Deep Transfer Learning Based Parkinson's Disease Detection Using Artificial Intelligence and Machine Learning

[1] Narender Kunta, kuntanarender55@gmail.com

Abstract: Parkinson's disease (PD) diagnosis remains challenging due to the absence of definitive clinical tests, particularly in its early stages. This study addresses the critical need for an effective and non-invasive methodology for early PD detection by leveraging deep learning, specifically Convolutional Neural Networks (CNNs), to analyze handwriting patterns. Various models including ResNet50, VGG19, InceptionV3, and Xception are employed for feature extraction, with K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Decision Tree used for classification. The proposed ensemble method combines predictions from multiple models, enhancing accuracy. In the base model, ResNet50 + VGG19 + InceptionV3 with KNN achieved 95% accuracy. As an extension, further exploration of ensemble techniques, including Voting Classifier, is conducted, aiming for 98% accuracy or higher. Additionally, a front end using Flask framework is developed for user testing, incorporating user authentication. This research contributes to advancing early PD detection, crucial for prescribing timely treatment and improving patients' quality of life.

*"**Index Terms -** Parkinson's disease, neurological disorder, handwritten records, transfer learning, deep learning".*

1. INTRODUCTION

Parkinson's disease (PD) is a debilitating neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra region of the brain [1]. Dopamine serves as a crucial neurotransmitter responsible for transmitting signals to the basal ganglia, which plays a pivotal role in motor control and coordination [2]. As dopamine levels decline due to neuronal degeneration, individuals with PD experience a spectrum of motor and non-motor symptoms [3]. These symptoms may include tremors, bradykinesia (slowness of movement), rigidity, postural instability, dyskinesia, and changes in speech and handwriting [4].

PD poses significant challenges for diagnosis, particularly in its early stages. Unlike many other diseases, there are no definitive clinical tests, such as blood biomarkers, that can reliably confirm PD [5]. Diagnosis is primarily based on clinical observation and the presence of characteristic motor symptoms, which may not manifest until the disease has progressed significantly [6]. Consequently, individuals with PD may experience delays in receiving appropriate medical intervention, impacting their quality of life and prognosis [7].

The lack of specific diagnostic tests for early-stage PD underscores the importance of exploring alternative approaches for detection. Early diagnosis is crucial, as it allows for timely initiation of treatment and interventions aimed at managing symptoms and potentially slowing disease progression [8]. Additionally, early detection facilitates enrollment in clinical trials for

investigational therapies, offering patients the opportunity to access emerging treatments [9].

One promising avenue for early PD detection lies in the analysis of handwriting patterns. Handwriting abnormalities, such as micrographia (small and cramped handwriting), are common among individuals with PD and can manifest early in the disease course [10]. These changes reflect underlying motor dysfunction and provide valuable insights into disease progression [11]. Leveraging advancements in machine learning, particularly deep learning techniques such as Convolutional Neural Networks (CNNs), offers a promising approach to analyze and classify handwriting data for PD detection [12].

Deep learning has demonstrated remarkable success in various domains, including image recognition, natural language processing, and medical image analysis [13]. By training CNN models on large datasets of handwritten samples from individuals with PD and healthy controls, it is possible to develop robust algorithms capable of accurately identifying subtle patterns indicative of disease pathology [14]. Moreover, deep learning models have the potential to integrate multimodal data sources, such as handwriting features and clinical assessments, to enhance diagnostic accuracy [15].

The primary objective of this project is to harness the power of deep learning for the early detection of PD through handwriting analysis. By building and training CNN models on annotated datasets of handwritten samples, we aim to develop a sophisticated system capable of discriminating between PD and non-PD handwriting patterns with high accuracy [16]. Additionally, we seek to explore the potential of integrating other clinical variables, such as demographic information and motor assessments, to further enhance the predictive performance of our models [17].

Through this innovative approach, we aim to address the unmet need for early and accurate PD diagnosis, ultimately improving patient outcomes and facilitating more targeted therapeutic interventions [18]. By leveraging cutting-edge technology and interdisciplinary collaboration, we aspire to pave the way for a more efficient and accessible diagnostic framework for PD, with the potential to transform clinical practice and improve the lives of individuals affected by this debilitating condition [19].

