BIO Web of Conferences **232**, 06002 (2026) https://doi.org/10.1051/bioconf/202623206002
ICBBB 2026

Parkinson's Disease Detection Through Static Handwriting Analysis Using CNNs and SVM Ensemble

Myk Erolf Roble[1], *Heart Alvern* Sumicad[1], and *Christine* Bandalan[1]

[1]University of San Carlos, Gov. M. Cuenco Avenue, Talamban, Cebu City, Philippines

Abstract. Parkinson's Disease is a progressive neurological disorder marked by motor impairments, often reflected in handwriting anomalies such as tremors, micrographia, and stroke irregularities. The diagnosis often relies on subjective clinical evaluation. This study explores an objective detection framework using static handwriting samples, specifically spiral and wave drawings. The method employs an ensemble strategy that fuses deep features from two pre-trained Convolutional Neural Network (CNN) architectures, ResNet50 and DenseNet121, and classifies the feature representations using a Linear Support Vector Machine (SVM). To ensure rigor, a strict split-before-augmentation protocol was applied to a primary dataset of 204 subjects. The model achieved 97.28% accuracy on the primary test set. When independently verified using a local dataset, the system maintained strong generalization with an accuracy of 95.19%, 97.04% PD recall, and 93.33% specificity. Statistical analysis confirmed these results with narrow confidence intervals, validating the system's stability. The results confirm that dual-CNN feature fusion significantly improves detection stability and adaptability between datasets. This framework serves as an objective, scalable screening tool to complement neurological assessments.

1 Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects movement control, commonly manifesting as tremors, rigidity, slowness of movement (bradykinesia), and postural instability. Among its early and visible symptoms is micrographia, a condition where a person's handwriting becomes abnormally small and less legible over time [1]. While PD is currently incurable, early and accurate detection can significantly improve symptom management and quality of life.

Traditional diagnostic methods, including neurological examinations, the Unified Parkinson's Disease Rating Scale (UPDRS), and DaTscan imaging, are often costly, invasive, and highly dependent on clinical judgment [2-3]. These limitations make them less ideal for routine screening, especially in remote or resource-limited settings.

Handwriting analysis has emerged as a promising, non-invasive tool for detecting PD-related motor impairments. Research shows that handwriting tasks such as spiral and

---

[1] Corresponding author: 23102463@usc.edu.ph

wave drawings can reflect critical motor symptoms of PD, including tremors, stroke irregularities, and size reduction [4]. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated success in extracting meaningful features from handwriting images [5-6]. Despite their effectiveness, end-to-end CNN classifiers often require large datasets to generalize well and are prone to overfitting in small-sample scenarios, which are common in medical applications. To address this, researchers have proposed hybrid models that combine CNN-based feature extraction with a traditional machine learning classifier – Linear Support Vector Machines (SVMs). Linear SVMs are particularly effective in handling high-dimensional feature vectors, offer low computational complexity, and perform well on small datasets [7-8].

Given this context, the present study proposes a hybrid method for detecting Parkinson’s Disease using spiral and wave handwriting images. The approach uses pre-trained CNN models, ResNet and DenseNet, for deep feature extraction, followed by classification using a Linear SVM. The goal is to develop an accessible, scalable, and objective tool that complements traditional diagnostic procedures and enhances PD detection through handwriting-based analysis.

2 Related Works

2.1 Convolutional Neural Networks

Transfer learning (TL)-based convolutional neural network (CNN) architectures such as ResNet, VGG, and Inception have been extensively studied for medical imaging tasks, particularly in small-data scenarios. Kumar et al. [9] evaluated these models using 233 MRI scans for brain tumor detection and reported that VGG achieved high accuracy due to its relatively simple architecture and reduced parameter count, which helped limit overfitting. In contrast, ResNet provided a key advantage through residual connections, enabling the effective training of deeper networks without vanishing gradients or performance degradation. Inception, meanwhile, utilized parallel convolutional modules to efficiently manage network depth.

