ResNet-Based Parkinson's Disease Classification

Omar El Ariss and Kaoning Hu

***Abstract*—Parkinson's disease (PD) is a brain disorder that leads to shaking, stiffness, and difficulty with walking, balance, and coordination. The symptoms usually begin gradually and get worse over time. Early diagnosis is very important because treatments are more effective and easier to perform during the early stages of PD. However, early diagnosis is challenging because the symptoms start gradually, and at the early stages, they are not very noticeable. In this article, we propose a method that uses ResNet50, a residual network that has 50 layers, to help diagnosis PD. The data used are a collection of frequency features acquired by applying spectral analysis strategies to the speech recordings of the patient. We then convert the frequency features into a 2-D heat map. This heat map is passed to ResNet50, which predicts whether the patient has PD or not. We have conducted experiments and compared the accuracy with several state-of-the-art methods. The results have demonstrated the feasibility and robustness of the proposed method.**

***Impact Statement*— Parkinson's Disease (PD) is the most common movement disorder with more than 10 million PD patients worldwide. In the United States, around 60 000 are diagnosed with PD every year. Treatments are more effective at early stage, so early diagnosis is very important. A simple and effective classifier to diagnose PD patients is therefore crucial for doctors and patients. For the last decade, Convolutional Neural Networks (CNN) are the dominant approach for image classifications. Our proposed method benefits from the versatility, solid performance of pretrained CNN architecture by converting the frequency domain features of patients' speech recordings into heat maps. Our method achieved an accuracy of 90.7% in diagnosis of PD by only using Tunable Q-Factor Wavelet Transform (TQWT) features, outperforming other state-of-art methods. Therefore, the conversion from non-graphical features to heat maps provides simple, accurate deep network for PD diagnosis that is fast to train, and only requires TQWT features.**

***Index Terms*—Convolutional neural networks (CNNs), deep learning, diagnosis, frequency features, heat map, Parkinson's disease (PD), ResNet, speech recording, transfer learning.**

I. Introduction

PARKINSON'S disease (PD) is a neurodegenerative disorder that leads to shaking, stiffness, and difficulty in walking, balance, and coordination [1]. The symptoms usually begin gradually and get worse over time. In the United States, around 60 000 are diagnosed with PD every year [2]. An estimate of 572 out of 100 000 of age 45 or older is diagnosed with PD [3]. It is the second most common neurodegenerative disorder of aging and the most common movement disorder [4] with more than 10 million PD patients worldwide. According to Parkinson's Foundation, the medications alone cost an average of $2500 a year and therapeutic surgery can cost up to $100 000 per person [2].

Early diagnosis of PD is very important because medications are more effective when administered at the early stages. This can be seen from a comparative study that was conducted by dividing patients into two groups. One group was treated with Rasagiline, while the other was treated with Placebo. After 6–8 months, both groups were treated with the same amount of Rasagiline. After several years, the results showed that the early started patients benefited from obvious long-term improvement in terms of the unified PD rating scale scores [5]–[7]. It is also shown that nonpharmacologic treatments, such as increased exercise, are also easier to perform in the early stages of PD and may help slow down the disease progression [8]. However, early diagnosis is a challenging task. First, at the early stages, the symptoms are not significant and not very noticeable. Second, many other neurologic disorders also have similar symptoms, such as multiple system atrophy, Dementia with Lewy bodies, progressive supranuclear palsy, and Corticobasal degeneration. [9] Third, tests are not always easy to perform. Although PD may be detected by performing neurotests and scanning of the brain, these tests are expensive and inconvenient to the patients [10].

In recent years, instead of scanning the brain, scientists began to extract features from different data acquired from the patients, such as electroencephalography (EEG) [11], rate of finger tapping [12], pace of walking [10], and speech [13]. With the aid of machine learning algorithms, the experiments showed the improvement of accuracy in the diagnosis of PD at the early stage [14].

The common symptoms of PD include tremor, muscle stiffness, slowness of movement, and impaired balance [9]. Researchers have also found that the muscle stiffness alters muscular control of the phonatory subsystem, leading to the voice change of the patients [15]. It is suggested that speech impairment can be detected as early as 5 years prior to the diagnosis [16].

Although speech impairment is a common symptom in other neurodegenerative disorders related to dementia [17], it is distinguishable between PD and other types of dementia [18], [19]. Dementia is defined as the loss of cognitive functioning, such as thinking, remembering, and reasoning, to such an extent that it interferes with a person's daily life and activities [20]. Researchers have observed five most common forms of

Manuscript received 4 March 2022; revised 8 May 2022; accepted 25 June 2022. Date of publication 25 July 2022; date of current version 22 September 2023. This paper was recommended for publication by Associate Editor Mihail Popescu upon evaluation of the reviewers' comments. *(Corresponding author: Kaoning Hu.)*

The authors are with the Department of Computer Science and Information Systems, Texas A&M University - Commerce, Commerce, TX 75428 USA (e-mail: omar.el.ariss@tamuc.edu; khu1@binghamton.edu).

Digital Object Identifier 10.1109/TAI.2022.3193651

dementia [20]: Alzheimer's Disease (the most common dementia), Lewy body dementia (LBD), vascular dementia, frontotemporal dementia, and mixed dementia (a combination of two or more types of dementia). Among these forms of dementia, PD is associated with one type of LBD [21]. However, PD patients develop cognitive symptoms slower than the other type of LBD, and not all PD patients develop dementia [21]. Researchers have also compared Alzheimer's disease and PD, the two most common neurodegenerative disorders, and have discovered that "dementia of the Alzheimer type (DAT) produced significantly greater language disturbances, including anomia, decreased information content of spontaneous speech, and diminished word list generation. PD patients had significantly decreased phrase length, impaired speech melody, dysarthria, and agraphia" [18]. Some other researchers compared the voice and language problems of PD with other common neurodegenerative disorders. They have found no significant difference between PD patients and healthy people in figurative language comprehension, a domain that is commonly impaired in other neurodegenerative diseases [19].

In this article, we propose a novel approach to diagnose PD at an early stage using speech signals acquired from the patients. There are several reasons. First, the voice change, known as the speech motor impairment, is one of the early symptoms of PD [15], [16], [22], [23]. Second, the speech motor impairment caused by PD can be distinguished from other common neurodegenerative disorders [18], [19]. Third, the acquisition of speech signals is easy, simple, and also noninvasive [13].

We use ResNet50, a residual neural network of 50 layers, as a classifier. To process the speech signal, we extract the features from the speech recordings of the patient, and then convert the features into a 2-D heat map. Then, the heat map is passed to the ResNet50 model, which predicts whether the patient has PD or not. The heat map used is composed of rows of frequency-domain data processing algorithms. Each column in a row is a feature value generated by that algorithm. It is, therefore, a 2-D matrix of frequency-domain features of the speech signals.

The ResNet50 architecture that we have trained to process the heatmap is a deep artificial neural network (ANN) that has convolutional layers and pooling layers, with the skip connections that can skip over some layers. The convolutional layers can extract useful patterns from the differences (i.e., relative values) between frequency-domain feature values. This is important because every patient's base voice is different. Therefore, we assume the relative values between the frequency-domain feature values contain more useful patterns than the absolute values of these features. Our assumption has been verified by the results of the experiment. The pooling layers provides downsampling of the patterns extracted by the convolutional layers. With sufficient training, they can eliminate redundant patterns while enhancing useful multiscale patterns. The skip connections, i.e., shortcuts to jump over some layers, are used to avoid the problem of vanishing gradients, making the training process more effective.