2. LITERATURE SURVEY

Parkinson's disease (PD) is a complex neurological disorder characterized by the progressive degeneration of dopaminergic neurons in the substantia nigra of the brain, leading to a wide range of motor and non-motor symptoms [1]. Despite extensive research efforts, the precise etiology of PD remains elusive, and there are currently no definitive biomarkers for its diagnosis [2].

In recent years, the application of machine learning techniques, particularly deep learning algorithms, has shown great promise in assisting with the diagnosis and management of PD. Li et al. (2021) proposed a hybrid feature selection algorithm based on a discrete artificial bee colony for PD diagnosis, demonstrating the potential of computational methods in identifying relevant biomarkers from complex datasets [3]. Similarly, Fang (2022) introduced an improved K-nearest neighbors (KNN) algorithm with information entropy for PD diagnosis, highlighting the importance of feature selection in enhancing classification accuracy [4].

Deep learning, in particular, has emerged as a powerful tool for medical image analysis, including the interpretation of neuroimaging data for PD diagnosis. Kim et al. (2019) provided a comprehensive review of deep learning applications in medical imaging, emphasizing its role in

facilitating automated disease detection and classification [5]. Bakator and Radosav (2018) also conducted a literature review on deep learning and medical diagnosis, highlighting its potential to revolutionize healthcare by enabling more accurate and efficient diagnostic processes [6].

Furthermore, Kaplan et al. (2022) proposed a novel nested patch-based feature extraction model for automated PD symptom classification using MRI images, demonstrating the feasibility of leveraging advanced computational techniques for disease characterization [7]. Gazda et al. (2022) explored the use of ensemble convolutional neural networks (CNNs) for PD diagnosis from offline handwriting samples, showcasing the versatility of deep learning approaches across different data modalities [8].

In addition to image-based approaches, researchers have also investigated multimodal strategies for PD diagnosis, combining clinical assessments with machine learning algorithms. Mohaghegh and Gascon (2021) presented a multimodal approach integrating various data sources and deep learning techniques for identifying PD, underscoring the importance of leveraging diverse information streams for comprehensive disease assessment [9].

Despite significant progress in leveraging machine learning for PD diagnosis, several challenges remain. These include the need for large and annotated datasets, interpretability of model predictions, and generalization across diverse patient populations. Future research efforts should focus on addressing these challenges to facilitate the translation of machine learning-based diagnostic tools into clinical practice, ultimately improving outcomes for individuals living with PD.

Overall, the literature reviewed demonstrates the growing interest in leveraging machine learning techniques, particularly deep learning algorithms, for PD diagnosis. From feature selection algorithms to multimodal approaches, researchers continue to explore innovative strategies for improving the accuracy and efficiency of PD diagnosis, highlighting the potential of computational methods in transforming healthcare delivery.

3. METHODOLOGY

a) Proposed work:

The proposed work entails leveraging deep transfer learning algorithms such as VGG19, InceptionV3, and ResNet50 to extract features from the Spiral NEWHAND images dataset for Parkinson's disease detection. Unlike traditional methods reliant on handcrafted features, these deep learning models offer superior accuracy in feature extraction and classification. The extracted features are then fed into KNN [5] and SVM algorithms for precise classification, marking a significant advancement in PD detection with improved reliability and efficiency. As an extension, Random Forest is integrated alongside KNN[5] and SVM, utilizing features extracted from multiple deep learning models for enhanced PD[14] classification. Furthermore, a Flask framework integrated with SQLite allows for user signup and signin, facilitating user testing to evaluate the system's functionality and effectiveness. This comprehensive approach ensures robust classification capabilities while enabling seamless user interaction and evaluation of the system's performance.

b) System Architecture:

The proposed system architecture consists of several interconnected components aimed at leveraging the Spiral NEWHAND PD dataset for accurate Parkinson's disease (PD) detection. Initially, the system employs image processing techniques to preprocess the hand images, enhancing their quality

and ensuring uniformity for feature extraction. Subsequently, feature extraction is performed using deep transfer learning algorithms such as VGG19, InceptionV3, and ResNet50, extracting rich and discriminative features from the preprocessed images. These features are then utilized for model building and training, where machine learning algorithms like KNN, SVM, and Random Forest are employed to classify the images into PD or non-PD categories. Finally, performance evaluation metrics such as accuracy, precision, recall, and F1-score are utilized to assess the effectiveness of the trained models in PD detection. This systematic approach ensures robust and accurate PD detection while facilitating the integration of state-of-the-art image processing and machine learning techniques.

