CLASSIFICATION OF PARKINSON DISEASE FROM MRI SAMPLES USING MACHINE LEARNING

[1]Dr. Srinivasu Polinati,[2]Priyanka Galla, [3]Srivalli Choppalli,[4]Pavitra Dasari, [5]Hyndavi Gorle
[1]Associate Professor,[2,3,4,5]Final Year Students,[1,2,3,4,5]Department of Electronics and Communication Engineering, Vignan Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

The project introduces a precise method for detecting Parkinson's disease (PD) using MRI, a non-invasive imaging technique. PD is a debilitating neurodegenerative disorder with no definitive cure, emphasizing the critical need for early diagnosis and intervention to optimize treatment outcomes. While brain MRI can reveal structural changes associated with PD, manual interpretation presents challenges. To address this, machine learning techniques are leveraged to automate PD classification using MRI data, potentially enhancing diagnosis and healthcare management. This study highlights the potential of Principal Component Analysis (PCA) as a feature reduction technique in conjunction with machine learning for early PD diagnosis. By employing PCA, we effectively reduce the dimensionality of brain MRI data, improving computational efficiency and enhancing model performance. The collected datasets are utilized to evaluate the proposed method, demonstrating its accuracy and effectiveness in distinguishing PD patients from healthy controls. This research underscores the role of machine learning, particularly when coupled with PCA, in facilitating early PD detection and improving patient outcomes.

Keywords: Parkinson's disease (PD), MRI, machine learning, PCA, classification, accuracy, metrics, diagnosis.

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder [1]. With aging populations, its prevalence is expected to increase. The motor deficits of PD are caused by degeneration of the dopamine-producing (dopaminergic) neurons in the substantia nigra region of the brain. Pathological changes are present in other neuronal populations as well, explaining the development of various non-motor impairments in PD [2]. Most diagnoses are based on clinical detection of motor signs—the presence of two or more of tremor, rigidity, bradykinesia, or postural impairment [3]. Confirmatory evidence can be provided by dopamine transporter scanning, though this test is not widely available across the world. Other medical imaging modalities lack sensitivity. There is a need for biomarkers that can, with high reliability, recognize PD before overt motor signs appear. Parkinson's disease (PD) progresses through stages, though it's important to note that the progression can vary significantly among individuals. The commonly recognized staging system used by professionals is known as the Hoehn and Yahr scale. Stage 1: Symptoms are mild and typically unilateral (affecting one side of the body). Tremors or other movement issues might be present but generally don't interfere with daily activities. Stage 2: Symptoms worsen, affecting both sides of the body. Balance problems might arise, but individuals can still live independently[4,5].Stage 3: Considered mid-stage PD. Balance and coordination worsen, making daily activities more challenging. Falls become more common, but individuals are still capable of living independently with assistance. Stage 4: Symptoms are severe and debilitating. Individuals may require assistance to perform daily activities, and walking may be impaired. Stage 5: This is the most advanced stage, where individuals are typically wheelchair-bound or bedridden. Round-the-clock care and assistance are usually necessary[6].

In human beings, MRI plays a pivotal role in understanding PD's underlying brain changes, aiding diagnosis, disease tracking, and research advancements. While not a treatment per se, MRI's value lies in its ability to provide detailed brain images, allowing clinicians to accurately diagnose PD by detecting specific structural brain alterations associated with the condition. Additionally, MRI helps monitor disease progression, enabling healthcare professionals to assess treatment effectiveness and make informed adjustments. Furthermore, in research, MRI contributes significantly by unravelling PD's complex neurobiology, potentially leading to the discovery of new therapeutic targets or approaches[7]. While MRI itself doesn't directly treat PD, its indispensable role in guiding diagnosis, monitoring, and research contributes substantially to refining treatment strategies and improving overall patient care and outcomes.

