

Received January 17, 2020, accepted February 6, 2020, date of publication February 14, 2020, date of current version February 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974008

A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease

LAIBA ZAHID¹, MUAZZAM MAQSOOD¹, MEHR YAHYA DURRANI¹, MAHEEN BAKHTYAR², JUNAID BABER², HABIBULLAH JAMAL³, IRFAN MEHMOOD⁴, AND OH-YOUNG SONG⁵

¹Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan

²Department of Computer Science and Information Technology, University of Balochistan, Quetta 87300, Pakistan

³Faculty of Engineering Sciences, Ghulam Ishaq Khan Institute, Topi 23460, Pakistan

⁴Department of Media Design and Technology, Faculty of Engineering & Informatics, University of Bradford, Bradford BD7 1AZ, U.K.

⁵Department of Software, Sejong University, Seoul 05006, South Korea

Corresponding authors: Muazzam Maqsood (muazzam.maqsood@cuiatk.edu.pk) and Oh-young Song (oysong@sejong.edu)

This work was supported in part by the MSIT (Ministry of Science and ICT), South Korea, through the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) under Grant IITP-2019-2016-0-00312, and in part by the Faculty Research Fund of Sejong University in 2019.

ABSTRACT Parkinson's disease is a neural degenerative disease. It slowly progresses from mild to severe stage, resulting in the degeneration of dopamine cells of neurons. Due to the deficiency of dopamine cells in the brain, it leads to a motor (tremor, slowness, impaired posture) and non-motor (speech, olfactory) defects in the body. Early detection of Parkinson's disease is a difficult chore as the symptoms of disease appear overtime. However, different diagnostic systems have contributed towards disease detection by considering gait, tremor and speech characteristics. Recent work has shown that speech impairments can be considered as a possible predictor for Parkinson's disease classification and remains an open research area. The speech signals show major differences and variations for Parkinson patients as compared to normal human beings. Therefore, variation in speech should be modeled using acoustic features to identify these variations. In this research, we propose three methods- the first method employs a transfer learning-based approach using spectrograms of speech recordings, the second method evaluates deep features extracted from speech spectrograms using machine learning classifiers and the third method evaluates simple acoustic feature of recordings using machine learning classifiers. The proposed frameworks are evaluated on a Spanish dataset pc-Gita. The results show that the second framework shows promising results with deep features. The highest 99.7% accuracy on vowel \o\ and read text is observed using a multilayer perceptron. Whereas 99.1% accuracy observed on vowel \i\ deep features using random forest. The deep feature-based method performs better as compared to simple acoustic features and transfer learning approaches. The proposed methodology outperforms the existing techniques on the pc-Gita dataset for Parkinson's disease detection.

INDEX TERMS Parkinson disease, classification, deep features, speech signals, transfer learning.

I. INTRODUCTION

Parkinson's disease (PD) is a slowly progressing neural degenerative disease. The main origin of the disease is still unknown [1]. However, researchers have found that certain hereditary and environmental factors aid in the cause of Parkinson's disease. Research has shown that nearly 100-250 persons out of 100,000 are suffering from Parkinson's

disease [2]. The disease is common among people age 50 or above. However, early symptoms have also been observed at the age of 30-50. The disease primarily affects the central nervous system by causing the degeneration of dopamine-producing neuron cells. Dopamine is a chemical produced by substantial nigra (basal ganglia) which is responsible for transferring signals within the brain. Loss of dopamine-producing cells results in movement disorder in PD patients. The symptoms of Parkinson's disease are classified into the motor and non-motor symptoms. Motor symptoms are

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Farouk.

associated with movements and are more perceptible as compared to non-motor symptoms [3]. In motor symptoms, the patient suffers from slowness of movement referred to as bradykinesia, rigidity, postural instability and tremor. Non-motor symptoms are evident at a particular interval; include sleep disorder, speech, and swallowing problem and olfactory disorder (loss of sense of smell) [1], [3]. The effect of Parkinson's disease on speech is characterized as phonation, articulation, and prosody. Phonation is the use of vocal folds for speech and articulation refers to the use of special tissues in speech production. While prosody is related to amplitude, loudness and pitch to produce sound. Most of the work in Parkinson disease detection considered phonation that includes the pronunciation of vowels \a\, \e\, \i\, \o\, \u\ [4].

Speech signals are usually considered as one of the main methods to diagnose Parkinson's disease. In [5] authors performed experimentation on speech signal recordings of three languages identifying speech pronunciation problems in it. It is observed that speech pronunciation like vowels, sentences, words are affected by this disease. Therefore, speech is considered as a major predictor of PD disease. In [6] authors accounted that articulation, intelligibility, prosody features of speech signal shows promising results in the detection of PD. In [7] the author presented that somehow the age factor also contributes to disease. They assessed that the speech recordings of young speakers show significant defects in speech pronunciation tasks. The researchers also observed the monitoring of skype calls using normal sentences shows significant errors in the pronunciation of PD patients [8]. Traditionally, acoustic features are considered in most of the recent works along with SVM for PD detection. Most of the recent work has performed disease detection using gait [9], handwriting [10], [11] and speech datasets. Furthermore, the literature study shows Gaussian based model, several machine learning techniques, convolution neural networks [12] are contributing to PD diagnosis. Parkinson's disease detection using most suitable speech impairment features is imperative and still is an open research area.

This research work contemplates speech recordings using spectrograms and acoustic features. In our work, all recordings are transformed into short-time Fourier transform (spectrograms) that are used in the transfer learning method. We proposed a simple acoustic features based method and also considered a pre-trained convolution neural network architecture [13], [14] Alexnet model for deep feature extraction and detection of PD. For fair comparison, we have also used transfer learning-based classification. To evaluate the performance of the proposed methods, Parkinson's disease speech recordings from PC-GITA [15] dataset are used. The results show that the deep features based technique produced better results. The main contributions of our research for PD detection using speech signals are the following:

1. We propose a spectrogram based approach to extract deep feature to distinguish PD patients from healthy

2. We propose an acoustic-phonetic based approach for the detection of PD disease
3. We conducted a comparison between a proposed deep feature-based approach with simple acoustic features and transfer learning-based methods

The rest of the research paper is organized as follows. Section 2 explains the literature review, section 3 explain the proposed methodology, section 4 shows experimental setup and dataset details, section 5 gives detail view of results and simulations, followed by the conclusion and future work.

II. LITERATURE REVIEW

This section explains the existing techniques on Parkinson's disease detection using the Spanish speech dataset. The techniques presented by different researchers are grouped into two categories that are explained in detail in the following subsections.

A. MACHINE LEARNING BASED METHODS

In the past few years, machine learning-based disease classification is widely used in the medical field and has acquired remarkable significance [16], [17]. A Gaussian based density was considered by Moro-Velazquez *et al.* [18] from four to five different PD corpora. They exploited phonetic text-dependent utterances that require vocal tract features of speech signals. They assessed words, sentences, monologues and vowels in three corpora with more male patients (Czech dataset). They showed better results as compared to other (Spanish) datasets and reported 81% accuracy. Rueda *et al.* [6] used a wrapper feature selection method for vowel \a\ and words \pa-ta-ka\ (Articulation, phonation, Diadochokinetic features) from pc-Gita recordings and achieved 70% accuracy. Pérez-Toro *et al.* [19] considered classical features like term frequency and bag of words of monologues only. They used pc-Gita recordings and reported that language pronunciation contains enough information for PD classification and achieved 72% accuracy results. Karan *et al.* [20] assessed inherent and decomposition-based features from vowel \a\, \o\ only from two datasets pc-Gita and Saarbrucken dataset. They have reported 96% accuracy on the Spanish dataset using random forest and support vector machine. Kacha *et al.* [21] showed that use of PCA for articulation features from sentences of Spanish pc-Gita speech spectrograms (STFT) is efficient for the detection of Parkinson's disease. Parra-Gallego *et al.* [4] considered articulation and intelligibility features from the words like \pa-ta-ka\ from Spanish pc-Gita recordings. The authors reported 88% accuracy in their work when classifiers were trained using intelligibility and articulation-based features.

Vasquez-Correa *et al.* [22] proposed a novel approach by considering the on-off state of vocal folds (i.e. on when the candidate starts speaking and off when they stop speaking). It includes words, vowels, monologues, and sentences from pc-Gita recordings. They reported 94.9% accuracy in speech signals for PD classification. Garcia *et al.* [23]

assessed the appropriateness of i-vectors to classify Parkinson's disease using articulation, prosody, phonation features for words, vowels and 10 sentences from pc-Gita dataset. They computed cosine difference that confers for disease detection contributing 78% accuracy using articulation features. Klumpp *et al.* [24] developed a model that considered the voice signal during phone calls and also considered syllable \pa-ta-ka\. They evaluated the severity and onset of PD disease in their work. Arias-Vergara *et al.* [7] assessed the aging factor in their research work. They considered articulation, prosody features, and age factors from pc-Gita Spanish speech vowels.