Additional studies involving EfficientNet, GoogLeNet, and XceptionNet reported strong performance in multiclass disease classification tasks, including COVID-19 and monkeypox. However, ResNet variants consistently demonstrated robust and stable performance across datasets, with ResNet-101 achieving the highest accuracy in multiclass monkeypox classification, highlighting its strong generalization capability even with limited data [10]. Similarly, K and M [6] reported that DenseNet-201 achieved 94% classification accuracy on spiral drawing datasets and outperformed other pre-trained models. These findings support the selection of ResNet and DenseNet as reliable architectures for transfer learning in medical image classification.

2.2 Machine Learning Classifiers in Medical Imaging

Among traditional classifiers, Support Vector Machines (SVM) are particularly valued in medical and biomedical applications for their resilience to small or imperfect data. A 2023 survey by Guido, Ferrisi, Lofaro, and Conforti [11] emphasizes that SVM’s capacity to define optimal decision margins makes it ideal in settings with limited or imbalanced datasets. This resilience was further confirmed in an empirical study evaluating six classifiers on 18 small medical datasets, where SVM ranked as the third most robust, only behind AdaBoost and Naïve Bayes [7].

As evidenced by the work of Shanbehzadeh et al. (2022) [12], ensemble algorithms represent a more robust alternative to traditional ML methods, specifically in their capacity to accurately categorize high-risk individuals and serve as a practical screening tool for early breast cancer detection. Similarly, Guido, Ferrisi, Lofaro, and Conforti [11] emphasized that extracting relevant features significantly enhances SVM performance in healthcare applications. Thus, combining a Convolutional Neural Network (CNN) like ResNet with a Support Vector Machine (SVM) is a strong methodological choice for medical image classification, particularly in data-scarce environments. ResNet excels at extracting deep, high-level features from complex handwriting images, effectively converting unstructured input into structured representations. These optimized features can then be classified by an SVM, specifically Linear SVM as it is often favored for classification, as the features extracted are typically high-level and linearly separable, This makes Linear SVM both computationally efficient and effective, particularly in small-sample medical datasets [8].

2.3 Ensemble Learning and Hybrid Models

Ensemble learning has been shown to improve the generalization performance of deep learning models by combining multiple architectures or training strategies. Prior studies indicate that deep ensembles outperform single models in various tasks, including medical image classification, object detection, and sentiment analysis [13]. In medical diagnosis, ensemble CNN approaches have achieved improved detection accuracy for conditions such as skin cancer, COVID-19, and diabetic retinopathy [14–16].

3 Technical Background

This study uses a hybrid architecture that combines deep convolutional neural networks (CNNs) for feature extraction with a Linear Support Vector Machine (SVM) for final classification. CNNs are widely used in image analysis due to their ability to automatically learn hierarchical visual patterns through convolution, activation, and pooling layers. These layers transform handwriting images into high-dimensional feature vectors that capture both global and fine-grained stroke characteristics.

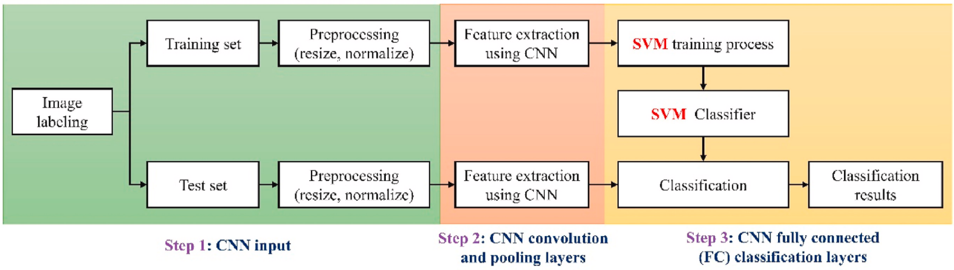

Fig. 1. Resnet-50 Model architecture, retrieved from Mukherjee, 2022[17]

Two pretrained CNNs—ResNet50 and DenseNet121—serve as the feature extractors. ResNet50 employs residual ("skip") connections [18] to address vanishing gradients and enable deeper, more stable learning, making it effective for capturing broad structural handwriting distortions such as tremors and irregular stroke pressure. DenseNet121 uses dense connectivity, where each layer receives inputs from all previous layers, promoting feature reuse and preserving spatial detail. This allows DenseNet to detect subtler handwriting abnormalities, including micrographia and jitter [19]. Their combined representations offer complementary strengths for handwriting-based PD detection.

