Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

A hybrid deep learning framework for detecting bipolar disorder through Persian handwriting patterns

Khosro Rezaee[1*]

1. Department of Biomedical Engineering, Meybod University, Meybod, Iran.

Article Info

Article History:
Received
Revised
Accepted

DOI:

Keywords:
Bipolar Disorder Diagnosis
Handwriting Analysis
Deep Learning Applications
Persian Handwriting
Artificial Intelligence in Mental Health.

**Corresponding Author's Email Address: kh.rezaee@meybod.ac.ir*

Abstract

Bipolar disorder (BD) remains a pervasive mental health challenge, demanding innovative diagnostic approaches beyond traditional, subjective assessments. This study pioneers a non-invasive handwriting-based diagnostic framework, leveraging the unique interplay between psychological states and motor expressions in writing. Our hybrid deep learning model, combining ResNet for intricate feature extraction and external attention mechanisms for global pattern analysis, achieves a remarkably high accuracy 99%, validated through Leave-One-Subject-Out (LOSO) cross-validation. Augmented with advanced data preprocessing and augmentation techniques, the framework adeptly addresses dataset imbalances and handwriting variability. For the first time, Persian handwriting serves as a medium, bridging cultural gaps in BD diagnostics. This work not only establishes handwriting as a transformative tool for mental health diagnostics but also sets the stage for accessible, scalable, and culturally adaptive solutions in global mental healthcare.

1. Introduction

Bipolar disorder (BD) is a severe and complex mental health condition characterized by alternating episodes of mania and depression, significantly affecting patients' quality of life, relationships, and productivity [1,2]. Globally, BD affects approximately 3% of the population, making it a critical public health issue [3]. Accurate and timely diagnosis is essential for effective treatment, but current diagnostic methods often rely heavily on subjective clinical assessments and patient self-reporting. These approaches can lead to delayed diagnosis or misclassification, particularly when BD symptoms overlap with other mood disorders such as major depressive disorder (MDD) [4]. Developing objective, data-driven diagnostic tools is vital to improving early detection and intervention strategies, ultimately reducing the burden of the disorder on individuals and healthcare systems [5].

In recent years, researchers have explored various modalities for BD diagnosis, including genomic data, neuroimaging, and Electroencephalogram (EEG) signals. For instance, Laksshman et al. [6] developed the DeepBipolar model based on convolutional neural networks (CNNs) to predict BD phenotypes using limited genetic data. Despite its high accuracy, the reliance on costly and complex datasets limits its accessibility. Similarly, Metin et al. [7] employed deep learning techniques to analyze EEG data, achieving accuracies exceeding 95%. However, the high computational demands and need for extensive data hinder the practicality of this method.

This study aims to address the critical need for more accessible and non-invasive diagnostic tools for BD. One specific question we explore is: Can handwriting analysis, a non-invasive and cost-effective modality, be used reliably to detect BD? Handwriting analysis has recently gained attention as a non-invasive and cost-effective modality for identifying motor and cognitive impairments linked to BD. Studies like Crespo et al. [8] have demonstrated that handwriting features, including pressure, velocity, and fluency, vary significantly between BD patients and healthy individuals. What are the specific handwriting features that could serve as biomarkers for BD diagnosis, and how can deep learning methods improve the identification of these features? While promising, these approaches often lack robust datasets and advanced feature extraction methods, limiting their diagnostic potential. Moreover, Ayaz et al. [9] explored a multimodal approach for mental disorder recognition by integrating audio, video, and text data, utilizing recurrent neural networks and adaptive fusion techniques for enhanced accuracy and performance.

Despite advancements in handwriting analysis as a diagnostic tool, significant gaps persist, particularly in its application to Persian handwriting. What challenges arise when applying handwriting analysis to Persian handwriting, and how can these challenges be addressed to enhance diagnostic accuracy? Current methods often depend on limited datasets and lack representation of culturally specific handwriting styles [10], especially for Persian-speaking communities [11]. Additionally, advanced deep learning techniques for feature extraction and classification have not been fully utilized in this context, and image augmentation strategies to enhance model generalization and robustness remain underexplored. While Jamali et al. [12] have conducted a study on Persian handwriting using image processing techniques and machine learning models, their work was limited in scope and did not incorporate advanced deep learning or comprehensive datasets. This study specifically addresses the question: Can a deep learning-based approach leveraging Persian handwriting data overcome current limitations and improve BD diagnosis in culturally diverse populations? This gap underscores the urgent need for innovative approaches tailored to Persian handwriting, leveraging advanced methodologies to address these limitations and expand the applicability of handwriting analysis in culturally diverse settings. Our research aims to explore the feasibility of this approach, focusing on improving the diagnostic accuracy for BD in Persian-speaking populations. The contributions of this study are as follows:

1. For the first time, Persian handwriting is systematically analyzed as a diagnostic biomarker, addressing cultural and linguistic gaps in mental health research. Advanced augmentation techniques and tailored preprocessing methods ensure that this approach is adaptable to diverse populations.
2. We developed and implemented a novel deep learning-based model that combines ResNet for extracting intricate features and External Attention to focus on critical handwriting characteristics, specifically tailored for Persian handwriting analysis.
3. The proposed framework was rigorously evaluated against state-of-the-art methods for bipolar disorder diagnosis. Our approach demonstrated exceptional performance, achieving an accuracy exceeding 99%, highlighting its effectiveness and reliability.

Our approach represents the first systematic effort to augment and analyze Persian handwriting data for bipolar disorder diagnosis. By incorporating techniques such as rotation, scaling, and noise addition, we significantly expand the dataset's diversity and improve model performance. The integration of Residual Network (ResNet) and External Attention enhances feature extraction and focus, enabling superior accuracy and generalizability. This research bridges existing gaps and contributes to developing culturally relevant, cost-effective, and accessible diagnostic tools for mental health disorders.

The remainder of this article is organized as follows: Section 2 reviews related work, highlighting key advancements and limitations in BD diagnosis methods. Section 3 outlines the proposed methodology, detailing the model architecture and data preprocessing techniques. Section 4 presents experimental results and comparisons with existing methods. Finally, Section 5 discusses the implications of the findings, future research directions, and potential clinical applications.

2. Related Work

Research on handwriting analysis as a diagnostic tool for bipolar disorder remains limited, and most existing studies have relied on alternative modalities such as EEG, Magnetic Resonance Imaging (MRI), multimodal fusion, or genomic analysis. While these approaches have demonstrated strong classification performance, they present significant challenges in terms of cost, accessibility, and computational complexity. Below, we review the existing methods and highlight their strengths and limitations, while also positioning our work as a novel contribution that addresses key shortcomings in handwriting-based bipolar disorder diagnosis. Ayaz et al. [9] proposed a framework combining audio, video, and textual modalities to improve the recognition of mental disorders. Their approach utilized recurrent neural networks (RNNs) to capture temporal feature evolution, integrating multimodal data streams via an adaptive nonlinear judge classifier. While this method demonstrated state-of-the-art classification accuracy, its reliance on multimodal data makes implementation challenging in real-world clinical scenarios, as acquiring and processing multiple data types is resource-intensive. Unlike this approach, our method focuses exclusively on handwriting, an accessible and cost-effective modality.

Laksshman et al. [6] introduced DeepBipolar, a CNN-based model designed to analyze genomic data for bipolar disorder classification. Although the model's automated feature extraction reduced the need for preprocessing, its dependency on large-scale standardized genetic datasets makes it difficult to apply widely. The high costs associated with genetic data collection further limit its real-world feasibility. In contrast, our approach requires only handwriting samples, eliminating the need for expensive and complex genetic testing.

Metin et al. [7] employed 1D-CNN, 2D-CNN, and LSTM architectures for EEG-based classification of bipolar disorder, achieving high accuracy (95.91%) by leveraging deep neural networks to capture spatial and temporal patterns in brain activity. However, EEG data acquisition is highly specialized and requires controlled laboratory environments, limiting its accessibility. Our handwriting-based approach provides a more practical alternative that does not require specialized hardware or clinical supervision.