The authors also accounted for gender factors and reported that the age factor plays a very important role in classifying Parkinson's disease. Their research evaluated that young speaker signals are more contributing towards PD classification process than the old age speaker signals. They modeled binary and multi-class support vector machines and compared their results with neural networks and reported 95% accuracy. Moro-Velázquez *et al.* [25] presented a new approach for classifying a speech signal into Parkinson's patient or healthy patient. They proposed a phonological feature-based method in which the speech signal of words, monologues and read text from the Spanish pc-Gita dataset. The authors reported that this approach is quite useful in the assessment of Parkinson's disease patients in clinics. Moro-Velázquez *et al.* [25] used traditional machine learning methods considering articulation and phonological features for Parkinson's disease detection. They assessed kinetic features for /pa-ta-ka/, two read sentences and a sustained vowel /a/ from Spanish pc-Gita dataset from Parkinson disease patient (PDP) speech signals. They performed classification tasks using Gaussian mixture modeling and i-vectors and reported 87% accuracy. Orozco-Arroyave *et al.* [26] proposed an open-source software for Parkinson's disease. The authors have used phonation, articulation, prosody, and intelligibility dimensions of speech signal from pc-Gita dataset for vowels using conventional machine learning to identify Parkinson. They designed a system that can be easily adopted by clinicians to assess different voice diseases. Arias-Vergara *et al.* [8] proposed a model for assessment of Parkinson's disease using individual speaker speech signal analysis. They assessed phonation, articulation, and prosody to model recordings of spontaneous speech and a read text from the Spanish pc-Gita dataset from different channels (mobile phone calls, online calls like skype). In this work, authors observed that skype speech signals were effective in distant observation of Parkinson's disease patients. They performed evaluation using Gaussian mixture modeling and i-vectors and obtained a 0.77% correlation. Vásquez-Correa *et al.* [27] presented an improved version of m-FDA for PD detection. They considered phonation, articulation, prosody, and intelligibility features from Spanish vowel /a/, sentences and words. Orhan *et al.* [28] considered vowel recordings of freely available datasets [29]. They incorporated statistical pooling for increasing features and used ReliefF for selecting the best features and achieved

91% accuracy using SVM. El Maachi *et al.* [30] evaluated the gait physionet dataset for PD diagnosis. The author assessed Parkinson disease from gait using a deep 1-D neural network. They achieved 98.7% accuracy using deep neural networks and 85.3% accuracy in finding the severity of PD. Turker and Dogan [31] proposed an octopus based multiple pooling method (comprising of eight pooling method) for feature extraction. They evaluated the vowels dataset and achieved 99.2% accuracy using SVM for PD and gender classification. Diogo *et al.* [32] presented an approach for early diagnosis of PD using three distinct database consisting of vowel pronunciation in different language (Portuguese, uci data etc). They evaluated acoustic and phonetic characteristics of speech. Their work achieved 99.94% of highest accuracy using random forest.

B. CONVOLUTION NEURAL NETWORK-BASED METHODS

Recent studies show that neural networks immensely contribute to speech classification [13], [33]. Trinh and Darragh [12] proposed a convolution neural network-based approach for two PD datasets- Saarbrücken voice databases and pc-Gita dataset and achieved 96.7% accuracy for a pc-Gita dataset. Naranjo *et al.* [29] proposed a convolution neural network model for feature extraction and classification process. They considered articulation from /pa-ta-ka/ words, sentences and read text from pc-Gita and observed start, stop signal of speech. They extracted features from spectrograms and achieved 89% accuracy using Gaussian mixture modeling and i vector. Teixeira *et al.* [34] concentrated on discretizing a neural network and its information sources by utilizing quantization and weight scaling. They applied linear homomorphic encryption batching technique on the Spanish pc-Gita dataset. They observed the time of 1.4ms rather than the original approach where prediction took 4.5s to compute results. Arias-Vergara *et al.* [35] considered phonation, articulation, and prosody data from monologue recordings of extended version of the Spanish language dataset. They used support vector machine and convolution neural network-based model to extract the most suitable features and support vector machine for classification. The achieved 84% accuracy and showed prosody features are effective than the others.

III. PROPOSED METHODOLOGY

In this work, we proposed spectrogram and acoustic feature-based frameworks for Parkinson's disease classification. The first framework employs a transfer learning approach for speech spectrograms. In our second proposed method, we evaluated deep learning-based feature extraction from speech spectrograms while the third method considers the simple acoustic feature method for Parkinson disease detection. Deep learning has seen tremendous results in many fields such as computer vision, image processing and speech signal recognition [14]. In our work, we used pre-trained convolution neural network architecture Alexnet for feature extraction from speech signal and speech spectrograms. Fig 1 shows PD classification using the transfer learning

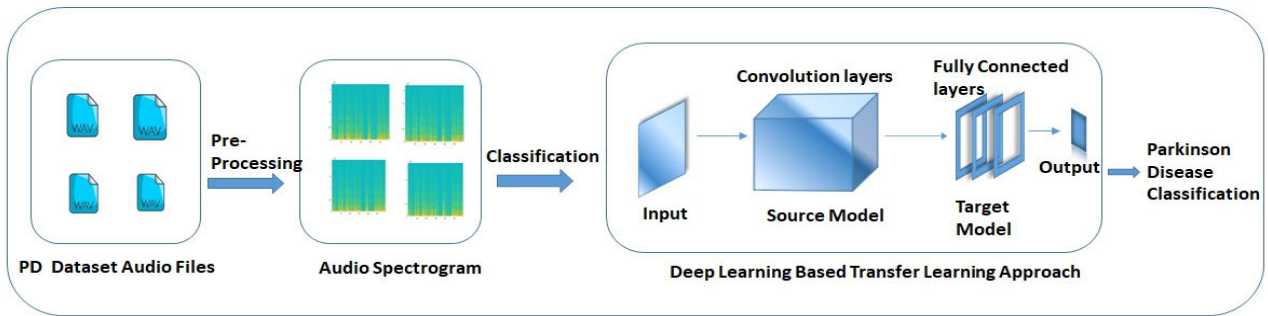


FIGURE 1. Spectrogram based transfer learning model for PD classification.

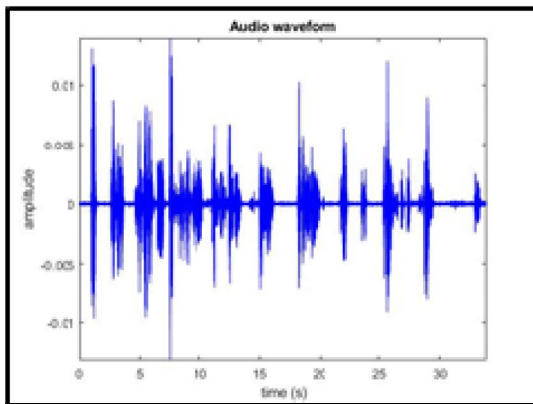


FIGURE 2. Waveform of monologue of healthy candidate.

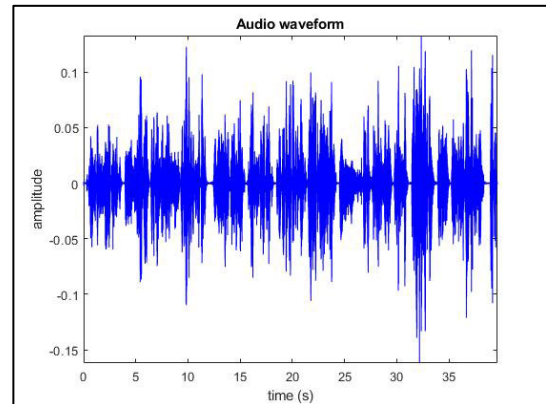


FIGURE 3. Waveform of monologue of parkinson patient.

method. In both models, the first step is signal preprocessing followed by deep feature extraction using Alexnet and hand-crafted feature extraction. We have performed classification using transfer learning and machine learning classifiers. All steps of the methodology section are explained in detail in following subsections.