Both CNNs are initialized with ImageNet weights to leverage transfer learning, improving generalization and reducing training time—especially important for this study's relatively small handwriting dataset.

Instead of relying on the CNNs' softmax classifier, the fused feature vectors are passed to a Linear SVM. SVMs construct a maximum-margin hyperplane between classes and are well suited for small, high-dimensional datasets [22]. This hybrid CNN–SVM design has been widely used in medical image classification and has demonstrated improved generalization compared to softmax-based CNNs, particularly in studies involving limited training data[20-21].

4 Design and Methodology

This section describes the research environment, datasets, preprocessing steps, model architecture, and evaluation methods used in developing the proposed handwriting-based Parkinson's Disease detection system.

4.1 Datasets and Research Environment

The study utilized two distinct datasets. The Public dataset comprises samples from 204 unique subjects, containing spiral and wave drawings from both Parkinson's Disease (PD) patients and healthy individuals. To verify generalization, a separate Local dataset was employed, consisting of handwriting samples from 30 unique subjects (15 PD, 15 Healthy) provided by Dr. Camara-Chua (2025). This Local dataset was kept completely isolated during the training and hyperparameter tuning phases to serve as a true hold-out verification set.

4.2 Preprocessing and Feature Extraction Pipeline

To prevent data leakage, dataset splitting was performed prior to augmentation. The unique subjects in the Public dataset were stratified into training (85%) and testing (15%) sets. Data augmentation was applied after this split, ensuring that no augmented variations of testing samples were present in the training set. This resulted in a robust training corpus while maintaining the integrity of the evaluation. All handwriting images were standardized through resizing, grayscale conversion, denoising, thresholding, centering, padding, and normalization. Data augmentation was applied to improve performance.

Preprocessed images were then passed through two pretrained convolutional neural networks—ResNet-50 and DenseNet-121—initialized with ImageNet weights. Outputs from both networks were flattened and concatenated into a single fused feature vector for each sample.

4.3 Model Training and Classification

Outputs from ResNet-50 (2048-D) and DenseNet-121 (1024-D) were flattened and concatenated into a single 3072-dimensional fused feature vector. Dimensionality reduction was explicitly omitted; the Linear SVM was trained directly on the raw high-dimensional features. This approach was chosen to prevent potential information loss associated with dimensionality reduction techniques, given that Linear SVMs are mathematically robust in high-dimensional spaces.

4.4 Conceptual Framework and System Architecture

The system follows a three-stage workflow:

- Preprocessing: Standardizes handwriting inputs.
- Dual-CNN Feature Extraction: Generates complementary high-level representations.
- Linear SVM Classification: Learns the optimal separating hyperplane for PD vs. Healthy.

4.5 Evaluation and Validation Procedure

Model performance was assessed on the unseen test split using metrics including accuracy, precision, recall, F1-score, and error rate. A confusion matrix was used to analyze diagnostic behavior and identify classification errors, with PD treated as the positive class. All metrics were computed using scikit-learn to ensure reproducibility.

Table 1. Target baseline of the proposed model.

Metric	Baseline Target
Accuracy	≥ 94%
Recall (PD Sensitivity)	≥ 92.5%
F1-Score	≈ 94%

5 Results and Discussion

This section presents the performance of the proposed dual CNN–SVM ensemble, including feature extraction results, model optimization, and evaluation on both the primary online dataset and the external local dataset.

