, which contains handwriting-derived motor features associated with psychiatric conditions. Preprocessing includes statistically robust feature selection using Recursive Feature Elimination (RFE) and cost-sensitive learning through class weighting to address class imbalance. Classification is performed using multiple machine learning algorithms, including XGBoost, Logistic Regression, Linear Discriminant Analysis (LDA), Gaussian Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines with linear and RBF kernels. Furthermore, a Voting Classifier integrating XGBoost and Logistic Regression is introduced to enhance predictive performance. To ensure transparency and clinical interpretability, Explainable Artificial Intelligence (XAI) based on SHAP is employed, while deployment is facilitated through a Flaskbased web application.

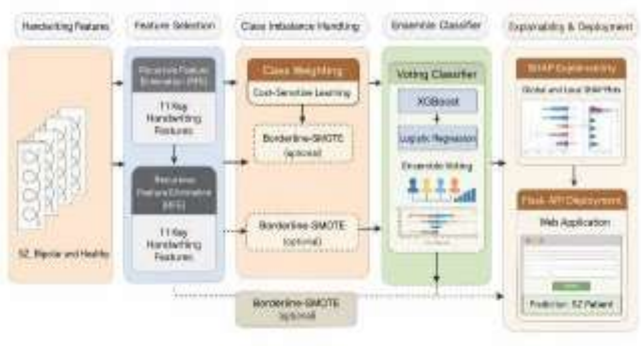

Fig -1: Proposed Architecture

The proposed system follows a structured, sequential, and modular pipeline, as illustrated in Fig. 1. The architecture consists of handwriting feature acquisition, exploratory analysis, feature selection, class imbalance handling, ensemble classification, explainability integration, and deployment. Each stage is carefully designed to ensure reliability, interpretability, and real-world clinical usability.

a) Dataset Collection (Handwriting Features Module)

The framework utilizes the SAV dataset obtained from Figshare, comprising 98 instances with 24 attributes. These features include demographic variables (age, sex), clinical measures (diagnostic category, medication dosage, PANSS scores), and handwriting-derived motor metrics such as velocity, acceleration, stroke length, pressure, peak count, fractal dimensions, entropy, and lacunarity. These attributes reflect motor dysfunction and cognitive impairment commonly associated with schizophrenia and bipolar disorder. The dataset serves as the foundational input to the architecture.

b) Data Visualization (Exploratory Analysis Stage)

Exploratory Data Analysis (EDA) is conducted before model training to understand dataset characteristics. Correlation matrices are generated to examine relationships among numerical features and detect multicollinearity. Outcome distribution plots are also used to visualize class imbalance among schizophrenia, bipolar disorder, and healthy control groups. This stage supports informed preprocessing decisions.

c) Feature Selection (Recursive Feature Elimination Module)

Recursive Feature Elimination (RFE) is employed to select the most informative features. RFE iteratively removes less significant variables based on model coefficients, retaining only those contributing most to classification performance. This process reduces redundancy, improves computational efficiency, and enhances interpretability. Ultimately, 11 key features are selected, capturing discriminative motor and behavioral patterns.

d) Class Imbalance Handling (Cost-Sensitive Learning Module)
Class imbalance is addressed using cost-sensitive learning through class weighting. Higher misclassification penalties are assigned to minority classes, ensuring balanced decision boundaries without generating synthetic samples. This method preserves data integrity and reduces overfitting risk. Borderline-SMOTE is optionally used during experimentation to analyze boundary-sensitive samples. Feature normalization using StandardScaler ensures consistent feature contribution.

e) Training and Testing (Model Validation Stage)
The dataset is split into training and testing sets using an 80:20 ratio. A fixed random seed ensures reproducibility. This validation strategy prevents data leakage and enables reliable performance comparison across classifiers.

f) Algorithms (Ensemble Classification Module)
Multiple machine learning models are trained using RFE-selected features and class weighting:

- **RFE–CW–XGBoost:** Captures complex non-linear relationships in handwriting data.
- **RFE–CW–Logistic Regression:** Provides interpretable linear classification.
- **RFE–CW–LDA:** Maximizes class separability efficiently.
- **RFE–CW–Gaussian Naïve Bayes:** Applies probabilistic modeling under independence assumptions.
- **RFE–CW–KNN:** Classifies samples based on similarity measures.
- **RFE–CW–SVM (Linear & RBF):** Constructs optimal decision boundaries for linear and non-linear separability.