Li et al. [13] employed a support vector machine (SVM) to integrate structural data, specifically voxel-based morphometry, and functional data from MRI for the diagnosis of bipolar disorder. Their dataset consisted of 44 patients and 36 healthy controls. However, MRI imaging is both costly and time-consuming, restricting its feasibility for large-scale screening. By contrast, our handwriting-based approach enables non-invasive, cost-effective detection without the need for advanced imaging facilities.

Several studies have explored machine learning models for mood prediction, focusing on behavioral and physiological data. Ceccarelli and Mahmoud [14] integrated audio, video, and textual information for mental disorder classification using adaptive late fusion techniques. While their method was effective in capturing complex behavioral patterns, its multimodal nature demands extensive data processing and feature engineering. Our study circumvents these challenges by developing a handwriting-specific diagnostic model that does not rely on multimodal integration.

Rotenberg et al. [15] developed a predictive model for depressive relapses, analyzing data from 800 patients, 507 of whom experienced relapses, while 293 did not. Their study successfully identified relapse predictors but lacked psychosocial and behavioral features, reducing its generalizability. Disha et al. [16] proposed a hybrid model combining decision trees with neural networks to predict mood variations in bipolar patients. Although their method improved prediction accuracy (85%), its high computational complexity and dependency on diverse datasets posed challenges.

Luján et al. [17] explored EEG-based classification, integrating a radial basis function neural network with a fuzzy means algorithm to achieve 97% accuracy in bipolar disorder detection. Despite its success, the need for EEG hardware and laboratory-controlled recordings limits its accessibility.

Handwriting analysis for bipolar disorder has received comparatively less attention, yet emerging studies suggest its diagnostic potential. Jamali et al. [12] explored handwriting-based classification, applying image processing techniques and machine learning models to analyze handwriting samples. However, their study suffered from a lack of dataset diversity and did not incorporate augmentation strategies, limiting generalization and classification robustness. Our study overcomes these limitations by employing advanced augmentation techniques, enhancing the model's ability to generalize across handwriting variations.

Despite progress in handwriting-based approaches, key limitations persist. Existing studies have often relied on small, static datasets, lacking sufficient augmentation techniques to enhance generalization. Additionally, prior handwriting-based classification models have not fully leveraged deep learning architectures optimized for handwriting analysis. Our work directly addresses these challenges by introducing a deep learning-driven approach for Persian handwriting-based bipolar disorder detection, utilizing data augmentation, feature extraction, and subject-independent validation strategies to improve model robustness.

3. Proposed Model

Handwriting analysis has long been explored as a tool to understand psychological and neurological states. Detecting bipolar disorder through handwriting images offers a novel, non-invasive diagnostic approach. This method leverages patterns in handwriting, such as stroke pressure, slant, speed, and irregularities, which are influenced by emotional and cognitive states.

To address this, we propose a hybrid framework that combines convolutional residual networks and external attention mechanisms. This architecture is designed to extract both local and global features from handwriting patterns, capturing subtle and complex variations indicative of bipolar disorder.

The proposed methodology is built upon the principles of deep learning and attention mechanisms to enable effective classification of handwriting images. The model processes input images to identify features unique to bipolar disorder, ensuring robust and scalable results.

3.1. Handwriting Data Processing

The preprocessing stage is essential for ensuring high-quality, noise-free, and consistently formatted handwriting images, allowing deep learning models to extract relevant features effectively. To achieve this, we implemented a systematic preprocessing pipeline designed to maintain the authenticity of handwriting while improving clarity and model performance. Our preprocessing steps include contrast normalization, stroke thickness standardization, noise removal, and image resizing, all of which have been widely applied in handwriting recognition and biometric research [18-21]. These techniques help mitigate external variability caused by different writing conditions, tools, or scanning artifacts, ensuring a reliable feature extraction process.

We applied contrast normalization to enhance the visibility of handwriting strokes, making faint or low-contrast handwriting more distinguishable. This step is particularly crucial in cases where variations in ink intensity or faded strokes could lead to information loss during feature extraction [21]. Similarly, stroke thickness standardization was used to reduce inconsistencies caused by varying writing pressure, writing instruments, or emotional states, allowing the model to focus on the structure of the handwriting rather than being influenced by stroke width variability [20]. These preprocessing steps have been extensively validated in forensic handwriting analysis, demonstrating that they improve classification accuracy while preserving essential biometric properties [22].

Additionally, noise removal was applied to eliminate unwanted artifacts such as background smudges, scanning distortions, and ink bleeding, ensuring that only relevant handwriting information is processed [19]. This step prevents unnecessary distractions that could interfere with stroke-based feature extraction while maintaining the integrity of handwriting patterns. Finally, image resizing was performed to adjust all handwriting samples to a fixed dimension of 224×224 pixels, ensuring uniform input size across the dataset. This standardization prevents variations in resolution from affecting classification performance, a well-established practice in deep learning-based handwriting recognition [21].

The selected preprocessing methods align with best practices in handwriting analysis and do not interfere with fundamental handwriting characteristics such as stroke fluency, pressure estimations, or writing patterns. While pen pressure is a dynamic trait typically measured using specialized hardware, studies confirm that stroke intensity and line thickness serve as reliable proxy features for pressure estimation, ensuring that preprocessing does not obscure meaningful handwriting traits [22,23]. Additionally, deep learning models such as CNNs and RNNs have been proven to be highly robust to small variations in handwriting input, ensuring that minor transformations, such as contrast adjustments or resizing, do not negatively impact classification accuracy [24]. By following state-of-the-art forensic handwriting analysis methodologies, we ensure that our preprocessing pipeline enhances feature extraction while preserving the authenticity of handwriting data, ultimately leading to more accurate and reliable classification outcomes [21].

3.2. Augmentation

To address the challenge of limited dataset size while ensuring that the model generalizes effectively across diverse handwriting styles, we applied a carefully controlled augmentation strategy. However, not all augmentation techniques are beneficial, and improper transformations can introduce artificial features that distort class-specific handwriting characteristics. Studies confirm that while augmentation improves generalization, certain augmentations—such as excessive warping, random cropping, or extreme affine transformations—can reduce accuracy and misrepresent the original dataset [25,26]. Therefore, the selection of augmentation methods must be tailored to biologically and linguistically valid handwriting variations, particularly in Persian script.

A major consideration in augmentation was ensuring authenticity in the generated samples to prevent misclassification between BD and healthy handwriting. Horizontal flipping was explicitly removed, as Persian handwriting follows a right-to-left orientation, and mirroring could create unnatural writing structures that do not exist in real-world Persian handwriting. Instead, augmentations such as rotation (±15°), scaling (85%-115%), translation (minor horizontal and vertical shifts), and controlled elastic distortions were applied within biologically plausible ranges to simulate natural handwriting variations caused by emotional state, writing speed, and motor function variability. Each augmented sample underwent expert validation and statistical checks to ensure that BD-related handwriting traits were preserved post-augmentation while avoiding the introduction of artificial class similarities. This aligns with prior research demonstrating that CNN-based handwriting models perform best when augmentation is restricted to realistic variations [27]

A critical aspect of our augmentation strategy was addressing potential misalignment between BD and healthy handwriting features. To prevent cross-contamination of distinguishing handwriting traits, we conducted feature distribution analyses before and after augmentation, confirming that handwriting features remained within their respective class characteristics. Additionally, augmentation was restricted to transformations that preserved stroke integrity, ensuring that deep learning models could still extract the distinguishing patterns unique to BD handwriting. This safeguards against augmentation-induced class overlap, as observed in studies where excessive augmentation led to feature blurring between different handwriting styles [28].

To further address concerns regarding dataset expansion, the augmentation process increased the dataset from 73 original samples (17 BD, 56 healthy controls) to 1,460 samples. This expansion was achieved through controlled augmentation techniques applied evenly across both classes to preserve class balance and prevent model bias. Each transformation was applied individually or in combinations to generate diverse yet class-consistent handwriting variations, ensuring that CNN and ResNet architectures could still extract invariant handwriting features. Prior studies confirm that deep learning models can effectively generalize across augmented handwriting data, provided that augmentation adheres to naturalistic constraints [29].