5.1 System Overview

The final model integrates features extracted from two pretrained CNNs—ResNet-50 and DenseNet-121—which transform each handwriting image into high-dimensional embeddings (2048-D and 1024-D, respectively). After preprocessing and resizing (224×224), the features are concatenated into a 3072-dimensional vector and standardized before being classified using a Linear SVM. This hybrid design maximizes representational richness while preventing overfitting on small medical datasets.

Model optimization was performed using a stratified 85/15 train-test split and five-fold cross-validation on the training portion. A grid search over $C=\{0.01,0.1,1,10,100\}$ identified $C = 0.01$ as the best configuration, achieving a mean cross-validated weighted F1-score of 0.9606.

5.2 Performance on Primary Dataset

Testing on the unseen 15% of the primary dataset produced the following results:

Table 2. Result of the 15% testing on the Primary Dataset.

Metric	**Value**	**Margin of Error**	**95% Confidence Interval**
Accuracy	97.28%	±2.35%	94.93% – 99.63%
Sensitivity (PD Recall)	100%	N/A	Approaches 100%, lower bound ~96.0%
Specificity (Healthy Recall)	94.6%	±4.63%	89.94% – 99.20%

The reported 95% confidence intervals indicate that the model's performance is both robust and statistically reliable, particularly the perfect sensitivity for PD cases. This supports the generalization capability of the proposed ensemble on unseen data.

Table 3. Comparison of the target baseline and the testing result.

Metric	**Goal**	**Achieved**	**Result**
Accuracy	≥ 94%	97.28%	Exceeded
Recall (PD)	≥ 92.5%	100%	Exceeded
Weighted F1-Score	~94%	97.28%	Exceeded

The confusion matrix showed 179 correctly classified out of 184 samples, with perfect recall for PD and 94.6% recall for Healthy samples. These outcomes surpass the baseline targets established in Chapter 4 and exceed the accuracy range (81–87%) reported in an earlier study using various machine learning approaches [22].

5.3 Verification on Local Dataset

The trained ensemble was verified on the isolated Local dataset. The model achieved an accuracy of 95.19% across this augmented set, confirming its ability to generalize to new subjects and variations in handwriting styles.

Table 4. Result on the Local Dataset Verification

Metric	**Value**	**Margin of Error**	**95% Confidence Interval**
Accuracy	95.19%	±2.55%	92.64% – 97.74%
Sensitivity (PD Recall)	97.04%	±2.85%	94.19% – 99.89%
Specificity (Healthy Recall)	93.33%	±4.21%	89.12% – 97.54%

5.4 Discussion of findings

The proposed CNN–SVM ensemble demonstrated excellent performance across both datasets. On the primary dataset, accuracy reached 97.28%, with 100% PD recall, ensuring no tremor-affected samples were missed—an essential quality for screening applications. The confusion matrix and cross-validation consistency indicate a stable decision boundary with low risk of overfitting.

A key strength of the system is its consistently high sensitivity to PD-related handwriting characteristics. This, combined with the margin-based stability of Linear SVM and the complementary strengths of ResNet and DenseNet feature extraction, validates the hybrid approach as an effective method for handwriting-based PD detection.

Verification using the local dataset confirmed robust generalization, with 95.19% accuracy and 97.04% PD recall under real-world variability. Although healthy-class recall decreased slightly, the overall weighted F1-score remained high (0.9527), demonstrating reliability even in less controlled conditions.

Previous studies utilizing standard deep learning classifiers have established a performance ceiling around 93-94%. For instance, Alniemi and Mahmood [5] reported a 93.33% accuracy using a custom CNN architecture on spiral and wave drawings. Similarly, K. & M. [6] achieved 94.0% accuracy using a fine-tuned DenseNet-201 model.

In contrast, the proposed Dual-CNN + SVM ensemble achieved 97.28% accuracy, representing a performance increase of approximately 3.3% to 4.0% over these single-architecture end-to-end baselines. This margin is significant in medical screening, where sensitivity is paramount.

