# An Approach to Detect Parkinson's Disease Employing Pre-trained Deep Neural Network on Hypokinetic Dysarthria Symptoms

**Samiappan Dhanalakshmi** ( dhanalas@srmist.edu.in )
SRM Institute of Science and Technology https://orcid.org/0000-0002-6970-2719

**Harisudha Kuresan**
SRM Institute of Science and Technology

**Arathy Jeevan**
SRM Institute of Science and Technology

**Sukirti Gupta**
SRM Institute of Science and Technology

**Research Article**

**Keywords:** Parkinson Disease (PD), Dysarthria, Inceptionv3, ResNet50, MobileNetv2, CNN, LSTM

**Posted Date:** June 6th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-864459/v1

# Abstract

Parkinson's disease (PD) is a long-term central nervous system disease that causes damage in nerve cells and consequently affects mobility due to the lack of dopamine in the brain. PD is similarly effective in so-called dysarthria and comprises different pitches, lengthy pauses, monotonous and slow or slurred speech. The datasets containing several vocal features, gait features, and handwriting features of PD patients were analyzed in the proposed investigation and healthy control groups. This data package is taken as an input to study specific Machine Learning Algorithms and neural networks, such as ANN and LSTM, Random Forest Classifier, Naive Bayes, Support Vector Machines (SVM), Linear Regression, K-Nearest Neighbor (KNN). The second phase was followed by voice input from PD and healthy patients. The input is converted to Mel spectrograms, and various Transfer Learning models, such as Inceptionv3, ResNet50, and MobileNetv2, are examined. Inception-V3 gives 96.88 % accuracy compared with all other variants.


# 1. Introduction

Parkinson's disease is widely prevalent globally, with an estimate of around 11 million of the world’s population suffering from this disease. The Nervous system is severely damaged in PD [1]. The condition builds up gradually; early symptoms start with tremors, stiffness, slowed movement, and change in voice. The patients also suffer from behavioral changes, sleep deprivation, depression, difficulty in recollecting, and fatigue. Parkinson's disease affects nerves and muscles, affecting the Voicebox, throat, respiratory muscles, and facial muscles, top of mouth, tongue, and lips. 89% of Parkinson-affected patients have Dysarthria symptoms, manifesting mainly in articulation, phonation, pitch, speech fluency, speeds of speech [2].

Speech disorder caused due to weakened muscles is known as dysarthria. The affected speech characteristics are mainly respiration, resonance, pitch, prosody, vocal quality, loudness, and articulation [3]. Hypokinetic dysarthria is linked to basal ganglia dysfunctions. Basal ganglia are the structures found in the brain that controls cognitive function. It also provides support in motor movements by regulating the muscle tone [4, 22].

J.I.Godino-Llorente [5] in his work, the various new biomarkers for detecting Parkinson Disease. A Diadochokinetic test has been identified for multidimensional speech analysis. Based on kinetic behavior, the new biomarkers are related to articulatory indicators. In the proposed work authors compare the existing baseline biomarkers with the new ones. The results gave an accuracy of around 85%. In this work [6], the relevant information has been extracted using different speech processing algorithms. Tunable Q-Wavelet Transform better resolution in frequency than Discrete Wavelet Transforms. Various voice recordings were collected, and feature sets were extracted, which were fed to different classification and prediction models.

The proposed work shows that MFCC and TWQT give the best accuracy. The proposed work models the handwriting dynamics into a time series, taking these loaded into a Convolutional Neural Network [7, 23, 24]. Images with various resolutions and different size training sets were taken to check whether the patient had Parkinson's or not. A smartpen with a vibration sensor attached was used to get the raw data for handwriting features, excluding the images. CNN on images had a better result than the feature set extracted from the smartpen. Movement disorders can be assessed with wearable sensors.

The paper [8, 25] collects data from various Parkinson patients using Inertial Measurement Units. Attaching wearable on patients collect the different motor movement measurements, which is then input to ML and CNN pipeline to get accurate results? The results obtained for Deep Learning outperform ML algorithms with a classification rate of 4.6%. The proposed work validates the efficiency of supervised learning algorithms and deep neural networks to detect Parkinson’s disease [9, 16]. With pathology as base ground, peak accuracy obtained was 85%, exceeding the other diagnostic test results for motor features. Gradient Boosted Decision Tree is the beat classifier with high precision, F-1, and recall scores. The AUC gives around 0.924 on selected AVEC features. Luca Parsi [10] approached the early diagnosis of Parkinson's disease with a hybrid algorithm.

The performance was increased with feature reduction using Artificial Neural Networks. Feature selection was made using the weights from MLP. Hence, around 35% reductions in the number of features were fed into a Varangian Support Vector Machine. The scaled Filter bank technique differentiates those who have Parkinson's disease (PD) from healthy subjects. This article [11, 26] uses the Mel frequency cepstral coefficients (MFCC) as an expression used to discriminate between the two groups, taken from both PD and healthy people's speech samples. After looking at the spectrum of healthy and PD groups of people, the standard Mel frequency filter was created and changed.

For the region of interest in the Mel mid-frequency zone, the spectrum range was roughly 300 Hz – 1700 Hz. The patients with PD fall within the low-middle frequency range regarding classification accuracies acquired using a radian based network. The system's performances employing the two Mel filter banks have been assessed and adjusted. There was an increase in classification accuracy of 6.3 % with the new filter bank [11, 21].