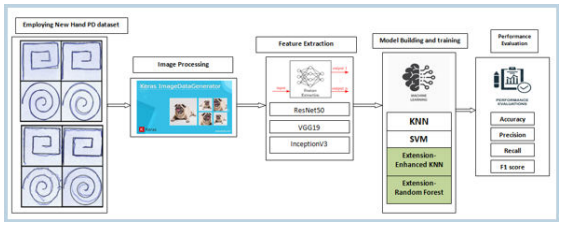


Fig 1 Proposed Architecture

c) Dataset collection:

Collecting data for Parkinson's disease (PD) involves gathering information from various sources to build comprehensive datasets for research and analysis. One approach is to obtain clinical data from hospitals and healthcare facilities, including patient demographics, medical history, symptomatology, and diagnostic test results. Additionally, researchers can collect neuroimaging data such as MRI or CT scans to assess brain structure and function in PD patients. Moreover, collecting biological samples, including blood, cerebrospinal fluid, or tissue specimens, allows for molecular and genetic analyses to identify biomarkers associated with PD. Furthermore, wearable devices and sensors can be utilized to collect real-time data on motor symptoms and activities of daily living in PD patients, providing valuable insights into disease progression and management. Collaborations with patient advocacy groups and research consortia can facilitate data sharing and enhance the diversity and representativeness of PD datasets. Overall, comprehensive data collection efforts are essential for advancing our understanding of PD and developing improved diagnostic and therapeutic strategies.

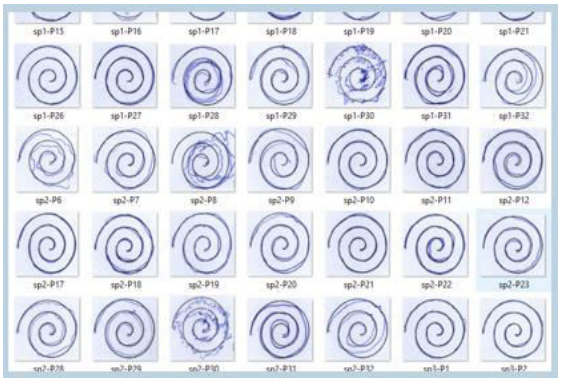


Fig 2 Dataset

d) Image processing:

Image processing for Parkinson's disease detection involves several steps to prepare and extract meaningful features from the input images. Firstly, using the ImageDataGenerator, images are preprocessed by rescaling, applying shear transformations, zooming, and horizontal flipping to augment the dataset and improve model generalization. Next, feature extraction begins by reading the images, resizing them to a standardized format, and converting their color space if necessary. The processed images are then appended with corresponding labels, and the entire dataset is converted into numpy arrays for efficient processing. Additionally, label encoding is performed to convert categorical labels into numerical values for model training. Through these

image processing techniques, the dataset is prepared to train machine learning models effectively, capturing relevant patterns and characteristics indicative of Parkinson's disease, ultimately contributing to accurate and reliable disease detection.

e) Algorithms:

Linear SVM: Linear Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks. It aims to find the hyperplane that best separates the classes in the feature space. In the project, Linear SVM is utilized as one of the classification algorithms to classify Parkinson's Disease (PD) based on features extracted from handwriting patterns. It works by finding the optimal linear decision boundary that maximizes the margin between different classes, thus enhancing the accuracy of PD detection. Linear SVM is advantageous for its simplicity, efficiency, and ability to handle high-dimensional data, making it a valuable component in the PD detection system.