A. SIGNAL PREPROCESSING

To input data into a classifier, the speech signals are first converted to spectrograms. A spectrogram is a visual representation of the signal spectrum that changes over time [36]. In our work, we converted Parkinson disease speech signal into a spectrogram. In the time domain, digitally sampled data is divided into segments that overlap and form Fourier transform that calculate the spectral amplitude of each segment. Each segment corresponds to vertical line in image. Table 1. Shows signal processing parameters, Fig 2. shows waveform of monologue pronounced by healthy candidate and Fig 3. shows waveform of Parkinson disease.

B. FEATURE EXTRACTION: DEEP FEATURES USING ALEXNET MODEL

In our work, we used deep learning and machine learning methods for the classification process. In order to input data into our classifiers, deep features were extracted [37], [38] from our speech signal dataset. We used deep learning convolution model Alexnet for extracting deep features from

TABLE 1. Signal processing parameters.

Sr.no	Parameter	Value
1.	Hamming Window	512
2.	No overlap	256
3.	NFFT	1024

the pc-Gita dataset. Our dataset consists of spectrograms of vowels \a, \e, \i, \o, \u, monologues and read text. Alexnet is a deep neural network-based model that consists of 8 layers. The first five layers form the convolution layer whereas the last three layers combine to form fully connected layers. Fig 4. shows feature extraction process using Alexnet. To input data into Alexnet model, we scaled spectrograms to fit into the model. Spectrograms obtained from signal preprocessing are of size 224×224 whereas Alexnet accepts input images of size 227×227 .

Thus, all images are scaled accordingly. Alexnet model has a convolution layer, hidden layers, and classification layer at the end. First five convolution layers of the network is trained on ImageNet data and last three fully connected layer are replaced with target Parkinson disease data. In this work,

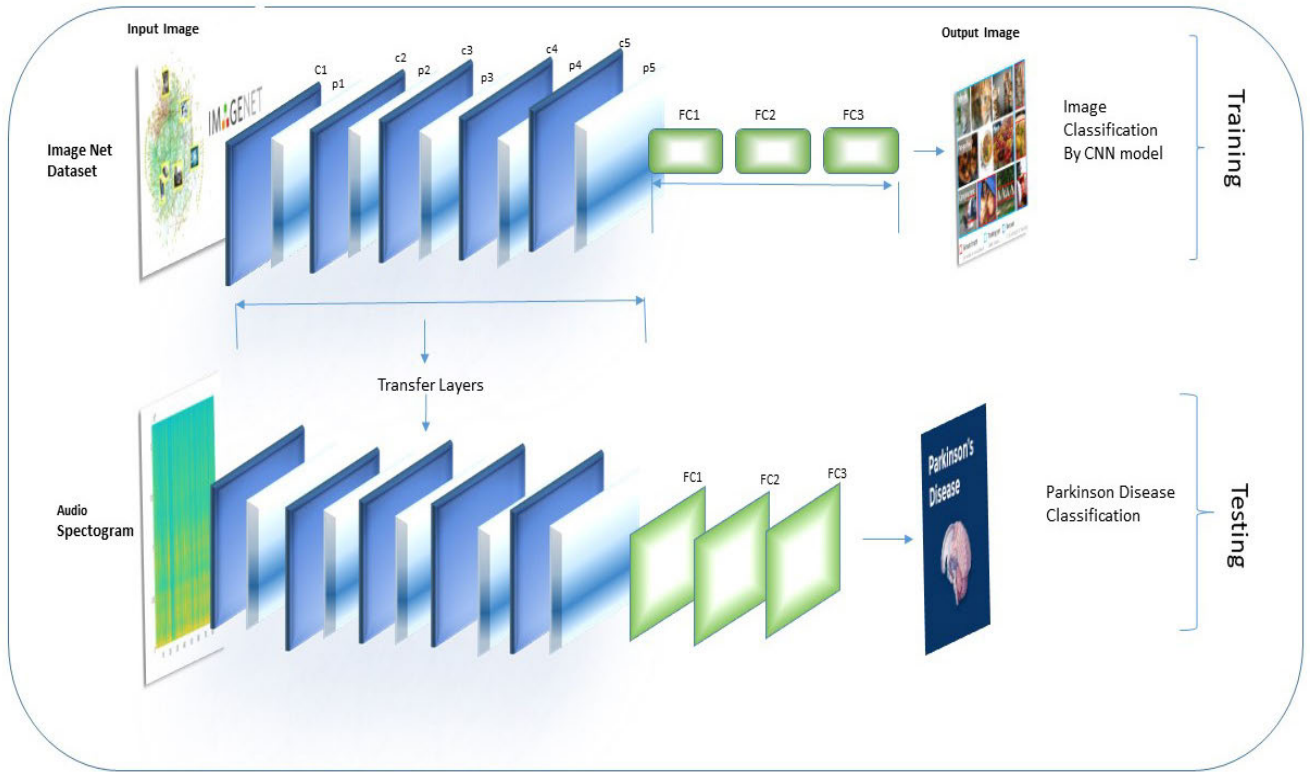


FIGURE 4. Transfer learning model using Alexnet.

TABLE 2. Deep features extracted using Alexnet.

Sr.no	Dataset	No of features
1.	Vowels	\a\ 98
		\e\ 150
		\i\ 150
		\o\ 150
		\u\ 150
2.	Monologue	50
3.	Read Text	50

we extracted features from the first five convolution layers thus fully connected classification layers are not used in this architecture. The model extracts deep shallow features from individual data vowels, monologues and reads text. Table 2. shows a total number of features extracted from each set of Parkinson’s Spanish speech data. The generic transfer learning architecture is presented here:

a: INPUT LAYER

In the transfer learning model, the first layer accepts input that is basically images of size 227×227 . We input RGB spectrograms (vowels, monologues, read the text) and each of them is given as a separate input.

$$Input = N_i \times W_i \times H_i \times D \quad (eq.1)$$

In eq.1, N_i represents a number of images, W_i is with for input image i , H_i is height D is the depth.

b: PRE-TRAINING USING ALEXNET

Initially, Alexnet model is trained on ImageNet data that consist of 1.2 million images. This model accepts images of size 227×227 so all RGB images are resized from 224×224 size to 227×227 to fit into the model. The first five layers of Alexnet is trained using ImageNet data. The fully connected layers $fc1$, $fc2$, $fc3$ are shown in Fig 4, are replaced by the pc-Gita Spanish speech dataset.

c: CONVOLUTION LAYER

This layer performs all computation and transfer it to the next fully connected layers. This layer generates convolution feature map $c1$, $c2$, $c3$, $c4$, $c5$ of input images and refers each feature map of the previous layer to the next layer.

d: POOLING LAYER

This layer reduces all computations and parameters to reduce network complexity. It reduces the dimensionality of input data by mixing the output of the previous layer with the input of the next layer.

e: FULLY CONNECTED LAYER

A fully connected layer connects neurons in a layer to each neuron in another. In principle, it is identical to the traditional multilayer perceptron. The dashed matrix passes through fully connected layers to classify data.

TABLE 3. Handcrafted features extracted using Mir toolbox.

Feature	No of features	Details
Root mean square(rmse)	6	Counts the square root of magnitude of speech recordings
Spectrum	6	Measures the difference between the highest and the lowest values
Zero crossing rate(zcr)	6	Identifies the signal at point zero
Pitch	6	Represents number of times vocal folds causes vibration while producing sound
Entropy	6	Measures the energy value of the signal
Mel- Frequency cepstral (mfcc)	85	Transforms the signal into small frames and find stft of all the divided frames and mfcc values of each signal
Statistical features	6	Slope, standard deviation, variance, mean etc.
Total	122	

TABLE 4. Varying parameter values in deep learning-based transfer learning approach.

Sr.no	Parameter	Value
1.	Epoch size	5-10
2.	Weighted learn rate	30-70
3.	Bias learn rate factor	50
4.	Batch Size	5-10
5.	Initial learn rate	1e-4

f: REPLACEMENT OF LAST LAYERS

To perform classification, we initially trained our Alexnet model on ImageNet data images embedded in it. To test our model's accuracy we replaced layers with our target data pc-Gita Spanish speech recordings' spectrograms.

g: NETWORK TRAINING

We trained our network for each set of dataset i-e vowels, the weighted learn rate is also varied at different points from 30 to 70 to assess accuracy at different learn rates. The bias factor is varied from 40 to 70 and batch sizes 5 to 10. The initial learn rate is set to 1e-4. Table 4 show complete details of the parameters.

C. FEATURE EXTRACTION: HANDCRAFTED FEATURE-BASED MODEL

In this feature extraction model, we extracted simple acoustic features from Spanish speech recordings. Separate sets of features were extracted for both Parkinson's disease patient

and healthy patient for each set of data monologues, vowels, read text and words. We extract simple acoustic features [39] of each recording. These features include spectral features and statistical features. Table 3 shows the details of features extracted for each recording and derivative of each feature.