In [12, 17,18], the authors investigated a pre-trained model using RNN transducer-based end-to-end speech recognition alignments. The enormous memory and intricate neural structure make RNN-T training problematic. The pre-trained encoder only uses cross-entropy (CE) loss to pre-train the RNN-T encoder. The approaches presented can also minimize the delay in the time of RNN-T modeling greatly. Authors[13, 19] use voice functionality in this study to diagnose and remote monitor PD patients at an early stage of computer-assisted diagnosis. The key contribution of this research is a remarkable reduction in accuracy and the number of voice characteristics picked for PD sense while using the latest and largest public dataset available.

Nelson Yalta et al [14] have introduced a CNN-based end-to-end speech recognition system in the home environment, where it is seen that a variety of sounds and distortions are widespread in casual or daily

conversations, adversely affecting the performance of automated voice-recognition systems.This multichannel end-to-end voice recognition technology is used to defeat this convolutionary neural network (CNN). The system includes a focused neural network of the encoder/decoder that directly creates a text from a sound input as an output. The experiments have shown that the word error rate is reduced by 8.5 and 0.6 % absolute from the single-channel end-to-end and the best baseline (LF-MMI TDNN) on the CHiME-5 corpus. Studies by authors [15, 20, and 27] test the support vector machine (SVM) approach's ability to tell Parkinson's disease patients from from healthy people.

In this study, a sample of Parkinson's ailment identification focused on a cepstral analysis had been produced after employing WT signal transformation to implement this sample on a database of sound and disease recordings. The wavelet Daubechies was employed with third-level approximation to compressed voice sounds. In the classification of two linear and RBF kernels, the cepstral coefficients are used. They found afterward that the RBF kernel is more precise than the linear kernel, and superior results than the Acoustic Signal technique were obtained utilizing the wavelet. The SVM MFCC extraction without using a Wavelet; 7% higher with the linear SVM kernel and 14% higher with the RBF SVM kernel.

The composition of the paper is as follows: Sect. 2 summarizes the proposed methodology employed for PD detection. Section 3 entails the quantitative analysis of experimental results and discussion in terms of various performance metrics. The paper is concluded in Sect. 4 with the future scope of the proposed work.

# 2. Methodology

The methodology used in our work is divided into two phases. In the first phase, various datasets containing various vocal features [16], Gait features [17], and handwriting features [18][19] of PD patients as well as healthy control groups are examined. This dataset has been taken as an input to analyze with some machine learning algorithms such as Random Forest Classifier, Naive Bayes, Support Vector Machines (SVM), Linear Regression, K-Nearest Neighbor (KNN) and Neural Networks such as ANN and LSTM. The step for machine learning is shown in Fig. 1.

In the second phase, Mel-spectrograms of these voice recordings is used as input to CNN. Pre-trained models such as ResNet-50, Inception, and MobileNet-V2 are analyzed to classify input spectrograms into Parkinson Group or Healthy Group. The step for the deep learning model is shown in Fig. 2. The dataset consists of PD and normal subjects whose features are extracted from their voice, gait, and handwriting. Those samples are given as input to the deep learning model. For voice recording consisting of sustained phonations of /a/e/o/ sounds are converted into mel spectrograms. Then it is given to the pre-trained model where the weights are randomly initialized initially, and the model has trained again with the dataset provided. The output layers are customized and modified as per the need. If the data size is small, some of the initial layers can be frozen, and the remaining layers can be used to train the model. The performance of the system is analyzed using testing and training accuracies.

# 3. Dataset

The Voice dataset for the study includes PD's classification dataset from UCI Machine Learning Repository combined with the data collected from the SRM Medical College Hospital and Research center. Around 246 sample instances are considered for analyses. MFCC and DWT were used for feature extraction on the voice dataset. For Phase 1, the features are extracted from the three datasets - voice, gait, and handwriting have been listed in the below table. For phase 2, the recording of all the subjects consisting of 'a', 'e', 'o', phonations are considered. The proposed methodology contains the sounds /a/e/o/ for further analysis.

A Mel spectrogram sample is given in Figs. 3 and 4, for healthy and PD patients after necessary preprocessing was performed on the recorded data. Figures 3 and 4 show the mel spectrograms of /a/ phonation for PD and healthy individuals, which were maintained for 2 seconds. The PD subject's signal strength is diminished because of their verbal slurring and poor loudness. The hot pink patches signify a large volume, while the cool violet and black dots indicate quiet. For the patients in the PD group, the frequency is between 0 and 1024 Hz. The spectrograms of patients and healthy individuals show changes in frequency across time and amplitude.

Table 1
Feature extracts from various datasets

| Dataset | Features |
|---|---|
| Voice Dataset | ● Jitter<br>● Shimmer<br>● Noise to Harmonic Ratio<br>● Harmonic to Noise Ratio<br>● Pitch Periodic Entropy(PPE)<br>● Recurrence Period Density Entropy<br>● Detrended Fluctuation Analysis |
| Gait Dataset | ● Hoehn and Yahr Rating Scale<br>● Unified Parkinson Disease Rating Scale (UPDRS)<br>● Unified Parkinson Disease Rating Scale Motor Score(UPDRSM) |
| Handwriting Dataset | ● Number of Strokes<br>● Speed (x, y, z-direction)<br>● Jerk (x, y, z-direction)<br>● Acceleration (x, y, z-direction) |

# 4. Machine Learning Algorithms

An application of AI (artificial intelligence), Machine learning can learn and improve automatically from previous experience without being particularly programmed. It accesses the data and then uses it to know itself, which is a development to the computer program. The two types of techniques that machine learning uses are Supervised knowledge can train the model according to input and output data to predict the outputs of future or upcoming output, and unsupervised learning can determine patterns that are hidden or structures that are intrinsic in the input data.