LITERATURESURVEY

The research paper titled "Sensor Fusion for identification of Freezing Gait Episodes Using Machine Learning Techniques" by Syed Aziz Shah et al. The focus of this research seems to be on identifying freezing of gait (FOG), a debilitating symptom of Parkinson's disease that disrupts a person's ability to walk smoothly. The image might showcase specific sections of the paper, highlighting details on the methodologies used, such as the type of machine learning techniques employed or the type of sensors involved in data collection. Perhaps, data visualizations like graphs or charts could also be glimpsed on the screen, providing insights into the findings related to FOG detection. Another research paper titled "Clinical Utility of Opicapone in the Management of Parkinson's Disease: A Short Review on Emerging Data and Place in Therapy" by Linda Azevedo et al. This paper delves into the clinical effectiveness of opicapone, a medication used to manage Parkinson's disease. The image might display the abstract or introduction section, outlining the paper's objective of critically evaluating available data on opicapone's use in treating Parkinson's. It's possible that the screen shows key points discussed in the review, such as opicapone's mechanism of action, its effectiveness compared to other treatments, or potential side effects.

The survey compares two related papers. The first paper is titled "Sensor Fusion for identification of Freezing Gait Episodes Using Machine Learning Techniques" by Syed Aziz Shah et al. [1]. This paper looks at identifying freezing of gait (FOG), a motor impairment that can effect a Parkinson's disease sufferer's ability to complete daily activities (ADLs). The second paper is titled "Clinical Utility of Opicapone in the Management of Parkinson's Disease: A Short Review on Emerging Data and Place in Therapy" by Linda Azevedo et al. [2]. This paper looks at the use of opicapone, a medication used to treat Parkinson's disease.

EXISTINGMETHOD

Existing deep learning approaches for Parkinson's disease detection, like those utilizing MRI samples, often rely on manually engineered features. This method requires domain expertise to identify and extract relevant characteristics from the data, which can be a time-consuming process. Additionally, these approaches might utilize raw, unprocessed data, leading to high dimensionality and potentially increasing model complexity. This can hinder performance and training efficiency.

PROPOSEDMETHOD

We propose a novel approach for Parkinson's disease detection using a Convolutional Neural Network (CNN) coupled with Principal Component Analysis (PCA), hereafter referred to as CNN-PCA. This method leverages the strengths of both techniques to achieve superior classification performance.The proposed CNN-PCA method follows a multi-stage processing pipeline. Initially, the raw data undergoes preprocessing steps to ensure consistency and compatibility for analysis. This involve techniques like normalization. Subsequently, the preprocessed data is fed into the CNN architecture. CNNs are particularly adept at extracting relevant features from complex data, such as images used in Parkinson's disease detection. In this context, the CNN is tasked with automatically learning discriminative features that potentially capture crucial characteristics associated with Parkinson's disease.

However, a large number of extracted features can introduce complexity and potentially hinder model performance. To address this challenge, the feature set is subjected to dimensionality reduction using PCA[8]. PCA acts as a data compression technique, effectively eliminating redundant features while preserving the most informative ones. This not only streamlines the data but also reduces the computational burden on the subsequent classification stage. Finally, the reduced feature set is utilized to train a classifier, enabling it to distinguish between Parkinson's disease and healthy control cases.The core strength of CNN-PCA lies in the complementary nature of its components. CNNs eliminate the need for manual feature engineering by automatically learning informative features directly from the data. PCA, on the other hand, effectively reduces data complexity while retaining crucial information, potentially leading to a more robust and accurate classification model. This combined approach offers promising potential for applications like Parkinson's disease detection, as evidenced by its superior performance compared to other benchmark models evaluated in this study.

