

An optimized RNN-LSTM approach for parkinson's disease early detection using speech features

Hadeel Ahmed Abd El Aal¹, Shereen A. Taie², Nashwa El-Bendary³

^{1,2}Faculty of Computers and Information, Fayoum University, Fayoum, Egypt

³College of Computing and Information technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Aswan, Egypt

Article Info

Article history:

Received Nov 28, 2020

Revised Mar 2, 2021

Accepted Aug 19, 2021

Keywords:

Adaptive moment estimation

Long short-term memory

Parkinson's disease

Recurrent neural network

ABSTRACT

Parkinson's disease (PD) is the second most common neurodegenerative disorder disease right after Alzheimer's and the most common movement disorder for elderly people. It is characterized as a progressive loss of muscle control, which leads to trembling characterized by uncontrollable shaking, or (tremors) in different parts of the body. In recent years, deep learning (DL) models achieved significant progress in automatic speech recognition, however, limited studies addressed the problem of distinguishing people with PD for further clinical diagnosis. In this paper, an approach for the early detection of patients with PD using speech features was proposed, a recurrent neural network (RNN) with long short-term memory (LSTM) is applied with the batch normalization layer and adaptive moment estimation (ADAM) optimization algorithm used after the network hidden layers to improve the classification performance. The proposed approach is applied with 2 benchmark datasets of speech features for patients with PD and healthy control subjects. The proposed approach achieved an accuracy of 95.8% and MCC=92.04% for the testing dataset. In future work, we aim to increase the voice features that will be worked on and consider using handwriting kinematic features.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Hadeel Ahmed Abd El Aal

Department of Computer Science

Faculty of Computers and Information, Fayoum University

Motol, Etsa, Fayoum, Egypt

Email: ha_fci_2012@yahoo.com

1. INTRODUCTION

Parkinson's disease (PD) is a complex neurological illness that, being classified as a degenerative, chronic, and progressive disease that affects a person's movements [1], [2]. Most people are diagnosed during their 70s, although 15% of cases occur among people who are under 50 years of age. Its expansion rate is estimated to be 1.5% approximately for people aged over 65 years [3]. The Clinic pathological studies show that up to 25% of the patients with PD are diagnosed incorrectly [4], The accuracy of clinical diagnosis can reach approximately 90% within a period of 2 years and 9 months [5]. Diagnosing PD is rather difficult, up till now there is no blood test that can reveal whether a person has a PD or not. Such illness is usually diagnosed through clinical exams and brain scans. These methods are quite costly, sometimes erroneous, and need an elevated level of professional expertise.

Machine learning (ML) is a technique for analyzing data, it automatically learns the information and attitudes of a system and perceives the complexity of patterns with ease [6]. Deep learning (DL) is considered a great evolution of machine learning. It is inspired by brain operationality; it uses a programmable neural

network [7] that authorizes the machines to make accurate decisions without needing interference from humans.

A neural model with appropriate generalization can provide precise answers even when testing it with inputs that have never been experienced before in the training set [8], also DL offer high prediction performance compared to other ML methods such as support vector machine (SVM) and random forest (RF) [9]. In recurrent neural networks (RNN) with long short-term memory (LSTM), the impermanent correlations of the input data can be learned [10], which consists of blocks of memory that allows retaining input information for a long period [9]. The optimizer is a method to adjust the varied parameters of the model. optimizing the neural network is very beneficial for increasing the accuracy and reducing the loss. Instead of mapping inputs to outputs alone, the RNN-LSTM network has the capability of learning a mapping function from inputs to outputs over time. An explicit set of observations need not be pre-specified. The main contributions of this paper are:

- Proposing an enhanced approach based on deep learning through using RNN-LSTM for early detection of PD using voice features.
- Applying the proposed RNN-LSTM approach with a batch normalization layer after each hidden layer to standardize the outputs of the hidden layers.
- Applying the adaptive moment estimation (ADAM) optimization algorithm for training the network by updating the weights of the network iteratively based on the training data while training.

The rest of this paper is organized as; section 2 presents state-of-the-art studies for PD detection, section 3 describes the phases of the proposed approach, section 4 presents and discusses the obtained experimental results, section 5 presents conclusions and future work.

