A Hybrid Deep Learning Approach for Early Parkinson's Detection from Handwriting

C Manoj kumar[1], Gudivada Lokesh[2], V Sridhar[3]
[1]*PG Scholar,* [2]*Associate Professor,* [3]*Assistant Professor*
Dept. of Computer Science and Engineering,
VEMU Institute of Technology, P. Kothakota, Chittoor, Andhra Pradesh, India.

Abstract - The Hybrid Deep Learning Approach study presents an enhanced Parkinson's Disease (PD) detection model using deep transfer learning and genetic algorithm-based feature optimization. Unlike traditional approaches relying on error-prone handcrafted features, the proposed method employs VGG19, InceptionV3, and ResNet50 to extract robust features from NEWHANDPD spiral handwriting images. These features are further refined using a genetic algorithm to improve accuracy. K-Nearest Neighbour (KNN) is then used for classification. While Support Vector Machine (SVM) showed limited accuracy, KNN outperformed significantly. As an extension, hyperparameter tuning was applied to KNN, achieving improved accuracy of 96–98%, compared to the standard KNN's 92–95%. This hybrid and optimized approach enhances early PD detection, supporting clinical decision-making with higher reliability.

Key Words: Disease, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Parkinson's Disease, Deep Learning, Convolutional Neural Networks (CNNs).

1. INTRODUCTION

Parkinson's Disease (PD) is a progressive and incurable neurological disorder primarily caused by the depletion of dopamine—a neurotransmitter essential for regulating movement and coordination. As dopamine-producing cells in the brain's basal ganglia deteriorate, individuals begin to experience symptoms such as tremors, muscle stiffness, speech difficulties, and postural instability. Early signs like finger tremors often affect handwriting, leading to "micrographia"—a condition where writing becomes small and cramped. This subtle change can serve as a key indicator for early detection. However, diagnosing PD in its initial stages remains challenging due to the lack of definitive clinical tests. With the evolution of artificial intelligence, particularly deep learning, the potential for early and accurate diagnosis has significantly improved. Convolutional Neural Networks (CNNs) and transfer learning methods now enable the automated analysis of medical data, including handwriting patterns. These approaches have shown exceptional accuracy and promise in identifying PD, offering hope for better symptom management and timely intervention. In our approach, we tried to select a limited number of features using domain knowledge about Parkinson's disease and applied the method of machine learning used in competitive works in order to emphasize the main novelty of this work.

2. LITERATURE SURVEY

Fang (2022) - Improved K-Nearest Neighbors (KNN) with Entropy: Fang proposed an enhanced version of the KNN algorithm by incorporating entropy-based weighting to improve classification accuracy for PD detection. Using the UCI dataset, the framework emphasized the role of information entropy in distinguishing relevant features from noise, improving decision boundaries. The method utilized a 5-fold cross-validation to validate the model's reliability. This framework advanced the understanding of feature importance in machine learning, highlighting entropy as a vital factor for enhancing classification precision[1-5].

Kaplan et al. (2022) - Handcrafted Feature Engineering with Integrated Multiview (IMV) Learning: .Kaplan and colleagues explored PD detection through MRI scan analysis using handcrafted features and multiple feature selectors. The framework adopted patch-based learning combined with IMV analysis to classify PD symptoms, including clinical stages, dementia status, and motor impairment. This hybrid approach demonstrated the importance of integrating clinical knowledge with machine learning techniques to improve the accuracy and robustness of medical diagnostic tools[6].

Gazda et al. (2022) - Ensemble of Convolutional Neural Networks (CNNs): Gazda et al. implemented an ensemble learning framework using five CNN models for detecting PD through handwriting analysis, utilizing the PaHaW and NewHandPD datasets. This theoretical model leveraged transfer learning to enhance generalization and reduce overfitting. By combining multiple CNN architectures, the ensemble approach aimed to increase diagnostic accuracy and stability across different handwriting tasks, demonstrating the benefits of model diversity[7].

Mohaghegh and Gascon (2021) - Vision Transformer (ViT) for Handwritten Data: This framework introduced a Vision Transformer model utilizing spiral and meander drawings for PD detection. The ViT architecture incorporated a multi-layer perceptron classifier and was pre-trained using self-supervised learning techniques (DINO) on ImageNet. The entropy-based dropout layers improved generalization by reducing overfitting. This study illustrated the effectiveness of transformer models in medical image classification, offering a new avenue for neurological disorder detection[8-12].