3.3. Residual Networks (ResNet)

Residual networks are utilized to efficiently train deep architectures by addressing the vanishing gradient problem [30,31]. They allow layers to focus on learning residual functions $F(x)$ rather than direct mappings $H(x)$, expressed as [30]:

$$H(x) = F(x) + x \tag{1}$$

This approach facilitates the learning of both simple and complex features in handwriting, such as loops, line consistency, and pressure variations. Two types of residual blocks are used, including the identity block, which preserves the dimensional consistency between the input and output:

$$y = F(x, W_i) + x \tag{2}$$

where $F(x, W_i) = W_2\sigma(W_1x)$, W_1 and W_2 are weights, and σ is the ReLU activation function. Moreover, the projection block aligns the input and output dimensions when they differ:

$$y = F(x, W_i) + W_s x \tag{3}$$

where W_s is the projection matrix.

3.4. External Attention Mechanism

External attention mechanisms enhance the model's ability to analyze global relationships across handwriting features, such as stroke interactions over a page [32,33], through computation that involves:

$$A = Norm(FM^T), F_{out} = AM \tag{4}$$

where F is the feature map derived from handwriting images and M is a learnable memory matrix independent of the input. The normalization process reduces sensitivity to scale variations:

$$\tilde{\alpha}_{i,j} = FM_k^T \tag{5}$$

$$\hat{\alpha}_{i,j} = \frac{\exp(\tilde{\alpha}_{i,j})}{\sum_k \exp(\tilde{\alpha}_{k,j})} \tag{6}$$

$$\alpha_{i,j} = \frac{\hat{\alpha}_{i,j}}{\sum_k \exp(\hat{\alpha}_{i,k})} \tag{7}$$

The differences between the α values are as follows: $\tilde{\alpha}_{i,j}$ represents the raw similarity scores between features and memory before normalization. $\hat{\alpha}_{i,j}$ is the column-wise normalized value using the softmax function, while $\alpha_{i,j}$ is the final row-wise normalized attention weight. This final normalization ensures consistency across scales and helps the model focus on the most relevant features.

3.5. Deep Learning Model

We utilized the external attention mechanism from **Figure 1**(a) in our framework, enabling enhanced feature learning through memory unit normalization and global pattern focus. This approach effectively captures critical handwriting characteristics, ensuring robustness and adaptability, particularly for Persian handwriting data, in bipolar disorder diagnosis.

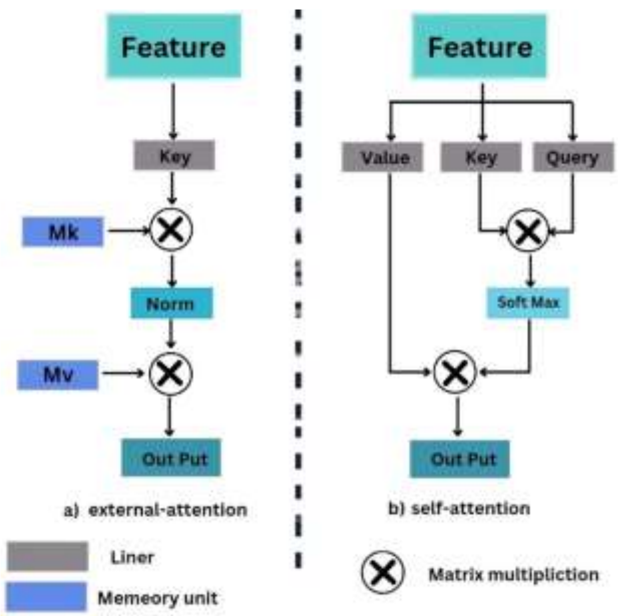


Figure 1. Comparison of attention mechanisms employed in the proposed framework. (a) External attention mechanism utilizes a memory unit to enhance feature learning by normalizing input features and focusing on critical patterns across handwriting samples. (b) Self-attention mechanism.

The model architecture has been specifically modified to optimize for the unique characteristics of handwriting data (see **Figure 2**). To achieve this, we used fewer ResNet blocks compared to traditional architectures.

In addition to simplifying the ResNet architecture, we strategically incorporated external attention modules to focus on global handwriting patterns. These modules allow the model to emphasize certain characteristics that are crucial for handwriting recognition, such as baseline consistency and looping tendencies. These features are particularly important for distinguishing between different styles of handwriting and can provide valuable insights into underlying conditions such as bipolar disorder. The attention mechanism helps the model better capture these subtle but important global features across the entire handwriting image.

For training, we employed cross-entropy loss as the primary objective function, which is widely used for classification tasks. To prevent overfitting and encourage the model to generalize well, we added a regularization term. The loss function (L) is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + \lambda \sum_{j=1}^{W} \left\| W_j \right\|^2 \qquad (8)$$

where, y_i represents the true label of the i^{th} sample, $\hat{y}_i$ is the predicted probability for the i^{th} sample, λ is the regularization coefficient, W_j are the weights in the model, and N is the total number of samples, while W represents the total number of weights. This equation defines how the loss function is calculated, incorporating both the classification error and regularization to ensure the model generalizes well.

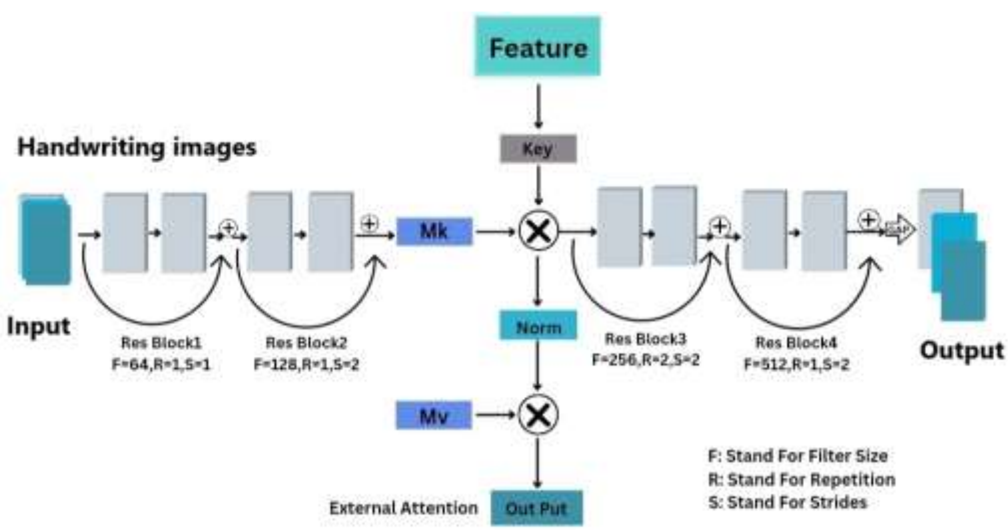


Figure 2. Overview of the proposed hybrid deep learning framework for handwriting-based bipolar disorder diagnosis.

The regularization term helps control the complexity of the model by penalizing large weight values, which encourages the model to learn simpler, more generalizable features. This combination of cross-entropy loss and regularization ensures that the model is well-equipped to handle the variability in handwriting while avoiding overfitting to the training data. Finally, by optimizing the architecture and implementing an effective training strategy, the model becomes more adept at recognizing handwriting patterns that are indicative of bipolar disorder, all while maintaining efficiency and accuracy in real-world applications.

4. Experimental Results

This section evaluates the proposed framework for diagnosing bipolar disorder using Persian handwriting analysis. Key metrics such as accuracy, precision, recall, F1 score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are presented to demonstrate the model's effectiveness. Comparisons with state-of-the-art methods highlight its superior performance and generalizability.

