

Identifying Parkinson’s Disease Through the Classification of Audio Recording Data

James Bielby
De Montfort University
School of Computer Science
and Informatics
Leicester, United Kingdom
P15182608@my365.dmu.ac.uk

Dr. Stefan Kuhn
De Montfort University
Software Technology Research Laboratory
Leicester, United Kingdom
stefan.kuhn@dmu.ac.uk

Dr. Simon Colreavy-Donnelly
De Montfort University
Institute of Artificial Intelligence
Leicester, United Kingdom
simon.colreavy-donnelly@dmu.ac.uk

Dr. Fabio Caraffini
De Montfort University
Institute of Artificial Intelligence
Leicester, United Kingdom
fabio.caraffini@dmu.ac.uk

Dr. Stuart O’Connor
De Montfort University
Cyber Technology Institute
Leicester, United Kingdom
stuart.oconnor@dmu.ac.uk

Dr. Zacharias A. Anastassi
De Montfort University
Institute of Artificial Intelligence
Leicester, United Kingdom
zacharias.anastassi@dmu.ac.uk

Abstract—Developments in artificial intelligence can be leveraged to support the diagnosis of degenerative disorders, such as epilepsy and Parkinson’s disease. This study aims to provide a software solution, focused initially towards Parkinson’s disease, which can positively impact medical practice surrounding degenerative diagnoses. Through the use of a dataset containing numerical data representing acoustic features extracted from an audio recording of an individual, it is determined if a neural approach can provide an improvement over previous results in the area. This is achieved through the implementation of a feed-forward neural network and a layer recurrent neural network. By comparison with the state-of-the-art, a Bayesian approach providing a classification accuracy benchmark of 87.1%, it is found that the implemented neural networks are capable of average accuracy of 96%, highlighting improved accuracy for the classification process. The solution is capable of supporting the diagnosis of Parkinson’s disease in an advisory capacity and is envisioned to inform the process of referral through general practice.

Index Terms—Parkinson’s disease, recurrent neural network, audio processing, pre-diagnostic tools.

I. INTRODUCTION

Since the beginnings of Medicine, developments in technology have aided doctors with making diagnosis [1], [2], disseminating prevention schemes [3], and prescription of therapies [4]. The vertiginous technological growth of the last few decades made it possible for the introduction of Computer-Aided Diagnosis (CAD) and other devices in medicine [5], [6]. CAD systems have presented themselves in many formats, from simple tools for e.g. monitoring blood pressure and heart rate [7], to more advanced embedded platforms capable of image recognition, as done e.g. for brain scans [8]. A common CAD system referred to as a “Dopamine Transporter Scan”, aka DaTScan, is used to observe the movement of dopamine through the body to further strengthening Parkinson’s disease (PD) diagnoses [9]–[11], thus identifying pieces of information that may go unseen otherwise [12], [13]. Machine learning

and deep learning has been used here in a variety of formats from image analysis with convolutional neural network (CNN) technologies to semantic data mining analysis of extensive patient databases [14]. An important field of CAD is image analysis. “U-Net” [15] is a typical example for this, using a CNN to distinguish regions in medical images.

Recent advances in AI are being applied to diagnose and timely deal with degenerative disorders [16] such as e.g. epilepsy, Parkinson’s disease, Alzheimer’s disease, multiple sclerosis, and ischaemic brain stroke by analysing EEG signals [17], MRI images [18], and other data types [19]. This investigation focuses on Parkinson’s Disease and proposes a method based on the recurrent neural network approach to its early detection through the analysis of audio signals.

II. PARKINSON’S DISORDER AND ITS DIAGNOSIS

It is known that Parkinson’s disease is caused by a loss of nerve cells, specifically, those within the Substantia Nigra (see Fig. 1), which is responsible for housing the majority of the brain’s dopamine neurons. Since the latter is attributed to functions, like motor control and movement [4], [6], [20], [21], its predominant symptoms are tremors of the body, slow or sluggish movement, and stiff/rigid muscles. These symptoms only show in a relatively late stage, thus being able to detect this illness at the early stages, i.e. before the appearance of such symptoms, could increase the quality of life of patients and provide better treatment options.