Based on the Bayes theorem, NB classifier classifies certain objects according to features of the group. LR classifier predicts the value based on an object dependency on another independent object. SVM, a binary classifier, draws all given data into the 'n' dimension, where the hyperplane separates the datasets into groups. Based on the majority of neighbors KNN classifier predicts the group. Random Forest generates multiple classification trees based on the various rules system. The vote is calculated for each object, and the highest vote for prediction is considered final.

# 5. Neural Networks

An artificial neuron network may be understood as the structure of a human neural network, and the function may be referred to as a computer model of a neural network. It learns based on the information provided in input and output data given which affects the structure of the ANN. RNN's particular type is LSTM. LSTM overcomes the problem of long-term dependency. Generally, neural network repeating modules are linked to form a chain-like structure for the RNN. In LSTM, a chain-like structure of the repeating module is seen. An LSTM cell state is one of the vital things which runs through the entire chain where information can be added/removed from the cell states using gates. There are three types of gates. The control gate that controls parts of the long-term state that has to be deleted is controlled by f(t). The input gate contains the amount of g(t) to be added, which is owned by g(t), and the output gate will control part of the long-term state this is to be read and output to be sent to y(t) and h(t) is controlled by o(t).

Once the training and test data are split as 70:30, the features are given to the Machine Learning algorithms: NB, LR, SVM, and RFC. ANN and LSTM under Neural Networks are used to learn from the features mentioned in Table 1 to predict whether the features correspond to Parkinson's or Healthy Group.

# 6. Convolution Neural Network

Convolutional Neural Networks take in images as input, while a set of weights are multiplied with it. A 2-D array of weights is known as the kernel. The pooling layer is a nonlinear downsampling activity that is normally applied after a convolution layer. A pooling layer reduces the computational complexities required to process the huge amount of data connected to an image provided. Fully connected are generally found towards the end of CNN structures and are used to upgrade destinations.

Many ML models have a final layer that is completely linked, which assembles the information obtained from earlier layers and produces the output. A cost function is a prediction error referred to as a loss function. The gradient it uses to update weights in a neural network model is its primary focus. The technique of gradient descent is used to either train neural networks, in which case it iteratively refreshes the trainable limits (i.e., kernels and weights) to minimize loss, or else it's used to update a network's values so as to minimize loss.

The input Mel Spectrograms are multiplied with a set of weights. The output of this is known as the feature map. These features are mapped into the ReLU activation function. The function gives the input as output if positive and zero if negative. The CNN model has 4 Convolutional and Max Pooling layers. Output layer with a single neuron provides 0 for healthy and 1 for Parkinson. A sigmoid activation function makes the model output lie between 0 and 1. The summary of the CNN model is shown in Fig. 5.

# 7. Transfer Learning Models

Transferring the learning from pre-trained models helps in training various complex neural networks with considerably fewer data and is provided in Fig. 6. In this approach, the data was only examined by using the weights of the pre-trained models, enabling the dataset to be trained.

## (i) Inception-V3

IncetionV3 is a deep 48-layer network containing symmetrical and asymmetrical blocks with turmoil, average pools, full pools, concatenations, dropouts, and ultimately linked layers. On the ImageNet dataset, the model reached more than 78.1% precision [20]. In 20,000 categories, ImageNet is a significant visual basis comprising over 14 million images. The primary inception model is shown in Fig. 7. In the proposed work, InceptionV3 pre-trained model is used as the base model. The network architecture is made of the following -

1. Factorized Convolutions: It reduces the number of parameters without affecting the efficiency. It involves replacing large convolutions with smaller ones and using asymmetric convolutions to replace single n x n convolutions.

2. Auxiliary classifier: Insertion of small CNN layers during training. It acts as a regulator.

3. Grid size reduction: Pooling operations are responsible for grid size reduction. It helps reduce computational costs.

## (ii) ResNet-50

ResNet50 has 48 layers, along with one MaxPool and one intermediate pool layer. Figure 8 shows the simple residual block. The difference between residual block input and output is the residual function.

The network is initiated with convolutions and max-pooling using kernel sizes of seven and three. The first stage consists of three Residual blocks with three layers. The Kernel sizes are 64, 64, and 128 for the first stage. There is an identity connection for retaining the status. The last step consists of a pooling layer followed by a fully linked neuron layer.

$$y = F(x, \{Wi\}) + x$$

$$F(x, \{Wi\}) = W2(W1x)$$

where;

y is the output function,

x is the input to the residual block

F(x, {Wi}) is the residual block.

Wi where 1 ≤ i ≤ number of layers in a residual block.

## (iii) MobileNetV2

MobileNetV2 model works very efficiently on mobile devices. It is a 53 layer deep network that uses depthwise separable convolutions, Width multiplier, Linear bottlenecks, and shortcut connections [22]. The basic building block of MobileNet V2 is shown in Fig. 9. A residual block is having stride 1 and stride 2are the two blocks available. Both blocks have three layers, each with 1x1 convolution, the second has depthwise convolution, and the last layer consists of 1x1 convolution.

# 8. Simulation Results And Discussion

Python using Keras library functions is used for deep learning models. Various performance metrics considering Confusion Matrix that gives TP, TN, FP and FN values are taken for analysis. The metrics include accuracy, sensitivity, specificity, F-1 Score, and Matthew Correlation Coefficient (MCC).