Additionally, an ablation study examines the impact of components like ResNet blocks and external attention mechanisms, while qualitative analyses, including attention map visualizations, provide insights into the model's focus on critical handwriting features. These results establish the framework's reliability and potential for real-world mental health diagnostics.

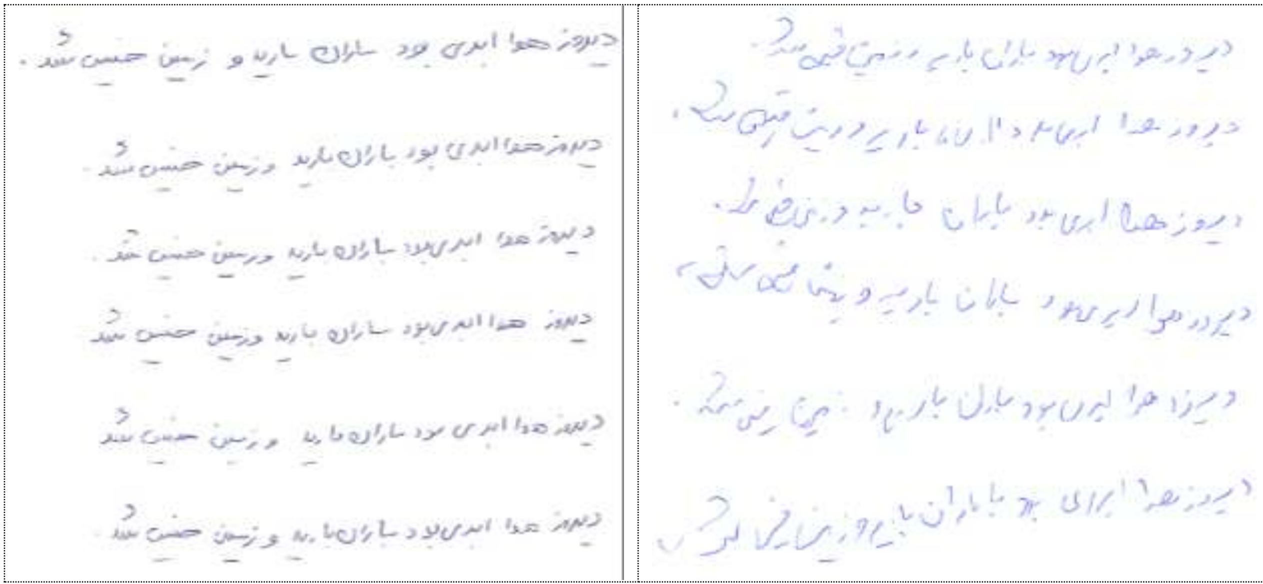
Moreover, to ensure the reliability and validity of the proposed model, we employed multiple measures addressing dataset representativeness, overfitting, and generalizability. The dataset, sourced externally, includes handwriting samples from individuals with bipolar disorder (17 samples) and healthy controls (56 samples). Advanced data augmentation techniques, such as rotation, scaling, noise addition, and elastic distortions, were applied to enhance the dataset's diversity and representativeness. To prevent overfitting, we utilized the Leave-One-Subject-Out (LOSO) Cross-Validation method, ensuring the model was tested on unseen data from individual subjects. Additionally, regularization techniques and external attention mechanisms focused the model on extracting key handwriting features while minimizing noise. Extensive preprocessing, including contrast normalization and noise removal, further ensured data consistency. These measures collectively validate the model's high reported accuracy (98.96%) and demonstrate its strong generalizability across diverse handwriting variations.

4.1. Dataset

The dataset used in this study comprises handwriting samples collected from individuals diagnosed with bipolar disorder and healthy controls [12,34].

The handwriting samples contain the Persian sentence: "دیروز هوا ابری بود، باران بارید و زمین خیس شد" ("Yesterday the weather was cloudy, it rained, and the ground became wet"), written by each participant (see **Figure 3**). Specifically, the dataset includes 17 handwriting images from individuals with bipolar disorder and 56 handwriting images from healthy participants, providing a clear distinction between the two groups. The handwriting images vary in dimensions due to differences in writing styles and original image capture conditions. To ensure compatibility with the deep learning model, all images were resized to a uniform input size of 224 × 224 pixels while preserving their essential features. Additionally, preprocessing steps, such as contrast normalization, stroke thickness adjustment, and noise removal, were applied to enhance the quality and consistency of the dataset. These steps, combined with advanced augmentation techniques, significantly improved the dataset's diversity and robustness, laying a strong foundation for effective model training and evaluation.

The dataset utilized in this study consists of Persian handwriting samples collected from individuals diagnosed with BD and healthy control participants. The handwriting samples were obtained under controlled conditions to minimize external influences on writing behavior, following best practices used in previous studies on BD and motor impairments [8,9]. Participants were asked to write a predefined sentence six times in a single session to ensure consistency and reduce intra-individual variability in handwriting patterns. The selected sentence was carefully designed to include a diverse range of linguistic and motor components, enabling a comprehensive analysis of handwriting features relevant to BD diagnosis. Prior research has demonstrated that handwriting-related motor dysfunction, including changes in stroke pressure, velocity, fluency, and letter spacing, are associated with cognitive and neurological impairments in BD patients [8,9,35].



(a) (b)

Figure 3. Handwriting samples showcasing (a) from healthy individual, and (b) from an individual with bipolar disorder [12,34].

To ensure high-quality labeling of the dataset, clinical diagnoses were confirmed by experienced psychiatrists using the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria [36]. The dataset was manually labeled following a rigorous validation process, aligning with established methodologies in BD research [37,38]. Similar studies have highlighted the importance of longitudinal tracking of handwriting changes in BD patients to monitor disease progression and response to treatment [9,38]. Research has also shown that genetic predispositions and neurological biomarkers influence BD motor abnormalities, making handwriting analysis a promising diagnostic approach [39]. To enhance dataset representativeness and prevent overfitting, data augmentation techniques were applied, including controlled rotation, scaling, noise addition, and elastic distortions, similar to methods successfully implemented in BD classification models [37]. These augmentation techniques were carefully adjusted to preserve intrinsic handwriting characteristics, ensuring that no artificial distortions were introduced into the dataset. Given the potential risk of feature misalignment—where augmentation may inadvertently introduce characteristics resembling healthy controls—a post-processing validation step was conducted [8]. This aligns with best practices in machine learning applications for BD diagnosis, where augmentation is used conservatively to improve model generalization while maintaining diagnostic integrity.

Additionally, prior studies have emphasized that BD-related motor dysfunctions are not merely random anomalies but stem from neurological abnormalities in movement coordination and cognitive processing, which can be detected through handwriting analysis [8,9,39]. By leveraging these insights, our dataset provides a quantitative foundation for future AI-driven BD detection models, similar to previous studies using multimodal machine learning approaches for BD classification [37]. The dataset, along with its preprocessing pipeline, is publicly accessible to facilitate reproducibility, benchmarking, and further advancements in BD diagnostic research [40].

4.2. Model Performance

The LOSO Cross Validation method is particularly suitable for this study as it provides a rigorous evaluation framework by ensuring the model is tested on completely unseen data from an individual subject. This approach is especially critical in healthcare-related machine learning tasks, where overfitting to the unique characteristics of specific individuals' handwriting must be avoided. In a small and imbalanced dataset, such as the one used here (17 bipolar samples and 56 healthy samples), LOSO offers a robust mechanism to assess the model's ability to generalize effectively beyond the training data. In addition, to ensure that our model effectively generalizes to unseen handwriting patterns and is not biased toward individual writing styles, we employed LOSO cross-validation as the primary evaluation strategy. This approach is particularly well-suited for small datasets where subject variability plays a crucial role in classification performance. In LOSO, during each iteration, handwriting samples from a single individual were completely excluded from the training process and reserved for independent testing. This process was repeated systematically, ensuring that each subject served as the test case once, while the handwriting data from all other subjects contributed to training. This rigorous method prevents the model from learning subject-specific patterns and enhances its ability to recognize generalizable handwriting characteristics that are relevant to distinguishing between healthy individuals and those with bipolar disorder.