The main contributions of this work can be summarized as follows. First, converting frequency values of voice recordings to color encoded 2-D heat map images are an effective representation. The use of heat maps maintains the values of features, their magnitude, location, and the variation between different features. Second, to the best of our knowledge, this is the first ResNet50 architecture that uses heat maps to represent frequency-domain vocal features for PD classification. Finally, experimental results on a publicly available dataset demonstrates that our proposed approach outperforms previous research work.

The rest of this article is organized as follows. In Section II, we will review the related work. In Section III, the residual neural network and transfer learning will be explained. Then, we will demonstrate how the heat map is generated and used with ResNet50 to diagnose PD in Section IV. In Section V, we will show the results and compare ours with the state-of-art methods. Finally, Section VI concludes this article.

II. Related Work

In the recent decade, researchers have used computer-aided diagnosis to detect PD at early stage. Various data collected from the patients have been used in the diagnosis of PD. Adams [12] proposed a method to use the rate of finger tapping to diagnose PD. They have collected keystroke timing information from 103 subjects, including 32 with mild PD severity and the remainder non-PD, as they typed on a computer keyboard over an extended period. However, the slowness of finger tapping (bradykinesia) is a common symptom in many movement-related disorders. Although the accuracy to identify the slowness of finger tapping is high, it will be difficult to distinguish PD from other movement-related disorders. Rehman et al. [10] used the manner and pace of walking to assess 93 PD patients and 103 non-PD persons. However, the pace of walking may be affected by many diseases, including common arthritis. Oh et al. [11] used EEG as the input to a 13-layer convolutional neural networks (CNNs) to diagnose PD. However, the acquisition of EEG is not easy. The research placed 64, 128, or 256 EEG electrodes around the head. The quality of EEG is affected by blinks and eye movements of the test subject as well as the power line noise generated by any nearby electric equipment. Sometimes, researchers have to manually spot artifacts during the data acquisition [24]. Besides finger tapping, gait, and EEG, many other data, such as magnetic resonance imaging (MRI) [25] and handwriting [26], have also been used to diagnose PD.

Comparing with other data, the acquisition of speech signal is simple and noninvasive [13]. Meanwhile, 90% of PD patients exhibit some form of vocal disorders at the earlier stages of the disease [27]. Although vocal cord dysfunction (VCD) can also cause similar symptoms, it is diagnosed via fiberoptic laryngoscopy and continuous laryngoscopy during exercise [28]. In addition, VCD mainly affects children and young adults until the age of 40 [29], [30]. In recent years, speech characteristics have been successfully used to evaluate PD and to monitor its evolution after medical treatment [31]. Therefore, we have also chosen to use speech signals to diagnose PD at the early stage.

With the aid of computers and the advance of artificial intelligence algorithms, researchers have increasingly benefited from machine learning algorithms to diagnose PD. Senturk [32] used support vector machines with recursive feature elimination

to process the speech signals of 23 PD patients and 8 healthy people. Their work achieved 93.84% accuracy. Parisi et al. [33] proposed a hybrid algorithm of multilayer perceptron and Lagrangian support vector machine [34]. Their method was tested on a dataset of 20 PD patients and 20 healthy people and achieved 99.29% accuracy. Sakar et al. [27] applied tunable Q-factor wavelet transform (TQWT) [35] to the speech signals for feature extraction and feed the extracted features to several different classifiers to diagnose PD. To test the effectiveness of their algorithm, they collected voice recordings from 252 individuals, including 188 patients and 64 healthy people. Their test result showed a maximum accuracy of 86%. They have also made their dataset publicly available. Finally, Tracy et al. [36] used paralinguistic voice features to classify patients with mild PD from those that do not have the disease. They collected 15 227 voice recordings from 2289 individuals, 246 of those were PD patients. Then, they applied logistic regression, random forest, and gradient boost, which gave the best results.

As deep neural networks achieved success in multiple tasks of artificial intelligence, people also applied deep neural networks, such as CNNs [37], [38], to help diagnose PD [14]. For example, Gunduz [31] proposed a combination of two CNN frameworks, each with nine layers, to process the frequency-domain features of the speech signal. They have tested their algorithm using Sakar's dataset and achieved 86.9% accuracy. Shivangi et al. [39] proposed a framework that combined two networks: a six-layer CNN to process the gait data and a four-layer ANN to process the speech data. They tested their framework on 91 subjects, including 48 healthy subjects and 43 PD patients, and achieved 88.17% accuracy. Sivaranjini and Sujatha [25] used AlexNet, an eight-layer CNN, to diagnose PD. They tested their method on the MRI data of 182 subjects, including 82 healthy subjects and 100 PD patients, and achieved 88.9% accuracy. Kollia et al. [40] proposed a new loss function for their deep neural networks and adopted a combination of transfer learning, k-means clustering, and k-nearest neighbor classification of deep neural network learned representations of the MRI and DaTscan (a visualization of the dopamine transporter levels in the brain) data. They have tested their method on different datasets and achieved the accuracies from 81.1% to 98.9%. Finally, Xu et al. [41] used 176 speech signals of 40 people, where 20 of those are PD patients. The authors transformed the voice recording into spectrograms and then applied a ResNet50 model to diagnose PD. The authors then used a deep convolutional generative adversarial network to generate spectrograms to augment the model with more data and were able to raise their accuracy results to 91%.

Deep neural networks are capable of modeling complex nonlinear relationships between the features and the class label due to their architecture and large set of parameters, specifically the weights of the neurons. They can also handle redundant features because parameters are automatically learned. The deep neural networks can also work unsupervised. However, the deep neural networks still have some limitations. First, to learn the values of the parameters, a large training set is required. The training process will also be extremely time-consuming. Second, to learn the parameters of the hidden layers, gradient-based learning methods and backpropagation are used. However, the side effect is the vanishing gradient problem, i.e., there is a possibility that after some iterations of training, the gradient will become too small (vanishing), and the parameters will effectively stop updating. To alleviate the first limitation, we have utilized transfer learning [42], [43]. Instead of training a deep neural network from scratch, we will train a deep neural network upon an existing deep neural network, which was trained for a similar task. To alleviate the second limitation, we have adapted the residual neural network (ResNet) [44]. ResNet is able to avoid vanishing gradient by using skip connections. As a result, ResNet architecture can be built into very deep neural networks, for example, a deep neural network of 50 layers, known as ResNet50. In Section III, we will introduce transfer learning and ResNet50.

III. Transfer Learning and ResNet50

In recent years, very large ANNs (deep neural networks) are successfully used in machine learning to solve various problems. This method is called deep learning. Due to its capability of modeling complex nonlinear relationships, deep learning has become a universal approximator. However, a major requirement of a deep neural network is the need of a huge training dataset, which contains tens of thousands or more of training samples. The huge training set also makes the training process of deep neural networks extremely time-consuming. In addition, in some classification problems, such as PD detection, it is very difficult to find or create large datasets due to the low number of PD patients. The need for a large data can be alleviated by the proper use of transfer learning.