2. RELATED WORK

Classification techniques based on ML and DL would be a convenient tool for an accurate diagnosis to differentiate healthy people from individuals with PD. Zham *et al.* [11] used a naïve bayes (NB) algorithm on handwriting tasks and spiral drawing, different measures were used for each task. The fourth task has achieved the best classification accuracy with 83.2%. Taleb *et al.* [12] used a feature selection technique on handwriting tasks based on statistical tests and the SVM classifier. The feature giving the highest classification performance is picked up firstly. Features were provided separately one by one as an input to the SVM classifier. The highest classification accuracy obtained of a solitary feature was 87.5%. Then, features were fed continuously one after another until they get 86 features. The best classification accuracy of a group of features was 96.875% for N=12 features. Drotár *et al.* [4] compared three different classifiers: K-nearest neighbors (K-NN), ensemble AdaBoost classifier, and SVM on parkinson's disease handwriting based on pressure and kinematic features using (PaHaW) dataset. SVM obtained the best result of all three classifiers with an accuracy of 81.3%. Also, Drotár *et al.* [13] used SVM on handwriting features to classify the PD patients, the accuracy was 88.1% for 162 handwriting features.

Moreover, in Drotár *et al.* [14] they used SVM classifier for measuring the in-air and on-surface kinematic variables of the handwriting features of the PD patients. The achieved accuracies were 84% for in-air movement, 78% for on-surface movement, and 85% for both in the air + on surface movement. On the other hand, in [15]. Afonso *et al.* used the optimum-path forest (OPF), deep-hierarchical OPF (dOPF), and k-means algorithms for the identification of parkinson's disease on the handwriting of spiral and meander features, the best result was for the K-means algorithm with an accuracy=84.17%. Pereira *et al.* [16] applied a convolutional neural network (CNN) on spiral and meander hand drawing features of PD patients, the accuracy for 128*128 meander images was 87.14% and the accuracy for 128*128 spiral images was 77.92%.

Also, Pereira *et al.* [17] used three classifiers NB, OPF, and SVM on the handwriting of spiral drawing, the NB classifier obtained the best result with accuracy=78.9%. Heremans *et al.* [18] used handwriting features to estimate the quality of writing in PD patients with and without freezing of gait (FOG). The writing qualities were severely affected by patients with FOG. Grover *et al.* [19] in this survey used deep neural network (DNN) on UCI's voice dataset with three layers: input, hidden and output layer. The classification accuracy was 94.4% for training and 62.7% for testing. Saikia *et al.* [20] used an artificial neural network to classify PD patients from healthy controls in addition to providing the different progression stages of the disease based on the Electroencephalogram and the Electromyogram features. In [21] proposed a model for detecting the PD disease via smell signature using two sensors to analyze the sweat components and comparing these components between the PD and non-PD individuals. In [22] compared the classification accuracies of five different classifiers, the SVM, NB, KNN, DT, and the LDA, relying on gait dynamics. The average accuracy of the first three classifiers was 96.8% and 93.5% for the last two classifiers.

Shinde *et al.* [23] used the rate of eye blinking per minute to determine parkinsonism, where if the rate is higher than ten blinks per minute the individual is considered as having PD. In order to enhance the

detection of patients with PD, in this paper, we proposed a RNN with LSTM and ADAM optimizer based on different voice features. Despite that LSTM requires some memory, RNN with LSTM can deal with large datasets without increasing the size of the model. Also, LSTM is more effective in comparison to the traditional time series models as it learns long-term dependencies that use former time proceedings to inform the next ones, so it allows information to persist and achieves best results.

The proposed model overcomes the disadvantage of existing models with respect to the limited dataset and features that seriously affect the accuracy of PD prediction. In addition to emphasizing the benefit of accumulation, as traditional neural networks applying direct feedforward appears shortcoming, meanwhile, RNN with LSTM is considered as a loop network that learns long-term dependencies, which enhance the prediction. Different measures were used to validate the model.

3. RESEARCH METHOD

The proposed model embraces three main phases listed is being as; preprocessing phase, optimization phase, and classification phase. The framework of the proposed model for diagnosing parkinson's diseases based on speech features is illustrated in Figure 1.