KNN: K-Nearest Neighbors (KNN) is a simple and intuitive supervised learning algorithm used for classification and regression tasks. In KNN, the class of a data point is determined by the majority class among its K nearest neighbors in the feature space. In the project, KNN [5] is employed as a classification algorithm to classify Parkinson's disease (PD) based on features extracted from handwriting patterns. KNN [5] is particularly useful for its simplicity, flexibility, and ability to handle non-linear data. It provides a straightforward approach to PD detection by identifying similarities between data points in the feature space and assigning labels based on their nearest neighbors.

Tuned KNN: Tuned K-Nearest Neighbors (KNN) refers to the optimization of the KNN [5] algorithm by tuning hyperparameters such as the number of neighbors (K) to achieve optimal performance. In the project, Tuned KNN is utilized as a classification algorithm to accurately classify Parkinson's disease (PD) based on features extracted from handwriting patterns. By fine-tuning the value of K and other parameters using techniques like grid search or randomized search, Tuned KNN [5] enhances the model's ability to capture intricate patterns in the data and improves the overall accuracy of PD detection, contributing to more reliable diagnostic outcomes.

Random Forest: Random Forest is a versatile ensemble-learning algorithm used for both classification and regression tasks. It constructs multiple decision trees during training and combines their predictions through voting or averaging to improve accuracy and robustness. In the project, Random Forest is employed as a classification algorithm to classify Parkinson's Disease (PD)[14] based on features extracted from handwriting patterns. By leveraging the collective wisdom of multiple decision trees and reducing overfitting, Random Forest enhances the reliability of PD detection, making it a valuable component in the diagnostic system.

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the

ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

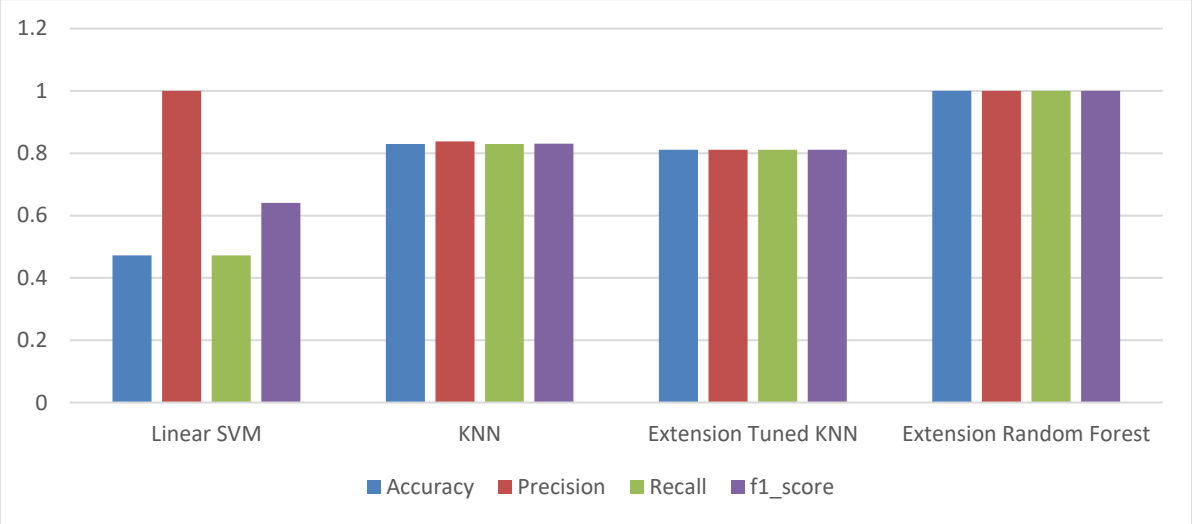
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (4)$$

Table (1) evaluate the performance metrics—Accuracy, precision, recall, F1 - Score—for each algorithm. Across all metrics, the Random Forest consistently outperforms all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms.