D. PARKINSON DISEASE CLASSIFICATION

After signal preprocessing and feature extraction, we performed classification. To classify PD patients, we utilized deep learning convolution neural network architecture and machine learning models; support vector machine, random forest, and multilayer perceptron. The following subsection explains methods for classification in detail.

1) TRANSFER LEARNING APPROACH.

The transfer learning method is the most widely used technique in the deep learning model. We trained model on base data and utilized it to learn features and transfer it to target data [40], [41] using Alexnet model [42], [43].

2) MACHINE LEARNING BASED APPROACH

In this work, we utilized machine learning classifiers for Parkinson disease detection. We used simple and deep features that were extracted using feature extraction shown in Fig 5 using Alexnet. We performed 5 cross-validation for each set of our data. The following sections explain classifiers in detail with varying parameters used to train and test data.

a: SUPPORT VECTOR MACHINE

The support vector machine has been widely used in different fields like computer-aided diagnostic system, recognition and vision system [37], [38]. It is the most popular used machine learning model for binary classification due

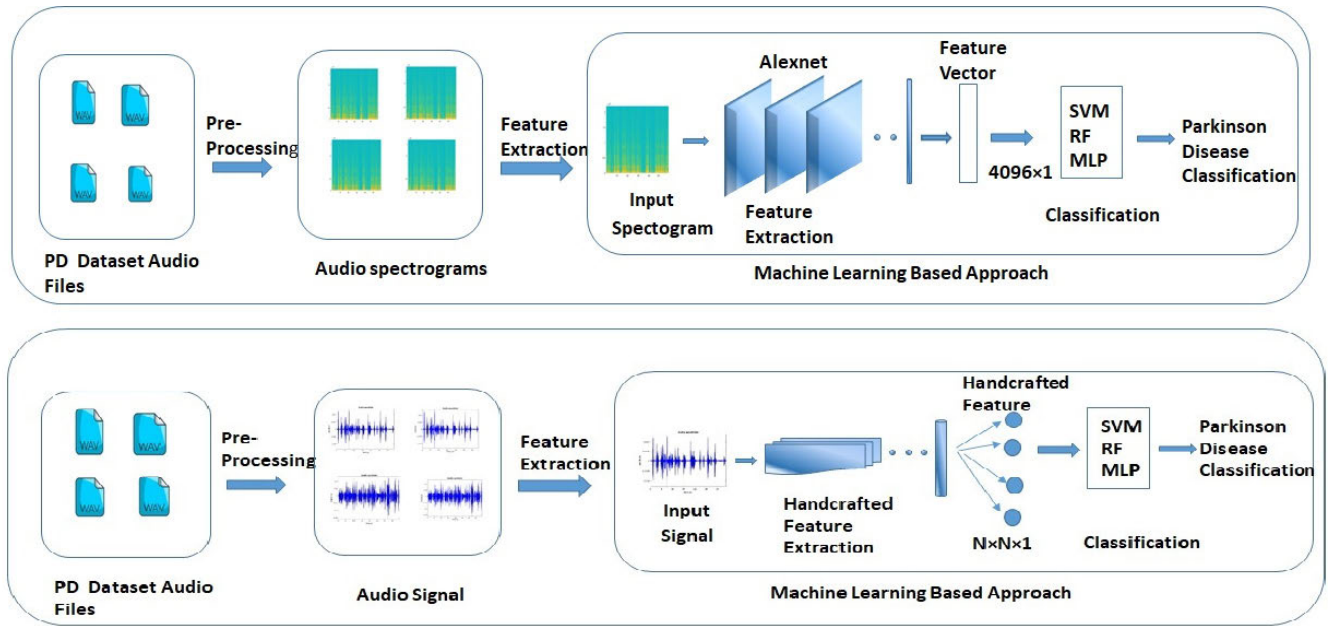


FIGURE 5. Machine learning-based approach for PD classification using deep features and handcrafted features.

to its generalizability. It generates different support vectors of the given input. It identifies the linear and non-linear surfaces in the input support vectors by constructing hyperplane, which later classifies the data. The complexity parameters in this model build hyperplane from the class label. The hyperplane that computes the largest distance from the training data has the highest classification result. The parameters like gamma rate are set to 0.01, complexity value is 1.0, the degree value is 3 and the coefficient *coef0* is 1. We have used the linear in our research. In the linear kernel, the projection for input task is considered by dot product for the input value *y* and the support vector *y_i* is calculated.

$$f(y) = B(0) + \text{sum}(z_i(y, y_i)) \tag{eq.2}$$

In eq.2, coefficient B (0) and *z_i* is estimated for each input value *y* and evaluated by learning a data learning algorithm.

b: RANDOM FOREST

An ensemble method widely used in the classification processes that make use of different decision trees for classifying data [44]. It builds the bootstrap templates from the random forest original data and grows a raw classification or regression tree for each bootstrap template. It considers each node instead of choosing the only best disclosure from all predictors. It performs a random selection of predictors and chooses the best split between them as shown in Fig 6. In our research, we utilized default parameters for random forest.

c: MULTILAYER PERCEPTRON

A multilayer perceptron is an artificial neural network that are broadly utilized in speech, image, and vision recognition

system. It has been observed from recent research that a multilayer perceptron is extensively used in the medical diagnostic field [43], [44]. It is a feedforward neural network that is made up of an input layer, an output layer and in between, there is a hidden layer. The input layer accepts the input value whereas the hidden layer sends information from the input layer to the output layer as shown in Fig 6. A hidden layer consists of a number of neurons, each of the hidden layer neuron has information for its input influencing it by growing them by their connection weights. The output of each neuron is defined as

$$w_i = f(\sum x_{ii}z_i) \tag{eq.3}$$

In eq.3, *f* is defined as an activation function, which is proportional to input weights. It is mostly some threshold value a simple sigmoid or a hyperbolic tangent function. The learning rate for the multilayer perceptron ranges from 0 to 1 where 0.3 is set as a default value.

IV. EXPERIMENTAL SETUP

A. DATASET

We utilized PC-GITA [15] Spanish language dataset. The dataset consists of a Spanish speech signal recording of 50 people that are PD patients and 50 HC people as shown in Table 5. The dataset includes recordings of 25 male and 25 female persons. The dataset belongs to Spanish language Table 5 depicts dataset details. The dataset consists of the recording of vowels, monologues and read the text. Each recording consists of different voice features that are discussed below.

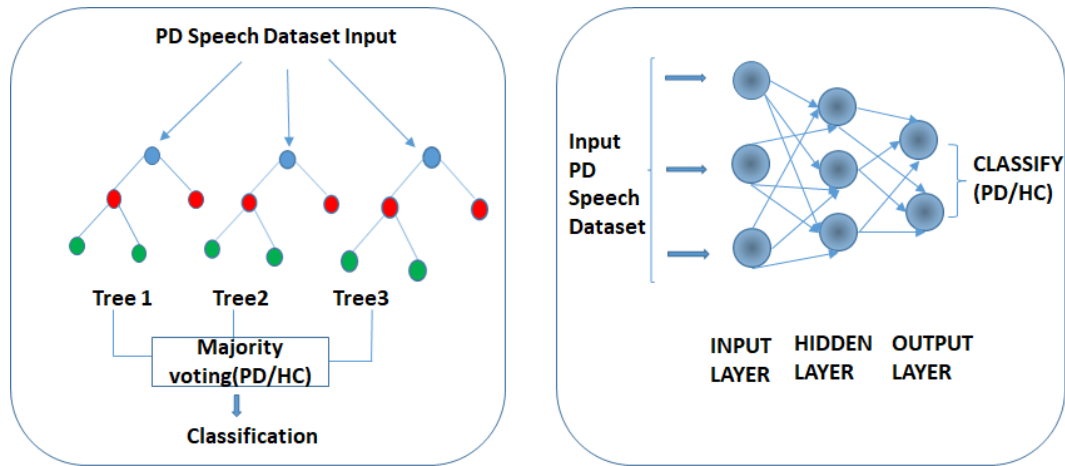


FIGURE 6. Random forest and multilayer perceptron model.

1) PHONATION

The phonation analysis in continuous speech is performed by extracting voiced segments from the utterance. The feature set includes seven descriptors such as jitter and shimmer. The first and second derivatives of F0, long term perturbation features such as the amplitude perturbation quotient, the pitch perturbation quotient, and the energy.