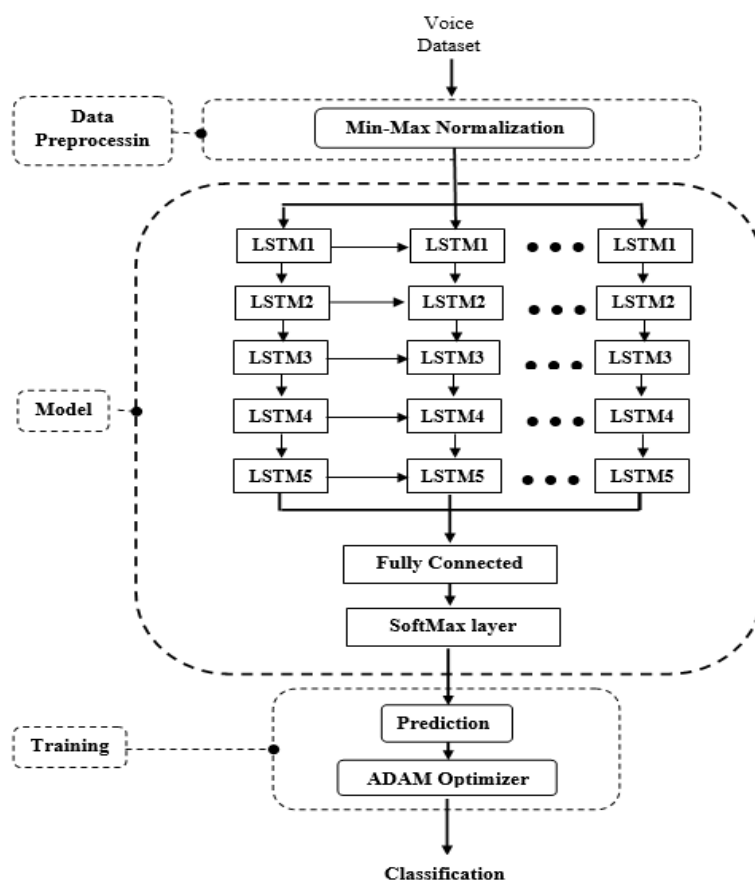


Figure 1. The proposed approach

The proposed model structure consists of seven layers (input layer, 5 hidden layers, and the output layer). LSTM input layer contains 27 neurons a neuron for each feature, five LSTM hidden layers, a 27 neurons dense layer followed by a two-neuron dense layer as an output layer. Each LSTM layer is appended by a dropout and a batch normalization layer. The dropout regularizes the input and the recurrent connections to the LSTM units by excluding some inputs from activation (drops them out) based on statistical calculations. The batch normalization layer standardizes the outputs of the hidden layer by normalizing the values coming from the previous layer. The batch normalization layer reduces the overfitting as it has a slight regularization effect which improves the performance of the model.

Finally, a 27 neuron dense layer followed by a fully connected dense layer, where all neurons in the previous layer are connected to that layer, the last dense layer works as the output layer. The following subsections illustrate the details of each phase.

3.1. Preprocessing phase

This phase worked to collect and prepare the data for the following phases to improve the results and suppress the effect of outliers in it. Min-max normalization was applied to make every datapoint have the same range of values so each feature is equally important. This is done via (1). This process helps to have small standard deviations, which can suppress the effect of outliers.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

3.2. Optimization phase

The main goal of deep learning and machine learning is reducing the diversity between the actual output and the predicted output. This is known as the cost function or loss function. To assure adequate generalization of an algorithm and to diminish the cost function by detecting the optimized value of the weights appears the urge of using optimization via training the neural network. This makes a better prediction for the data that was not seen before.

In the proposed model two different optimizers were used, the commonly known SGD optimizer and the most widely used optimizer for deep learning models the ADAM optimizer. The ADAM optimizer has achieved the best performance, and this will be displayed in 3.2.4. subsection. ADAM optimizer [24], [25] is one of the most recommended optimization techniques, it is essentially combining the advantages of the stochastic gradient descent (SGD) with momentum algorithm and the root mean square (RMS). The advantages of ADAM could be pointed out in the following points:

- The ADAM algorithm doesn't need high memory requirements.
- The ADAM algorithm makes use of the average of the second moments of the gradients not only adapting the learning rates based on the average of the first moments. The first moment is mean, and the second moment is uncentered variance.
- The ADAM algorithm works very well even with a little regulation of hyperparameters.