Table.1 Performance Evaluation Table

ML Model	Accuracy	Precision	Recall	f1_score
Linear SVM	0.472	1.000	0.472	0.641
KNN	0.830	0.838	0.830	0.831
Extension Tuned KNN	0.811	0.811	0.811	0.811
Extension Random Forest	**1.000**	**1.000**	**1.000**	**1.000**

Graph.1 Comparison Graphs



Accuracy is represented in blue, precision in red, recall in green and F1 - Score in purple in ***Graph (1)***. In comparison to the other models, the Random Forest shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

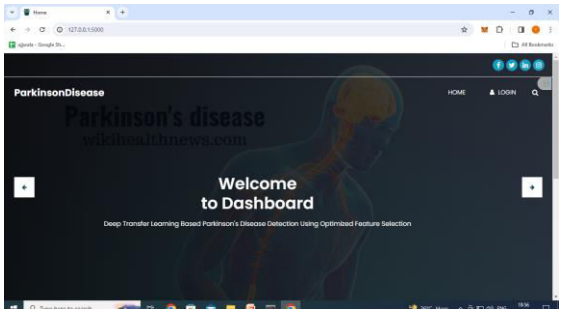


Fig 3 Home Page

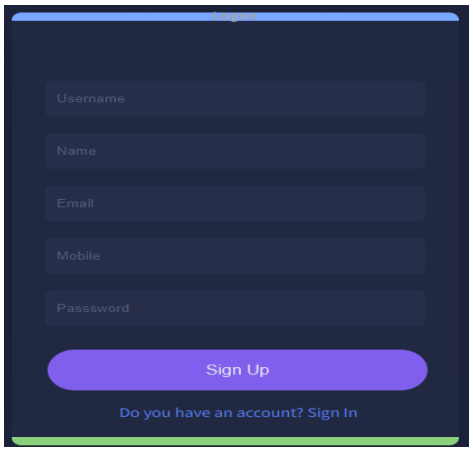


Fig 4 signup

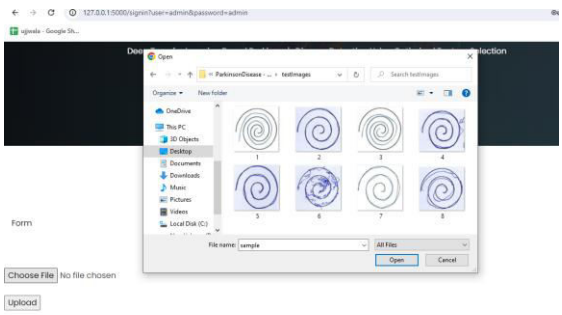


Fig 5 upload input image

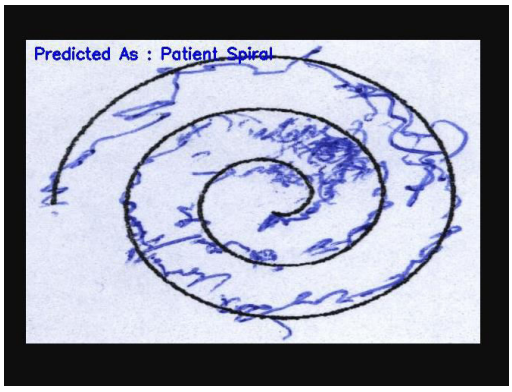


Fig 6 Predicted Result

5. CONCLUSION

In conclusion, the project represents a significant advancement in Parkinson's disease (PD) detection through the integration of deep learning models such as VGG19, InceptionV3, and ResNet50 with traditional approaches like KNN, SVM, and Random Forest. By combining these methodologies, the system achieves superior accuracy in identifying subtle indicators of PD, leading to more precise and timely diagnosis. The detailed features extracted by the deep learning models contribute to the system's enhanced performance, particularly when utilized in conjunction with Random Forest. Moreover, the user-friendly front-end streamlines the input, preprocessing, and prediction processes, making the system accessible and practical for users seeking early PD detection. Overall, the project offers enhanced detection capabilities, underscoring its potential as a valuable tool for healthcare professionals and individuals alike in the quest for accurate and timely diagnosis of Parkinson's disease.