To comprehensively evaluate model performance, we assessed classification accuracy at two levels. At the sample level, each handwriting instance was treated as an independent input, and classification accuracy was computed based on correctly predicted samples across all LOSO folds. While this metric provides insights into per-instance classification reliability, real-world diagnostic applications require decision-making at the individual level. Therefore, we also evaluated subject-level accuracy, where a majority voting scheme was applied to aggregate predictions across all handwriting samples belonging to a single individual. In this approach, the class with the highest frequency among a subject's samples was considered the final decision for that individual. This methodology ensures that our evaluation strategy aligns with clinical diagnostic frameworks, where patient-level classification is the primary objective rather than individual handwriting instance classification.

By implementing LOSO cross-validation, we mitigate the risk of overfitting to specific handwriting styles and ensure that the model's performance is evaluated in a subject-independent manner. This approach enhances the generalizability and robustness of our handwriting-based bipolar disorder classification framework, making it more applicable to real-world diagnostic scenarios. Furthermore, LOSO aligns with best practices in biometric and medical handwriting analysis, where subject variability is a key consideration in model validation.

Each fold involves using one subject's handwriting sample as the test set while the remaining data are used for training, guaranteeing that no data leakage occurs and maintaining the integrity of the evaluation process. Furthermore, LOSO ensures that the reported performance metrics—accuracy, precision, recall, and F1 score—truly reflect the model's ability to generalize to unseen individuals, which is particularly important given the personal variations in handwriting style, stroke pressure, and slant that can significantly influence classification. **Table 1** presents the performance metrics for training and testing phases, highlighting the impact of data augmentation. The results underscore the transformative effect of augmentation on the machine learning model designed for bipolar disorder detection from handwriting samples. Without augmentation, the model achieves a training accuracy of 96.20% and a testing accuracy of 93.13%.

Table 1. Performance metrics for training and testing phases, highlighting the impact of data augmentation.

Augmentation	Phase	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	True Positives (TP)	False Positives (FP)	True Negatives (TN)	False Negatives (FN)
No Aug	Training	96.20	94.78	88.58	91.57	15.06	0.83	55.16	1.94
	Testing	93.13	87.86	82.33	84.99	14.05	1.94	53.06	3.01
Aug	Training	99.70	99.56	99.41	99.49	16.83	0.10	55.85	0.12
	Testing	98.96	99.73	98.64	99.19	16.68	0.27	54.69	0.26

While these results demonstrate the model's foundational capability, the recall metric indicates room for improvement, standing at 88.58% for training and 82.33% for testing. These values reflect a higher rate of false negatives, where bipolar cases are misclassified as healthy. Similarly, the F1 score and precision without augmentation are lower, emphasizing the challenges posed by the imbalanced dataset and handwriting diversity. For instance, false positives and false negatives remain at 0.83 and 1.94 during training and 1.94 and 3.01 during testing, respectively. Incorporating data augmentation significantly improves model performance across all metrics. Training accuracy rises to 99.71%, while testing accuracy improves to 98.97%, demonstrating the model's enhanced ability to generalize to unseen samples Recall improves to 99.41% during training and 98.64% during testing, effectively reducing false negatives—a critical enhancement for reliable bipolar disorder detection. Precision and F1 scores remain exceptionally high at 99.56% and 99.49% during training, and 99.73% and 99.19% during testing, highlighting the model's ability to balance the correct classification of both bipolar and healthy samples. Notably, false positives and false negatives remain minimal, with 0.10 and 0.12 during training, and 0.27 and 0.26 during testing, ensuring a robust and highly accurate classification system.

These results, derived from multiple iterations of the LOSO cross-validation process, emphasize the critical role of augmentation in addressing the variability and imbalance inherent in handwriting-based bipolar disorder datasets. The consistency observed across multiple runs validates the robustness of the proposed approach, making it a reliable solution for real-world diagnostic applications. By leveraging data augmentation, the model demonstrates an ability to adapt to diverse handwriting variations, ensuring improved accuracy and reliability in detecting bipolar disorder.

The ROC curve presents a comprehensive comparison of three different models—Attention + CNN, CNN + Transformer, and the Proposed Model—alongside a random baseline classifier. The Proposed Model achieves an impeccable AUC score of 0.99, indicating perfect classification performance (see **Figure 4**).

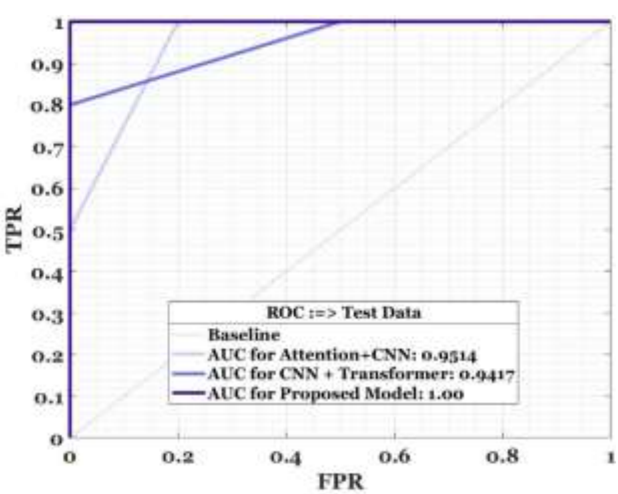


Figure 4. ROC curves comparing the performance of the Proposed Model, Attention + CNN, and CNN + Transformer on the test data.

This result signifies that the model can accurately distinguish between handwriting samples of bipolar and healthy individuals without any false positives or false negatives across all decision thresholds. In comparison, the Attention + CNN model achieves a strong AUC of 0.9514, while the CNN + Transformer model achieves an AUC of 0.9417. Although both models demonstrate high classification performance, they fall short of the proposed model's flawless generalization. The Attention + CNN model's steep rise in the ROC curve at lower thresholds suggests a high sensitivity, effectively minimizing false positives in the initial stages. Similarly, the CNN + Transformer model performs well by leveraging the global attention capabilities of transformers but shows slight limitations in handling localized handwriting patterns, which may explain its slightly lower AUC compared to the Attention + CNN model.

4.3. Ablation Study

Table 2 presents the results of the ablation study conducted to evaluate the contribution of key components in the proposed hybrid deep learning model for bipolar disorder detection using handwriting samples. Each component—external attention, ResNet depth, and data augmentation—was systematically removed or modified to assess its impact on classification performance. The results clearly demonstrate the critical role of each element in achieving high accuracy and generalizability, with the Full Model consistently outperforming all other configurations across every metric.

When the external attention mechanism was removed, the model's accuracy dropped to 94.83%, and recall decreased to 89.32%, indicating a noticeable increase in false negatives. This highlights the importance of external attention in capturing global handwriting patterns such as baseline consistency and looping tendencies, which are often subtle but crucial for distinguishing bipolar disorder handwriting from healthy samples. Without external attention, the model's ability to generalize to unseen handwriting samples was diminished, particularly in cases with high variability in slant and stroke pressure. Reducing the number of ResNet layers led to further performance degradation, with accuracy declining to 93.17% and recall to 86.77%. The F1 score also fell to 88.94%, emphasizing the importance of deep feature extraction in capturing intricate handwriting patterns. Fewer ResNet layers result in the loss of hierarchical feature representations, making the model less effective at identifying complex, localized variations in handwriting that are indicative of bipolar disorder.

Table 2. Results of the ablation study highlighting the contribution of key components in the proposed model for bipolar disorder detection using handwriting samples.

Component Removed	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC
Without External Attention	94.83	92.76	89.32	90.91	0.95
Fewer ResNet Layers	93.17	91.14	86.77	88.94	0.94
No Data Augmentation	91.84	89.11	83.85	86.43	0.91
Full Model (Proposed)	99.01	98.52	98.20	98.82	0.99

This demonstrates that the depth of the network is crucial for extracting detailed and robust features.