The ADAM optimizer works according to the following steps:

- a. Initiate the 1st moment $m_0 = \text{Zero}$, initiate the 2nd moment $n_0 = \text{Zero}$, and initialize the first time period $T = \text{Zero}$.
- b. Update the bias of the 1st and 2nd moments, this is shown in (2), (3).

$$m_t = \beta_1 \otimes m_{t-1} \oplus (1 - \beta_1) \otimes dw \quad (2)$$

$$n_t = \beta_2 \otimes n_{t-1} \oplus (1 - \beta_2) \otimes dw^2 \quad (3)$$

- c. Calculate bias-corrected of the 1st and 2nd moments, as shown in (4), (5).

$$m_{tcorr} = m_t / (1 - \beta_{1t}) \quad (4)$$

$$n_{tcorr} = n_t / (1 - \beta_{2t}) \quad (5)$$

- d. Update the parameters P and S, see (6), (7).

$$P = P - \varepsilon \otimes m_{tcorr} / (\sqrt{n_{tcorr}} \oplus \varepsilon) \quad (6)$$

$$S = S - \varepsilon \otimes m_{tcorr} / (\sqrt{n_{tcorr}} \oplus \varepsilon) \quad (7)$$

Where; β_1 and β_2 are hyperparameters with default values of 0.9 and 0.999 respectively. ε is the learning rate $\varepsilon = 10^{-3}$. The ADAM optimizer is shown in Figure 2.

3.3. Classification phase

The proposed model applied RNN with LSTM for classifying healthy individuals from PD patients and used the ADAM optimizer to update the weights of the network iteratively, this will be illustrated in more details in the next subsections.

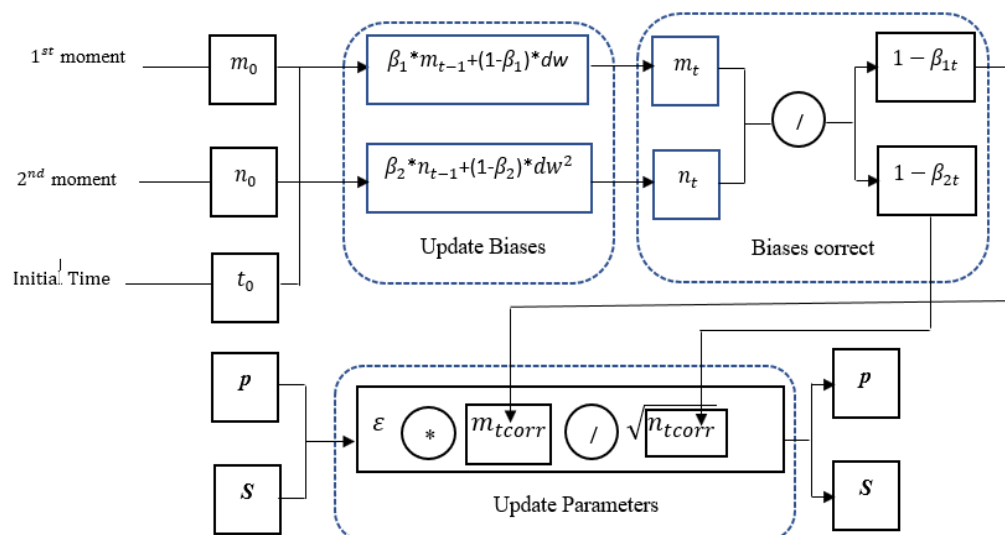


Figure 2. ADAM optimizer

3.3.1. Recurrent neural networks

RNN is a generalization of a feedforward neural network that contains an internal memory. In RNN the output of the current input relies on the prior computation. After getting the output, it is copied and sent back into the recurrent network. For making a decision, RNNs use the internal memory to operate on a series of inputs where all the inputs are associated with each other.

3.3.2. Long short-term memory

LSTM uses back-propagation for training. LSTM network has mainly three gates. input gate, forget gate, and the output gate. The input gate uses a sigmoid function to decide which values from the input shall be activated and modify the memory. The forget gate determines what details from the previous state could be discarded from the block. Finally, the output gate controls the output.

3.3.3. Regularization with dropout

In general, the most common problem that neural network models suffer from is overfitting. Overfitting could be explained as that the model has a good performance with the training dataset but does not perform very well with the test dataset. To overcome this problem, the proposed model applied the dropout regularization technique. The dropout is carried out on both the training and testing states. The dropout parameter value used was 0.2.

3.3.4. The recurrent neural networks model with adam optimizer

The RNN model comprises an Input layer, then passed to five LSTM hidden layers, and the last layer is the output layer. Now, elaborating on the application of the ADAM optimizer on the proposed Recurrent Neural Networks model in more detail. The dataset is loaded and all the data is normalized into values between 0 and 1. The training data is processed for a batch size of 104 sample records and 10 epochs. The training data is compiled with the ADAM optimizer which updates the weights of the network iteratively, using sparse_categorical_crossentropy loss function with learning rate=0.001 and decay=1e-4. The network structure of the proposed model is shown in Table 1.