A. Transfer Learning

Transfer learning is a method in machine learning where a model that is trained for one task is used as the starting point of a model for another task. Without transfer learning, we would typically initialize the parameters (weights and biases of the neurons) of a neural network with random values and train it using a very large training set. The parameters will be updated after the neural network learns from each training sample. This process is extremely time-consuming. However, if we have a neural network that has already been trained on a similar task, we will start our training process from this neural network using its parameters as our initial parameters. The parameters of the neural network will be updated after our training set is learned.

Researchers have discovered that a previously trained neural network is able to learn faster from the training data than a randomly initialized neural network. This not only reduces the time of training but also makes it possible to train a neural network with a much smaller training set. Yosinski et al. [45] quantified the transferability of features from each layer of a neural network. They have analyzed both positive and negative effects of transfer learning. Despite of negative effects, such as optimization difficulties and the specialization of higher layer features to the original task, they have found that features transferred from another similar task or another distant task are both better than random initial parameters. Eventually, they concluded that transfer learning could be a generally useful technique for improving deep neural network performance. In other

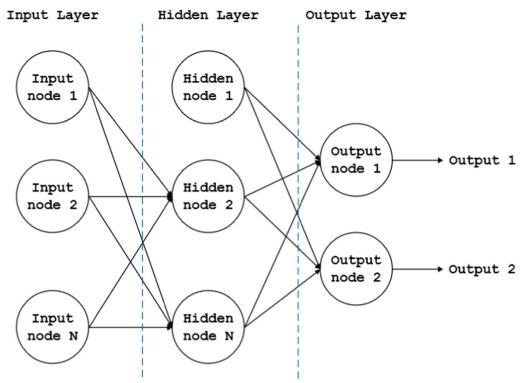

Fig. 1. Layers of an ANN.

words, the knowledge learnt from one task is useful in another task. In [42], the transferability regardless of the similarity of the tasks was verified again as the authors successfully transferred a CNN trained for the classification of nonmedical images to the task of medical image classification. In [46], transfer learning's ability to reduce the size of the training set is verified again as the authors successfully transferred a pretrained generic CNN to the task of emotion recognition using a small training set.

B. ResNet50

ResNet50 is the abbreviation of residual network with 50 layers. It is developed based on the deep convolution neural networks. In a general deep CNN, many layers are stacked to construct the network. Each layer is constructed by a set of neurons. Each neuron is a basic computation unit that takes its input from some of the neurons in its previous layer, multiplies the input by its weights, and then passes the sum through the activation function to some of the neurons in the next layer, as shown in Fig. 1.

Deep CNNs have been successfully used in many tasks of digital image processing and computer vision. The input of a deep CNN is typically an image. In the input layer, which is the first layer of a CNN, each neuron processes a small rectangular patch of the input image by performing a convolution operation, which is the dot product of the image patch and a filter. The filters can be high-pass filters or low-pass filters, which extract high-frequency information or low-frequency information, respectively. The output of one layer is passed to the neurons of the next layer, which may either use it for another convolution operation or downsample it (pooling) to reduce the data size and represent the data at a relatively coarser scale for a potential multiscale analysis.

To reduce the computation cost and overfitting in training, in a deep CNN, layers are not fully connected. In other words, each convolutional neuron only takes input from its receptive field in the previous layer. It does not take input from all neurons in the previous layer. This pattern applies to all layers except the last layer, which is fully connected, as shown in Fig. 2.

In a deep CNN, there are usually many layers. For example, the VGG network has 19 layers [47], and the GoogleNet [48] has 22 layers. However, increasing the number of layers does not guarantee the improvement of the performance because very deep neural networks are hard to train due to the problem of vanishing gradients. After some iterations of training, there is a possibility that the gradient will vanish into a very small value. As a result, when the parameters (weights) no longer update after each iteration, i.e., the neural network stops learning.

To alleviate the problem, instead of learning convolutional features, ResNet learns the residual. This is done by skip connection. Without skip connection, a layer in a neural network only sends its output to the neurons in its next layer. With skip connections, a layer may also "skip" some intermediate layers and send its output to the neurons several layers ahead as shown in Fig. 3. For example, in ResNet, the input of a convolution layer can be sent directly to the layer after the convolution layer. Therefore, the layer after the convolution layer will receive both the data before the convolution and the data after the convolution.

As a result, ResNet can be trained more effectively and can be built with more layers. In terms of the accuracy, ResNet is able to surpass VGG, GoogleNet, and many other classifiers by a significant margin [49].

IV. Proposed Method

The dataset used in this research is publicly available and can be found at the UCI's machine learning repository website [50]. The original source of the data is from Sakar's work [27] and was collected by the Department of Neurology in Istanbul University. There are 756 records of voice recording of the vowel *a* repeated three times by 252 participants. The recording of the participant's speech has been originally converted by the authors from time domain to frequency domain. In this article, we only use a subset of the original features since our proposed work converts numeric data to images. Therefore, we wanted the frequency-domain features to be of similar size. Here are the following frequency-domain algorithms that we use.

TQWT features: TQWT is a discrete-time wavelet transform. Its parameters are directly related to the Q-factor of the transform, so it can be tuned according to the oscillatory behavior of the signal to which it is applied [35]. In our work, we have used the following TQWT functions:

1) Energy;
2) Shannon's entropy;
3) entropy (sum of logarithms only);
4) mean of Teager–Kaiser energy operator (TKEO);
5) standard deviation of TKEO;
6) median values;
7) mean values;
8) standard deviation;
9) minimum values;
10) maximum values;
11) skewness (measure of symmetry) values;
12) Kurtosis (measure of the "tailedness") values.

In the dataset that we used, each of the above function has 36 feature values. As a result, the TQWT features will make a 12×36 matrix (432 values). The original authors used an application for speech analysis, such as Praat, to properly tune the features to match the domain characteristics of the speech signal.

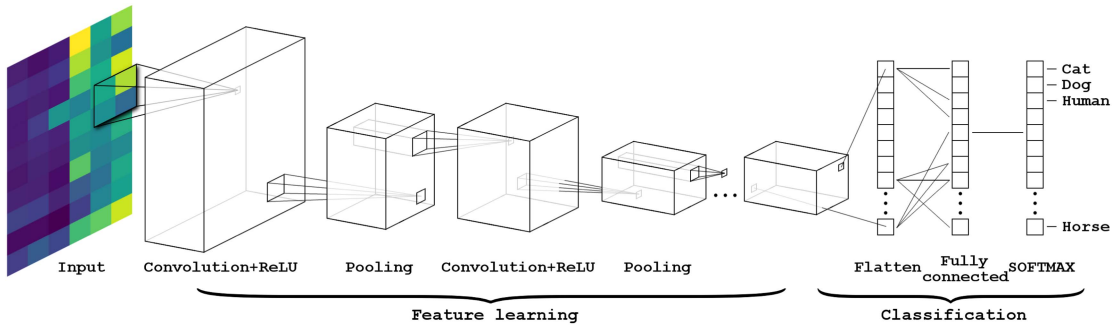

Fig. 2. Convolutional neural network.