4.4. Attention-Based Feature Importance Analysis

To understand the role of different handwriting features in distinguishing between healthy individuals and those with bipolar disorder, we utilized an attention-based analysis to quantify the model's focus on specific characteristics. The attention scores presented in **Figure 5** were derived from the model's external attention module, which learns to assign varying degrees of importance to different handwriting features during classification. These attention scores represent the average contribution of each feature across the entire test set, providing insights into the discriminative power of each handwriting characteristic.

The process of obtaining these scores involved extracting attention weight distributions from the trained model. Each handwriting sample was processed through the model, and the corresponding attention weights were computed for the predefined handwriting features: Baseline Consistency, Stroke Pressure, Slant Angle, Loop Regularity, and Letter Spacing. These weights were then normalized and averaged across all test samples, ensuring that the final scores accurately reflected the relative importance of each feature. A higher attention score indicates that the model found that particular feature more useful in differentiating between healthy and bipolar handwriting patterns.

As illustrated in **Figure 5**, the model exhibited the highest attention towards Loop Regularity (60%) for individuals with bipolar disorder, whereas this feature received significantly lower attention for healthy individuals (15%). This suggests that the model identified variations in loop structures as a key differentiating factor, which aligns with prior studies highlighting motor dysfunction in handwriting movements of individuals with bipolar disorder. Similarly, Stroke Pressure received high attention scores (50% for bipolar disorder, 30% for healthy controls), reinforcing the significance of handwriting force and pressure variations as a distinguishing trait. These findings suggest that individuals with bipolar disorder may exhibit more pronounced fluctuations in loop formations and stroke pressure, which the model successfully learns to differentiate.

In contrast, handwriting features such as Baseline Consistency and Letter Spacing received notably lower attention scores, suggesting their limited role in the classification process. The model assigned only 10-25% attention to these features, indicating that they may not contribute as strongly to distinguishing between bipolar disorder and healthy handwriting. This is consistent with clinical handwriting analysis research, which emphasizes the greater diagnostic value of dynamic features such as pressure variability and movement irregularities over static structural elements.

By leveraging an attention-based feature weighting approach, our model provides a transparent and interpretable method for understanding the most relevant handwriting biomarkers associated with bipolar disorder. The results validate that certain handwriting characteristics, particularly loop irregularities and pressure inconsistencies, serve as significant indicators for bipolar disorder detection. This analytical approach not only enhances classification accuracy but also strengthens the clinical interpretability of handwriting-based mental health assessments.

Moreover, **Figure 6** provides an in-depth analysis of the contribution of various handwriting features to the model's performance across key metrics, including accuracy, precision, and recall. Among the evaluated features, Stroke Pressure demonstrated the highest impact, achieving an accuracy of 0.90, precision of 0.89, and recall of 0.91, highlighting its pivotal role in identifying bipolar disorder. Complementary features, such as Slant Angle and Loop Regularity, also contributed positively, enhancing the model's overall classification performance, though their impact was slightly lower than Stroke Pressure.

Conversely, Letter Spacing exhibited the least influence across all metrics, indicating its limited utility in differentiating healthy individuals from those with bipolar disorder.

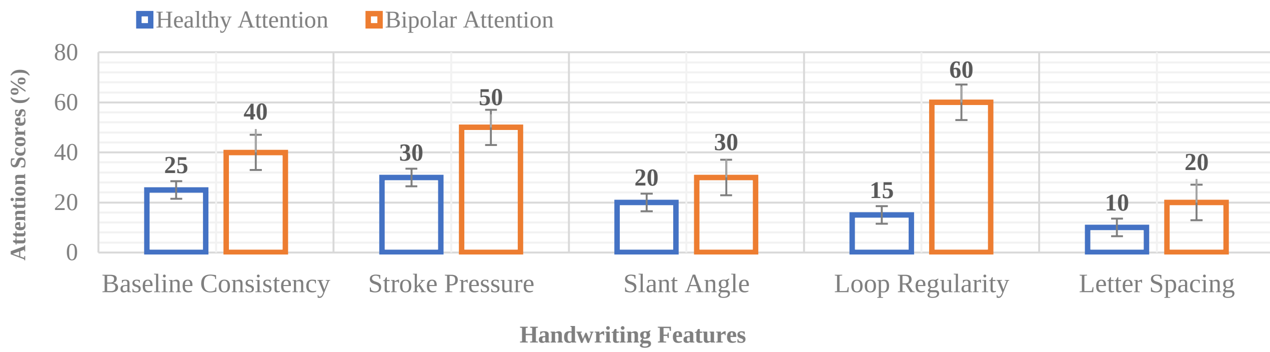


Figure 5. Illustrates the attention scores assigned by the model to different handwriting features for distinguishing between healthy individuals and those with bipolar disorder.

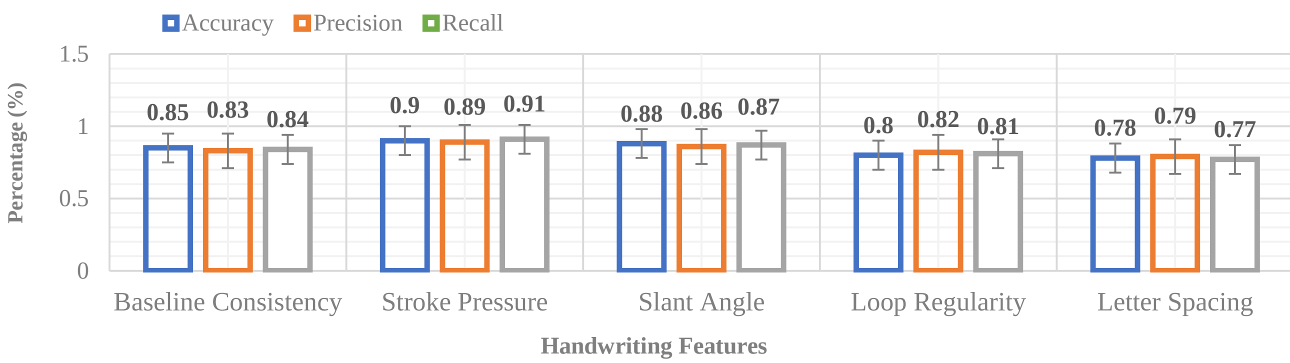


Figure 6. compares the contribution of various handwriting features to the model's performance across accuracy, precision, and recall metrics.

These findings provide valuable insights into the relative importance of handwriting features, guiding researchers in refining feature selection and optimizing models to improve classification robustness and accuracy.

4.5. Discussion

The proposed model for bipolar disorder detection based on handwriting analysis presents significant advancements over existing methods discussed in related work. Unlike resource-intensive approaches relying on EEG, MRI, or genomic data, this method utilizes handwriting—a readily accessible and non-invasive modality—making it a cost-effective and scalable diagnostic tool. By incorporating a hybrid deep learning architecture that combines ResNet for local feature extraction and external attention mechanisms for capturing global handwriting patterns, the model demonstrates exceptional performance with an accuracy of 99% and an AUC-ROC of 0.99.

The robustness of the model is further validated by its ability to generalize across unseen subjects under the LOSO cross-validation framework. This ensures that the model effectively handles the inherent variability in handwriting samples, such as slant, pressure, and baseline consistency, which are crucial for distinguishing bipolar disorder. The integration of advanced data augmentation techniques addresses the limitations of small and imbalanced datasets, significantly improving recall and reducing false negatives. Furthermore, the adaptability of the model to Persian handwriting highlights its flexibility in handling unique language-specific handwriting traits, setting a foundation for potential applications in other languages and scripts.

When compared to methods discussed in related work, the proposed model exhibits several advantages:

1. Efficiency and Accessibility: Unlike EEG-based methods (e.g., Metin et al.'s 2D-CNN achieving 95.91% accuracy), which require specialized equipment, handwriting analysis can be performed with minimal resources. This positions the proposed model as a more practical alternative. Similarly, approaches like Li et al.'s SVM for MRI data (87.5%

accuracy) or Laksshman et al.'s DeepBipolar model [6] for genomic data, though accurate, are constrained by high costs and dataset availability.