Table 1. Network structure

Description	Value
Number of network layers	7
Number of hidden layers	5
Learning rate	0.01
decay	1e-4
Patch size	104
Number of epochs	10
Loss function	sparse_categorical_crossentropy
Activation function	SoftMax
Number of training samples	1040
Number of test samples	168
optimization	ADAM optimizer

Table 2 shows the proposed model performance with ADAM optimizer and the performance of the typical RNN “RNN with stochastic gradient descent (SGD) optimizer”. From Table 2 the ADAM optimizer has improved the accuracy of the proposed model by approximately 15.6% more than the typical RNN.

Table 2. Performance of the RNN with ADAM and SGD optimizers

Measurements	RNN with ADAM	RNN with SGD
Accuracy	95.8%	80.2%
Recall	100%	78.8%
Precision	92.3%	87.8%
F-score	96%	83.05%

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we discuss the obtained results through presenting the used datasets with brief details about the features of each dataset, the experimental settings, the measures used to validate the model performance. Also, we present a comparison between the proposed model and the model presented by Grover *et al.* [19] that addresses the same problem based on the accuracy performance and the structure of the two models. Moreover, we examine the accuracies and some validation measures of the different ML algorithms such as RNN with ADAM optimizer, RNN with SGD, SVM, and K-NN that we applied on the two datasets in order to highlight the best model for detecting PD. Finally, we show a performance comparison between the proposed model and other related works.

4.1. Datasets and experimental setting

In our experiment, we work with Python programming language along with TensorFlow and Keras libraries. The proposed model implemented a RNN with LSTM along with ADAM optimizer and a sparse_categorical_crossentropy loss function. We also consider the presented model of [19] that used a feed-forward neural network with three hidden layers. Two benchmark datasets of speech features are used in this study. The first PD dataset (DS1) is the parkinson’s telemonitoring voice dataset from the UCI public repository of datasets [26]. This dataset consists of 1040 samples for training and 168 samples for testing with 27 voice features.

The second dataset (DS2) is created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, this dataset contains 195 samples 130 samples for training, and 65 samples for testing with 22 voice features [27]. When applying the second dataset, we modified the number of neurons in the hidden layers of the network to be 22 neurons according to the number of the voice features and kept the same network structure. Details of the features of both dataset’s are listed in Table 3.

4.2. Results

We used different measures to validate our model, these measures are accuracy, recall, precision, and F-score. Where true positive (TP), true negatives (TN), false positive (FP), and false negatives (FN) as shown in (8)-(11).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{F-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

The accuracy of a model is a method to measure how the model correctly classifies the data. It is the ratio between the correctly predicted samples to the whole number of the prediction samples. Precision is the ratio of the rightfully predicted as positive by the model to all positives, in other words, precision clarifies how many predicted PD patients are actually PD. Recall measures how correctly the model identifies true positives, in the proposed model the recall shows how many PD patients are correctly predicted. F-score is the average of the recall and precision. The obtained classification accuracy of our model on the first dataset was 95.8%, in comparison to the proposed methodology by Grover *et al.* [19], which was 62.7%. This shows that our proposed model has the discrimination of 33.1% for the classification accuracy over the

methodology presented in [19]. Table 4 presents a brief comparison between the structure of the two models and the accuracy performance of each model.