The ratio of perfectly predicted patients who have PD to a total number of patients gives the accuracy. The proportion of PD patients predicted positive gives the sensitivity. The proportion of healthy patients predicted negative gives the specificity. F1 gives a relation between the metrics sensitivity and specificity. Matthews Correlation Coefficient (MCC) considers confusion matrix values to give a correlation coefficient. The value is always between + 1 and − 1. The formulae for the performance metrics are given below:

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

Sensitivity = $\frac{TP}{TP+FN}$ (2)

Specificity = $\frac{TN}{TN+FP}$ (3)

F1 Score = $\frac{2TP}{2TP+FP+FN}$ (4)

MCC = $\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ (5)

Table 2
Performance metrics for ML and Neural networks for Voice Dataset

| Model / NN | Accuracy (%) | Sensitivity | Specificity | F-1 score (%) | MCC |
|---|---|---|---|---|---|
| **NB** | 88.88 | 0.8 | 1.00 | 88.88 | 0.8 |
| **LR** | 77.77 | 0.75 | 0.80 | 75.00 | 0.55 |
| **SVM** | 77.77 | 0.75 | 0.80 | 75.00 | 0.55 |
| **KNN** | 59.15 | 0.58 | 0.60 | 65.06 | 0.17 |
| **RFC** | 77.77 | 0.666 | 1.00 | 80.00 | 0.63 |
| **ANN** | 90.00 | 0.85 | 1.00 | 92.30 | 0.80 |
| **LSTM** | 70.00 | 0.75 | 0.5 | 80.00 | 0.21 |

In Table 2, ANN provides an accuracy of 90% compared to other ML and LSTM algorithms for the voice dataset.

Table 3
Performance metrics for ML and Neural networks for Gait Dataset

| Model / NN | Accuracy (%) | Sensitivity | Specificity | F-1 score (%) | MCC |
|---|---|---|---|---|---|
| NB | 99.00 | 1.0 | 1.0 | 73.74 | 1.0 |
| LR | 96.21 | 1.0 | 0.92 | 66.66 | 0.92 |
| SVM | 96.21 | 1.0 | 0.92 | 66.66 | 0.92 |
| KNN | 94.45 | 1.0 | 0.88 | 64.86 | 0.88 |
| RFC | 96.00 | 1.0 | 0.92 | 96.15 | 0.92 |
| ANN | 96.07 | 0.63 | 0.94 | 96.87 | 0.91 |
| LSTM | 71.42 | 0.96 | 0.78 | 66.66 | 0.41 |

The Naïve Bayes classifier achieved an accuracy of 99% when compared to the other classifiers presented in Table 3 when applied to the gait dataset.

Table 4
Performance metrics for ML and Neural networks for Handwriting Dataset

| Model / NN | Accuracy (%) | Sensitivity | Specificity | F-1 score (%) | MCC |
|---|---|---|---|---|---|
| NB | 75.00 | 0.88 | 0.42 | 83.33 | 0.34 |
| LR | 66.66 | 0.78 | 0.2 | 69.76 | -0.01 |
| SVM | 70.83 | 0.77 | 0.1 | 73.91 | -0.15 |
| KNN | 75.00 | 0.80 | 0.33 | 75.55 | 0.11 |
| RFC | 83.33 | 0.82 | 1.0 | 80.85 | 0.40 |
| ANN | 86.95 | 0.84 | 1.0 | 91.42 | 0.69 |
| LSTM | 73.91 | 0.76 | 0.5 | 84.21 | 0.84 |

ANN provides an accuracy of 86.95% for the handwriting dataset of PD patients compared to other ML and LSTM algorithms and is listed in Table 4. The Parkinson's voice dataset performed equivalently well compared to the motor symptom (gait) of the PD subjects. These shows that speech can be used as an early diagnosis of Parkinson's disease. In Fig. 9, the ROC curves have been plotted for all the ML and ANN algorithms for the voice dataset of PD. An area under the curve (AUC) of 0.92 is achieved for the ANN algorithm.

Table 5 shows the PD speech dataset implemented for various deep learning methods. Inception-V3 provides an accuracy of 96.88%. Figures 10 and 11 shows the accuracy and loss values for Inception-V3 and ResNet-50 algorithms.

Table 5
Performance metrics for CNN and Transfer Learning models for Voice Dataset

| Model | Accuracy (%) | Sensitivity | Specificity | F-1 score (%) | MCC |
|---|---|---|---|---|---|
| **CNN** | 84.14 | 0.80 | 0.88 | 85.39 | 0.68 |
| **Inception-V3** | 96.88 | 0.89 | 1.00 | 92.30 | 0.55 |
| **ResNet-50** | 93.75 | 0.85 | 0.92 | 64.86 | 0.55 |
| **MobileNetV2** | 85.94 | 0.89 | 0.88 | 83.01 | 0.178 |

The accuracy and loss curve in Figs. 10 and 11 gives us an idea about how the validation dataset has been trained. The training loss shows how well the training dataset fits, while the validation loss shows how well the test dataset fits. Here we can see that the loss decreases as the epoch increases and accuracy increases, which means the model is learning over time. Figure 10 shows the accuracy plot for the training and the validation data. The validation accuracy of 96.88% is achieved for the Inception V3 model. Similarly, for ResNet 50, the validation accuracy of 93.75% is achieved.

# 9. Conclusion

Naïve Bayes gives 88% accuracy among ML models, and ANN gives 90% accuracy among Neural Networks for voice datasets. Similarly, for the gait dataset, Naïve Bayes and ANN models outperform other models. Phase 1 results indicate that gait analysis and voice dataset features set are helpful for Parkinson Detection. Analyzing the performance metrics of three PD-related datasets shows that for Naive Bayes algorithm in ML algorithms and ANN in Neural Network has better efficiency than other classification algorithms.