Summary Table

Author	Year	Method	Dataset	Positives	Negatives
Fang	2022	Improved KNN (entropy)	UCI Dataset	Improved accuracy with entropy weighting	Traditional methods showed lower accuracy
Kaplan et al.	2022	Handcrafted Feature Engineering & IMV	MRI Scans	Outstanding performance in PD stage detection	Limited generalization due to specific dataset
Gazda et al.	2022	Ensemble of CNNs	PaHaW & NewHandPD	Competitive results with ensemble model	Computational cost of ensemble models
Mohaghegh and Gascon	2021	Vision Transformer (ViT)	Spiral and Meander Handwriting Data	High accuracy with vision transformers	Complexity in model implementation
Fratello et al.	2021	Linear SVM, Medium KNN	Handwriting Data	Moderate accuracy for different classifiers	Limited dataset size

3. PROBLEM STATEMENT

Existing methods for Parkinson's disease detection lack accuracy due to ineffective feature extraction techniques. Addressing this, our project aims to develop a robust deep learning model for precise early diagnosis.Traditional Parkinson's Disease detection methods rely heavily on clinical observations and handcrafted features derived from patient data, such as handwriting samples. Machine learning algorithms like Support Vector Machines (SVM) have been applied to identify patterns in handwriting, but their accuracy is often limited due to manual feature extraction, which can be inconsistent and error-prone. These systems require expert intervention to select relevant features, making them less adaptable to diverse datasets. Additionally, the absence of deep learning integration restricts their performance in complex classification tasks, resulting in lower accuracy and delayed diagnosis, especially in early stages of Parkinson's Disease.

4. EXISTING SYSTEM

Parkinson's disease (PD) is one of the chronic neurological diseases whose progression is slow and symptoms have similarities with other diseases. Early detection and diagnosis of PD is crucialto prescribe proper treatment for patient's productive and healthy lives. The disease's symptoms are characterized by tremors, muscle rigidity, slowness in movements, balancing along with other psychiatricsymptoms. The dynamics of hand written records served as one of the dominant mechanisms which support PD detection and assessment. Several machine learning methods have been investigated for the earlydetection of this disease. But most of these handcrafted feature extraction techniques predominantly sufferfrom low performance accuracy issues. This cannot be tolerable for dealing with detection of such a chronicailment.

Disadvantages:

- Limited handwriting sample availability.
- Challenges in standardized data collection.

- Dependency on manual data.
- Insufficient training data variability

5. PROPOSED SYSTEM

In propose work this system leverages deep transfer learning and genetic algorithms to enhance the accuracy of Parkinson's Disease detection using spiral handwriting images. Instead of relying on handcrafted features, it uses pre-trained CNN models like VGG19, InceptionV3, and ResNet50 to extract high-level image features. These features are then optimized through a Genetic Algorithm, selecting only the most relevant ones for classification. Finally, a K-Nearest Neighbour (KNN) classifier is used to predict outcomes. As an extension, hyperparameter tuning is applied to KNN, further boosting its accuracy to 96–98%. This hybrid method offers a powerful, automated, and scalable solution for early PD detection.

In propose work to improve detection rate author utilizing many machine and deep learning algorithms which are explained in below points

Deep Transfer Learning Algorithms: Now-a-days deep CNN algorithm has proven its accuracy in almost all fields for accurate features extraction and classification. All existing techniques were dependent on Hand Crafted features obtained from human expertise but this hand crafted features are error prone so author employing VGG19, IncpetionV3 and Resnet50 deep learning algorithms for features extraction. This algorithm get trained on Spiral NEWHAND images dataset and then extract features using trained model.

Genetic Algorithm for features optimization: Extracted features from deep model will be input to Genetic Algorithm which will analyse each features using various iterations such as Population, mutation crossover and selection. While iterating GA will select features which are highly accurate in enhancing prediction and then output selected accurate features as optimized features.

KNN Classification: KNN is one of simplest classifier which can enhance its prediction accuracy if input features are more optimized and accurate. So author employing KNN to predict Parkinson disease from optimized features.

Existing SVM Algorithm: as existing author has used SVM whose accuracy is not much as good as KNN .The SVM (Support Vector Machine) algorithm is a supervised learning algorithm used for both classification and regression tasks. It finds the optimal hyperplane that separates data points into different classes, maximizing the margin between them. SVMs are known for their robustness and ability to handle both linear and non-linear data.

Advantages:

- Enhanced early PD detection accuracy.
- Efficient feature optimization process.
- Improved diagnostic ability.
- Efficient utilization of deep learning.

6. SYSTEM ARCHITECTURE

Most of the people in the world are affected by the disease. So, to find a better solution to cure the disease and predict the disease in the initial stage the machine learning algorithm is carried out. We can spot the unwellness of the people and we analyse the body parameters. The medical data are uploaded in the database. The data can be gathered and analysed and formulated using the Supervised Machine Learning Technique. This method can provide an effective outcome to treat the disease. The simulation has been carried which is tested by the neural region which is shown in below mentioned diagram is an outline of the proposed block diagram and work flow.