Table 3. Datasets features

Data Features for (DS1)				Data Features for (DS2)			
Feature	Feature description	Feature	Feature description	Feature	Feature description	Feature	Feature description
Jilter (local)	Several of variation in fundamental frequency	Median pitch	Pitch parameters	MDVP: Fo (Hz)	Average vocal fundamental frequency	HNR	Measures of ratio of noise to tonal components in the voice
Jitter (local absolute)	Several of variation in fundamental frequency	Mean pitch	Pitch parameters	MDVP: Flo (Hz)	Minimum vocal fundamental frequency	RDPE	nonlinear dynamical complexity measures
Jitter (rap)	Several of variation in fundamental frequency	Standard deviation	Pitch parameters	MDVP: Jitter (Abs)	Several measures of variation in fundamental frequency	D2	nonlinear dynamical complexity measures
Jitter (ppq5)	Several of variation in fundamental frequency	Minimum pitch	Pitch parameters	MDVP: RAP	Several measures of variation in fundamental frequency	spread1	nonlinear measures of fundamental frequency variation
Jitter (ddp)	Several of variation in fundamental frequency	Maximum pitch	Pitch parameters	MDVP: RAP	Several measures of variation in fundamental frequency	spread2	nonlinear measures of fundamental frequency variation
Shimmer (local)	Several measures of variation in amplitude	Number of pulses	Pulse Parameters	MDVP: Jitter (%)	Several measures of variation in fundamental frequency	PPE	nonlinear measures of fundamental frequency variation
Shimmer (local, dB)	Several measures of variation in amplitude	Number of periods	Pulse Parameters	MDVP: PPQ	Several measures of variation in fundamental frequency	DFA	Signal fractal scaling exponent
Shimmer (apq3)	Several measures of variation in amplitude	Mean period	Pulse Parameters	Jitter: DDP	Several measures of variation in fundamental frequency	MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
Shimmer (apq5)	Several measures of variation in amplitude	Standard deviation of period	Pulse Parameters	MDVP: Shimmer	Several measures of variation in amplitude		
Shimmer (apq11)	Several measures of variation in amplitude	Fraction of locally unvoiced frames	Voicing parameters	MDVP: Shimmer (dB)	Several measures of variation in amplitude		
Shimmer (dda)	Several measures of variation in amplitude	Number of voice breaks	Voicing parameters	Shimmer : APQ5	Several measures of variation in amplitude		
AC	Harmonicities Parameters	Degree of Voice Breaks	Voicing parameters	MDVP: APQ	Several measures of variation in amplitude		
NTH	Harmonicities Parameters	UPDRS	Voicing parameters	Shimmer : DDA	Several measures of variation in amplitude		
HTN	Harmonicities Parameters			NHR	Measures of ratio of noise to tonal components in the voice		

From Table 4 the proposed approach had a higher accuracy than the approach of Grover *et al.* [19] by approximately 33%. Different ML algorithms were applied to find out the best model for predicting the possibility of having parkinson's disease, these algorithms are the RNN with ADAM optimizer, RNN with

SGD, SVM, and K-NN. The accuracies of the applied models on the two voice datasets (DS1) and (DS2) are shown in Figure 3.

Table 4. DNN and RNN comparison

Neural network	DNN	RNN with ADAM optimizer
Number of hidden layers	3	5
Type of Neural Network	Feed forward neural network	RNN
Memory	-	LSTM
Data normalization	Min-max normalization	Min-max normalization
Optimizer	-	ADAM optimizer
Loss function	-	sparse_categorical_crossentropy
Number of neurons in hidden layers	10,20,10	27
Measurements	Accuracy	accuracy, recall, precision, F-score
Testing accuracy	62.7335%	95.8%

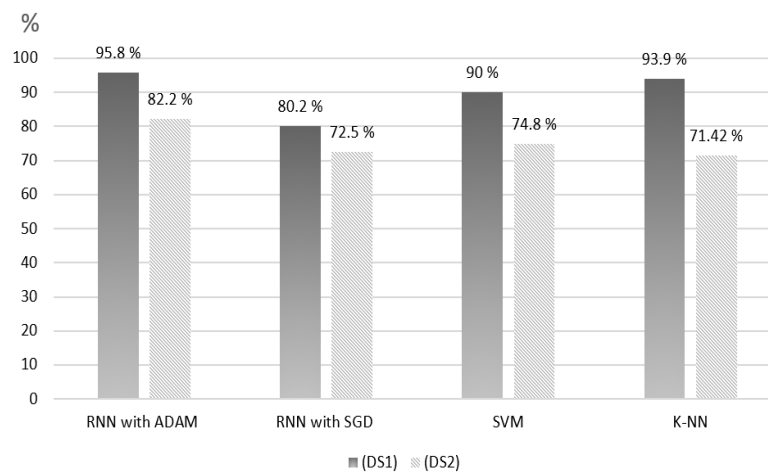


Figure 3. Average accuracies of the different models

Figure 3 shows that the RNN model with ADAM optimizer on the first dataset (DS1) increased the accuracy of the classification by 15.6% in comparison to the RNN with SGD, achieved better classification accuracy by 5.8% than the SVM algorithm, and improved the accuracy by 1.9% than the K-NN.

Also, Figure 3 illustrates that the RNN model with ADAM optimizer has maintained the best accuracy performance on the second dataset (DS2) with a difference of 9.7%, 7.4%, and 10.7% versus the RNN with SGD, SVM, and the KNN models respectively. These results have shown that the RNN model with ADAM optimizer has achieved the best classification result on both voice datasets. Table 5 shows the performance of these models on the two datasets based on the recall, precision, and the F-score.