6. FUTURE SCOPE

The future scope of this study encompasses a comprehensive exploration of feature selection techniques aimed at enhancing the accuracy and efficiency of Parkinson's disease (PD) detection. This includes leveraging deep transfer learning methods, such as fine-tuning pre-trained neural networks like VGG19, InceptionV3, and ResNet50, to extract high-level features from relevant datasets. Additionally, the project investigates optimized feature selection strategies to identify the most discriminative and informative features for PD classification. These techniques may involve advanced algorithms such as genetic algorithms, recursive feature elimination, or hybrid approaches that combine multiple selection methods. By focusing on feature selection, the project aims to

streamline the input data, reduce dimensionality, and improve the generalization performance of the classification models. Ultimately, the scope of feature selection in this project encompasses refining and optimizing the input features to enhance the accuracy and robustness of PD detection using deep transfer learning techniques.

REFERENCES

[1] B. R. Bloem, M. S. Okun, and C. Klein, Parkinsons disease, Lancet, vol. 397, no. 10291, pp. 22842303, 2021.

[2] H. Li, C.-M. Pun, F. Xu, L. Pan, R. Zong, H. Gao, and H. Lu, A hybrid feature selection algorithm based on a discrete arti cial bee colony for Parkinsons diagnosis, ACM Trans. Internet Technol., vol. 21, no. 3, pp. 122, Aug. 2021.

[3] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H. J. Bae, and N. Kim, Deep learning in medical imaging, Neurospine, vol. 16, no. 4, p. 657, 2019.

[4] M. Bakator and D. Radosav, Deep learning and medical diagnosis: Areview of literature, Multimodal Technol. Interact., vol. 2, no. 3, p. 47, Aug. 2018.

[5] Z. Fang, Improved KNN algorithm with information entropy for the diagnosis of Parkinsons disease, in Proc. Int. Conf. Mach. Learn. Knowl. Eng. (MLKE), Feb. 2022, pp. 98101.

[6] E. Kaplan, E. Altunisik, Y. E. Firat, P. D. Barua, S. Dogan, M. Baygin, F. B. Demir, T. Tuncer, E. Palmer, R.-S. Tan, P. Yu, J. Soar, H. Fujita, and U. R. Acharya, Novel nested patch-based feature extraction model for automated Parkinsons disease symptom classi cation using MRI images, Comput. Methods Programs Biomed., vol. 224, Sep. 2022, Art. no. 107030.

[7] M. Gazda, M. Hires, and P. Drotar, Ensemble of convolutional neural networks for Parkinsons disease diagnosis from of ine handwriting, Dept. Comput. Inform., Intell. Inf. Syst. Lab, Tech. Univ. Kosice, Koice, Slovakia, Tech. Rep. 9, 2022.

[8] M. Mohaghegh and J. Gascon, Identifying Parkinsons disease using multimodal approach and deep learning, in Proc. 6th Int. Conf. Innov. Technol. Intell. Syst. Ind. Appl. (CITISIA), Nov. 2021, pp. 16.

[9] M.Fratello, F. Cordella, G. Albani, G. Veneziano, G. Marano, A. Paf , and A. Pallotti, Classi cation-based screening of Parkinsons disease patients through graph and handwriting signals, Eng. Proc., vol. 11, no. 1, p. 49, 2021.

[10] A. Gold, Understanding the mann-whitney test, J. Property Tax Assessment Admin., vol. 4, no. 3, pp. 5557, 2007.

[11] H. W. Loh, C. P. Ooi, E. Palmer, P. D. Barua, S. Dogan, T. Tuncer, M. Baygin, and U. R. Acharya, GaborPDNet: Gabor transformation and deep neural network for Parkinsons disease detection using EEG signals, Electronics, vol. 10, no. 14, p. 1740, Jul. 2021.

[12] S. Chakraborty, S. Aich, J. Seong-Sim, E. Han, J. Park, and H.-C. Kim, Parkinsons disease detection from spiral and wave drawings using convolutional neural networks: A multistage classi er approach, in Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT), Feb. 2020, pp. 298303.

[13] S. Nõmm, S. Zarembo, K. Medijainen, P. Taba, and A. Toomela, Deep CNNbasedclassi cation of the archimedes spiral drawing tests to support diagnostics of the Parkinsons disease, IFAC-PapersOnLine, vol. 53, no. 5, pp. 260264, 2020.