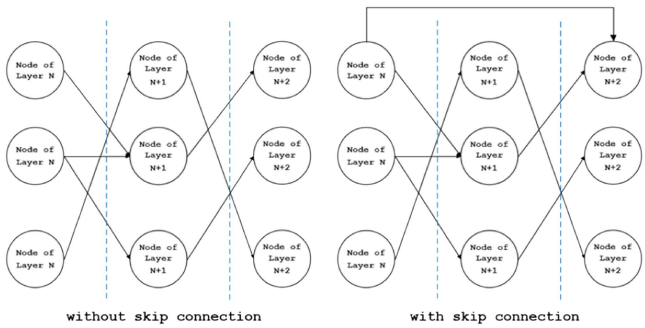

Fig. 3. Skip connection of ResNet.

Therefore, the number of feature values of each TQWT function as well as the number of feature values of other functions listed below were predetermined by the researchers who captured the data [27].

Discrete wavelet transform (DWT) features: It includes detail wavelet coefficients *(DET)* and approximation wavelet coefficients (*APP*) with and without log transformation (*LT*):

1) Shannon's entropy and sum of logarithms of DET before LT;
2) TKEO means and standard deviations of DET before LT;
3) Shannon's entropy and sum of logarithms of APP before LT;
4) Shannon's entropy and sum of logarithms of DET after LT;
5) TKEO means and standard deviations of DET after LT;
6) Shannon's entropy and sum of logarithms of APP after LT;
7) TKEO means and standard deviations of APP after LT.

Each of the above function has 20 values. Therefore, the DWT features will make a 7×20 matrix.

Mel frequency cepstral coefficients (MFCCs): MFCCs mimic the characteristics of human ear and have been used in different tasks, such as speaker recognition [31]. In our work, we used the following functions:

1) mean values of the log-energy of the signal and the original 13 MFCCs;
2) first derivatives of the means;
3) second derivatives of the means;
4) standard deviations of the log-energy of the signal and the original 13 MFCCs;
5) first derivatives of the standard deviations;
6) second derivatives of the standard deviations.

Each function has 14 feature values. Therefore, the MFCC features will make a 6×14 matrix.

In order to benefit from the strengths of ResNet and transfer learning, we need the data to be image-based while still maintaining the main characteristics of the different frequency algorithms that were collected. This is because that ResNet is a neural network used as a backbone for many computer vision tasks and was the winner of ImageNet challenge in 2015. We also want to benefit from the strengths of CNN in image classification and apply it to domain frequency data. In order to do that, we converted each feature set (TQWT, DWT, and MFCC) into heat map images. The main goal here is through the use of a heat map, the classifier can observe several characteristics of the speech. The convolutional layers can extract useful patterns from the differences (relative values) between frequency-domain feature values. In addition, the CNN can also identify the relative values in the heat map not only at different feature levels in one signal processing algorithm but also on multiple algorithms at once.

The heat map is fed to a residual neural network of 50 layers (ResNet50) for the task of early PD diagnosis. The convolution layers in ResNet50 can extract high-frequency patterns and low-frequency patterns from the training data as well as the relative values between frequency-domain feature values. Our assumption that relative values contain more useful patterns is verified by the experimental results in Section V. The pooling layers provide the downsampling of the patterns extracted by the convolutional layers. With sufficient training, they can eliminate redundant patterns while enhancing multiscale patterns.

We created heat maps from the frequencies, where each row is a frequency-domain algorithm, while each column is a frequency-domain feature value. Fig. 4 shows an example of a heat map with PD, and Fig. 5 is a heat map of a healthy person's voice recording. In both examples, the heat maps are constructed using MFCC features. There are six rows in each heat map, each representing a different function. Each row in the heatmap has 14 different columns that represent the frequency feature values.

In our experiments, the ResNet50 is tuned and trained using the following settings.

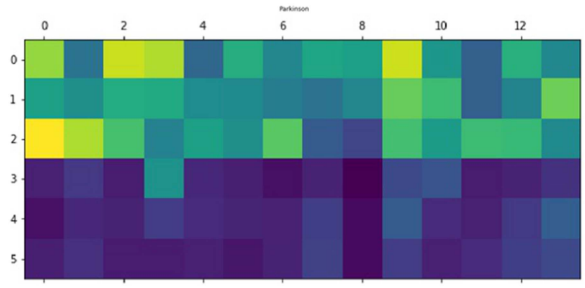

Fig. 4. Heat map of a voice recording with PD.

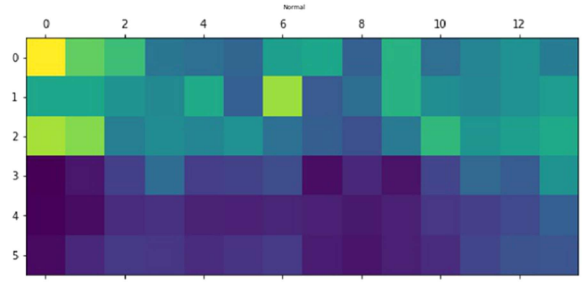

Fig. 5. Heat map of a healthy person's voice recording.

1) A ResNet50 model that is pretrained on ImageNet is used.
2) The last layer of the pretrained model is trained for 15 epochs using one-cycle policy. Discriminative learning is used with a default range with a maximum learning rate of 0.3.
3) The model is fine-tuned using one-cycle policy. The entire model (all layers) is retrained using ten epochs. Discriminative learning is also used with a default range with a maximum learning rate of 0.03. This step sometimes improved the classification results, while other times it did not. In this work, fine tuning will be discarded if the fine-tuned results do not show improvements.

The detailed experiments are explained in Section V.

V. Evaluation Results

In this section, we evaluate the effectiveness of our proposed ReNet50 model. Our evaluation is performed using Sakar's dataset [27]. The dataset contains 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1 ± 10.9). The control group consists of 64 healthy individuals (23 men and 41 women) with ages varying between 41 and 82 (61.1 ± 8.9). The sustained phonation of the vowel /a/ was collected from each subject with three repetitions. A total of 756 recordings, 564 have PD, while 192 recordings are normal. A total of 80% of the recordings are used for training, while 20% are used for evaluation.

Two evaluation metrics are used to measure the performance of the deep learning model: accuracy and F-measure. The accuracy of the model is calculated using the following:

$$\text{accuracy} = \frac{\text{correctly classified experiments}}{\text{total number of experiments}}. \quad (1)$$

The F-measure metric is the harmonic mean of the precision and recall. Precision is used as the measure of correctness of the classifier (2), while recall stands for its completeness (3)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

where

TP experiments that were correctly classified as PD;
FP experiments that were misclassified as PD;
FN experiments that were misclassified as normal;
TN experiments that were correctly classified as normal.

Since the data we are working with are unbalanced, we also calculated the 95% confidence interval (CI) and the balanced accuracy (average of sensitivity and specificity) using the following formula:

$$\text{Balanced Accuracy} = \frac{\left(\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{TN+FP}}\right)}{2}. \quad (4)$$

To properly evaluate our proposed approach, we analyze our ResNet50 architecture and the use of heatmaps to diagnose PD on six different research questions.