Table 5. Validation measurements of the different models on (DS1) and (DS2)

ML Algorithm	Dataset	Recall	Precision	F-score
RNN With ADAM	(DS1)	100%	92.3%	96%
	(DS2)	99%	82.2%	90.24%
SVM	(DS1)	85%	100%	92%
	(DS2)	74%	100%	85%
K-NN	(DS1)	92%	99%	95%
	(DS2)	71%	100%	83%
RNN with SGD	(DS1)	78.8%	87.8%	83.05%
	(DS2)	79.2%	72.5%	75.35%

From Table 5 the RNN model with ADAM optimizer has achieved high results on the different datasets by various validation measures. These results highlight the benefits of the LSTM along with the ADAM optimizer.

The achieved result of the different models applied on the second dataset (DS2) could have lower performance due to the small number of samples in comparison to the first dataset (DS1). Table 6 compares the validation performance between previous surveyed studies with different models and datasets with the performance of the proposed approach for detecting PD.

Table 6. Different models performance comparison

Work	Classifier	Dataset	Accuracy	Recall	Precision
[11]	NB	Hand Writing	83%	83.2%	83.2%
[14]	SVM	Hand Writing	85%	85.2%	85.9%
[17]	NB	Hand Writing	78.9%	91%	24%
[19]	DNN	Voice	62.7%	-	-
Proposed approach	RNN with ADAM	Voice (DS1)	95.8%	100%	92.3%
		Voice (DS2)	82.2%	99%	82.2%

Moreover, the matthews correlation coefficient (MCC) of the proposed model with the first dataset (DS1) was calculated, and it gives 92.04%. MCC considers all the TP, FP, TN, and FN values, and the high value of the MCC (near to 1) means that the two classes were properly predicted, even in case one of the two classes is disproportionately represented. MCC can be calculated from (12).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (12)$$

The elapsed time for the whole process was 20 minutes with 104 epochs. Each epoch takes approximately 11 seconds.

5. CONCLUSION

In this paper, we presented a model with the aim to diagnose parkinson's disease with less human interference and in a much cheaper and more efficient way. A RNN with LSTM and ADAM optimizer was used with sparse_categorical_crossentropy loss function and the SoftMax activation function. The model was applied in two different voice datasets, and multiple measures were computed to evaluate the model performance. The achieved accuracy on the first dataset is 95.8%, the recall is 100%, the precision is 92.3%, and the F-score is 96%. For the second dataset, the proposed approach obtained an accuracy of 82.2%, 99% for recall, 82.2% for precision, and 90.24 % for F-score. For future work, we will work on considering more voice features with other kinematic features like handwriting features.

REFERENCES

- [1] S. Sveinbjornsdottir, "The clinical symptoms of Parkinson's disease," *journal of neurochemistry*, vol. 139, no. 1, pp. 318-342, 2016, doi: 10.1111/jnc.13691.
- [2] R. M. Sadek, *et al.*, "Parkinson's Disease Prediction Using Artificial Neural Network," *International Journal of Academic Health and Medical Research (IAHMR)*, vol. 3, no. 1, pp. 1-8, 2019.
- [3] J. Mucha *et al.*, "Fractional Derivatives of Online Handwriting: A New Approach of Parkinsonic Dysgraphia Analysis," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, pp. 1-4, doi: 10.1109/TSP.2018.8441293.
- [4] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and F.-Z. Marcos, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial Intelligence in Medicine*, vol. 67, pp. 39-46, 2016, doi: 10.1016/j.artmed.2016.01.004.
- [5] L. M-Velazquez, *et al.*, "Phonetic relevance and phonemic grouping of speech in the automatic detection of Parkinson's Disease," *Scientific Reports*, vol. 9, pp. 1-16, 2019, doi: 10.1038/s41598-019-55271-y.
- [6] M. Nawir, A. Amir, N. Yaakob, and O. Bi Lynn, "Effective and efficient network anomaly detection system using machine learning algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 46-51, 2019, doi: 10.11591/eei.v8i1.1387.
- [7] D. Kollias, A. Tagaris, A. Stafylopatis, S. Kollias and G.Tagaris, "Deep neural architectures for prediction in healthcare," *Springer*, vol. 4, pp. 119-131, 2018, doi: 10.1007/S40747-017-0064-6.
- [8] C. H. Goay, A. Abd Aziz, N. S. Ahmad, and P. Goh, "Progress in neural network based techniques for signal integrity analysis—a survey," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 1, pp. 276-282, 2019, doi: 10.11591/eei.v8i1.1405.
- [9] B. K. Reddy and D. Delen, "Predicting Hospital Readmission for Lupus Patients: An RNN-LSTM based Deep-Learning methodology," *Computers in Biology and Medicine*, vol. 101, pp. 199-209, 2018, doi: 10.1016/j.compbiomed.2018.08.029.