6 Summary, Conclusions, and Recommendations

6.1 Summary of Findings

This study developed and evaluated a dual CNN–SVM ensemble for detecting Parkinson's Disease from static handwriting images. The system utilized two pretrained convolutional neural networks—ResNet-50 and DenseNet-121—to extract complementary feature representations, which were then classified using a Linear Support Vector Machine. The approach was designed to perform reliably even on small datasets while maintaining strong sensitivity to tremor-related handwriting patterns.

On the primary handwriting dataset, the model achieved 97.28% accuracy, a weighted F1-score of 0.9728, and 100% recall for PD, successfully identifying all tremor-affected samples. These outcomes surpassed all baseline targets established in Chapter 4.

Further assessed using a locally collected dataset. The ensemble achieved 95.19% accuracy, 97.04% PD recall, and a weighted F1-score of 0.9527. Although accuracy decreased slightly, the system continued to detect PD samples with high sensitivity. The minor performance drop is attributed to natural handwriting variations, differences in lighting or paper quality, and irregularities sometimes present in healthy handwriting.

Overall, the findings show that the proposed CNN–SVM ensemble performs exceptionally well under structured conditions and remains robust when applied to real-world handwriting samples. The results support the feasibility of using static handwriting as an accessible, reliable screening tool for tremor detection.

6.2 Conclusion

The study concludes that the proposed CNN–SVM ensemble is an effective and reliable method for detecting Parkinson's Disease from static handwriting samples. The combination of deep feature extraction (ResNet-50 and DenseNet-121) and linear SVM classification enabled the model to achieve excellent performance on the primary dataset and strong generalization on the local dataset.

High PD recall across both evaluations demonstrates the system's suitability for screening applications, where identifying true cases is critical. The results confirm that handwriting contains detectable motor signatures associated with Parkinson's Disease and that a hybrid CNN–SVM pipeline can capture these patterns with high accuracy. This work establishes static handwriting analysis as a promising, non-invasive, reliable approach for PD-presenting handwriting assessment.

6.3 Recommendations

To further improve and extend this work, several recommendations are proposed:

1. Expand the Dataset
   Incorporating a larger and more diverse set of handwriting samples—varying in age, writing conditions, paper type, pen pressure, and cultural writing differences—would improve the model and reduce false classification.
2. Enhance Preprocessing Techniques
   Refinements such as advanced stroke isolation, adaptive noise reduction, or contrast enhancement may help reduce misclassification of healthy samples that contain natural irregularities.
3. Explore Dynamic Handwriting Signals
   Future work may include pen-trajectory data, time-based stroke measurements, or tablet-collected handwriting, which capture fine motor fluctuations not visible in static images.
4. Experiment with Alternative Models
   Testing other ensemble configurations, additional CNN architectures, or different classifiers (e.g., RBF-SVM, Random Forest, XGBoost) may uncover performance improvements on more challenging datasets.

By addressing these areas, future research can strengthen diagnostic reliability and move toward practical deployment of handwriting-based neurological screening tools.

References

1. N. Yang, J. Liu, D. Sun, J. Ding, L. Sun, X. Qi, W. Yan. Motor symptoms of Parkinson’s disease: Critical markers for early AI-assisted diagnosis. Front. Aging Neurosci. 17, 1602426 (2025). https://doi.org/10.3389/fnagi.2025.1602426

2. S. B. Dias, A. Grammatikopoulou, J. A. Diniz, K. Dimitropoulos, N. Grammalidis, V. Zilidou, L. J. Hadjileontiadis. Innovative Parkinson’s Disease Patients’ Motor Skills Assessment: the I-PROGNOSIS paradigm. Front. Comput. Sci. 2, 20 (2020). https://doi.org/10.3389/fcomp.2020.00020

3. H. Suominen, M. Manocha, J. Desborough, A. Parkinson, D. Apthorp. Finger tapping measures for Parkinson’s disease: Preliminary evaluation of an Android application for data collection in Australia. Stud. Health Technol. Inform. (2021). https://doi.org/10.3233/shti210775