A Voting Classifier combining XGBoost and Logistic Regression is implemented as the final predictive model, leveraging complementary strengths to improve stability and diagnostic accuracy.

g) Explainability and Deployment (XAI & Web Application Module)
Explainable Artificial Intelligence is integrated using SHAP to provide global and local interpretability. SHAP visualizations (summary, force, and waterfall plots) highlight feature contributions, ensuring transparency and clinical trust.

Experimental Results

To evaluate the effectiveness of the proposed handwriting-based classification framework, several standard performance metrics are employed: Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC). These metrics provide a comprehensive assessment of model performance, particularly in medical diagnostic applications where both false positives and false negatives have critical consequences.

(1) Accuracy
Accuracy measures the overall ability of the model to correctly classify patients and healthy individuals. It represents the proportion of correctly predicted instances out of the total number of samples evaluated.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Where:

- **TP** = True Positives
- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

(2) Precision
Precision evaluates the proportion of correctly predicted positive cases among all cases predicted as positive. It reflects the reliability of positive predictions and helps assess how many identified patients truly belong to the target class.

$$\text{Precision} = \frac{TP}{TP + FP}$$

(3) Recall
Recall, also known as Sensitivity, measures the model's ability to correctly identify all actual positive cases. In medical diagnosis, high recall is crucial to minimize false negatives and avoid missed diagnoses.

$$\text{Recall} = \frac{TP}{TP + FN}$$

(4) F1-Score
The F1-Score provides a balanced measure by combining precision and recall into a single metric. It is particularly useful when dealing with imbalanced datasets.

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

(5) Matthews Correlation Coefficient (MCC)
The Matthews Correlation Coefficient (MCC) is a robust metric that considers all four confusion matrix components (TP, TN, FP, FN). It provides a balanced evaluation even when class sizes are unequal and ranges from −1 to +1, where +1 indicates perfect classification, 0 indicates random prediction, and −1 indicates total disagreement.

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

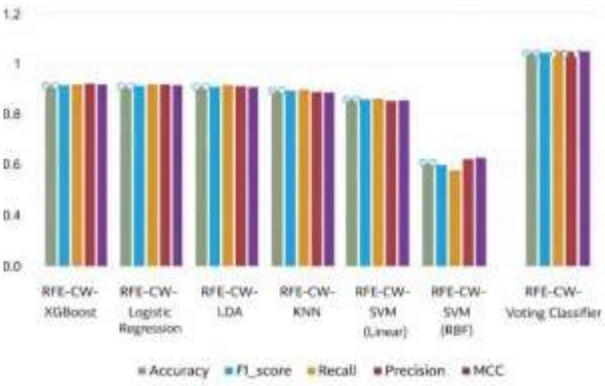

Fig. 2 Comparison Graph

Fig. 2 presents the comparative performance of multiple machine learning classifiers evaluated using five metrics: accuracy (green), F1-score (sky blue), recall (orange), precision (brown), and Matthews
Correlation Coefficient (MCC) (purple). The results demonstrate that models trained using Recursive Feature Elimination (RFE) and cost-sensitive learning (CW) achieve consistently strong performance across all metrics. Among the evaluated approaches, the RFE–CW–Voting Classifier integrating XGBoost and Logistic Regression achieves the highest scores for all evaluation metrics, indicating superior robustness, balanced classification, and improved generalization compared to individual classifiers. The comparative analysis highlights the effectiveness of ensemble learning in enhancing diagnostic performance for handwriting-based detection of schizophrenia and bipolar disorder.

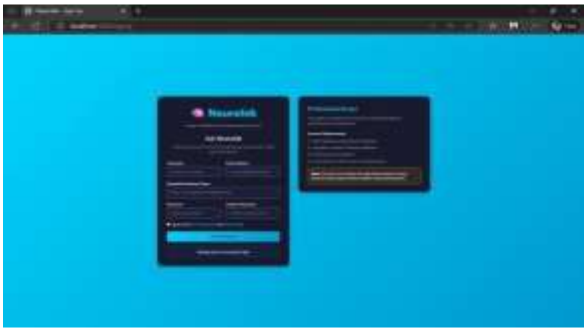

Fig. 3 Upload Input Data

Fig. 3 shows the web-based user interface where handwriting-derived features are provided as input for schizophrenia and bipolar disorder detection.