- [10] R. Zhao, J. Wang, R. Yan and K. Mao, "Machine health monitoring with LSTM networks," *2016 10th International Conference on Sensing Technology (ICST)*, 2016, pp. 1-6, doi: 10.1109/ICSensT.2016.7796266.
- [11] P. Zham, S. P. Arjunan, S. Raghav and D. K. Kumar, "Efficacy of Guided Spiral Drawing in the Classification of Parkinson's Disease," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1648-1652, Sept. 2018, doi: 10.1109/JBHI.2017.2762008.
- [12] C. Taleb, M. Khachab, C. Mokbel and L. Likforman-Sulem, "Feature selection for an improved Parkinson's disease identification based on handwriting," *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 2017, pp. 52-56, doi: 10.1109/ASAR.2017.8067759.
- [13] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal and M. Faundez-Zanuy, "Decision Support Framework for Parkinson's Disease Based on Novel Handwriting Markers," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508-516, May 2015, doi: 10.1109/TNSRE.2014.2359997.
- [14] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. F. -Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 405-411, 2014, doi: 10.1016/j.cmpb.2014.08.007.
- [15] L. C. S. Afonso, C. R. Pereira, S. A. T. Weber, C. Hook and J. P. Papa, "Parkinson's Disease Identification through Deep Optimum-Path Forest Clustering," *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2017, pp. 163-169, doi: 10.1109/SIBGRAPI.2017.28.
- [16] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa and J. P. Papa, "Deep Learning-Aided Parkinson's Disease Diagnosis from Handwritten Dynamics," *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 340-346, doi: 10.1109/SIBGRAPI.2016.054.
- [17] C. R. Pereira *et al.*, "A Step Towards the Automated Diagnosis of Parkinson's Disease: Analyzing Handwriting Movements," *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, 2015, pp. 171-176, doi: 10.1109/CBMS.2015.34.
- [18] E. Heremans, E. Nackaerts, S. Broeder, G. Vervoort, S. P. Swinnen, and A. Nieuwboer, "Handwriting impairments in people with parkinson's disease and freezing of gait," *Neurorehabilitation and neural repair*, vol. 30, no. 10, pp. 911-919, 2016, doi: 10.1177/1545968316642743.
- [19] S. Grover, S. Bhartia, A. A. Yadav, and K. R. Seeja, "Predicting severity of parkinson's disease using deep learning," *Procedia Computer Science*, vol. 132, pp. 1788-1794, 2018, doi: 10.1016/j.procs.2018.05.154.
- [20] A. Saikia, V. Majhi, M. Hussain, A. R. barua, S. Paul and J. K. Verma, "Machine Learning based Diagnostic System for Early Detection of Parkinson's Disease," *2020 International Conference on Computational Performance Evaluation (ComPE)*, 2020, pp. 275-279, doi: 10.1109/ComPE49325.2020.9200195.
- [21] S. Kulkarni, N. G. Kalayil, J. James, S. Parsewar and R. Shriram, "Detection of Parkinson's Disease through Smell Signatures," *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 808-812, doi: 10.1109/ICCSP48568.2020.9182283.
- [22] J. P. Félix *et al.*, "A Parkinson's Disease Classification Method: An Approach Using Gait Dynamics and Detrended Fluctuation Analysis," *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1-4, doi: 10.1109/CCECE.2019.8861759.
- [23] A. Shinde, R. Atre, A. S. Guleria, R. Nibandhe and R. Shriram, "Facial Features Based Prediction of Parkinson's Disease," *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1-5, doi: 10.1109/I2CT.2018.8529466.
- [24] D. P. Kingma and J. Lei Ba, "ADAM: A method for stochastic optimization," in *ICLR Conference*, 2015, pp. 1-15.
- [25] S. Bock, J. Goppold, and M. Weiß "An improvement of the convergence proof of the ADAM-Optimizer," *Conference paper at oth clusterKonferenz*, 2018, pp. 1-5.
- [26] A. Tsanas, M. A. Little, P. Mcsharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884-893, 2009, doi: 10.1109/TBME.2009.2036000.
- [27] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no.1, pp. 1-19, 2007, doi: 10.1186/1475-925X-6-23.