1) We have three candidate feature sets (TQWT, DWT, and MFCC). Which feature set will produce the best performance with ResNet50 model?
2) Does the use of more frequency-domain algorithms (rows) and more features (columns) increase the performance?
3) The ResNet50 architecture that we use was pretrained on ImageNet. How effective is transfer learning compared to training a fresh new ResNet50 model?
4) The ResNet50 architecture was pretrained on ImageNet using more than 1 million images of size 224×224. Should the size of the heat maps be adapted to the same or similar dimensions used by ImageNet, or can we use the images of different dimensions?
5) In each feature set, there are several frequency-domain algorithms. Each of these algorithms will become a row in the heat map. Each row will have many feature values, represented as columns. Does the order of the frequency-domain algorithms and their feature values have impact on the diagnosis of PD? If the order of the rows or columns is changed in both the training set and the test set, will the PD diagnosis stays the same?
6) How does the performance of our proposed approach compare to previous work that used the same dataset?

A. Research Question 1

Section IV discussed three different frequency-domain feature sets that were applied to the original voice recording. It is important to determine which feature set give better prediction results. We constructed three different types of heat map figures, where each type used a different combination of algorithms. There is only one restriction of the construction of different types of heat maps. Each row in the heat map must have the

TABLE I
THREE HEAT MAP TYPES ORDERED ACCORDING TO SIZE (LARGEST TO SMALLEST)

Heat Map Type	Columns	Rows	Image Dimension	Image Size
Type 1 (TQWT)	36	12	707×263	58KB – 70KB
Type 2 (DWT)	20	7	665×258	28KB – 36KB
Type 3 (MFCC)	14	6	548×258	22KB – 29KB

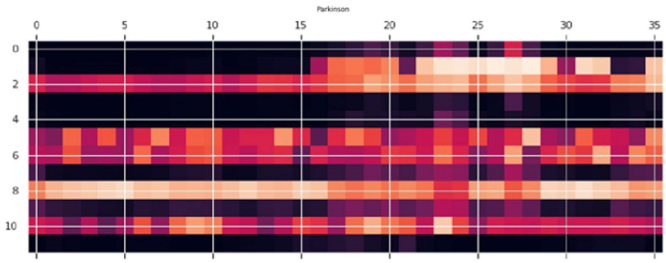

Fig. 6. Type 1 heat map – 12 rows × 36 columns.

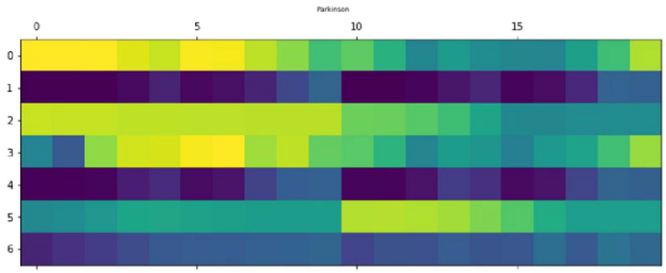

Fig. 7. Type 2 heat map – 7 rows × 20 columns.

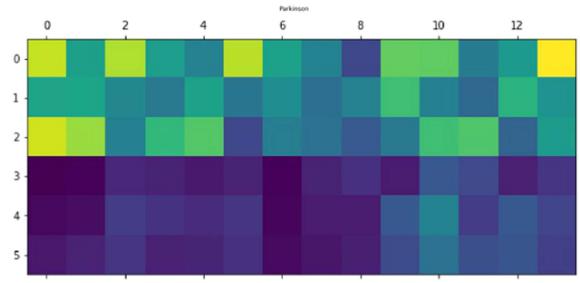

Fig. 8. Type 3 heat map – 6 rows × 14 columns.

TABLE II
EVALUATION RESULTS FOR THE THREE HEAT MAP TYPES

Model	Fine Tuning	Accuracy (CI 95%)	Balanced Accuracy	F-Measure
TQWT	No	**0.901 (0.0475)**	**0.875**	**0.934**
DWT	No	0.795 (0.0642)	0.607	0.877
MFCC	Yes	0.834 (0.059)	0.663	0.901

The best accuracies and F-measures are marked using bold fonts.

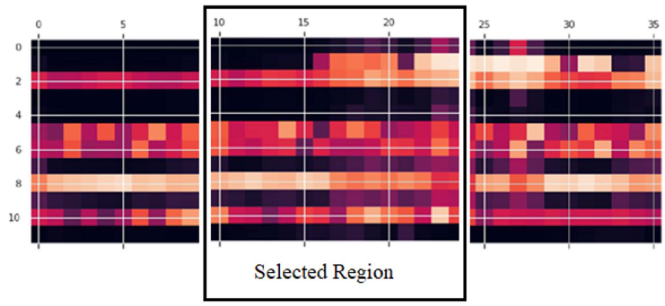

Fig. 9. Example of a cropped heat map.

same number of columns. Otherwise, there will be empty regions in the map, which will be problematic for the model. As a result, we only select the frequency algorithms (functions) that produce the same number of feature values. For example, the TQWT heat map in Table I uses 12 different frequency functions, each produces 32 feature values. Heat maps of this type have a dimension of 707 × 263 pixels and a size that ranges between 58 and 70 kB. The different algorithms (functions) used in each row of TQWT, DWT, and MFCC heat maps have been described in Section IV in details.

Fig. 6 shows an example of the TQWT heat map for the voice recording of a PD patient. Figs. 7 and 8 show the heatmaps for DWT and MFCC, respectively. The image labels are for demonstration purpose and are not a part of the heat map.

The dataset used for each heatmap type is composed of 756 heat map images, where 192 of these heat maps do not have PD. The dataset is split into 80% training and 20% testing with a fixed random seed. For each type (TQWT, DWT, and MFCC) of heat map, the ResNet50 architecture is first pretrained on ImageNet, the last layer of the pretrained model is then trained using the training set, then the architecture is evaluated using the test set. Table II presents that the TQWT features gave the best result, a 93% F1-measure value. On the other hand, DWT gave the lowest prediction results and the lowest balanced accuracy. MFCC gave better accuracy F-measure results when all architecture layers are trained, but it still gave a low balanced accuracy.

B. Research Question 2

The results from Table II show that using a heat map with more frequency-domain algorithms (rows) and more features (columns) increases the performance of the ResNet50 model. This is because the TQWT heat maps have the greatest number of features and functions. Yet, the results are insufficient to conclude that using heat map with more features give better results since the algorithms used in the three types of heatmap are different. In order to evaluate the impact of the number of features in a heat map on the deep learning model, we train and evaluate the architecture on a subset of the heat map instead of the entire heat map. This is done by only taking the middle region of the heat map and discarding the rest, as shown in Fig. 9. Fig. 10 shows the cropped image for each heat map type.

Table III presents the comparison of the accuracy as well as the balanced accuracy of the model before and after the cropped size is used instead of the original heat map. TQWT heat map model showed a −6.7% drop in accuracy when fewer features are used to train and evaluate the model. On the other hand, DWT showed a 2% slight improvement in accuracy, and MFCC showed a 0.7% slight improvement. TQWT had a drop in accuracy since

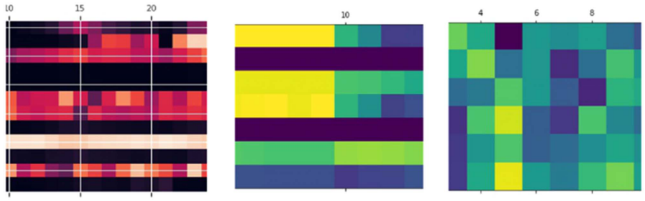

Fig. 10. Cropped image for heatmap types 1, 2, and 3.