Fig. 4 – Predicted Results

Fig. 4 presents the system output for abnormal handwriting input, displaying the prediction result as “Abnormal handwriting detected”, indicating potential psychiatric motor irregularities.

Fig. 5 Upload Parameter’s information

Fig. 5 illustrates an parameter-wise reading of dielectric

pen which user will feed, Fig. 6 Shap- Waterfall Plot

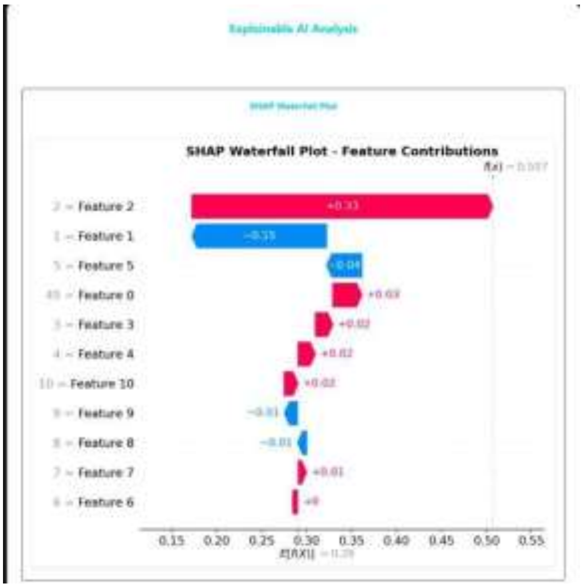

Fig. 6 displays the system with a graph of feature contributions in the analysis,

3. CONCLUSIONS

The proposed system demonstrates that handwriting-based analysis can serve as a reliable, non-invasive, and accessible alternative for detecting schizophrenia (SZ) and bipolar disorder. By addressing the limitations of costly and less practical imaging techniques such as MRI, the framework offers a scalable and clinically feasible diagnostic support solution. Using the DataRepository SAV dataset from Figshare, statistically robust feature selection was performed through Recursive Feature Elimination (RFE) to identify the most informative handwriting-derived and clinical features. Class imbalance was effectively managed using cost-sensitive learning through class weighting, ensuring balanced model performance without altering the original data distribution.

Multiple machine learning algorithms—including XGBoost, Logistic Regression, Linear Discriminant Analysis (LDA), Gaussian Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines (linear and RBF kernels)—were implemented and evaluated. To enhance predictive stability and performance, a Voting Classifier integrating XGBoost and Logistic Regression was employed, achieving superior results across accuracy, precision, recall, F1-score, MCC, and confusion matrix analysis. The integration of SHAP-based Explainable Artificial Intelligence ensured feature-level interpretability, strengthening transparency and supporting trustworthy clinical decision-making. The final optimized model was successfully deployed through a Flask-based web application, enabling real-time, user-friendly predictions.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the faculty members of the Department of Artificial Intelligence and Machine Learning for their continuous guidance, constructive feedback, and academic support throughout the completion of this research work. Their insights into machine learning methodologies, statistical modeling, and healthcare data analytics significantly contributed to the development and refinement of the proposed handwriting-based diagnostic framework.

The authors also extend their heartfelt appreciation to their project mentor for valuable technical discussions and methodological recommendations that enhanced the clarity, experimental design, and performance evaluation of the classification system. Their guidance played an essential role in strengthening the robustness and interpretability of the proposed model.

Furthermore, the authors acknowledge the open-source research community and the developers of essential tools and libraries, including Scikit-learn, XGBoost, SHAP, Pandas, NumPy, and Flask, which supported the implementation, experimentation, and deployment of this system. The authors also sincerely thank the contributors of the DataRepository SAV dataset available on Figshare, which formed the foundation of this research.

Finally, the authors express their gratitude to their peers and colleagues for their assistance during data analysis, validation, and system testing, which helped refine and strengthen the overall framework.

REFERENCES

[1] 1 Jamali, A., Kargar, R., Alipour, S., & Rostamy-Malkhalifeh, M. (2025). Machine learning approach for bipolar disorder analysis and recognition based on handwriting digital images. AUT Journal of Mathematics and Computing, 6(3), 223-239.

[2] Faundez-Zanuy, M., Mekyska, J., & Impedovo, D. (2021). Online handwriting, signature and touch dynamics: tasks and potential applications in the field of security and health. Cognitive Computation, 13(5), 1406-1421.