2) PROSODY

The prosody features are based on duration, the F0 contour, and the energy contour. We computed 13 features per utterance including the average, standard deviation, and maximum value of F0.

3) ARTICULATION

The articulatory capability of the patients is evaluated with information from the onset/offset transitions to model the difficulties of patients to start/stop the movement of the vocal folds. The set of features extracted from the onset and offset includes 12 Mel-Frequency Cepstral Coefficients (MFCCs) with their first and second derivatives.

B. EVALUATION METRICS

The results obtained after classification from deep and machine learning models are evaluated using the following evaluation metrics explained in the below subsection.

1) ACCURACY

It is defined as the total number of samples that are truly classified and the total number of negatively classified results.

$$Accuracy = \left(\frac{TP + TN}{Total} \right) \quad (eq.4)$$

In this TP is true positive and TN is true negative, and total represents total number of class predictions.

2) SENSITIVITY

Sensitivity is the ability of a test to correctly identify those with the disease (true positive rate)

$$Sensitivity = \frac{TP}{TP + TN} \quad (eq.5)$$

3) SPECIFICITY

is the ability of the test to correctly identify those without the disease (true negative rate).

$$Specificity = \frac{TN}{TN + TP} \quad (eq.6)$$

4) F1 SCORE

It is defined as a ratio or numerical average of precision and recall values from the classification result.

$$f1\ score = 2 \times \frac{precision \times recall}{precision + recall} \quad (eq.7)$$

V. RESULTS

This section explains in detail the results obtained transfer learning, deep feature-based, and machine learning approach.

A. TRANSFER LEARNING BASED APPROACH RESULTS

In our first framework, the initial step comprises the conversion of Spanish speech recordings into spectrograms.

All speech recordings are transformed into their relative spectrograms by using parameters described in signal preprocessing. In transfer, the learning approach model is trained using a source dataset, which is replaced by our target dataset [6]. Once our model is trained on the source dataset, we replaced it with our target dataset speech recordings spectrograms. The spectrograms of each set of data i.e. vowels, monologues, read text and words are individually considered by varying different parameters. The model is trained by varying parameters discussed in the network training section discussed earlier. The corresponding results obtained by varying parameters are shown in Table 6. The results depicted in

TABLE 5. Dataset table consists of details of data with total number of recordings and age range of Parkinson disease patient recordings and healthy candidate for each gender (M/F).

Data		PD	HC	GENDER	AGE (Male/Female)
Vowels	A	150	150		PDP 62.2±11.2M
	E	150	150	75 M 75 F	60.1±7.8 F
	I	150	150		
	O	150	150		HC 61.2±11.3 M 60.7±7.7 F
	U	150	150		
Words	Apto	50	50	25 M 25 F	
	Atelta	50	50	25 M 25 F	
		50	50	25 M 25 F	

the vowel dataset that vowel \a\ shows the highest accuracy of 73.7% while vowel \e\ shows the highest accuracy of 75%. In the case of vowels, the accuracy achieved for \i\ is 81.4% and vowel \o\ achieved maximum accuracy of 72.6%. Vowel \u\ attained 82.3% on epoch size 9. The read text dataset showed the highest accuracy of 91% on epoch size 8 where the weighted learn rate is 30. For monologues, 86.36% highest accuracy observed on weighted learn rate 30 and epoch size 7.

For words, /apto/ we achieved 77.2 % accuracy on epoch size 7 and weighted learn rate 60, whereas for \atelta\ we achieved 73.7% accuracy on epoch size 9 and weighted learn rate 40. It is observed from the above results that the highest accuracy is observed in reading text 91% that is a major contribution using transfer learning in this research. Fig 7 shows the training process for reading text data. Figure 7 depicts training done using epoch size 6, blue lines depict the training process that is smoothed, light blue lines show the actual training process and black lines show the accuracy achieved on validation data. The accuracy achieved in this figure is 77.27% on epoch size 7 and weighted learn rate 40. Fig 8 shows the loss observed in read text data on epoch size 6.

B. MACHINE LEARNING BASED APPROACH RESULTS

1) DEEP FEATURES BASED RESULTS

In our second framework, we evaluated our dataset Spanish speech signal dataset using a machine learning model. The initial step comprises of deep feature extraction. The Spanish speech recordings spectrograms obtained after signal preprocessing steps are used to extract the speech features. In this approach, spectrograms are used as input into the feature extraction Alexnet model. The model extracts deep features from spectrograms as shown in the transfer learning architecture model in the above sections. The total number of features extracted for the vowel recordings dataset is 150. The 50 features are extracted for monologues, read text and word dataset. The extracted features for each of the data

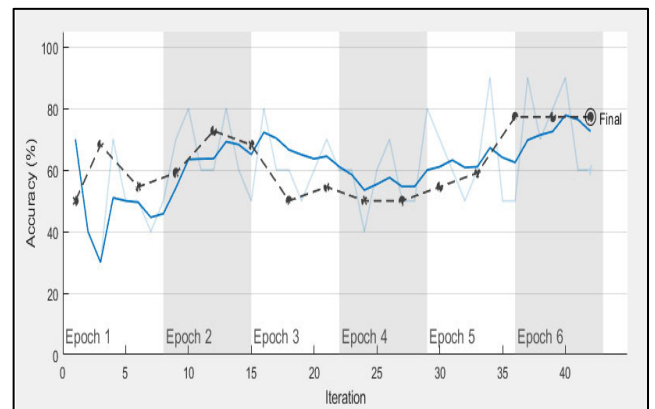


FIGURE 7. Accuracy observed during training process.

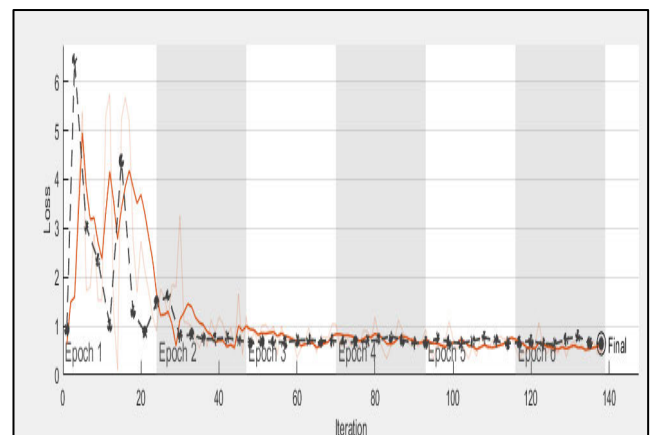


FIGURE 8. Loss observed during training process.

separately input into different machine learning classifiers. The support vector machine, random forest and multilayer perceptron are separately validated on our Spanish speech recordings dataset. All of these classifiers used five-fold cross-validation. The support vector machine results show

TABLE 6. Accuracy obtained using transfer learning model by varying parameters, bias learn rate 50, initial learn rate is $1e-4$, epoch size vary from 6-10.

Sr. No	Dataset	Weighted Learn Rate	Epoch 6 %	Epoch 7 %	Epoch 8 %	Epoch 9 %	Epoch 10 %	Sensitivity	Specificity
1.	\a\	30	58.10	57.60	66.0	61.60	57.50	0.53±0.3	0.41±0.25
2.	\a\	40	55.40	63.1	64.50	64.70	58.80	0.5±0.4	0.42±0.23
3.	\a\	60	67.70	60.30	61.70	73.70	61.60	0.62±0.1	0.5±0.2
4.	\e\	30	60.30	61.2	64.20	71.20	67.10	0.441±0.1	0.49±0.27
5.	\e\	40	72.10	62.60	75.00	66.30	58.70	0.53±0.2	0.5±0.3
6.	\e\	60	66.20	64.10	70.10	61.30	66.40	0.6±0.1	0.6±0.2
7.	\i\	30	75.60	67.56	66.10	71.30	67.20	0.54±0.2	0.6±0.1
8.	\i\	40	73.40	58.6	67.40	81.40	63.70	0.52±0.21	0.45±0.2
9.	\i\	60	73.60	72.10	61.20	64.50	70.20	0.6±0.1	0.56±0.2
10.	\o\	30	64.10	67.30	63.99	69.10	63.40	0.63±0.2	0.6±0.2
11.	\o\	40	69.50	57.10	67.10	72.30	72.10	0.57±0.21	0.2±0.5
12.	\o\	60	60.59	72.60	61.10	69.90	66.10	0.6±0.12	0.5±0.2
13.	\u\	30	61.10	75.60	66.30	61.10	67.80	0.64±0.1	0.5±0.23
14.	\u\	40	76.10	54.00	63.40	61.0	67.90	0.6±0.1	0.5±0.2
15.	\u\	60	54.00	60.20	51.60	82.30	73.20	0.5±0.2	0.5±0.3
16.	Read Text	30	63.60	68.01	91.77	68.10	68.30	0.5±0.39	0.4±0.3
17.	Read Text	40	59.10	77.27	59.10	65.80	68.10	0.5±0.3	0.4±0.4
18.	Read Text	60	40.10	59.80	68.10	59.10	67.90	0.5±0.1	0.5±0.25
19.	Monologue	30	63.50	86.36	50.20	63.40	72.30	0.45±0.3	0.3±0.45
20.	Monologue	40	68.10	41.72	59.00	77.20	63.60	0.4±0.3	0.34±0.3
21.	/apto/	30	40.00	77.34	68.20	77.20	63.50	0.±40.23	0.4±0.2
22.	/apto/	40	36.10	54.20	63.00	63.10	68.10	0.30±.4	0.3±0.3
23.	/apto/	60	63.10	77.20	72.30	68.20	50.00	0.55±0.2	0.54±0.2
24.	/atelta/	30	72.00	45.67	63.40	59.10	63.20	0.4±0.3	0.4±0.2
25.	/atelta/	40	45.00	54.49	50.10	73.70	59.50	0.4±0.3	0.3±0.3
26.	/atelta/	60	68.00	54.32	54.50	59.10	54.20	0.5±0.1	0.42±0.2