TABLE III
EVALUATION RESULTS FOR THE THREE HEAT MAP TYPES

Model	Accuracy (Original)	Balanced Accuracy (Original)	Accuracy (Cropped)	Balanced Accuracy (Cropped)
TQWT	**0.901 (0.0475)**	0.875	0.834 (0.059)	0.752
DWT	0.795 (0.0642)	0.607	0.815 (0.0618)	0.650
MFCC	0.834 (0.059)	0.663	0.841 (0.0581)	0.677

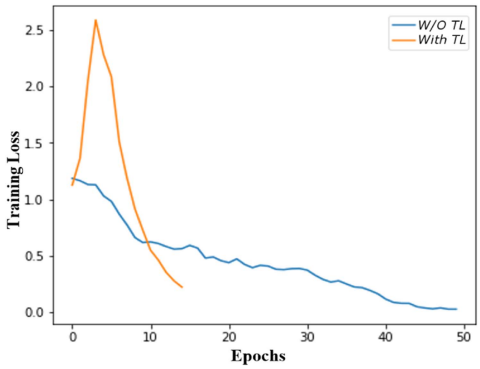

Fig. 11. Training loss of ResNet models with and without transfer learning after each epoch.

cropping the heat map images removed more features, while DWT and MFCC had lower number of features removed. This verifies our assumption that having more columns in heat maps allows the ResNet architecture to extract more image features to help PD prediction.

C. Research Question 3

In this section, we verify the effectiveness of transfer learning. We compare the performance of two ResNet50 models. The first model was trained from scratch. The second model was trained on a ResNet50 architecture that was pretrained on ImageNet using more than 1 million images of size 224 × 224. We randomly selected 80% of the samples from Sakar's dataset [27] for training and used the remaining 20% samples for test.

For both models, we have recorded the loss and the accuracy after each epoch of training. The results are displayed in Figs. 11 and 12. In Fig. 11, the horizontal axis displays the number of training epochs, while the vertical axis displays the loss after each training epoch. The blue curve shows the decrement of loss of the ResNet model without transfer learning, while the orange curve shows the decrement of loss of the Resnet model with transfer learning. It is observed that, with transfer learning, the training loss decreases significantly faster. Fig. 12 displays the

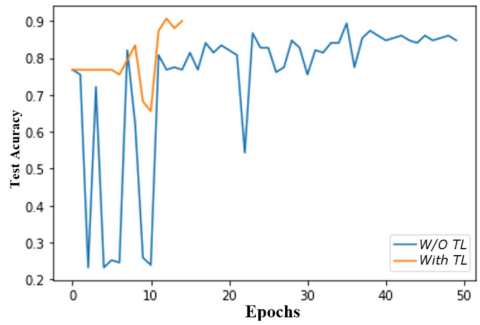

Fig. 12. Test accuracy of ResNet models with and without transfer learning after each epoch.

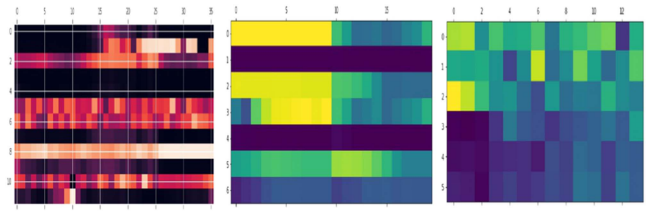

Fig. 13. Squished image for heat map types 1, 2, and 3.

TABLE IV
EVALUATION RESULTS USING HEAT MAPS

Model	Accuracy (Original)	Balanced Accuracy (Original)	Accuracy (Squished)	Balanced Accuracy (Squished)
TQWT	0.901 (0.0475)	0.875	**0.934 (0.0395)**	**0.887**
DWT	0.795 (0.0642)	0.607	0.762 (0.0672)	0.585
MFCC	0.834 (0.059)	0.663	0.854 (0.056)	0.766

increment of the test accuracy after each epoch. The blue curve represents the ResNet model without transfer learning, while the orange curve represents the Resnet model with transfer learning. It is also observed that, with transfer learning, the test accuracy increases much faster, and the performance gets stable much faster as well.

D. Research Question 4

The ResNet50 architecture was pretrained on ImageNet using more than 1 million images of size 224 × 224 pixels. In Section V-A, the last layer was trained on images of incompatible size (refer to Table I). Next, we are going to train and evaluate our CNN models on heat maps of the same size with ImageNet. For each heat map type, we resized ("squished") the original size of the heat map to a compatible square sized image of 224 × 224 pixels. Fig. 13 shows an example for each heat map after resizing. The rows and columns still represent the same feature values, but the images' aspect ratios have been changed.

Table IV presents the performance of the three types when the squished images are used instead of the original ones. TQWT and MFCC heat map types showed good increase in both accuracy and balanced accuracy, while MFCC heat map type showed a decrease with the squished images. TQWT had a 3% of improvement in accuracy, while MFCC had a 10% increase

TABLE V
EVALUATION RESULTS USING HEAT MAPS

Model	Accuracy (CI 95%)	Balanced Accuracy	F-Measure
ResNet50 (TQWT)	0.901 (0.0475)	0.875	0.934
Columns shuffled, same order	0.914 (0.0446)	0.854	0.945
Rows shuffled, same order	0.921 (0.0429)	0.878	0.949
Rows shuffled, random order	0.781 (0.0657)	0.588	0.870

in balanced accuracy. The results show that using resized heat maps of the same size as the pretrained architecture instead of the original heat maps does not have a significant impact to the overall performance.

E. Research Question 5

We believe that the order of the frequency-domain algorithms and their feature values on the heat map has little impact to the diagnosis of PD, as long as the order of features in the training set and the test set is consistent. This is because the learning process of ResNet50 can adapt to the order of the rows and columns. Even if two relevant key features are placed far away from each other on the heat map, the pooling process can still get them close enough at a coarse scale so that they will be fed to some filter of a convolution layer. To verify our assumption, we conducted three experiments using the TQWT heat maps. In the first experiment, we shuffled the rows of the heat maps in both training set and test set into the same new order. In the second experiment, we shuffled the columns in both training set and test set into the same new order. Finally, we randomly shuffled the rows in training set and the rows in the test set into different new orders.

Table V presents a significant drop of 12% in accuracy and 28% drop in balanced accuracy when the rows are shuffled, and their order in the training and test is inconsistent. There is a slight increase of accuracy after we shuffle the rows or the columns as long as the order of features in the training set and test set is consistent. A 1.3% accuracy increase (2.1% decrease in balanced accuracy) is recorded when the columns are shuffled, and a 2% accuracy increase (0.3% increase in balanced accuracy) when the rows are shuffled. In general, we can conclude that the order of the frequency-domain algorithms and their feature values on the heat map does not have a significant impact to the diagnosis of PD.

F. Research Question 6

Gunduz [31] applied two different CNN frameworks, each with nine layers, and a support-vector machine (SVM) model on Sakar's dataset [27]. This is the same dataset we are using to evaluate our proposed approach. Therefore, we will compare their results to ours.