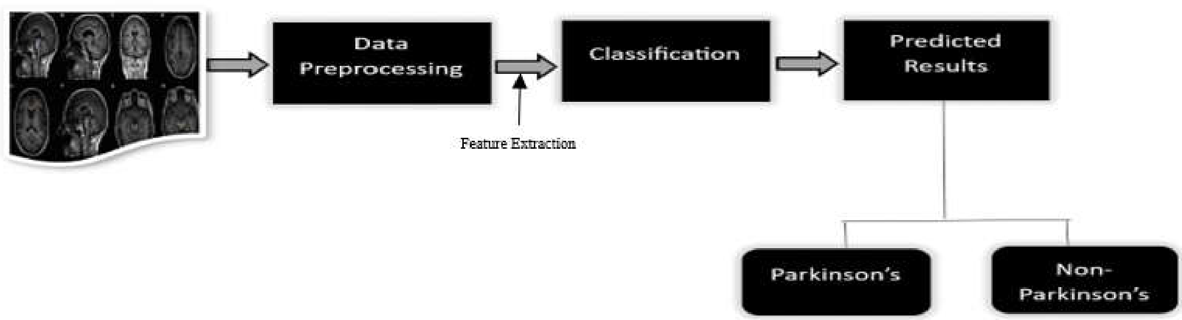

Fig.1: Block diagram of Proposed method

Datasets

For Parkinson's disease any type of datasets can be considered. We considered only type i.e., MRI images. We had collected the dataset from Kaggle website. The dataset is a record of 600 healthy and PD affected patients which is of image forms.