For further investigation, in the second phase, the analysis is done with Convolutional Neural Network on Mel-Spectrograms of healthy and PD. On a Mel-Spectrograms dataset, an accuracy of around 85% was obtained with CNN. By leveraging Transfer Learning, i.e., pre-trained models, ResNet-50 and InceptionV3 models give better accuracy than CNN, greater than 90%. Using the InceptionV3 pre-trained model, the efficiency is increased by 13%. In conclusion, Transfer Learning with speech processing can be utilized for the early diagnosis of PD. Further, the proposed work can be extended by increasing the size of the dataset and by implementing feature reduction techniques for improvement in accuracy.

# Declarations

### Conflicts of Interest

The authors declare no conflict of interest.

# Funding

The investigations presented in the manuscript are not funded by any external agencies.

# Acknowledgement

Authors gratefully thank SRM Medical College Hospital & Research Center

technical staff for their support in carrying out the observations reported here.

## References

[1] A. Lee and R. M. Gilbert, “Epidemiology of Parkinson's disease," Neurologic Clinics, vol. 34, no. 4, pp. 955–965, 2016.

[2] Schalling E, Johansson K, Hartelius L. “Speech and Communication Changes Reported by People with Parkinson's Disease”, Folia Phoniatr Logop;69(3):131-141, 2017.

[3] Kaitlin L. Lansford, Julie M. Liss, John N. Caviness, Rene L. Utianski, "A Cognitive-Perceptual Approach to Conceptualizing Speech Intelligibility Deficits and Remediation Practice in Hypokinetic Dysarthria", Parkinson’s Disease, vol. 2011;9 pages, 2011.

[4] M. K. MacPherson, J. E. Huber, and D. P. Snow, “The intonation-syntax interface in the speech of individuals with Parkinson’s disease” J Speech Lang Hear Res, vol. 54, no. 1, pp. 19–32, 2011.

[5] Godino-Llorente JI, Shattuck-Hufnagel S, Choi JY, Moro-Velázquez L, Gómez-García JA. “Towards the identification of Idiopathic Parkinson's Disease from the speech: New articulatory kinetic biomarkers” PLoS One, 2017.

[6] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q- factor wavelet transform”, Applied Soft Computing, 74, 255-263, 2019.

[7] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, “Deep learning-aided Parkinson’s disease diagnosis from handwritten dynamics” in 2016 29th SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI). IEEE, pp. 340–346, 2016.

[8] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz et al., “Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson’s disease assessment,” in 2016 38thAnnual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 655–658, 2016.

[9]13.T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H.Ghomi, “Parkinson’s disease diagnosis using machine learning and voice” in 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE, pp. 1–7, 2018.

[10] L. Parisi, N. RaviChandran, and M. L. Manaog, “Feature-driven machine learning to improve early diagnosis of Parkinson’s disease”, Expert Systems with Applications, vol. 110, pp. 182–190, 2018.

[11] Arko Chatterjee et al. “Diagnosis of Atherosclerotic Plaques in Carotid Artery using Transfer Learning”, 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp-1247-1251, 2020.

[12] H. Hu, R. Zhao, J. Li, L. Lu and Y. Gong, "Exploring Pre-Training with Alignments for RNN Transducer Based End-to-End Speech Recognition," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, pp. 7079-7083, 2020. DOI: 10.1109/ICASSP40776.2020.9054663.

[13] Solana-Lavalle, Gabriel & Galán-Hernández, Juan-Carlos & Rosas-Romero, Roberto. “Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features”. Bio cybernetics and Biomedical Engineering. 2020. DOI:40. 10.1016/j.bbe.2020.01.003.

[14] Samiappan, Dhanalakshmi et al, “Early diagnosis of Parkinson’s disease based on non-motor symptoms: a descriptive and factor analysis”, Journal of Ambient Intelligence and Humanized Computing. 1-15. 10.1007/s12652-021-02944-0

[15] Z. Soumaya, B. D. Taoufiq, B. Nsiri and A. Abdelkrim, "Diagnosis of Parkinson's disease using the wavelet transform and MFCC and SVM classifier," 2019 4th World Conference on Complex Systems (WCCS), Ouarzazate, Morocco, pp. 1-6, 2019. DOI: 10.1109/ICoCS.2019.8930802.

[16] Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O., “Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings”, IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013.

[17] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., & Stanley, H. E. “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals”, Circulation [Online]. 2002. DOI:101 (23), pp. e215–e220.

[18] Isenkul, M.E.; Sakar, B.E.; Kursun, O. “Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease”, The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014), pp. 171-175, 2014.

[19] Masunda, Sam et al. ‘Fusion of WPT and MFCC Feature Extraction in Parkinson’s Disease Diagnosis’. 1 Jan. 2019 : 363 – 372.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, 2018. DOI: 10.1109/CVPR.2018.00474.

[23] N. Yalta, S. Watanabe, T. Hori, K. Nakadai and T. Ogata, "CNN-based Multichannel End-to-End Speech Recognition for Everyday Home Environments*," 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, pp. 1-5, 2019. DOIi: 10.23919/EUSIPCO.2019.8902524.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

[25] Upadhya, S.S., Cheeran, A.N. & Nirmal, J.H. “Discriminating Parkinson diseased and healthy people using modified MFCC filter bank approach”, Int J Speech Technol 22, 1021–1029, 2019.

[26] Pauline SH et al, “Variable tap-length and non-parametric variable size NLMS adaptive filtering algorithm for acoustic echo cancellation”, Appl Acoust, 2020; 159.

[27] Monika et al, “Coefficient random permutation based compressed sensing for medical image compression”, Lecture Notes on Electrical Engineering, 443:529-536, 2018.