that the highest accuracy obtained for vowel \o\ is 93% whereas the least accuracy result is 83% on the monologue dataset. The highest accuracy in this model obtained is on vowel \e\ 99.4% whereas other vowel results also show an average accuracy of 99.1%.

The multilayer perceptron showed the highest accuracy of 99.7% on vowel \e\, \i\, \o\ whereas on the other data this model outperforms the other classifiers results. The results of the machine learning model clearly depict that vowels are sufficient in the classification of Parkinson's disease. Fig 9. depicts the accuracy obtained for each classifier. Minimum accuracies obtained using a machine learning approach is using support vector machine 83%. In the monologues dataset, the highest 99.7% accuracy is obtained using a multilayer perceptron.

2) HANDCRAFTED FEATURE-BASED RESULTS

Different acoustic features are used for Parkinson's disease detection. In this work, we consider handcrafted acoustic features from the Spanish speech dataset. This part of our research work performs a comparison with our deep feature-based machine learning model and transfer learning classification. Fig 10. depicts the results obtained using simple

acoustic features. Fig 10. depicts the results obtained for each set of data. It is observed that the highest accuracy is observed on vowel \e\ that is 84.6% using random forest, vowel \o\ showed 83.6% accuracy and vowel \a\ presents 83% accuracy while vowel \i\ show 82.8% accuracy using random forest. Least accuracy in the vowel dataset is observed using vowel \u\. In the case of the reading text dataset, the highest accuracy 73% is observed using the random forest. Monologues dataset showed a bad accuracy of 37% using a multilayer perceptron and 15% using random forest. Thus, it is observed that vowels are efficient in Parkinson disease detection when handcrafted features are utilized. However, in comparison to deep features, this accuracy is far less as the highest 99.7% accuracy is achieved. Thus, deep features based methods outperform other methods.

C. COMPARATIVE ANALYSIS

This section performs a brief comparison of recent existing techniques with the proposed technique in this research work shown in Table 7. In this research work, we evaluated deep learning and machine learning-based approaches. Each of these approaches considered a pc-Gita dataset that consists of Spanish speech recordings which include pronunciation of

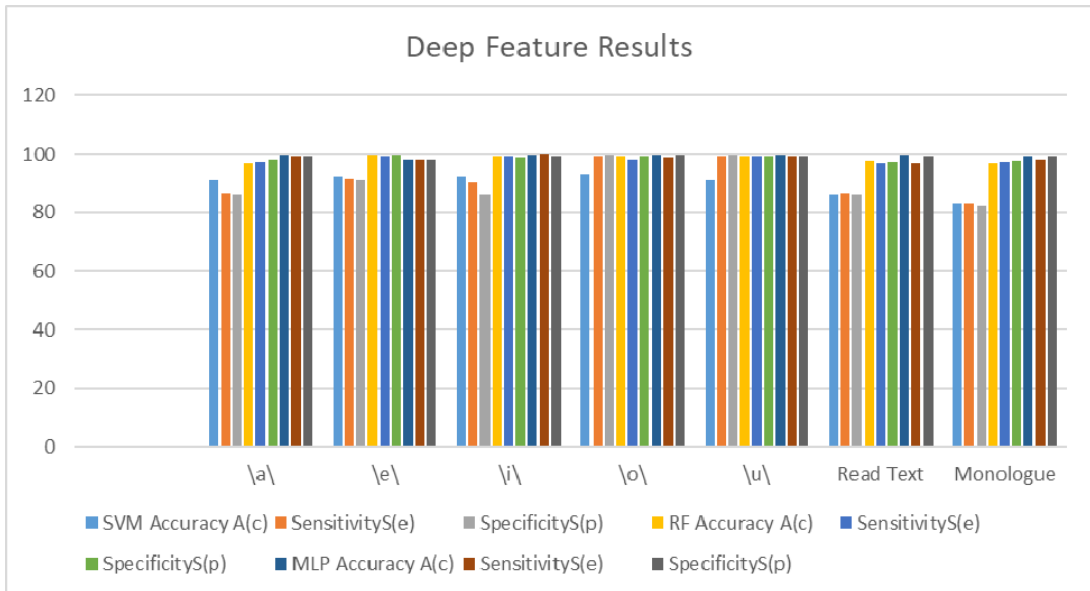


FIGURE 9. Accuracy observed using deep features.

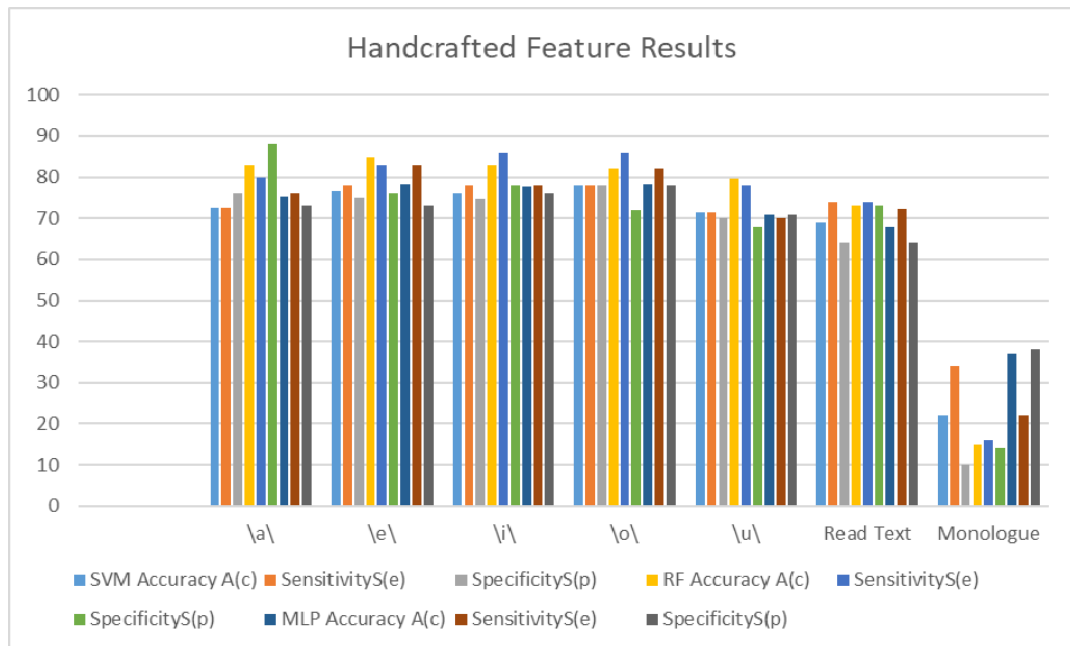


FIGURE 10. Accuracy observed using handcrafted features.

vowels, monologues, words and read the text. All of them are separately used as an input in both our model and their corresponding accuracies are recorded. The results clearly depict that deep feature-based method outperformed by presenting 98.3% of average accuracy for random forest and 99.3% for multilayer perceptron. In comparison to Karan *et al.* [20] incorporated inherent based features from two different sets of data observed 96% accuracy when classified using random forest and support vector machine. These results are better than the already published work for the same datasets which are presented in Table 7.