[14] T. Tuncer, S. Dogan, and U. R. Acharya, Automated detection of Parkinsons disease using minimum average maximum tree and singular value decomposition method with vowels, Biocybernetics Biomed. Eng., vol. 40, no. 1, pp. 211220, Jan. 2020.

[15] A. Das, H. S. Das, A. Choudhury, A. Neog, and S. Mazumdar, Detection of Parkinsons disease from hand-drawn images using deep transfer learning, in Proc. Congr. Intell. Syst. Singapore: Springer, Sep. 2020, pp. 6784.

[16] L. S. Bernardo, A. Quezada, R. Munoz, F. M. Maia, C. R. Pereira, W. Wu, and V. H. C. De Albuquerque, Handwritten pattern recognition for early Parkinsons disease diagnosis, Pattern Recognit. Lett., vol. 125, pp. 7884, Jul. 2019.

[17] A. Johri and A. Tripathi, Parkinson disease detection using deep neural networks, in Proc. 12th Int. Conf. Contemp. Comput. (IC), Aug. 2019, pp. 14.

[18] T. Tuncer and S. Dogan, A novel octopus based Parkinsons disease and gender recognition method using vowels, Appl. Acoust., vol. 155, pp. 7583, Dec. 2019.

[19] P. Khatamino, I. Canturk, and L. Ozyilmaz, A deep learning-CNN based system for medical diagnosis: An application on Parkinsons disease handwriting drawings, in Proc. 6th Int. Conf. Control Eng. Inf. Technol. (CEIT), Oct. 2018, pp. 16.

[20] C.-L. Liu, C.-H. Lee, and P.-M. Lin, A fall detection system using K-nearest neighbor classi er, Exp. Syst. Appl., vol. 37, no. 10, pp. 71747181, Oct. 2010.

[21] S. Katoch, S. S. Chauhan, and V. Kumar, A review on genetic algorithm: Past, present, and future, Multimedia Tools Appl., vol. 80, no. 5, pp. 80918126, 2021.

[22] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, Deep learning-aided Parkinsons disease diagnosis from handwritten dynamics, in Proc. 29th Conf. Graph., Patterns Images (SIBGRAPI), Oct. 2016, pp. 340346.

[23] C. R. Pereira, D. R. Pereira, F. A. Silva, J. P. Masieiro, S. A. Weber, C. Hook,andJ.P.Papa, Anewcomputervision-basedapproachtoaidthe diagnosis of Parkinsons disease, Comput. Methods Programs Biomed., vol. 136, pp. 7988, Nov. 2016.

[24] R. Ribani and M. Marengoni, A survey of transfer learning for convolutional neural networks, in Proc. 32nd SIBGRAPI Conf. Graph., Patterns Images Tuts. (SIBGRAPI-T), 2019, pp. 4757.

[25] S. Xu and Z. Pan, A novel ensemble of random forest for assisting diagnosis of Parkinsons disease on small handwritten dynamics dataset, Int. J. Med. Informat., vol. 144, Dec. 2020, Art. no. 104283.

[26] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, Assessing visual attributes of handwriting for prediction of neurological disorders A case study onParkinsons disease, Pattern Recognit. Lett., vol. 121, pp. 1927, Apr. 2019.

[27] A.Parziale, C. A. Della, R. Senatore, and A. Marcelli, A decision tree for automatic diagnosis of Parkinsons disease from of ine drawing samples: Experimentsandndings,inProc.Int.Conf.ImageAnal. Process.Cham, Switzerland: Springer, 2019, pp. 196206.

[28] J. P. Folador, M. C. S. Santos, L. M. D. Luiz, L. A. P. De Souza, M. F. Vieira, A. A. Pereira, and A. De Oliveira Andrade, On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people

with Parkinsons disease, Med. Biol. Eng. Comput., vol. 59, no. 1, pp. 195214, Jan. 2021.

[29] L. Parisi, D. Neagu, R. Ma, and F. Campean, Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinsons disease and COVID-19, Exp. Syst. Appl., vol. 187, Jan. 2022, Art. no. 115892.

Datasetlink:
https://www.kaggle.com/datasets/banilkumar20phd7071/handwritten-parkinsons-disease-augmented-data