As part of the broader medical technologies revolution, Parkinson’s disease is also an area of significant development [22]. Using systems like neural networks and fuzzy inference systems allows for more customisation and a deeper understanding of the underpinning issues. Implementing neural networks and similar deep learning systems is not always about improving the performance of pre-existing systems, but also facilitate access to the test. Another goal is to achieve the

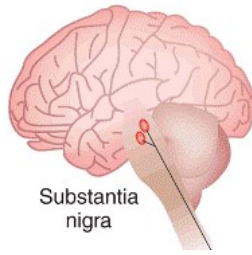


Fig. 1. Substantia Nigra position in the human brain.

same or similar results as a medical professional but provide a diagnosis in a fraction of the time. Medical professionals can take weeks, months or even years to provide an accurate diagnosis [23]. Deep learning and machine learning can assist in diagnosing based on subtle variations in the onset of initial symptoms. For this reason, a system that provides a reasonably accurate diagnosis within moments would be of great use.

Parkinson’s disease shares many symptoms with other tremor-related disorders. Because of this, the clinical diagnosis of Parkinson’s is a very long procedure and unfortunately still holds a significant misdiagnosis rate. Sources range from 19% to 24% [12], [24], [25]. The study in [26] describes an experiment where audio recordings were used to identify Parkinson’s within individuals. A Bayesian approach, a common data mining technique, was applied to this dataset, concluding that identification of individuals with Parkinson’s from audio recordings is possible with high accuracy (87.1%).

Another recent study shows good results for detecting Parkinson’s disease from audio recording [27]. Here, new recordings are done, using either a smartphone or professional recording equipment. Phonation (as in [26]) is used as well as speech. Furthermore, a different set of features is extracted. Therefore, the results are not directly comparable to those from [26] and from this paper. We still will include the results for comparison and discuss them in the Results section.

Further experiments have followed this study and several feature extraction methods [28] were employed to extract information from audio records. These often require tailored processing to manipulate data e.g. based on the gender of the individual. This premise relies on the fact that males and females have differing skeletal structure around the jaw. This should bridge the classification gap between genders. On the other hand, there are also publications arguing that no manipulation or pre-processing of this kind is needed at all [29]. We have decided to use the features extracted in [26] without changes, so this study focuses solely on the analysis step and not the feature extraction or data retrieval.

This paper aims to show those results still have room for improvement and that a machine learning approach can even beat human performance.

III. DATA AND METHODS

A. The data set

The study in [26] involved individuals conducting voice recordings. In particular, 80 individuals participated, 40 of which suffered from Parkinson’s disease (group 1) and 40 healthy individuals (group 2). Each group contains 31 females and 9 males individuals, with group 1 being formed by members of the “Regional Association Parkinson’s Disease” of Extremadura, who consented to provide 3 audio recordings per individual. This results in a database of 240 records. Each record is 5 seconds long and consists of a participant pronouncing the letter “a”. There are 48 variables per record [30]: ID of the patient, recording number (1 to 3), health status (Parkinsons patient or healthy), gender, 4 variables describing the “pitch local perturbation”, 5 variables for the “amplitude perturbation”, 5 “harmonic to noise ratio”, 3 nonlinear features, the “glottal to noise excitation ratio”, and 26 variables relating to “Mel frequency cepstral” coefficients (13 static and 13 time-sensitive).

As the dataset contains multiple records from the same individuals, there are several ways to judge the quality of the result. The system could be run on the full dataset (240 records), treating them as uniform records. We will call this the “uniform data” here. This provides the system with more knowledge but could result in some cases where one of the three recordings are classified differently from the other two. This would cause some serious doubts about the legitimacy of the outcomes. The alternative would be to use only 80 records, one from each user. This removes the issue of incoherent results but can suffer a decrease in accuracy due to less training material. For this study, the first option of using the full dataset was used, as it was done in [26]. To justify the result we look, in a second step, at results for all three recordings for each patient as an additional quality measure (called the “individualized data” hereafter). Here, coherency is our main concern. An incoherent result is defined as one where not all three recordings from an individual are classified into the same class, whereas a coherent result has three identical classifications.

B. Evaluation metrics

We used a Confusion Matrix to categorise the results. This takes the form of a two by two grid showing; True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These four values allow for the calculation of the four standard performance metrics: Accuracy, Recall (Sensitivity), Specificity, and Precision. The study in [26] uses these metrics to showcase the outcomes of their work, and calculating these values allows for a 1:1 comparison against the original work. The equations used to calculate these metrics are:

- Accuracy:

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

- Recall:

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

- Specificity:

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

- Precision:

$$\pi = \frac{TP}{TP + FP} \quad (4)$$

These metrics are commonly referred to in terms of classification. Precision is commonly referred to as positive classification, and this is because precision is a measure of correctly predicted positives to all positives that have been predicted.