In comparison to our transfer learning approach Trinh and Darragh [12] used a convolution neural network in their work. Tuncer and Dogan [31] assessed the extended version of Spanish recordings corpus using a convolution neural network. Trinh and Darragh [12] used Gaussian mixture modeling and i vector in their proposed work. The comparison shows that our proposed machine learning method outperformed existing techniques, however, the transfer learning approach compared with existing techniques does not prove to be as much as accurate than the convolution neural network research that has been already

TABLE 7. Comparative analysis of proposed method with existing techniques.

Sr.no	Author	Technique	Data Sample	Accuracy
1.	Karan <i>et al.</i> [20]	RF, SVM	Pc-Gita (Vowel \a\,o\)	96%
2.	L.Velazquez <i>et al.</i> [18]	FGMM	Pc-Gita (Vowel, monologues, Sentences, words, read text)	81%
3.	Reuda <i>et al.</i> [6]	RF, SVM	Pc-Gita (Vowel, words)	70%
4.	Para G. <i>et al.</i> [4]	SVM	Pc-Gita (words)	88%
5.	Vasquez <i>et al.</i> [27]	BRR	Pc-Gita (Vowel, words, sentences)	94.9%
6.	Arias <i>et al.</i> [35]	SVM, CNN	Pc-Gita (monologue)	84%
7.	Travieso <i>et al.</i> [44]	SVM	Pc-Gita (Vowel, monologues, Sentences, words, read text)	96%
8.	Orozco <i>et al.</i> [26]	SVM	Pc-Gita (Vowels)	81-97%
9.	Arias <i>et al.</i> [7]	SVM	Pc-Gita (Vowels)	95%
10.	Nam H. <i>et al.</i> [12]	CNN	Pc-Gita (Vowel, monologues, Sentences, words, read text)	96.7%
11.	Arias <i>et al.</i> [8]	GMM	Pc-Gita (read text, sentences)	0.77
	Proposed Research Work	Deep Features with RF and MLP	Pc-Gita (Vowel monologues, Sentences, words, read text)	98.3% 99.3%

presented [45], [46]. Parkinson disease shows several motor and non-motor symptoms. Gait, tremor and handwriting problem appear overtime. However, changes in the speech and handwriting start appearing early and are more evident. Changes in speech show a clear distinction between a healthy and PD patient. They suffer pause in pronouncing as well as jarring sound. Classification of PD is crucial; however, speech impairments are considered as an important biomarker for PD detection. However, it is essential to perform the speech collection of PD in a noise-free environment. Speech analysis using deep learning-based methods proved as a very good method because the variation in speaking can be modeled effectively. However, the symptoms of Parkinson can be observed in handwriting and gait. Therefore, the PD identification can be done in a more efficient way by combining all the above-mentioned biomarkers.

D. CONCLUSION

Parkinson's disease is one of the common diseases among people worldwide. Early diagnosis of the disease is an open research and many researchers have shown significant work in achieving the highest accuracy for its detection and diagnostic. In our work, we used Alexnet model for deep feature extraction and handcrafted feature extraction from the Spanish speech recordings dataset. For classification, we used transfer learning, deep feature and acoustic-phonetic feature-based methods. We proposed that deep features extracted using Alexnet are efficient to distinguish Parkinson patients from healthy patients. In our proposed models, random forest, multilayer perceptron, transfer learning achieved highest accuracy of 99%, 99.7%, 72% respectively. For vowels \o\ the same classifiers achieved 99%, 99.6, 76%. For read

text, the same classifiers achieved 97.8%, 99%, 91% whereas for monologues 97%, 99.3%, 86.36% accuracy is achieved. Thus, our work shows that speech analysis using deep features are efficient in distinguishing between healthy and Parkinson patient with high accuracy. Further, we also conclude that pronunciation recordings of vowels are enough in distinguishing patients from healthy. In future, the gait, tremor and other symptoms data can be assessed together using a deep feature method to identify that to what extent this approach is suitable with other scenarios. Furthermore, the feature selection process can be used for handcrafted feature method.

REFERENCES

- [1] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson disease epidemiology, pathology, genetics, and pathophysiology," *Clinics Geriatric Med.*, vol. 36, no. 1, pp. 1–12, Feb. 2020.
- [2] A. E. Lang and A. M. Lozano, "Parkinson's disease," *New England J. Med.*, vol. 339, pp. 1130–1143, 1998.
- [3] D. Georgiev, M. Domellof, K. Hamberg, L. Forsgren, and G. M. Hariz, "Sex differences, quality of life and non-motor symptoms in Parkinson's disease," *Neurology*, vol. 85, no. 1, Jul. 2015.
- [4] L. F. Parra-Gallego, T. Arias-Vergara, J. C. Vásquez-Correa, N. Garcia-Ospina, J. R. Orozco-Arroyave, and E. Nöth, "Automatic intelligibility assessment of Parkinson's disease with diadochokinetic exercises," in *Proc. Workshop Eng. Appl.*, 2018, pp. 223–230.
- [5] N. Hosseini-Kivanani, J. C. Vásquez-Correa, M. Stede, and E. Nöth, "Automated cross-language intelligibility analysis of Parkinson's disease patients using speech recognition technologies," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2019, pp. 74–80.
- [6] A. Rueda, J. C. Vásquez-Correa, C. D. Rios-Urrego, J. R. Orozco-Arroyave, S. Krishnan, and E. Nöth, "Feature representation of pathophysiology of parkinsonian dysarthria," in *Proc. Interspeech*, Sep. 2019, pp. 3048–3052.
- [7] T. Arias-Vergara, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, "Parkinson's disease and aging: Analysis of their effect in phonation and articulation of speech," *Cogn. Comput.*, vol. 9, no. 6, pp. 731–748, Dec. 2017.