# Figures

**Figure 1**

Flow diagram of Machine learning model

**Figure 2**

Steps for Deep learning models

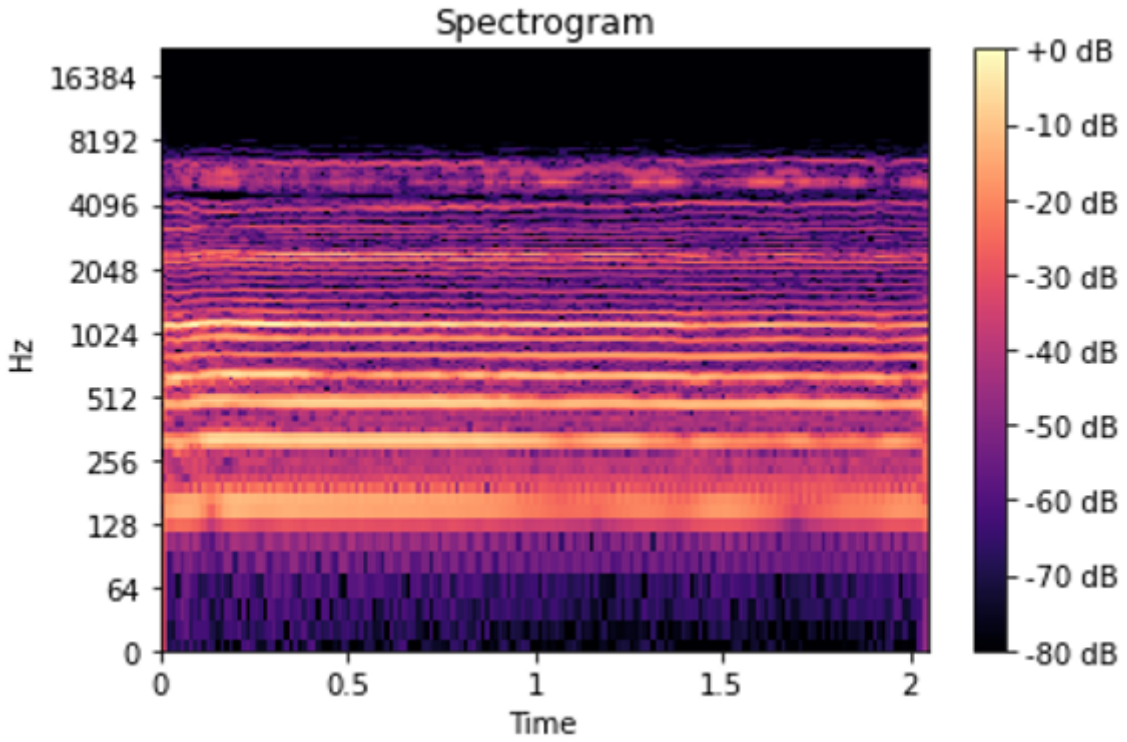


**Figure 3**

Mel-spectrogram of sustained /a/ phonation of the healthy subject

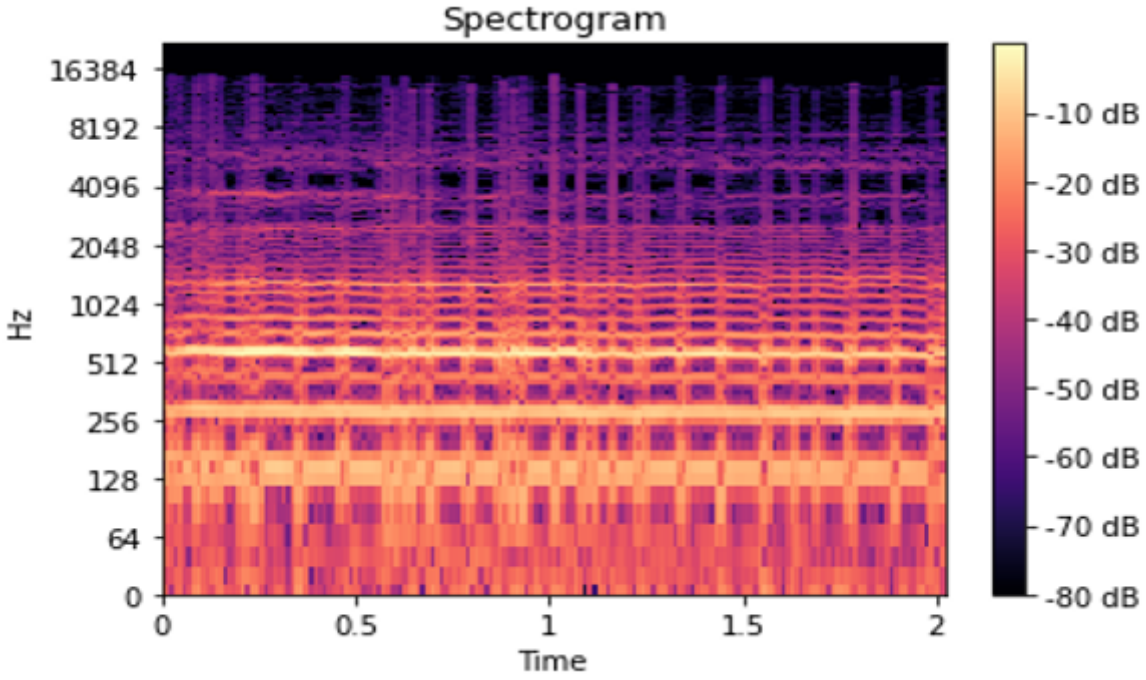