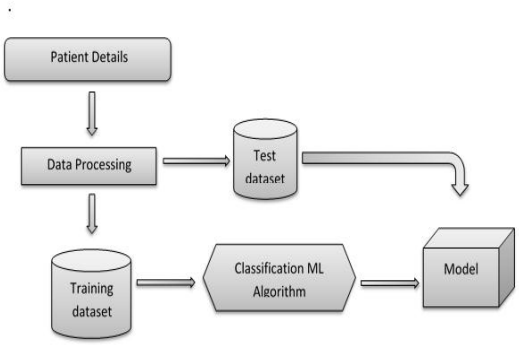


Fig -1: System Architecture

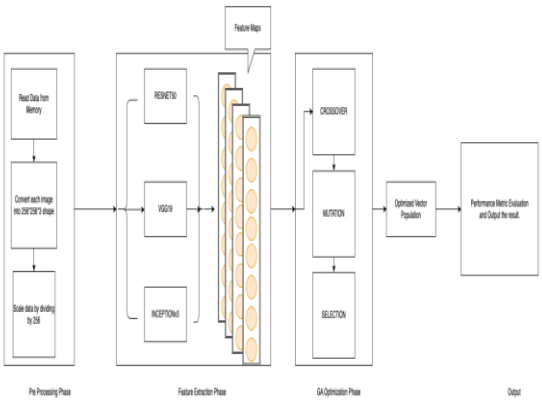


Fig -2: Work Flow

7. METHODOLOGY

Dataset Collection and Preprocessing

The NEWHANDPD dataset is used for this study, which contains spiral handwriting images from both healthy individuals and Parkinson's patients. These images are categorized into two folders—healthy and affected. Each image is processed to ensure uniformity in format by resizing, normalizing, and converting them into RGB format suitable for deep learning input. Data is also shuffled to eliminate any bias in the training process. Image preprocessing ensures noise reduction and improves the performance of the models.

Feature Extraction Using Deep Transfer Learning

Traditional handcrafted features often lack robustness and are prone to human error. Therefore, this work leverages powerful Convolutional Neural Networks (CNNs) through transfer learning. Pre-trained models such as VGG19, InceptionV3, and ResNet50 are utilized to extract deep features from the handwriting images. These models are originally trained on large-scale image datasets and can detect high-level image patterns that are not visible to the human eye. The last few layers of these networks are fine-tuned to adapt them to the specific task of Parkinson's detection.

Feature Optimization Using Genetic Algorithm

Although the deep learning models extract numerous features, not all of them contribute equally to classification accuracy. To address this, a Genetic Algorithm (GA) is applied to select the most relevant features. GA mimics natural selection processes such as crossover, mutation, and selection to evolve the feature set over generations. Features that yield higher classification accuracy are retained, while less significant ones are discarded. This optimization step not only reduces computational complexity but also improves the model's generalization.

Classification Using K-Nearest Neighbour (KNN)

The optimized feature set is then fed into the K-Nearest Neighbour (KNN) classifier. KNN is chosen for its simplicity and effectiveness in handling feature-based classification tasks. It predicts the class of a new sample based on the majority class of its 'k' nearest neighbors in the feature space. The accuracy of KNN improves significantly when fed with high-quality, optimized features.

Hyperparameter Tuning

To further enhance the performance of the KNN classifier, hyperparameter tuning is introduced. Parameters such as the value of 'k', distance metrics, and weighting functions are systematically varied to find the best combination. After tuning, the accuracy improves from an initial 92–95% to a range of 96–98%, showing a clear performance boost.

Evaluation Metrics

The model's effectiveness is evaluated using standard metrics like accuracy, precision, recall, F1-score, and ROC curves. Confusion matrices are also generated to visualize the classification results. Standard metrics like accuracy, precision, recall, F1-score, and ROC curves.

8. MODULES

1. Dataset Loading and Preprocessing

Dataset comprises images of spiral drawings collected from both healthy individuals and Parkinson's patients. The steps involved in dataset processing are as follows:

- Loading Dataset: Images are loaded from the provided dataset directory. Each image is resized to a standard size of 80x80 pixels and converted to a three-channel RGB format.
- Data Normalization and Shuffling: Image features are normalized to ensure uniformity in data distribution. The dataset is shuffled to randomize the order of images.

2. Deep Transfer Learning

Deep transfer learning is employed to extract high-level features from the spiral images. We utilize pre-trained convolutional neural network (CNN) models such as VGG19, ResNet50, and InceptionV3 for feature extraction. The steps include:

- Model Training: The selected CNN models are trained on the dataset to learn discriminative features from the spiral images.
- Feature Extraction: Features are extracted from the last convolutional layer of each CNN model. These features serve as representations of the input images.

3. Feature Optimization using Genetic Algorithm (GA)

To improve the efficiency of the feature set, a genetic algorithm is employed for feature selection. The key steps involve:

- Random Forest Classification: A Random Forest classifier is trained on the extracted features.
- Genetic Algorithm Optimization: The genetic algorithm optimizes feature selection by iteratively evolving a population of feature

subsets. Features that contribute most to classification performance are selected.

4. Model Training and Evaluation
Various machine learning algorithms are trained and evaluated using the optimized feature set. The algorithms include:

- Support Vector Machine (SVM): Trained using linear kernel to classify spiral images into healthy and Parkinson's classes.
- K-Nearest Neighbors (KNN): Initially trained on the deep model extracted features. Further extension involves tuning KNN parameters using grid search.

5. Performance Evaluation and Visualization
The performance of each algorithm is evaluated using metrics such as accuracy, precision, recall, and F1-score. Confusion matrices and Receiver Operating Characteristic (ROC) curves are plotted for visual analysis. Additionally, training accuracy and loss graphs are generated for the deep learning model.