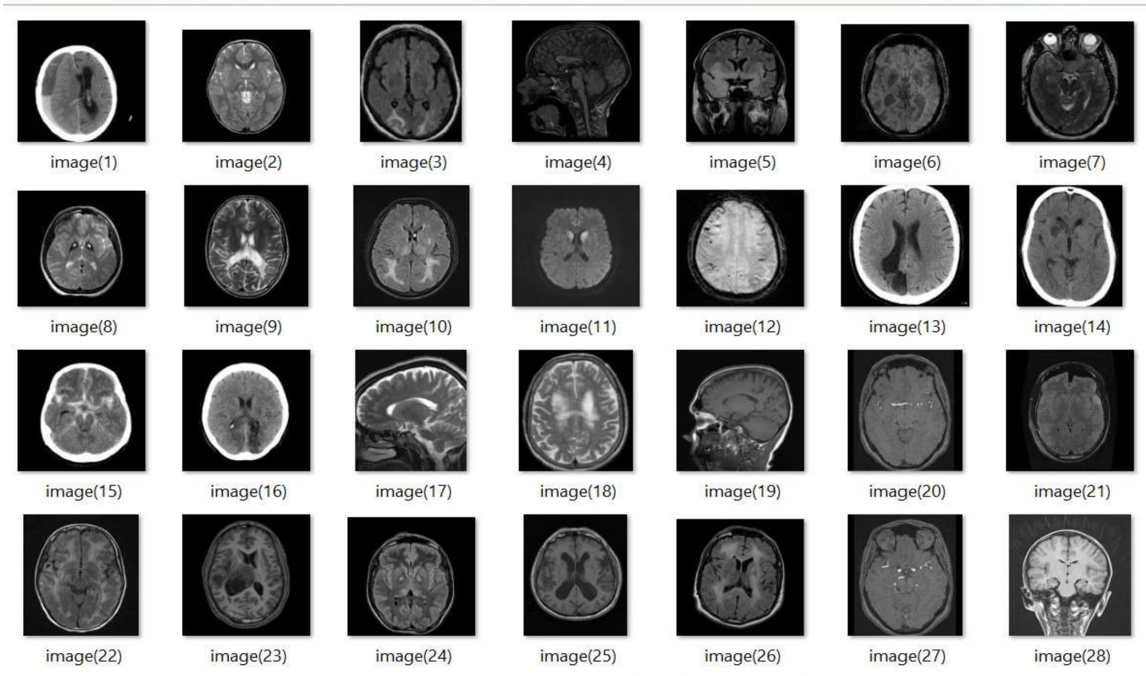

Fig.2: Health brain MRI Samples

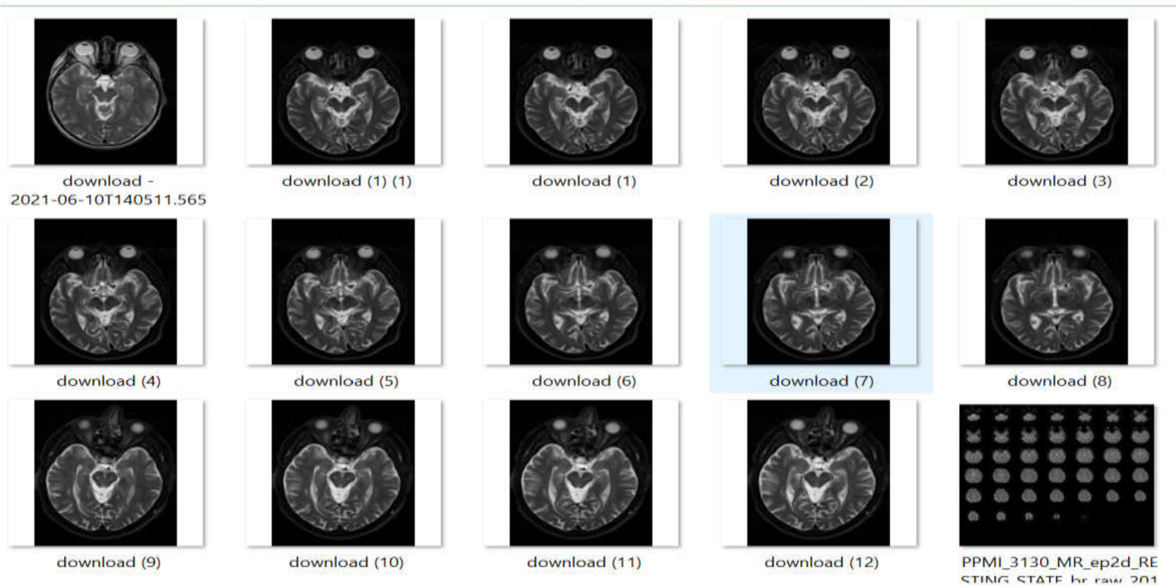

Fig.3: Parkinson disease brain MRI Samples

PRELIMINARIES

The Parkinson's disease is progressive neuro degenerative disorder that affects a lot only people significantly affecting their quality of life. It mostly affects the motor functions of human. The main motor symptoms are called "parkinsonism" or "parkinsonian syndrome". The symptoms of Parkinson's disease will occur slowly, the symptoms include shaking, rigidity, slowness of movement and difficulty with walking, Thinking and behavior change, Depression and anxiety are also common. There is a model for detecting Parkinson's using voice. The deflections in the voice will confirm the symptoms of Parkinson's disease. This project showed 73.8% efficiency. Confusion matrix is used to describe how well a classification system performs. A confusion matrix shows a classification algorithm's performance[9].

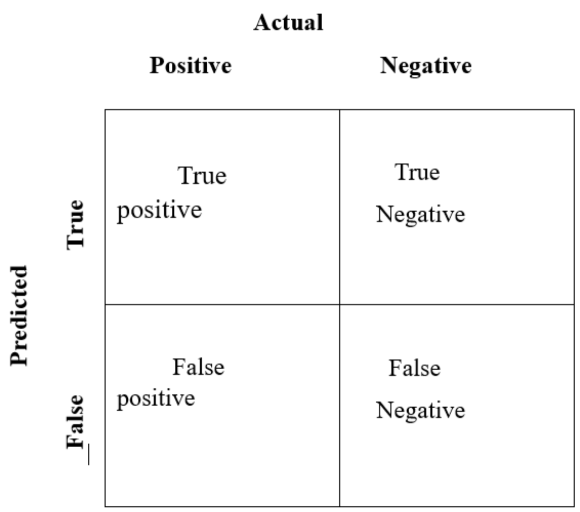

Fig.4: Representation of Confusion Matrix

- TP (True positive): The predicted value is positive and its actual value is positive.
- FP (False positive): Predicted is false but it’s actual positive. It’s known as type-1error.
- TN (True negative): Predicted is true but it’s actual negative. It’s known as type-2 error.
- FN (False negative): Predicted value is false and it’s actual negative.