Aside from these values, some additional information is also stored by the system. These values are used to identify if all three results from one individual are matching. For a competent system, it is expected that a high percentage of outcomes are matching when they originate from the same individual. This system simply helps to confirm this and reinforcing the validity of the results. Once the former has been developed, the neural networks can be also developed and tested using a consistent benchmark.

For evaluating the results, the 240 recordings were split into 168 training data and 72 test data. The records were chosen at random, there was an opportunity to ensure at least one record from each individual was included in the training data but this avoided due to the earlier decision to treat each record as an individual. The software used for the neural network experiments was Matlab.

C. Feed Forward Neural Network

To get a benchmark for neural networks, we firstly applied a conventional feed-forward neural network (FFNN). We use the scaled conjugate gradient as the activation function. Since the most important parameter to optimise in such a network is normally the number of neurons, we have done this systematically, using a network with three hidden layers. We have concluded that three hidden layers are the optimal number since experience shows that more hidden layers are rarely ever-improving the results. Pseudocode is given in Algorithm 1 was used to optimise the number of neurons. This results in 45 neurons in the first, 32 neurons in the second, and 6 neurons in the third hidden layer being optimal. Finally, the optimal configuration was run 100 times. The achieved average accuracy is 85%, which is similar to, but slightly worse than, the result from [26]. Since this network could be improved by more tuning, we can assume that a conventional, feed-forward neural network can perform similarly to the Bayesian approach in [26].

D. Layer Recurrent Neural Network

Once the FFNN has been optimised, it is useful to have a comparison. For this, a layer recurrent neural network (henceforth LRNN) was built. Compared to an FFNN, an LRNN is limited to a single hidden layer but is not limited to a single direction. The architecture of an LRNN is shown in

Algorithm 1 Iterative topology testing procedure for an FFNN with three hidden layers. This ensures that the number of neurons in layer $i = 1$ is never higher than in layer i .

```

for each integer  $i \in [2, 3, \dots, 45]$  do           ▷ size of layer 1
  for each integer  $j \in [2, \dots, i]$  do         ▷ size of layer 2
    for each integer  $k \in [2, \dots, j]$  do       ▷ size of layer 3
      Initialise an FFNN with  $i, j, k$  neurons
      Train FFNN
      If best result, store  $i, j, k$ 
    end for
  end for
end for

```

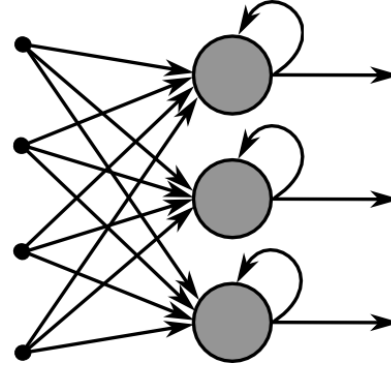


Fig. 2. The architecture of an LRNN [31].

Fig. 2 The data can be passed through the same layer as many times as is required, meaning the data is not going forward only. During training, this process is repeated until a stopping criterion is met. An LRNN is, therefore, mimicking an FFNN with an unlimited number of hidden layers. Along with this difference, layer recurrent neural networks are also able to make use of time-series data.

To improve performance, we also use “Bayesian Regularisation” during training. This aims to minimise training error with the inclusion of neuron weighting. This approach allows a system to generalise to a higher standard and avoid overfitting. Regularisation, and in particular Bayesian Regularisation, is commonly used with shallow neural networks, i. e. networks which contain one single hidden layer. Considering the LRNN technically only has a single layer (with numerous repetitions) these two choices go hand in hand.

As expected, this topology takes significantly longer, per iteration, than the FFNN to train. This is due to the increased complexity of the layer recurrent neural network. However, this complexity also allows for a greater level of customisation of the network. Several areas can be tweaked that will prioritise speed with only a small detriment to network accuracy. One consistent facet of Bayesian Regularisation is the relationship between training iterations and network accuracy. This training function tends to reach a rather high level of accuracy within the first iteration and make very minor improvements for each iteration thereafter.