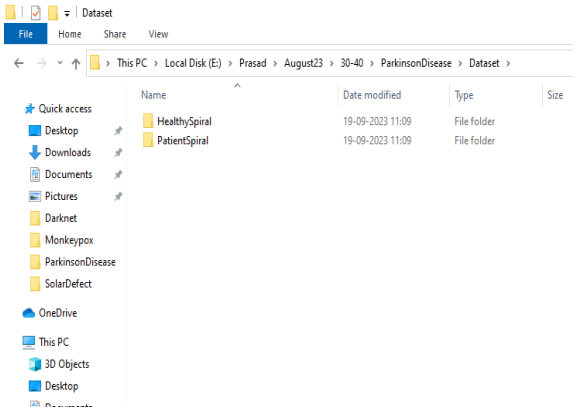
6. Disease Prediction on Test Images
A function is provided to predict Parkinson's disease using the trained model on test images. The function takes the path of the test image as input and displays the predicted class label along with the image.

Deep Transfer Learning Based Parkinson's Disease Detection Using Optimized Feature Selection
Parkinson's disease (PD) is one of the chronic neurological diseases whose progression is slow and symptoms have similarities with other diseases. Early detection and diagnosis of PD is crucial to prescribe proper treatment for patient's productive and healthy lives. The disease's symptoms are characterized by tremors, muscle rigidity, slowness in movements, balancing along with other psychiatric symptoms. The dynamics of handwritten records served as one of the dominant mechanisms which support PD detection and assessment. Several machine learning methods have been investigated for the early detection of this disease. But most of these handcrafted feature extraction techniques predominantly suffer from low performance accuracy issues. This cannot be tolerable for dealing with detection of such a chronic ailment.

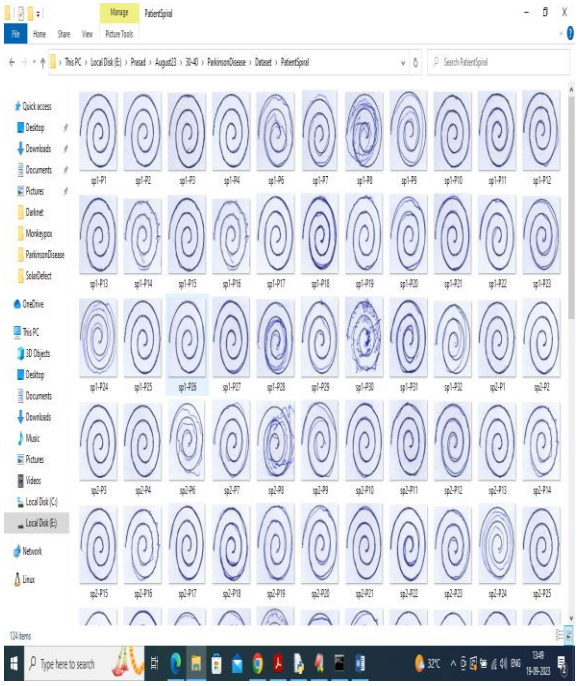
In propose work to improve detection rate author utilizing many machine and deep learning algorithms which are explained in below points

- Deep Transfer Learning Algorithms: Now-a-days deep CNN algorithm has proven its accuracy in almost all fields for accurate features extraction and classification. All existing techniques were dependent on Hand Crafted features obtained from human expertise. but this hand crafted features are error prone. This algorithm get trained on Spiral NEWHAND images dataset and then extract features using trained model
- Genetic Algorithm for features optimization: Extracted features from deep model will be input to Genetic Algorithm which will analyse each features using various iterations such as Population, mutation crossover and selection. While iterating GA will select features which are highly accurate in enhancing prediction and then output selected accurate features as optimized features
- KNN Classification: KNN is one of simplest classifier which can enhance its prediction accuracy if input features are more optimized and accurate. So author employing KNN to predict Parkinson disease from optimized features
- Existing SVM Algorithm: as existing author has used SVM whose accuracy is not much as good as KNN
- Extension Concept: In propose work author has not used any tuning functions to optimized KNN performance further so as extension we have tuned KNN with various Hyper Parameters and after tuning KNN giving accuracy between 96 to 98% where normal KNN accuracy between 92 to 95%..

To train and test each algorithm performance author has used NEWHANDPD dataset which can be download from below link
https://github.com/PaulLerner/deep_parkinson_handwriting#NewHandPD
From above link we have downloaded SPIRAL hand writing images and those image will look like below screen



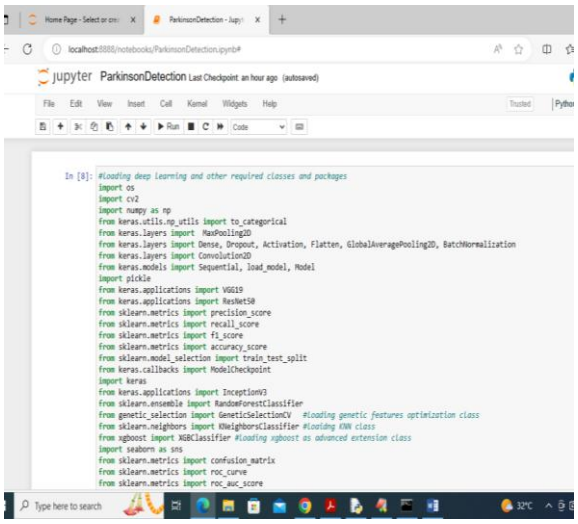
In screen dataset folder contains two folder for healthy and disease patients and just go inside any folder to view images like below screen