- [8] T. Arias-Vergara, J. C. Vázquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions," *Speech Commun.*, vol. 101, pp. 11–25, Jul. 2018.
- [9] E. Rastegari, S. Azizian, and H. Ali, "Machine learning and similarity network approaches to support automatic classification of parkinson's diseases using accelerometer-based gait analysis," in *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 4231–4242.
- [10] C. D. Rios-Urrego, J. C. Vázquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera, and J. R. Orozco-Arroyave, "Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features," *Comput. Methods Programs Biomed.*, vol. 173, pp. 43–52, May 2019.
- [11] R. Castrillon, A. Acien, J. R. Orozco-Arroyave, A. Morales, J. F. Vargas, R. Vera-Rodríguez, J. Fierrez, J. Ortega-García, and A. Villegas, "Characterization of the handwriting skills as a biomarker for parkinson disease," 2019, *arXiv:1903.08226*. [Online]. Available: <http://arxiv.org/abs/1903.08226>
- [12] N. Trinh and O. B. Darragh, "Pathological speech classification using a convolutional neural network," in *Proc. IMVIP, Ireland*, 2019, pp. 72–75.
- [13] R. Miikkulainen, J. Liang, E. Meyerson, A. Rawal, D. Fink, and O. Francon, "Evolving deep neural networks," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 293–312.
- [14] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
- [15] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proc. LREC*, 2014, pp. 342–347.
- [16] J. A. Nichols, H. W. Herbert Chan, and M. A. B. Baker, "Machine learning: Applications of artificial intelligence to imaging and diagnosis," *Biophys. Rev.*, vol. 11, no. 1, pp. 111–118, Feb. 2019.
- [17] H. R. Pereira and H. A. Ferreira, "Classification of patients with Parkinson's disease using medical imaging and artificial intelligence algorithms," in *Proc. Medit. Conf. Med. Biol. Eng. Comput.*, 2019, pp. 2043–2056.
- [18] L. Moro-Velázquez, J. A. Gomez-García, J. I. Godino-Llorente, J. Villalba, J. Ruz, S. Shattuck-Hufnagel, and N. Dehak, "A forced Gaussians based methodology for the differential evaluation of Parkinson's disease by means of speech processing," *Biomed. Signal Process. Control*, vol. 48, pp. 205–220, Feb. 2019.
- [19] P. Pérez-Toro, J. Vázquez-Correa, M. Strauss, J. Orozco-Arroyave, and E. Nöth, "Natural language analysis to detect Parkinson's disease," in *Proc. Int. Conf. Text, Speech, Dialogue*, 2019, pp. 82–90.
- [20] B. Karan, S. S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function based features from speech signal," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 249–264, Jan. 2020.
- [21] A. Kacha, F. Grenet, J. R. Orozco-Arroyave, and J. Schoentgen, "Principal component analysis of the spectrogram of the speech signal: Interpretation and application to dysarthric speech," *Comput. Speech Lang.*, vol. 59, pp. 114–122, Jan. 2020.
- [22] J. C. Vázquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of Parkinson's disease: A deep learning approach," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1618–1630, Jul. 2019.
- [23] N. García, J. R. Orozco-Arroyave, L. F. D'Haro, N. Dehak, and E. Nöth, "Evaluation of the neurological state of people with Parkinson's disease using i-vectors," in *Proc. Interspeech*, Aug. 2017, pp. 299–303.
- [24] P. Klumpp, T. Janu, T. Arias-Vergara, J. C. Vázquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Apkinson—A mobile monitoring solution for Parkinson's disease," in *Proc. Interspeech*, Aug. 2017, pp. 1839–1843.
- [25] L. Moro-Velázquez, J. A. Gómez-García, J. I. Godino-Llorente, J. Villalba, J. R. Orozco-Arroyave, and N. Dehak, "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's disease," *Appl. Soft Comput.*, vol. 62, pp. 649–666, Jan. 2018.
- [26] J. R. Orozco-Arroyave, J. C. Vázquez-Correa, J. F. Vargas-Bonilla, R. Arora, N. Dehak, P. S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva, H. Chinaei, A. Vann, N. Vogler, T. Bocklet, M. Cernak, J. Hannink, and E. Nöth, "NeuroSpeech: An open-source software for Parkinson's speech analysis," *Digit. Signal Process.*, vol. 77, pp. 207–221, Jun. 2018.
- [27] J. C. Vázquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth, "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease," *J. Commun. Disorders*, vol. 76, pp. 21–36, Nov. 2018.
- [28] O. Yaman, F. Ertam, and T. Tuncer, "Automated Parkinson's disease recognition based on statistical pooling method using acoustic features," *Med. Hypotheses*, vol. 135, Feb. 2020, Art. no. 109483.
- [29] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications," *Comput. Methods Programs Biomed.*, vol. 142, pp. 147–156, Apr. 2017.
- [30] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait," *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 113075.
- [31] T. Tuncer and S. Dogan, "A novel octopus based Parkinson's disease and gender recognition method using vowels," *Appl. Acoust.*, vol. 155, pp. 75–83, Dec. 2019.
- [32] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, "Automatic detection of Parkinson's disease based on acoustic analysis of speech," *Eng. Appl. Artif. Intell.*, vol. 77, pp. 148–158, Jan. 2019.
- [33] J. C. Vázquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Convolutional neural network to model articulation impairments in patients with Parkinson's disease," in *Proc. Interspeech*, Aug. 2017, pp. 314–318.
- [34] F. Teixeira, A. Abad, and I. Trancoso, "Privacy-preserving paralinguistic tasks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6575–6579.
- [35] T. Arias-Vergara, J. C. Vázquez-Correa, J. R. Orozco-Arroyave, P. Klumpp, and E. Nöth, "Unobtrusive monitoring of speech impairments of Parkinson's disease patients through mobile devices," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6004–6008.
- [36] M. Stolar, M. Lech, R. S. Bolia, and M. Skinner, "Acoustic characteristics of emotional speech using spectrogram image classification," in *Proc. 12th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2018, pp. 1–5.
- [37] J. I. Forcén, M. Pagola, E. Barrenechea, and H. Bustince, "Aggregation of deep features for image retrieval based on object detection," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2019, pp. 553–564.
- [38] F. Nazir, M. N. Majeed, M. A. Ghazanfar, and M. Maqsood, "Mispronunciation detection using deep convolutional neural network features and transfer learning-based model for Arabic phonemes," *IEEE Access*, vol. 7, pp. 52589–52608, 2019.
- [39] C. Castro, E. Vargas-Viveros, A. Sánchez, E. Gutiérrez-López, and D.-L. Flores, "Parkinson's disease classification using artificial neural networks," in *Proc. Latin Amer. Conf. Biomed. Eng.*, 2019, pp. 1060–1065.
- [40] M. Shu, "Deep learning for image classification on very small datasets using transfer learning," M.S. thesis, Iowa State Univ., Ames, IA, USA, 2019.
- [41] S. Afzal, M. Javed, M. Maqsood, F. Aadil, S. Rho, and I. Mehmood, "A segmentation-less efficient alzheimer detection approach using hybrid image features," in *Handbook of Multimedia Information Security: Techniques and Applications*. Springer, 2019, pp. 421–429.
- [42] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*. [Online]. Available: <http://arxiv.org/abs/1803.01164>
- [43] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O.-Y. Song, "Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans," *Sensors*, vol. 19, no. 11, p. 2645, Jun. 2019.
- [44] C. M. Travieso, J. B. Alonso, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, E. Nöth, and A. G. Ravelo-García, "Detection of different voice diseases based on the nonlinear characterization of speech signals," *Expert Syst. Appl.*, vol. 82, pp. 184–195, Oct. 2017.
- [45] K. Seppi, K. Ray Chaudhuri, M. Coelho, S. H. Fox, R. Katzenschlager, and S. Perez Lloret, "Update on treatments for nonmotor symptoms of Parkinson's disease—An evidence-based medicine review," *Movement Disorders*, vol. 34, pp. 180–198, Feb. 2019.
- [46] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019.



LAIBA ZAHID received the B.S. degree (Hons.) from the Comsats University Islamabad–Attock, in 2017, where she is currently pursuing the M.S. degree in computer science. Her research interests include machine learning and signal processing.



MUAZZAM MAQSOOD received the Ph.D. degree from UET Taxila, in 2017. He is currently an Assistant Professor with COMSATS University Islamabad–Attock, Pakistan. His research interests include medical imaging, machine learning’s, recommender systems, and image processing.



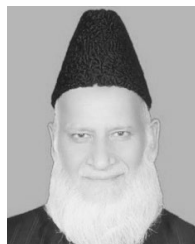
MEHR YAHYA DURRANI graduated in information technology. He received the master’s degree in computer science. He is currently pursuing the Ph.D. degree in pattern recognition. He is also an Assistant Professor with COMSATS University Islamabad–Attock, Pakistan.



MAHEEN BAKHTYAR received the master’s and Ph.D. degrees from the Asian Institute of Technology (AIT), Thailand, with one year of research experience from the National Institute of Informatics, Tokyo, Japan. She is currently working as an Assistant Professor with the Department of Computer Science and Information Technology, University of Balochistan, Pakistan. Her research interests mainly include information/knowledge management and retrieval, natural language processing, sentiment analysis, text processing, language understanding, question answering systems, and ontology processing.



JUNAID BABER received the M.S. and Ph.D. degrees in computer science from the Asian Institute of Technology, Thailand. He has spent one year as a Research Scientist with the National Institute of Informatics, Tokyo. He is currently working as a faculty member with the University of Balochistan, Quetta. His research interests lie in machine learning, high performance computing, and data analytics.



HABIBULLAH JAMAL received the B.Sc. degree in EE from the University of Engineering and Technology, Lahore, Pakistan, in 1974, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Canada, in 1979 and 1982, respectively.

He is currently a Professor of engineering sciences with the Ghulam Ishaq Khan Institute, Topi, Pakistan. He is the author of two textbooks and 132 research articles. His research interests include (but not limited to) signal processing, the design of microelectronic circuits and development of novel computer architectures for telecommunication, and national defense and other applications.

Dr. Jamal was a recipient of prestigious national level awards 8th TERADATA National IT Excellence Awards for Excellence in IT Education, in 12 April 2008; Performance Excellence in Engineering Awarded by The Institution of Engineers, Pakistan, in the Engineers Day 29 May 2007; 9th Pakistan Education Forum, National Education Award in 2003, and the National Book Council of Pakistan Award in 1991.



IRFAN MEHMOOD is currently a Senior Lecturer with the University of Bradford, U.K. His sustained contribution at various research and industry-collaborative projects gives him an extra edge to meet the current challenges faced in the field of multimedia analytics. Specifically, he has made significant contribution in the areas of video summarization, medical image analysis, visual surveillance, information mining, deep learning in industrial applications, and data encryption.



OH-YOUNG SONG received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering and Computer Science, Seoul National University, South Korea, in 1998, 2000, and 2004, respectively. He was a Postdoctoral Fellow at the School of Electrical Engineering and Computer Science, Seoul National University, from 2004 to 2006. He is currently an Associate Professor with the Department of Software, Sejong University, South Korea. His research interests include computer graphics, simulation, and machine learning. Especially, he has made a contribution in the areas of physics-based animation, human motion, numerical algorithms, VR/AR, medical image analysis, and deep learning.

• • •