Accuracy:

Accuracy is defined as the closeness of the measurements in relation to a given value is referred to as accuracy. Accuracy is also a term used to describe systematic error. Moreover, accuracy provides us with a measurement of statistical bias.

$$Accuracy = \frac{t_p+t_n}{t_P+t_n+f_P+f_n} \quad \ldots\ldots.(1)$$

Precision:

Also called as positive value is the bit of applicable cases. The perfection is intimately the capability of the classifier not to label a sample as positive if it's negative. High the value, better the classifier.

$$precision = \frac{t_p}{t_p+f_p} \quad \ldots\ldots..(2)$$

Recall (Sensitivity):

It represents the percentage of all relevant instances that were actually retrieved. The classifier's capacity to locate all the positive samples is known as recall. The accuracy increases as the value rises. positive samples. Higher the value higher is the accuracy.

$$recall = \frac{t_p}{t_p+f_n} \quad \ldots\ldots\ldots(3)$$

F1-Score:

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

$$\text{F1 Score} = \frac{2* precision * recall}{precision\ + recall} \quad \ldots\ldots\ldots.(4)$$

RESULTANALYSIS

An analysis of model performance reveals that CNN-PCA achieved the best results across all metrics (precision, recall, F1 score, AUC, and accuracy). MobileNet followed closely with an accuracy of 88.96%. Meanwhile, DenseNet121, VGG16, and ResNet50 exhibited identical accuracy of 61.36%. Inception V3 displayed the lowest performance with an accuracy of 50%.In essence, CNN-PCA stands out for its ability to accurately identify positive instances while minimizing false positives.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_3 (Dense)             (None, 512)               26112

 dropout_2 (Dropout)         (None, 512)               0

 dense_4 (Dense)             (None, 256)               131328

 dropout_3 (Dropout)         (None, 256)               0

 dense_5 (Dense)             (None, 2)                 514

=================================================================
Total params: 157954 (617.01 KB)
Trainable params: 157954 (617.01 KB)
Non-trainable params: 0 (0.00 Byte)
```

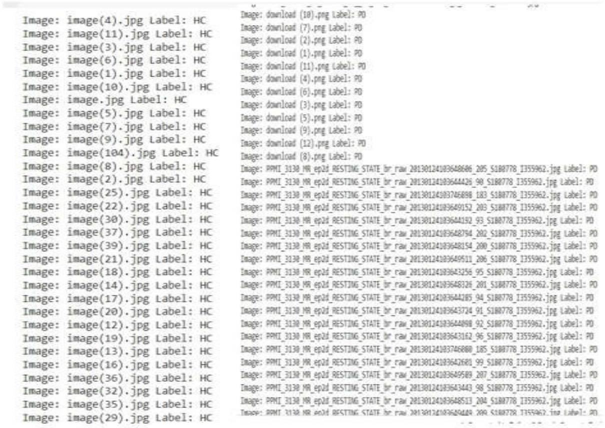

Fig.5: Architecture of CNN**Fig.6:** Classification of MRI samples

```
print(labels)