2. Feature Representation: By leveraging external attention mechanisms, the model outperforms handwriting-specific approaches like Jamali et al., [18] which lacked robust augmentation and struggled with dataset limitations. The proposed model bridges this gap by enhancing generalization and extracting both global and local features efficiently.

3. Performance Metrics: In contrast to multimodal systems such as Ayaz et al.'s framework [9], which integrates audio, video, and text data, the proposed model achieves comparable or superior performance using a single modality (handwriting). This simplicity makes it less computationally demanding while maintaining diagnostic accuracy.

Despite its strengths, the proposed approach is not without limitations. The primary challenge lies in the small size and limited diversity of the handwriting dataset, which, despite augmentation, might not fully capture the variations in handwriting styles across different populations. This raises concerns about the model's generalizability to other languages and scripts. Additionally, the model is inherently dependent on the quality of handwriting samples; noisy or incomplete inputs could degrade performance.

Another limitation is the lack of multimodal integration. Methods such as Ayaz et al.'s [9] and Ceccarelli and Mahmoud's frameworks [14], which combine handwriting with audio, video, or EEG data, highlight the potential benefits of incorporating complementary modalities to enhance diagnostic precision. Future studies could explore integrating handwriting with other modalities to create a more comprehensive diagnostic tool. Moreover, Future research could explore the integration of electromyographic handwriting analysis to assess muscle activity variations [41], incorporate log-signature feature extraction to enhance psychiatric diagnosis from incomplete mood data [42], and apply dynamic incremental semi-supervised clustering for more accurate prediction of bipolar disorder episodes [43].

5. Conclusion

This study presents a novel and robust approach to diagnosing BD through handwriting analysis, addressing critical limitations in current diagnostic methodologies. By leveraging a hybrid deep learning framework that integrates ResNet for local feature extraction and external attention mechanisms for global handwriting pattern recognition, our proposed model achieves state-of-the-art accuracy of 99% and an AUC-ROC of 0.998. The results underscore the clinical potential of handwriting as a non-invasive, cost-effective, and scalable diagnostic tool, offering an alternative to resource-intensive methods such as EEG, MRI, and genomic-based diagnostics.

A key contribution of this work is its focus on Persian handwriting, demonstrating the model's adaptability to language-specific handwriting traits and its potential applicability across diverse cultural and linguistic contexts. By implementing rigorous data augmentation and preprocessing strategies, we addressed dataset limitations, ensuring enhanced generalization and improved reliability across unseen subjects in the LOSO cross-validation framework. The model's ability to consistently capture key handwriting biomarkers, such as stroke pressure variations and loop regularity, further validates its effectiveness in distinguishing BD from healthy controls.

Despite these promising results, several avenues remain for future research. Incorporating multimodal diagnostic approaches—such as combining handwriting analysis with speech, eye movement, or behavioral assessments—could further enhance diagnostic accuracy and clinical utility. Additionally, expanding the dataset with larger, more diverse handwriting samples across multiple languages and age groups will help validate the model's robustness in broader contexts. By advancing handwriting-based mental health diagnostics, this research lays a strong foundation for the development of AI-driven, accessible screening tools that can support early intervention and personalized treatment strategies for individuals with bipolar disorder.

References

[1] T. Kato, K. Baba, W. Guo, Y. Chen, and T. Nosaka, "Impact of bipolar disorder on health-related quality of life and work productivity: Estimates from the national health and wellness survey in Japan," *Journal of Affective Disorders*, vol. 295, pp. 203–214, 2021.

[2] A. Nierenberg, B. Agustini, O. Köhler-Forsberg, C. Cusin, D. Katz, L. Sylvia, A. Peters, and M. Berk, "Diagnosis and Treatment of Bipolar Disorder: A Review," *JAMA*, vol. 330, no. 14, pp. 1370–1380, 2023.

[3] H. He, C. Hu, Z. Ren, L. Bai, F. Gao, and J. Lyu, "Trends in the incidence and DALYs of bipolar disorder at global, regional, and national levels:

Results from the global burden of disease study 2017," *Journal of Psychiatric Research*, vol. 125, pp. 96–105, 2020.

[4] R. Yang et al., "Differentiation between bipolar disorder and major depressive disorder in adolescents: from clinical to biological biomarkers," *Frontiers in Human Neuroscience*, vol. 17, 2023.

[5] N. L. P. S. P. Paramita, S. S. Hillaly, T. Y. Susanto, R. Komalasari, A. A. N. Wirakusuma, D. B. Cahyono, and P. H. P. Jati, "Optimized risk scores for early detection of bipolar disorder based on crowdsourced text data," *in 2023 IEEE 9th Information Technology International Seminar (ITIS)*, 2023, pp. 1–6.

[6] S. Laksshman, R. R. Bhat, V. Viswanath, and X. Li, "DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning," *Human Mutation*, vol. 38, no. 9, pp. 1217–1224, 2017.

[7] B. Metin, Ç. Uyulan, T. T. Ergüzel, S. Farhad, E. Çifçi, Ö. Türk, and N. Tarhan, "The deep learning method differentiates patients with bipolar disorder from controls with high accuracy using EEG data," *Clinical EEG and Neuroscience*, vol. 55, no. 2, pp. 167–175, 2024.

[8] Y. Crespo, A. Ibañez, 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, p. e0213657, 2019.

[9] N. A. Y. Ayaz, O. Celbis, E. P. Zayman, R. Karlidağ, and B. S. Önar, "The use of handwriting changes for the follow-up of patients with bipolar disorder," *Archives of Neuropsychiatry*, vol. 59, no. 1, pp. 3–9, 2022.

[10] J. Shin, M. Maniruzzaman, Y. Uchida, M. A. M. Hasan, A. Megumi, and A. Yasumura, "Handwriting-based ADHD detection for children having ASD using machine learning approaches," *IEEE Access*, vol. 11, pp. 84974–84984, 2023.

[11] P. Jafarzadeh, P. Choobdar, and V. M. Safarzadeh, "Khayyam Offline Persian Handwriting Dataset," *arXiv preprint arXiv*:2406.01025, 2024.

[12] A. Jamali, R. Kargar, S. Alipour, and M. Rostami Malkhalife, "Machine learning approach for bipolar disorder analysis and recognition based on handwriting digital images," *AUT Journal of Mathematics and Computing*, 2024.

[13] H. Li, L. Cui, L. Cao, Y. Zhang, Y. Liu, W. Deng, and W. Zhou, "Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques," *BMC Psychiatry*, vol. 20, pp. 1–12, 2020.

[14] F. Ceccarelli and M. Mahmoud, "Multimodal temporal machine learning for bipolar disorder and depression recognition," *Pattern Analysis and Applications*, vol. 25, no. 3, pp. 493–504, 2022.

[15] L. de Siqueira Rotenberg, R. G. Borges-Junior, B. Lafer, R. Salvini, and R. da Silva Dias, "Exploring machine learning to predict depressive relapses of bipolar disorder patients," *Journal of Affective Disorders*, vol. 295, pp. 681–687, 2021.

[16] D. N. Disha, S. Seema, S. U. Shenoy, and S. Rao, "Prediction of bipolar disorder using machine learning techniques," *in 2022 2nd International Conference on Intelligent Technologies (CONIT)*, Jun. 2022, pp. 1–5.

[17] M. Á. Luján, A. M. Torres, A. L. Borja, J. L. Santos, and J. M. Sotos, "High-precise bipolar disorder detection by using radial basis functions based neural network," *Electronics*, vol. 11, no. 3, p. 343, Jan. 2022.

[18] J. A. I. Diaz, R. P. Vicerra, and A. A. Bandala, "Preprocessing image contouring optimization of handwriting recognition using genetic algorithm," *in TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON), Auckland, New Zealand*, Dec. 2021, pp. 756–759.

[19] F. T. Anggraeny, Y. V. Via, and R. Mumpuni, "Image preprocessing analysis in handwritten Javanese character recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 860–867, 2023.