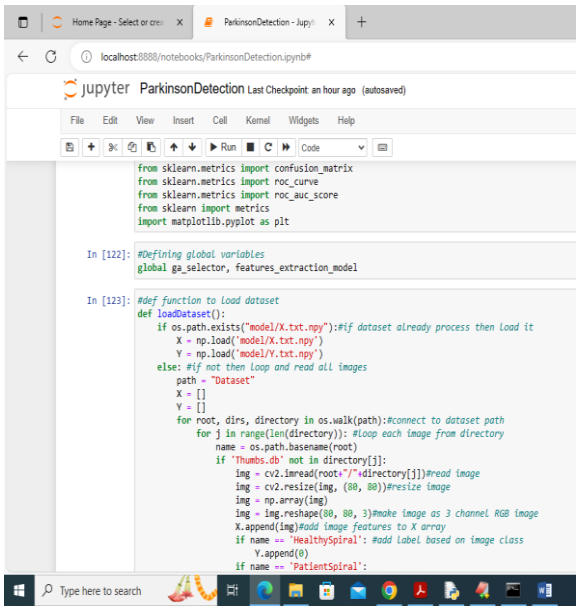
So by using above spiral images will train and test each algorithm. In above images we can see normal patients will have smooth drawing and Parkinson patient will have zigzag or jumbled drawing.

9. RESULTS

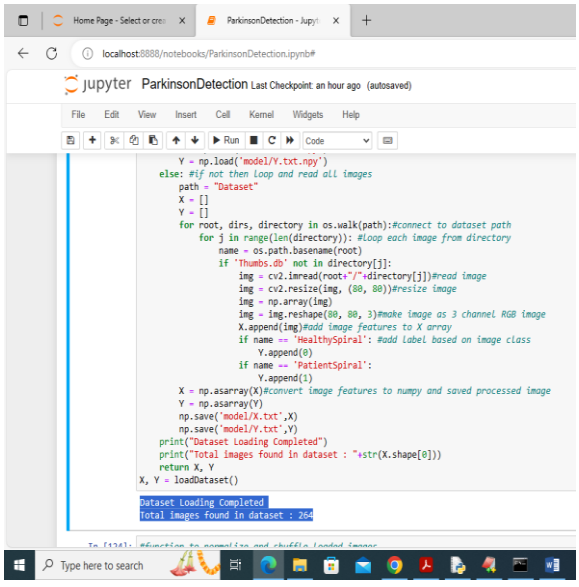
1. We have coded this project using JUPYTER notebook and below are the code and output screens with blue colour comments



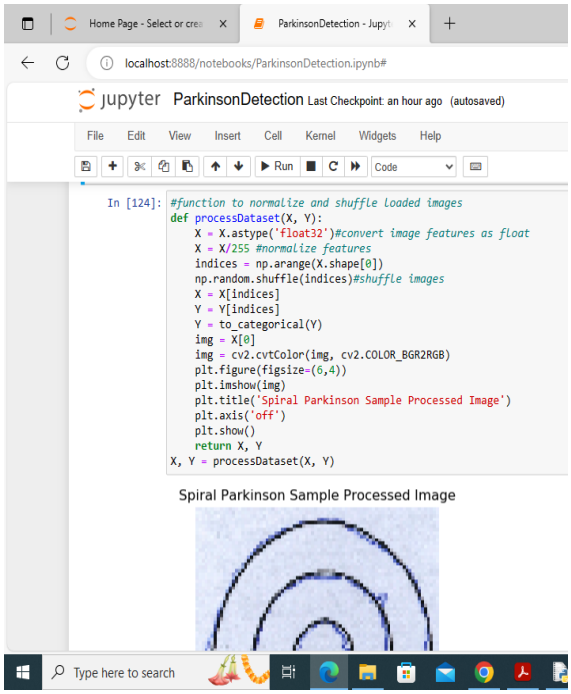
2. In above screen loading all required packages and classes



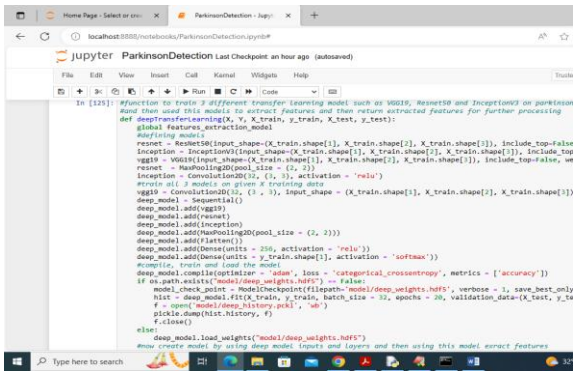
3. In above screen defining function to load images from dataset as 3 channel RGB images