4. A. Zemmar, A. Bennour, M. Tahar, D. Ghabban, M. Al-Sarem. Unveiling Parkinson’s: Handwriting symptoms with explainable and interpretable CNN model. Int. J. Pattern Recognit. Artif. Intell. 39 (2025). https://doi.org/10.1142/S0218001424400019

5. E. Alniemi, A. Mahmood. Convolutional neural networks for Parkinson’s disease detection based on handwriting analysis. Comput. Biol. Med. 162, 107185 (2023). http://doi.org/10.11591/ijeecs.v30.i1.pp267-275

6. S. K., N. T. M. Design of a deep fusion model for early Parkinson’s disease prediction using handwritten image analysis. Sci. Rep. 15, (2025). https://doi.org/10.1038/s41598-025-04807-6

7. A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, H. Kurdi. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Appl. Sci. 11, 796 (2021). https://doi.org/10.3390/app11020796

8. F. Prinzi, T. Currieri, S. Gaglio, S. Vitabile. Shallow and deep learning classifiers in medical image analysis. Eur. Radiol. Exp. 8, (2024). https://doi.org/10.1186/s41747-024-00428-2

9. N. Kumar, M. Gupta, D. Gupta, S. Tiwari. Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images. J. Ambient Intell. Humaniz. Comput. 14, 469–478 (2021). https://doi.org/10.1007/s12652-021-03306-6

10. A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, A. Mellit. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability 15, 5930 (2023). https://doi.org/10.3390/su15075930

11. R. Guido, S. Ferrisi, D. Lofaro, D. Conforti. An Overview on the advancements of support vector machine models in healthcare Applications: a review. Information 15, 235 (2024). https://doi.org/10.3390/info15040235

12. M. Shanbehzadeh, H. Kazemi-Arpanahi, M. B. Ghalibaf, A. Orooji. Performance evaluation of machine learning for breast cancer diagnosis: A case study. Inform. Med. Unlocked 31, 101009 (2022). https://doi.org/10.1016/j.imu.2022.101009

13. M.A. Ganaie, M. Hu, M. Tanveer, P.N. Suganthan, Ensemble deep learning: A review, Eng. Appl. Artif. Intell. 115, 105151 (2022). https://doi.org/10.1016/j.engappai.2022.105151

14. A. Aldwgeri, N. Abubacker. Ensemble of deep convolutional neural networks for skin lesion classification in dermoscopy images. In Proc. Int. Conf. Adv. Comput. Intell. (2019). https://doi.org/10.1007/978-3-030-34032-2_20

15. A. Narin, C. Kaya, Z. Pamuk. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal. Appl. 24, 1207–1220 (2021). https://doi.org/10.1007/s10044-021-00984-y

16. S. Sushith, M. Sathiya, V. Kalaipoonguzhali, V. Sathya. A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. Sci. Rep. 15 (2025). https://doi.org/10.1038/s41598-025-99309-w

17. Mukherjee, S. (2022, August 18). The annotated ReSNet-50 - TDS Archive - medium. Medium. https://medium.com/data-science/the-annotated-resnet-50-a6c536034758

18. M. Shaban. Deep Learning for Parkinson's Disease Diagnosis: A Short Survey. Computers 12, 58 (2023). https://doi.org/10.3390/computers12030058

19. Deepfake detection. J. Multidiscip. Knowl. 5, 63–74 (2025). https://doi.org/10.36676/jmk.v5.i2.76

20. Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J. H., Shoeibi, A., Gorriz, J. M., & Sani, Z. A. (2021). Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomedical Signal Processing and Control, 68, 102622. https://doi.org/10.1016/j.bspc.2021.102622

21. Ahlawat, S., & Choudhary, A. (2020). Hybrid CNN-SVM classifier for handwritten digit recognition. Procedia Computer Science, 167, 2554–2560. https://doi.org/10.1016/j.procs.2020.03.309

22. N. M. Ranjan, G. Mate, M. Bembde. Detection of Parkinson's Disease using Machine Learning Algorithms and Handwriting Analysis. J. Data Min. Manag. 8, 21–29 (2023). https://doi.org/10.46610/jodmm.2023.v08i01.004