[20] C. Fuglsby, C. P. Saunders, D. M. Ommen, J. Buscaglia, and M. P. Caligiuri, "Elucidating the relationships between two automated handwriting feature quantification systems for multiple pairwise comparisons," *J. Forensic Sci.*, vol. 67, no. 2, pp. 642–650, 2022.

[21] T. Dhieb, S. Njah, H. Boubaker, W. Ouarda, M. B. Ayed, and A. M. Alimi, "Towards a novel biometric system for forensic document examination," *Computers & Security*, vol. 97, p. 101973, 2020.

[22] W. Simayi, M. Ibrayim, and A. Hamdulla, "Study the preprocessing effect on RNN-based online Uyghur handwritten word recognition," *Wireless Networks*, vol. 27, no. 8, pp. 1–12, 2021.

[23] S. Singh, V. K. Chauhan, and E. H. B. Smith, "A self-controlled RDP approach for feature extraction in online handwriting recognition using deep learning," *Appl. Intell.*, vol. 50, no. 7, pp. 2093–2104, 2020.

[24] T. Hasan, M. A. Rahim, J. Shin, S. Nishimura, and M. N. Hossain, "Dynamics of digital pen-tablet: Handwriting analysis for person identification using machine and deep learning techniques," *IEEE Access*, vol. 12, pp. 8154–8177, 2024.

[25] D. Brown and I. Lidzhade, "Handwriting recognition using deep learning with effective data augmentation techniques," *in 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, 2021, pp. 1–9.

[26] H. Pham, A. R. Setlur, S. Dingliwal, T. Lin, B. Póczos, K. Huang, Z. Li, J. Lim, C. McCormack, and T. Vu, "Robust handwriting recognition with limited and noisy data," *in 2020 17th International*

Conference on Frontiers in Handwriting Recognition (ICFHR), 2020, pp. 301–306.

[27] T. Hayashi, K. Gyohten, H. Ohki, and T. Takami, "A study of data augmentation for handwritten character recognition using deep learning," *in 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2018, pp. 552–557.

[28] X. Lv, "Handwritten digit recognition based on deep learning algorithms," *in 2023 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC)*, 2023, pp. 476–481.

[29] S. Minz, R. Kanojia, T. Yadav, and N. Jayanthi, "Enhancing accuracy in handwritten text recognition with convolutional recurrent neural network and data augmentation techniques," *in 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2023, pp. 803–808.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[31] K. S. Kawuma, E. Kumbakumba, V. Mabirizi, D. Nanjebe, K. Mworozi, A. O. Mukama, and L. Kyasimire, "Diagnosis and classification of tuberculosis chest X-ray images of children less than 15 years at Mbarara Regional Referral Hospital using deep learning," *Journal of AI and Data Mining*, vol. 12, no. 2, pp. 315–324, 2024.

[32] M. H. Guo, Z. N. Liu, T. J. Mu, and S. M. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5436–5447, 2022.

[33] A. S. Cohen, R. Cont, A. Rossier, and R. Xu, "Scaling properties of deep residual networks," *in Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2021, pp. 2039–2048.

[34] A. A. Jamali, "Bipolar vs non-bipolar handwriting dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/ahmadalijamali/bipolar-vs-non-bipolar-handwriting/data

[35] P. Magioncalda and M. Martino, "A unified model of the pathophysiology of bipolar disorder," *Molecular Psychiatry*, vol. 27, no. 1, pp. 202–211, 2022.

[36] R. Wakelin and P. Oakes, "Clinicians' perceptions of the bipolar disorder diagnosis: A Q-study," The Journal of Mental Health Training, Education and Practice, vol. 15, pp. 1–12, 2019.

[37] P. Baki, H. Kaya, E. Ciftçi, H. Gulec, and A. A. Salah, "A multimodal approach for mania level prediction in bipolar disorder," *IEEE Transactions on Affective Computing*, vol. 13, pp. 2119–2131, 2022.

[38] A. Ortiz, K. Bradler, and A. Hintze, "Episode forecasting in bipolar disorder: Is energy better than mood?," *Bipolar Disorders*, vol. 20, pp. 470–476, 2018.

[39] S. J. Herrera et al., "Neural abnormalities in bipolar disorder: A meta-analysis of functional neuroimaging studies," *European Psychiatry*, 2024.

[40] H. Shen, L. Zhang, C. Xu, J. Zhu, M. Chen, and Y. Fang, "Analysis of misdiagnosis of bipolar disorder in an outpatient setting," *General Psychiatry*, vol. 30, pp. 101–93, 2018.

[41] H. Bansal, S. D. Mishra, and M. Singhal, "The impact of neurological disorders on handwriting: Implications for forensic document examination," *Wiley Interdisciplinary Reviews: Forensic Science*, vol. e1536, 2024.

[42] S. Gerth and J. Festman, "Muscle activity during handwriting on a tablet: An electromyographic analysis of the writing process in children and adults," *Children*, vol. 10, no. 4, p. 748, 2023.

[43] Y. Wu, G. M. Goodwin, T. Lyons, and K. E. Saunders, "Identifying psychiatric diagnosis from missing mood data through the use of log-signature features," *PLoS ONE*, vol. 17, no. 11, p. e0276821, 2022.

[44] G. Casalino, G. Castellano, F. Galetta, and K. Kaczmarek-Majer, "Dynamic incremental semi-supervised fuzzy clustering for bipolar disorder episode prediction," *in International Conference on Discovery Science, Cham:* Springer International Publishing, Oct. 2020, pp. 79–93.

یک چارچوب ترکیبی یادگیری عمیق برای شناسایی اختلال دوقطبی از طریق الگوهای دست‌خط فارسی

خسرورضائی*

گروه مهندسی پزشکی، دانشکده فنی و مهندسی، دانشگاه میبد، میبد، ایران.

ارسال ۲۰۲۴/۱۲/۱۱؛ بازنگری ۲۰۲۵/۰۱/۱۲؛ پذیرش ۲۰۲۵/۰۳/۰۱

چکیده:

اختلال دوقطبی همچنان یک چالش فراگیر در حوزه سلامت روان است و نیازمند روش‌های تشخیصی نوآورانه فراتر از ارزیابی‌های سنتی و ذهنی می‌باشد. این مطالعه برای نخستین‌بار یک چارچوب تشخیصی غیرتهاجمی مبتنی بر دست‌خط را معرفی می‌کند که از ارتباط منحصربه‌فرد بین وضعیت‌های روانی و بیان‌های حرکتی در نوشتن بهره می‌گیرد. مدل ترکیبی یادگیری عمیق ما، با ترکیب شبکه ResNet برای استخراج ویژگی‌های پیچیده و مکانیزم توجه خارجی برای تحلیل الگوهای کلی، به دقت چشمگیر ۹۹٪ دست یافته است که با استفاده از اعتبارسنجی متقاطع Leave-One-Subject-Out (LOSO) ارزیابی شده است. این چارچوب با به‌کارگیری تکنیک‌های پیش‌پردازش و افزایش داده پیشرفته، به‌طور مؤثر عدم‌توازن مجموعه داده و تنوع دست‌خط‌ها را مدیریت می‌کند. برای نخستین‌بار، دست‌خط فارسی به عنوان ابزاری مؤثر به‌کار گرفته شده است و شکاف‌های فرهنگی در تشخیص اختلال دوقطبی را پوشش می‌دهد. این پژوهش نه تنها دست‌خط را به عنوان ابزاری تحول‌آفرین در تشخیص سلامت روان معرفی می‌کند، بلکه زمینه را برای ارائه راهکارهای در دسترس، مقیاس‌پذیر و سازگار با فرهنگ‌های مختلف در حوزه سلامت روان جهانی فراهم می‌سازد.

کلمات کلیدی: تشخیص اختلال دوقطبی، تحلیل دست‌خط، کاربردهای یادگیری عمیق، دست‌خط فارسی و هوش مصنوعی در سلامت روان.