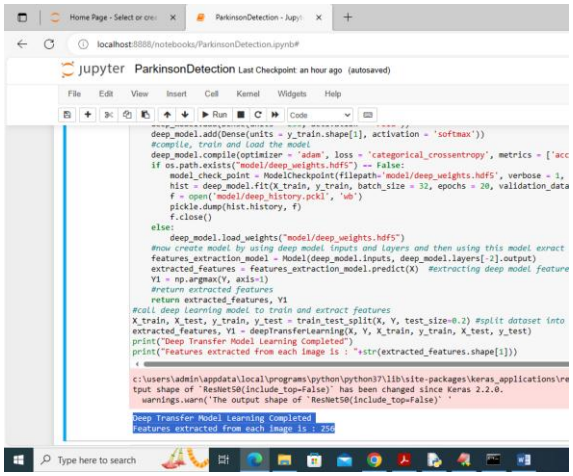
4. In above screen dataset loaded and it contains 264 images from both normal and Parkinson patients



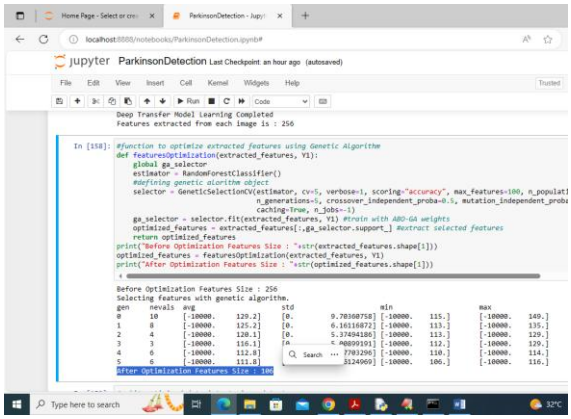
5. In above screen defining function to Pre-Process images such as normalize, shuffle and then displaying processed sample image



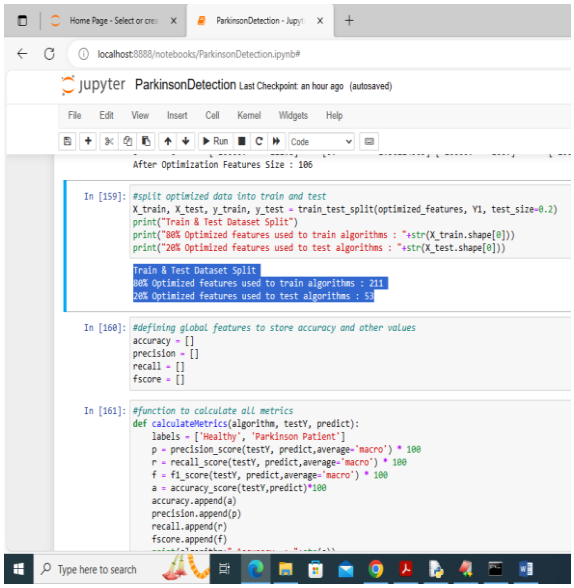
6. In above screen defining transfer learning deep models to trained on processed images and after training will get below screen with extracted number of features



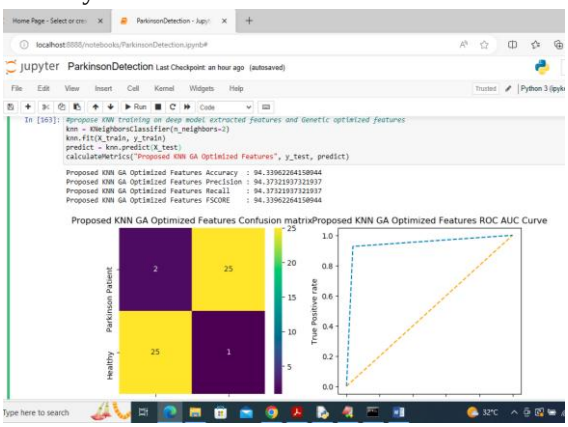
7. In above blue colour text we can see deep transfer learning model extracted 256 features from each image and this features will be input to Genetic Algorithm for further optimizations



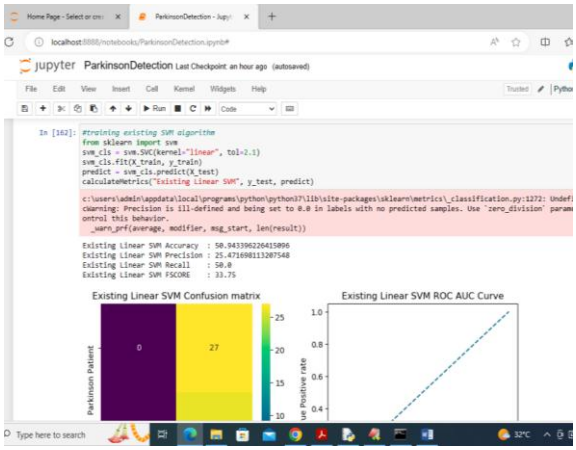
8. In above screen defining Genetic Algorithm function to optimized features and after optimization GA selected 106 features out of 256 which we can see in blue colour selected text