In their article, they split the data in a subjectwise way using leave-one-person-out cross validation. In each fold, they used 753 voice recordings of 251 participants for training and 3 voice recordings of the remaining participant for testing. The process was repeated 252 times, so each participant's data were used as the test sample once. To have a fair comparison, we followed their process. This is different from the previous experiments we have done where we split the training and testing into 80% and 20% without subjectwise consideration. The training and testing of our model using cross validation still followed the same process that we described in Section IV.

TABLE VI
RESNET50 VERSUS GUNDUZ'S MODELS USING TQWT AS FEATURES

Model	Accuracy	F-Measure
ResNet50 (Heat map Type 1 squished)	**0.885**	**0.926**
CNN_{M1}[31]	0.825	0.888
SVM[31]	0.829	0.894

TABLE VII
RESNET50 VERSUS GUNDUZ'S MODELS USING MFCC AS FEATURES

Model	Accuracy	F-Measure
ResNet50 (Heat map Type 3, squished)	**0.873**	**0.920**
CNN_{M1}[31]	0.782	0.864
SVM[31]	0.817	0.885

TABLE VIII
RESNET50 VERSUS GUNDUZ'S MODELS USING TQWT AS FEATURES

Model	Accuracy	F-Measure
ResNet50 (Heat map Type 1 squished)	**0.885**	**0.926**
CNN_{M2} (TQWT + MFCC + Concat) [31]	0.869	0.917
SVM (TQWT + MFCC) [31]	0.857	0.910
CNN_{M2} (TQWT + MFCC) [31]	0.849	0.902
CNN_{M1} (TQWT + Wavelet) [31]	0.845	0.902

Similar to our findings, in their article, the TQWT features gave the best results among the rest of the frequency-domain features. Table VI presents that our model outperforms the first CNN architecture and the SVM in both accuracy and F-measure.

We also compared our proposed model with theirs when the MFCC features were used. It can be seen from Table VII that our ResNet50 model outperforms both their CNN models and their SVM models. What is also interesting that our model had better accuracy (5% increase) when leave-one-person-out cross validation is used instead of 80%–20% split.

Gunduz improved on the performance of the models by concatenating additional features to the TQWT. As shown in Table VIII, the best performance they got was using the second CNN architecture on TQWT, MFCC, vocal fold, and time–frequency features. Our model still outperformed their best models by only using TQWT (Heat map type 1) as features for the ResNet50 architecture. Using less features means less effort to convert the speech recordings with different frequency-domain algorithms and less time to train the deep learning model.

VI. CONCLUSION

In this article, we proposed a method to use ResNet50, a residual network that has 50 layers, to process the subjects' speech recordings to help diagnosing PD. Comparing to the classic CNN, ResNet can be trained more effectively and can

be built with more layers because ResNet utilized skip connection to learn the residual features instead of the convolutional features. To take advantage of ResNet50, we converted the frequency-domain feature values into a 2-D heat map. These heat maps are passed as the only source of input to the ResNet50 architecture to predict whether the patient has PD.

The evaluation results verified our assumption that color encoded figures that represent frequency-domain features and their relative position contain key patterns that can be used to diagnose PD. By converting frequency-domain feature values into a 2-D heat map, we are able to use a deep neural network to extract the underlying pattern among the relative values of the features as well as the absolute values. By the aid of transfer learning, we are able to quickly train and tune a ResNet50 model that was pretrained for general image recognition task with a relatively small training set collected from PD patients and effectively use the new model in the diagnose of PD.

From the experimental results, we found the following discoveries.

1) Comparing to DWT features and MFCCs, the heat map of the TQWT features produced the best performance with ResNet50. The model achieved an accuracy of 93.4% (research question 4).
2) Manually reducing the size of the features significantly decreases the accuracy.
3) It is possible to resize the heat maps as if they are images to the size of images in ImageNet that were used to pretrain the ResNet50 architecture. Resizing did improve the accuracy of PD diagnosis when MFCC or TQWT is used.
4) The order of the features in the heat map does not have a significant influence on the accuracy as long as the training set and the test set have the same order. This validates our assumption that using a heat map allows the deep neural network architecture the freedom to decide which rows and columns are most advantageous to the classification task.
5) Our proposed method has outperformed recent state-of-the-art classifiers that were used to diagnose PD in both accuracy and F-measure. Similar to previous work, TQWT feature set outperformed the other feature sets.

In the future, we will study the combination of multiple frequency functions, such as the combination of MFCC and TQWT. In addition, using an existing dataset restricted us from experimenting with new features, therefore, we will also investigate new features from speech recordings. We plan to develop a method to combine the features generated by different frequency functions into one heat map and see whether the performance of the deep neural network models can be improved by using such a combined heat map. We also want to investigate the impact of color information on ResNet50 architecture for the diagnosis of PD.

REFERENCES

[1] National Institute on Aging, "Parkinson's disease," Accessed: Mar. 6, 2014. [Online]. Available: http://www.nia.nih.gov/health/parkinsons-disease

[2] Parkinson's Foundation, "Statistics," Accessed: Sep. 5, 2014. [Online]. Available: https://www.parkinson.org/Understanding-Parkinsons/Statistics

[3] C. Marras et al., "Prevalence of Parkinson's disease across North America," *NPJ Parkinsons Dis.*, vol. 4, no. 1, Jul. 2018, Art. no. 21, doi: 10.1038/s41531-018-0058-0.

[4] T. R. Mhyre, J. T. Boyd, R. W. Hamill, and K. A. Maguire-Zeiss, "Parkinson's disease," *Subcell Biochem.*, vol. 65, pp. 389–455, 2012, doi: 10.1007/978-94-007-5416-4_16.

[5] C. W. Olanow et al., "A double-blind, delayed-start trial of Rasagiline in Parkinson's disease," *New England J. Med.*, vol. 361, no. 13, pp. 1268–1278, Sep. 2009, doi: 10.1056/NEJMoa0809335.

[6] O. Rascol et al., "A double-blind, delayed-start trial of rasagiline in Parkinson's disease (the ADAGIO study): Prespecified and post-hoc analyses of the need for additional therapies, changes in UPDRS scores, and non-motor outcomes," *Lancet Neurol.*, vol. 10, no. 5, pp. 415–423, May 2011, doi: 10.1016/S1474-4422(11)70073-4.

[7] R. A. Hauser et al., "Long-term outcome of early versus delayed rasagiline treatment in early Parkinson's disease," *Movements Disorder*, vol. 24, no. 4, pp. 564–573, Mar. 2009, doi: 10.1002/mds.22402.

[8] F. L. Pagan, "Improving outcomes through early diagnosis of Parkinson's disease," *Amer. J. Managed Care*, vol. 18, no. 7, pp. S176–S182, Sep. 2012.

[9] National Institute of Neurological Disorders and Stroke, "Parkinson's disease: Hope through research," Accessed: Mar. 6, 2021. [Online]. Available: https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Parkinsons-Disease-Hope-Through-Research

[10] R. Z. U. Rehman et al., "Comparison of walking protocols and gait assessment systems for machine learning-based classification of Parkinson's disease," *Sensors (Basel)*, vol. 19, no. 24, Dec. 2019, Art. no. 5363, doi: 10.3390/s19245363.

[11] S. L. Oh et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 10927–10933, Aug. 2020, doi: 10.1007/s00521-018-3689-5.