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1]
```

```
Epoch 1/25
8/8 [==============================] - 0s 22ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 2/25
8/8 [==============================] - 0s 16ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 1.4311 - val_accuracy: 0.9839
Epoch 3/25
8/8 [==============================] - 0s 17ms/step - loss: 2.4702 - accuracy: 0.9959 - val_loss: 10.1649 - val_accuracy: 0.9839
Epoch 4/25
8/8 [==============================] - 0s 15ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 43.5877 - val_accuracy: 0.9839
Epoch 5/25
8/8 [==============================] - 0s 13ms/step - loss: 9.1886 - accuracy: 0.9959 - val_loss: 56.3514 - val_accuracy: 0.9839
Epoch 6/25
8/8 [==============================] - 0s 12ms/step - loss: 0.8066 - accuracy: 0.9959 - val_loss: 41.3691 - val_accuracy: 0.9839
Epoch 7/25
8/8 [==============================] - 0s 13ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 27.3271 - val_accuracy: 0.9839
Epoch 8/25
8/8 [==============================] - 0s 13ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 21.1833 - val_accuracy: 0.9839
Epoch 9/25
8/8 [==============================] - 0s 12ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 18.4300 - val_accuracy: 0.9839
Epoch 10/25
8/8 [==============================] - 0s 17ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 17.2262 - val_accuracy: 0.9839
Epoch 11/25
8/8 [==============================] - 0s 17ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 16.6992 - val_accuracy: 0.9839
Epoch 12/25
8/8 [==============================] - 0s 19ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 16.4687 - val_accuracy: 0.9839
Epoch 13/25
8/8 [==============================] - 0s 16ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 16.3679 - val_accuracy: 0.9839
Epoch 14/25
8/8 [==============================] - 0s 16ms/step - loss: 8.1300 - accuracy: 0.9959 - val_loss: 20.8470 - val_accuracy: 0.9839
Epoch 15/25
8/8 [==============================] - 0s 16ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 24.6703 - val_accuracy: 0.9839
Epoch 16/25
8/8 [==============================] - 0s 18ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 26.3650 - val_accuracy: 0.9839
```

Fig.7: Converting categorical labels to numerical values **Fig.8:** Epoch showing the learning rate of CNN model using PCA

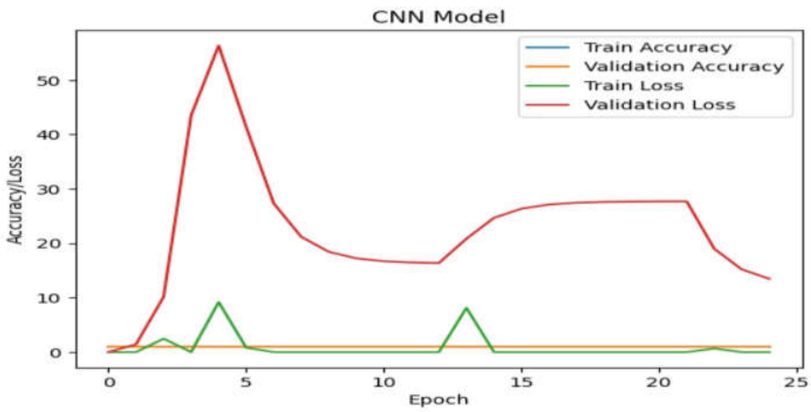

Fig.9: Plotting the graph CNN-PCA

Model	Precision	Recall	F1 Score	AUC	Accuracy
DenseNet121	0.6136	1.00	0.76	0.5	0.4253
VGG16	0.6136	1.00	0.76	0.5	0.6136
ResNet50	0.6136	1.00	0.76	0.5	0.6136
MobileNet	**0.897**	**0.925**	**0.911**	0.87	0.8896
InceptionV3	-	-	-	0.5	0.3864
CNN-PCA	**1.00**	**0.97**	**0.99**	**0.9857**	**0.9838**

Table 1: Comparison of Performance metrics

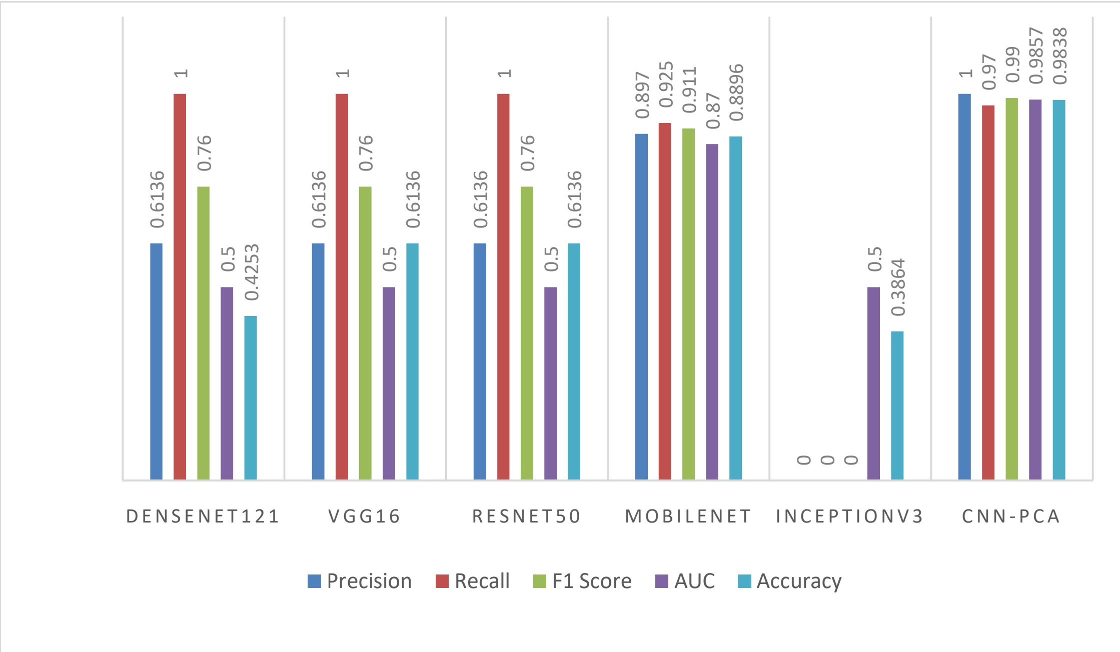

Fig.10: Comparison table in graphical representation