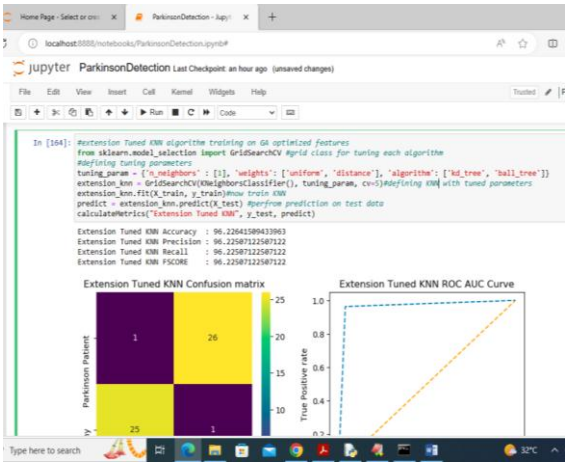
9. In above screen optimized features are splitting into train and test for training with KNN classifier for prediction and then defining function to calculate accuracy and other metrics



10. In above screen training proposed KNN on GA optimized features and after training KNN got 94.33% accuracy and can see precision and other metrics and in confusion matrix graph x-axis represents predicted Labels and y-axis represents true labels and yellow boxes contains correct prediction count and blue boxes contains incorrect prediction count which are very few. In ROC curve graph x-axis represents False Prediction Rate and y-axis represents True Prediction Rate and if blue line comes on top of orange line then all predictions are correct and if blue line goes below orange line then all predictions are incorrect.

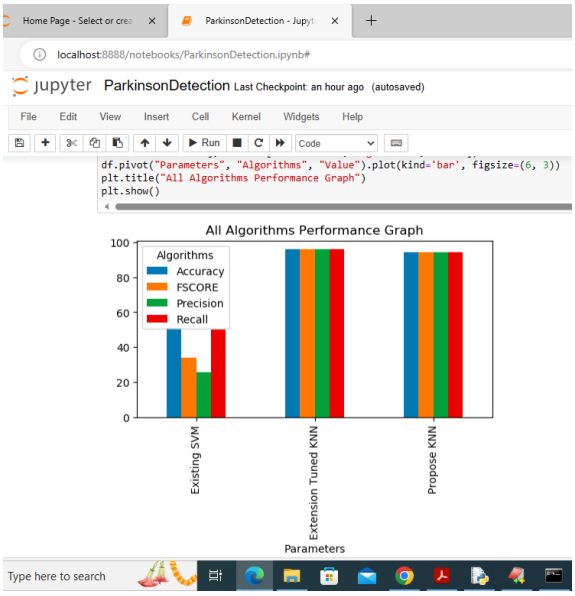


11. In above screen training existing Linear SVM and it got only 50% accuracy

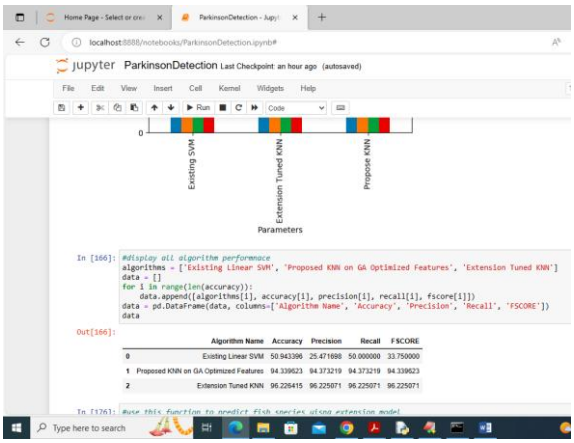


12. In above screen tuning KNN with hyper parameters as extension model and after training tuned Extension KNN got 96.22% accuracy which is higher than other algorithms

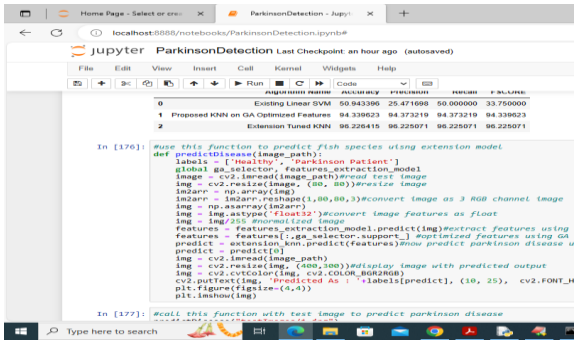
13. Above is the deep model training accuracy and loss graph where x-axis represents Iterations/EPOCH and y-axis represents accuracy and loss values. Green line is for accuracy and blue line for LOSS and with each increasing epoch accuracy got increase and reached closer to 1 and loss get decrease and reached closer to 0.