[12] W. R. Adams, "High-accuracy detection of early Parkinson's disease using multiple characteristics of finger movement while typing," *PLoS One*, vol. 12, no. 11, Nov. 2017, Art. no. e0188226, doi: 10.1371/journal.pone.0188226.

[13] S. S. Upadhya, A. N. Cheeran, and J. H. Nirmal, "Thomson Multitaper MFCC and PLP voice features for early detection of Parkinson disease," *Biomed. Signal Process. Control*, vol. 46, pp. 293–301, Sep. 2018, doi: 10.1016/j.bspc.2018.07.019.

[14] M. Y. Thanoun and M. T. Yaseen, "A comparative study of Parkinson disease diagnosis in machine learning," in *Proc. 4th Int. Conf. Adv. Artif. Intell.*, New York, NY, USA, Oct. 2020, pp. 23–28, doi: 10.1145/3441417.3441425.

[15] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguistics*, vol. 17, no. 6, pp. 439–453, Nov. 2004, doi: 10.1016/j.jneuroling.2004.06.001.

[16] B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study," *Brain Cogn.*, vol. 56, no. 1, pp. 24–29, Oct. 2004, doi: 10.1016/j.bandc.2004.05.002.

[17] B. Klimova and K. Kuca, "Speech and language impairments in dementia," *J. Appl. Biomed.*, vol. 14, pp. 97–103, Apr. 2016, doi: 10.1016/j.jab.2016.02.002.

[18] J. L. Cummings, A. Darkins, M. Mendez, M. A. Hill, and D. F. Benson, "Alzheimer's disease and Parkinson's disease: Comparison of speech and language alterations," *Neurology*, vol. 38, no. 5, pp. 680–684, May 1988, doi: 10.1212/wnl.38.5.680.

[19] S. Montemurro, S. Mondini, M. Signorini, A. Marchetto, V. Bambini, and G. Arcara, "Pragmatic language disorder in Parkinson's disease and the potential effect of cognitive reserve," *Front. Psychol.*, vol. 10, 2019, Art. no. 1220, Accessed: May 6, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.01220

[20] National Institute on Aging, "What is dementia? Symptoms, types, and diagnosis," Accessed: May 7, 2022. [Online]. Available: https://www.nia.nih.gov/health/what-is-dementia

[21] National Institute on Aging, "What is lewy body dementia? Causes, symptoms, and treatments," Accessed: May 7, 2022. [Online]. Available: https://www.nia.nih.gov/health/what-lewy-body-dementia-causes-symptoms-and-treatments

[22] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *J. Acoust. Soc. Amer.*, vol. 129, no. 1, pp. 350–367, Jan. 2011, doi: 10.1121/1.3514381.

[23] A. Ma, K. K. Lau, and D. Thyagarajan, "Voice changes in Parkinson's disease: What are they telling us?," *J. Clin. Neurosci.*, vol. 72, pp. 1–7, Feb. 2020, doi: 10.1016/j.jocn.2019.12.029.
[24] A. Puce and M. S. Hämäläinen, "A review of issues related to data acquisition and analysis in EEG/MEG studies," *Brain Sci.*, vol. 7, no. 6, May 2017, doi: 10.3390/brainsci7060058, Art. no. 58.
[25] S. Sivaranjini and C. M. Sujatha, "Deep learning based diagnosis of Parkinson's disease using convolutional neural network," *Multimed. Tools Appl.*, vol. 79, no. 21, pp. 15467–15479, Jun. 2020, doi: 10.1007/s11042-019-7469-8.
[26] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease," *Future Gener. Comput. Syst.*, vol. 117, pp. 234–244, Apr. 2021, doi: 10.1016/j.future.2020.11.020.
[27] C. O. Sakar et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Appl. Soft Comput.*, vol. 74, pp. 255–263, Jan. 2019, doi: 10.1016/j.asoc.2018.10.022.
[28] A. Fretzayas, M. Moustaki, I. Loukou, and K. Douros, "Differentiating vocal cord dysfunction from asthma," *J. Asthma Allergy*, vol. 10, pp. 277–283, Oct. 2017, doi: 10.2147/JAA.S146007.
[29] W. C. Chiang, A. Goh, L. Ho, J. P. L. Tang, and O. M. Chay, "Paradoxical vocal cord dysfunction: When a wheeze is not asthma," *Singap. Med. J.*, vol. 49, no. 4, pp. e110–e112, Apr. 2008.
[30] W. H. Ibrahim, H. A. Gheriani, A. A. Almohamed, and T. Raza, "Paradoxical vocal cord motion disorder: Past, present and future," *Postgraduate Med. J.*, vol. 83, no. 977, pp. 164–172, Mar. 2007, doi: 10.1136/pgmj.2006.052522.
[31] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564.
[32] Z. K. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Med. Hypotheses*, vol. 138, May 2020, Art. no. 109603, doi: 10.1016/j.mehy.2020.109603.
[33] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease," *Expert Syst. Appl.*, vol. 110, pp. 182–190, Nov. 2018, doi: 10.1016/j.eswa.2018.06.003.
[34] O. L. Mangasarian and D. R. Musicant, "Lagrangian support vector machines," *J. Mach. Learn. Res.*, vol. 1, pp. 161–177, Jan. 2001, doi: 10.1162/15324430152748218.
[35] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3560–3575, Aug. 2011, doi: 10.1109/TSP.2011.2143711.
[36] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. Hosseini Ghomi, "Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease," *J. Biomed. Informat.*, vol. 104, Apr. 2020, Art. no. 103362, doi: 10.1016/j.jbi.2019.103362.
[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 25, Jan. 2012, pp. 1097–1105, doi: 10.1145/3065386.
[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.
[39] Shivangi, A. Johri, and A. Tripathi, "Parkinson disease detection using deep neural networks," in *Proc. 12th Int. Conf. Contemporary Comput.*, Aug. 2019, pp. 1–4, doi: 10.1109/IC3.2019.8844941.
[40] I. Kollia, A.-G. Stafylopatis, and S. Kollias, "Predicting Parkinson's disease using latent information extracted from deep neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8851995.
[41] Z.-J. Xu, R.-F. Wang, J. Wang, and D.-H. Yu, "Parkinson's disease detection based on spectrogram-deep convolutional generative adversarial network sample augmentation," *IEEE Access*, vol. 8, pp. 206888–206900, 2020, doi: 10.1109/ACCESS.2020.3037775.
[42] H.-C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
[43] N. Ho and Y.-C. Kim, "Evaluation of transfer learning in deep convolutional neural network models for cardiac short axis slice classification," *Sci. Rep.*, vol. 11, no. 1, Jan. 2021, Art. no. 1839, doi: 10.1038/s41598-021-81525-9.
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[45] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
[46] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proc. ACM Int. Conf. Multimodal Interaction*, New York, NY, USA, Nov. 2015, pp. 443–449, doi: 10.1145/2818346.2830593.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2015, in *Proc. 3rd Int. Conf. Learn. Representations*, 2015.
[48] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[49] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," Apr. 2017, *arXiv:1605.07678*, Accessed: Sep. 5, 2021. [Online]. Available: http://arxiv.org/abs/1605.07678
[50] UCI Machine Learning Repository, Accessed: Sep. 6, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/index.php