CONCLUSION

In conclusion, the combination of Convolutional Neural Networks (CNNs) and Principal Component Analysis (PCA) offers a promising approach for analying MRI data containing healthy controls (HC) and patients with Parkinson's disease (PD). PCA efficiently extracts essential patterns and variations from the high-dimensional MRI data, which can be integrated into CNN architectures or used for preprocessing. This integration enhances the CNN's ability to distinguish between HC and PD samples. Furthermore, PCA aids in data augmentation and noise reduction, bolstering the robustness of the CNN model. By leveraging both techniques, researchers can develop hybrid models that capitalize on their strengths, leading to improved classification accuracy and efficiency.

REFERENCES

1. Wu wang, junho lee, Fouzi Harrou and Ying sun "Early detection of Parkinson's disease using machine learning" IEEE transactions on image processing (2020)
2. Fouzi Harrou,et al., "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning " ,2020Senthil Sivakumar M, Joy Vasantha Rani S P, An ADC BIST using on-chip ramp generationanddigital ORA, Microelectronics Journal, Vol. 81, 2018, pp.8-15.
3. Prabhavathi, K., Patil, S. (2022). "Tremors and Bradykinesia. In: Arjunan, S.P., Kumar, D.K. (eds) Techniques for Assessment of Parkinsonism for Diagnosis and Rehabilitation". Series in BioEngineering. Springer. 135–149 https://doi.org/10.1007/978-981-16-3056-9_9
4. K Simonyan and A Zisserman (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations,(pp. 1-14). ICLR, Available at: regular papers, 2005, Vol. 52,No.9, pp. 1732-1740.
5. S Ghodake and N Ranjan (2019). AI algorithm for health risk prediction, Journal of Emerging Technologies and Innovative Research,6(6), 600-607, Available at: http://www.jetir.org/papers/JETIR1906D01.pdf
6. R.V. Darekar, et al. (2023). A hybrid metaheuristic ensemble-based classification Available at:https://doi.org/10.1016/j.advengsoft.2023.103412.technique speech emotion recognition, Advances in Engineering Software.
7. Carneiro , T., Da N´obrega, R.V.M., Nepomuceno, T., Bian, G.B., De Albuquerque, V.H.C., Reboucas Filho, P.P.: Performance analysis of google colaboratory as a tool for accelerating deep learning applications. IEEE Access 6, 61677–61685 (2018)
8. F. Huang, H. Xu, T. Shen and L. Jin (2021), "Recognition of Parkinson's Disease Based on Residual Neural Network and Voice Diagnosis,"2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 381-386, doi:10.1109/ITNEC52019.2021.9586915.dataset --https://www.ppmi-info.org/access-data-specimens/download-data
9. Amalia Luque, Alejandro Carrasco, Alejandro Martín, Ana de las Heras, (2019), "the impact of class imbalance in classificationperformance metrics based on the binary confusion matrix", Pattern Recognition, Volume 91, Pages 216-231, ISSN 0031-3203,https://doi.org/10.1016/j.patcog.2019.02.023.