**Figure 4**

Mel-spectrogram of sustained /a/ phonation of Parkinson's subject

**Figure 5**

CNN Model Summary

**Figure 6**

Transfer learning from the pre-trained models

**Figure 7**

Base Inception module structure

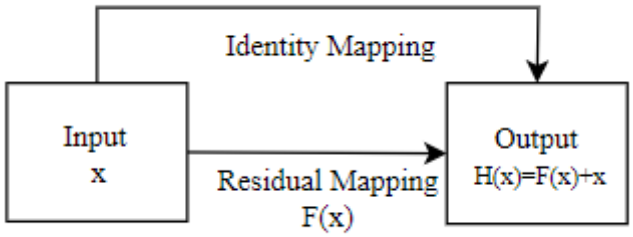


**Figure 8**

Residual Block

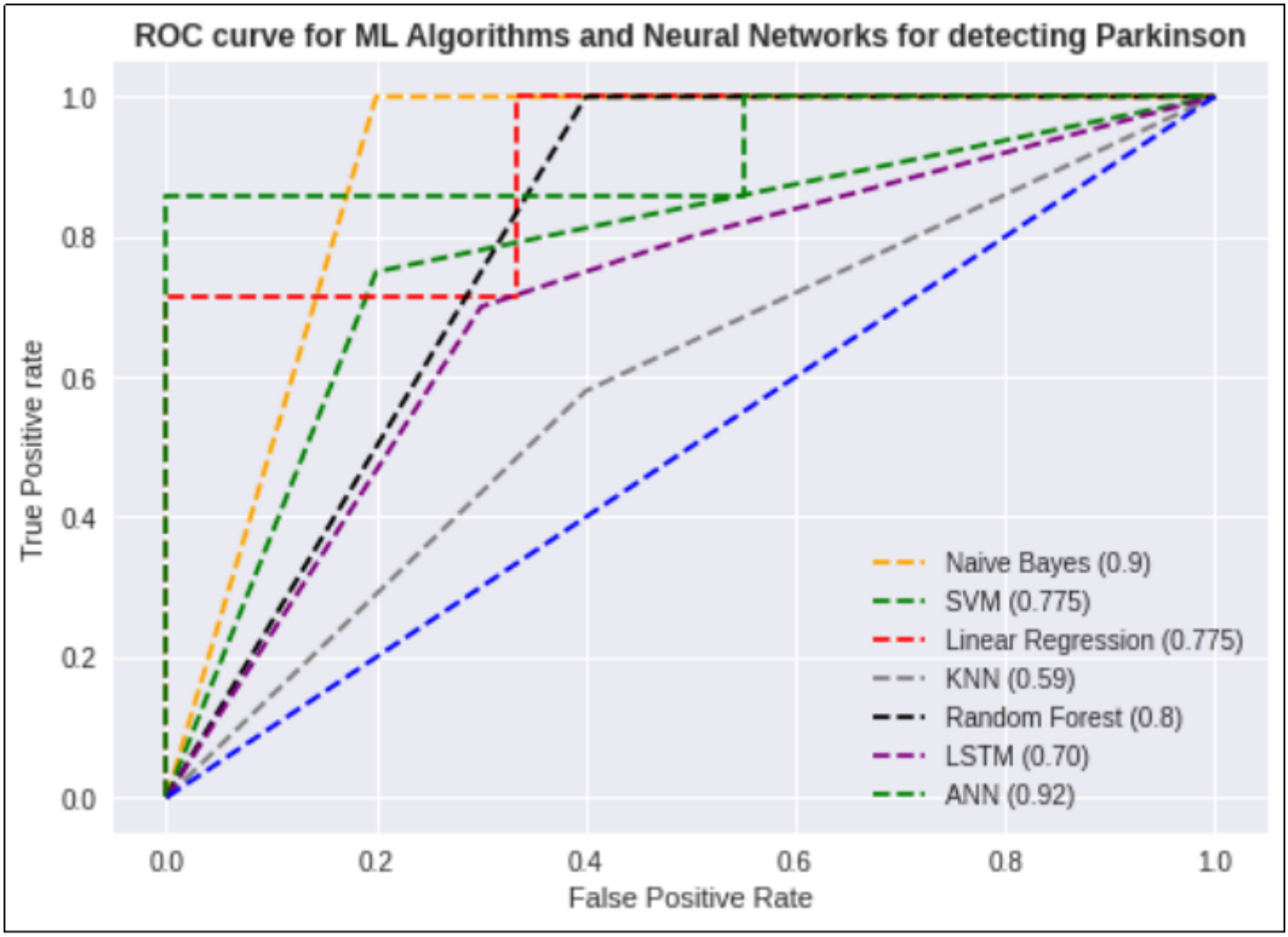