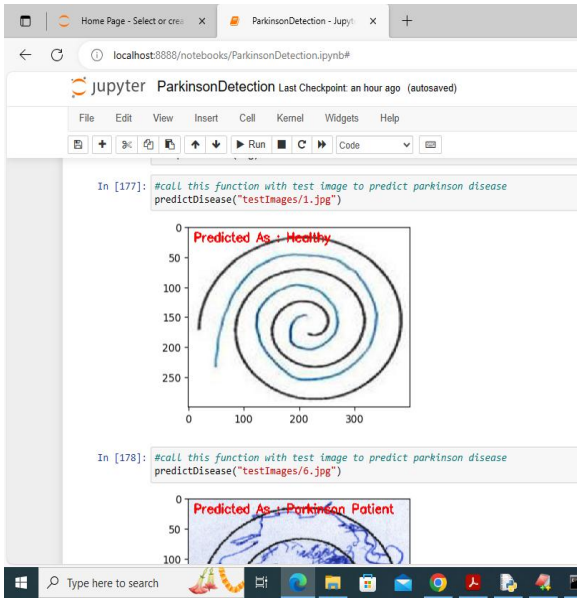
14. In above graph displaying all algorithm performance where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars and in all algorithms extension got high performance



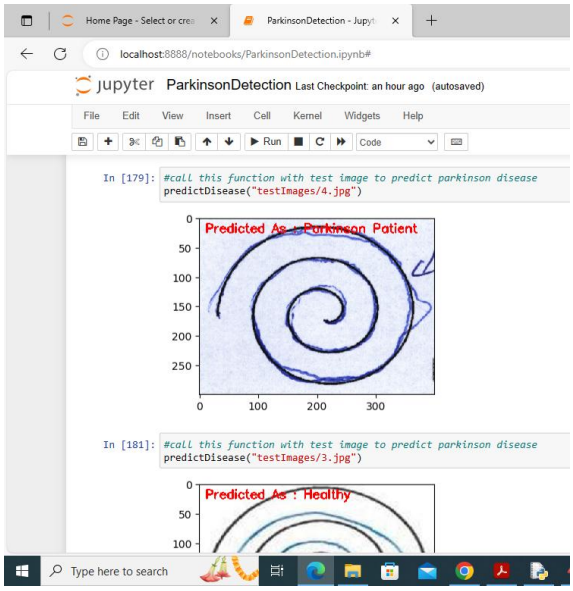
15. In above screen displaying all algorithm performance in tabular format



16. In above screen defining function to predict disease from test image and this function will read input image, process image, extract deep features, optimized features using GA and then predict disease using extension KNN object. All this operations you can see in blue colour comments



17. In above screen calling predict function with test image and then in red colour text we can see detected output as healthy or Parkinson patient



10 . CONCLUSIONS

This study presents an efficient and accurate system for early Parkinson's Disease detection using deep transfer learning and genetic algorithm-based feature optimization. By replacing traditional, error-prone handcrafted feature techniques with pre-trained CNN models like VGG19, InceptionV3, and ResNet50, the system effectively extracts high-quality image features from spiral handwriting data. These features are further refined using a Genetic Algorithm to improve classification performance. The optimized feature set is then classified using K-Nearest Neighbour (KNN), which, after hyperparameter tuning, achieves an impressive accuracy of up to 98%. Overall, the proposed method demonstrates a reliable and scalable approach that can support medical professionals in the early diagnosis of Parkinson's Disease, potentially leading to better treatment outcomes and quality of life.

11. FUTURE ENHANCEMENT

In propose work author not used any tuning functions to optimized KNN performance further so as extension we have tuned KNN with various Hyper Parameters and after tuning KNN giving accuracy between 96 to 98% where normal KNN accuracy between 92 to 95%.

12. REFERENCES

[1] B. R. Bloem, M. S. Okun, and C. Klein, ''Parkinson's disease,'' Lancet, vol. 397, no. 10291, pp. 2284–2303, 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 artificial bee colony for Parkinson's diagnosis,'' ACM Trans. Internet Technol., vol. 21, no. 3, pp. 1–22, 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: A review 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 Parkinson's disease,'' in Proc. Int. Conf. Mach.

Learn. Knowl. Eng. (MLKE), Feb. 2022, pp. 98–101.
[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 Parkinson’s disease symptom classification 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 Parkinson’s disease diagnosis from offline handwriting,’’ Dept. Comput. Inform., Intell. Inf. Syst. Lab, Tech. Univ. Kosice, Košice, Slovakia, Tech. Rep. 9, 2022.
[8] Mohaghegh and J. Gascon, ‘‘Identifying Parkinson’s disease using multimodal approach and deep learning,’’ in Proc. 6th Int. Conf. Innov. Technol. Intell. Syst. Ind. Appl. (CITISIA), Nov. 2021, pp. 1–6.
[9] M. Fratello, F. Cordella, G. Albani, G. Veneziano, G. Marano, A. Paffi, and A. Pallotti, ‘‘Classification-based screening of Parkinson’s 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. 55–57, 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 Parkinson’s 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, ‘‘Parkinson’s disease detection from spiral and wave drawings using convolutional neural networks: A multistage classifier approach,’’ in Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT), Feb. 2020, pp. 298–303.

BIOGRAPHIES

C Manoj Kumar,
234M1D5802,
Dept. of CSE,VEMU IT College,
P Kothakota, Andhra Pradesh.

Gudivada Lokesh,
Associate Professor,
Dept. of CSE,VEMU IT College,
P Kothakota, Andhra Pradesh.

V Sridhar,
Assistant Professor,
Dept. of CSE,VEMU IT College,
P Kothakota, Andhra Pradesh.