[3] Deepika, S., Suhitha, R., Alekhya, D., & VenkataVara Prasad, P. (2022, December). A Study on Borderline Personality Disorder using Machine Learning and Conventional Methods. In 2022 5th

International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1442-1446). IEEE.

[4] Casalino, G., Dominiak, M., Galetta, F., & Kaczmarek-Majer, K. (2020, May). Incremental semi-supervised fuzzy C-Means for bipolar disorder episode prediction. In 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS) (pp. 18). IEEE.

[5] Junaid, M., Ghergherehchi, M., & Lee, S. (2025). Multitask Deep Learning for Predicting Parkinson's Progression and Depression from Multimodal Time Series Data. IEEE Access.

[6] N. Sobahi, B. Ari, H. Cakar, O. F. Alcin, and A. Sengur, "A new signal to image mapping procedure and convolutional neural networks for efficient schizophrenia detection in EEG recordings," IEEE Sensors Journal, vol. 22, no. 8, pp. 7913–7919, 2022.

[7] Y. Zhao, Y. Liu, W. Lu, J. Li, P. Shan, C. Lian, X. Wang, C. Fu, C. Ma, and Y. Wang, "Intelligent IoT anklets for monitoring the assessment of Parkinson's diseases," IEEE Sensors Journal, 2023.

[8] Q. Wei, Y. Wang, Z. Zhou, R. Ma, and D. Li, "Classification of heart failure based on phase trajectory complexity of seismocardiogram," IEEE Sensors Journal, 2023.

[9] B. Wang, X. Zhang, L. Zhang, C. Li, T. Wu, and X. Wang, "Investigation of transducer distribution in compressive thermoacoustic tomography for breast cancer detection," IEEE Sensors Journal, vol. 24, no. 1, pp. 788–797, 2023.

[10] X. Wang, J. Zhang, S. Q. Xie, C. Shi, J. Li, and Z.-Q. Zhang, "Quantitative upper limb impairment assessment for stroke rehabilitation: A review," IEEE Sensors Journal, 2024.

[11] J. L. Winterburn, A. N. Voineskos, G. A. Devenyi, E. Plitman, C. de la FuenteSandoval, N. Bhagwat, A. Graff-Guerrero, J. Knight, and M. M. Chakravarty, "Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study," Schizophrenia Research, vol. 214, pp. 3–10, 2019.

[12] J. Walsh-Messinger, H. Jiang, H. Lee, K. Rothman, H. Ahn, and D. Malaspina, "Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning," Psychiatry Research, vol. 278, pp. 27–34, 2019.

[13] B. Gawda, "Density parameters of handwriting in schizophrenia and affective disorders assessed using the Raygraf computer software," in ICDAR 2021: 16th International Conference on Document Analysis and Recognition, pp. 380–394, Springer, 2021.

[14] C. R. Pereira, D. R. Pereira, F. A. da Silva, C. Hook, S. A. Weber, L. A. Pereira, and J. P. Papa, "A step towards the automated diagnosis of Parkinson's disease: Analyzing handwriting movements," in IEEE 28th International Symposium on Computer-Based Medical Systems, pp. 171– 176, IEEE, 2015.

[15] Y. Crespo, A. Ibanez, M. F. Soriano, S. Iglesias, and J. I. Aznarte, "Handwriting movements for assessment of motor symptoms in schizophrenia spectrum disorders and bipolar disorder," PLoS One, vol. 14, no. 3, 2019.

[16] Y. Crespo, M. Soriano, S. IglesiasParro, J. Aznarte, and A. Ibañez Molina, "Spatial analysis of handwritten texts as a marker of cognitive control,"

Journal of Motor Behavior, vol. 50, no. 6, pp. 643–652, 2018.

[17] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," IEEE Journal of Translational Engineering in Health and Medicine, vol. 7, pp. 1–10, 2019.

[18] Z. Rasheed, Y.-K. Ma, I. Ullah, T. Al Shloul, A. B. Tufail, Y. Y. Ghadi, M. Z. Khan, and H. G. Mohamed, "Automated classification of brain tumors from magnetic resonance imaging using deep learning," Brain Sciences, vol. 13, no. 4, p. 602, 2023.

[19] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and fuzzy AHP for heart failure risk prediction," Expert Systems with Applications, vol. 68, pp. 163–172, 2017.

[20] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," Journal of Machine Learning Research, vol. 18, no. 1, pp. 559–563, 2017.