**Figure 9**

ROC Curve for the ML Algorithms and Neural Networks on Voice Data

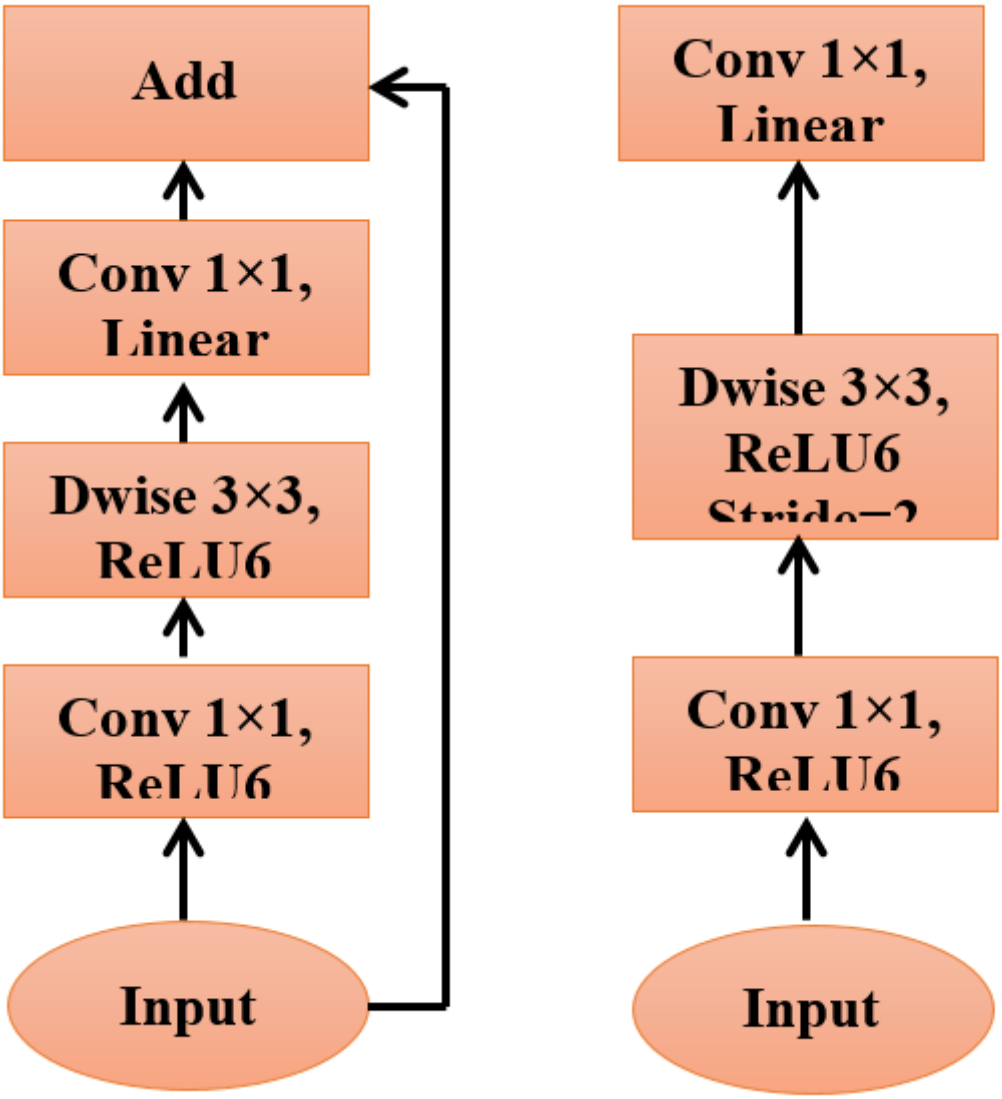


**Figure 10**

MobileNetV2 Building blocks

**Figure 11**

Accuracy and Loss graphs for InceptionV3 Model

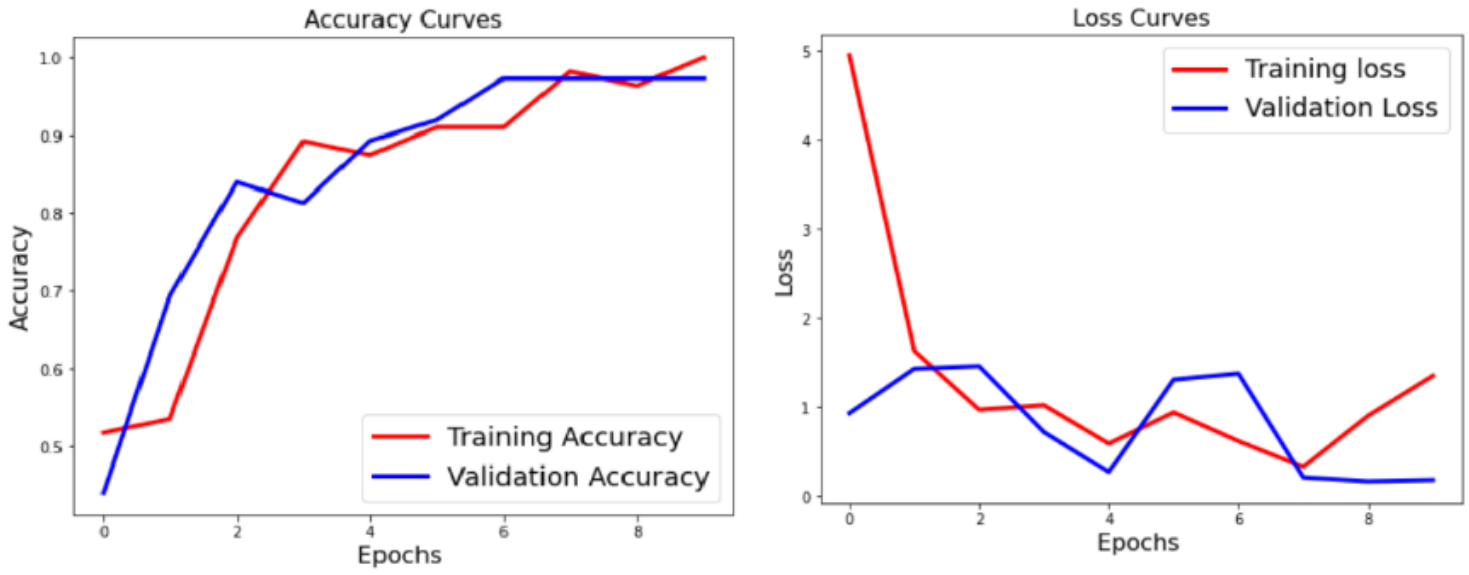


**Figure 12**

Accuracy and Loss graphs for ResNet 50 Model