Since LRNNs have one hidden layer only, the number of

neurons in the architecture is the only topological parameter to optimise. As for the FFNN, we wanted to find the optimal number of neurons as well as the number of training epochs. As the LRNN has a longer average run-time than simpler networks, it is not practical to attempt a comprehensive test. Therefore, a simple test was developed which provided a reasonable effort. To accomplish this we first tested the number of training iterations and secondly searched for a near-optimum number of neurons. Assuming training accuracy tapers to a point, there must be a limit to where time spent on training is no longer worth the increase of the accuracy. To observe this, the code in Algorithm 2 was used. Using the optimal number of epochs found there, we did a grid search for the number of neurons and found 47 neurons to be optimal. Finally, the optimal configuration was run again 100 times. The average accuracy of those runs is 96%.

Algorithm 2 Iterative epoch testing procedure for an LRNN

```

i = 1
while i ≤ 500 do
  Initialise an LRNN with 5 neurons
  Set training epochs to i
  Train NN
  If best result, store i
end while

```

E. Rapid Miner as benchmark

To get a wider benchmark for our results, we use Rapid Miner [32]. Rapid Miner is an integrated data mining tool. It allows the fast execution of many different techniques. At the same time, it is considered a market-leading product, for example, Gartner placed RapidMiner in the leader quadrant of its Magic Quadrant for data science & machine learning platforms for the sixth year in a row [33]. Therefore, the results can be considered a good estimate of how a particular method does on these data. Most likely a specialised tool with specialist knowledge will do better, but for a benchmark, the Rapid Miner results can be considered as valid, with caution.

The methods used from Rapid Miner are (abbreviations used in Table I): Naïve Bayes (NB), Generalised Linear Model (GLM), Logistic Regression (LR), Fast Large Margin (FLM), Deep Learning (DL), Decision Tree (DT), Random Forest (RF), Gradient Boosted Trees (GBT), and Support Vector Machine (SVM). Automatic feature selection is used with all of them. All chosen methods are suitable for a binary classification problem as we have got here. It should be noted that the Rapid Miner model called Deep Learning is a multi-layer feed-forward network and not one of the more advanced types like convolutional neural network or recurrent neural network. We have used the Auto Model feature of Rapid Miner and the default settings throughout, as chosen by Auto Miner, with no custom optimisation. Auto Miner has, for example, used an RBF kernel instead of the default linear kernel in the SVM.

TABLE I
STATISTICAL PERFORMANCE METRICS (FOR ABBREVIATIONS, SEE SECTION III-E, FOR EXPLANATION OF [27] SEE RESULTS)

Method	Accuracy	Recall	Specificity	Precision
FFNN	0.858	0.848	0.868	0.867
LRNN	0.960	0.964	0.955	0.956
[26]	0.871	0.842	0.900	0.894
[27] (AC)	0.9455	0.9455	0.9426	X
[27] (SP)	0.9294	0.9294	0.8921	X
NB	0.824	0.852	0.790	0.821
GLM	0.838	0.881	0.790	0.822
LR	0.766	0.795	0.743	0.776
FLM	0.693	0.943	0.448	0.637
DL	0.810	0.852	0.762	0.787
DT	0.838	0.848	0.824	0.860
RF	0.840	0.852	0.824	0.840
GBT	0.811	0.795	0.824	0.835
SVM	0.853	0.852	0.852	0.856

IV. RESULTS

A. Uniform data

Table I provides the results for the two neural networks designed here, and for comparison those from [26]. [27], and those produced using Rapid Miner. The clinical diagnosis of Parkinson’s, as mentioned previously, has a misdiagnosis rate of 19%-24%. Concerning every metric, the layer recurring neural network outperforms the other methods and its accuracy exceeds a clinical diagnosis. This is a remarkable result in itself. Also, the execution of the method is faster than a clinical diagnosis. The feed-forward neural network is in line with the benchmark results. The benchmark results vary in accuracy but are in line with what is to be expected from the various methods: Simpler methods like linear regression perform worse, whereas modern methods like the SVM perform similarly to the FFNN and [26]. Overall, the LRNN outperforms other neural networks as well as other methods, also those considered state of the art.

We have also included results from [27] in the table. As explained, these cannot be compared directly, since they use different but similar data. Therefore, they can serve as a benchmark. The figures quoted are for the smartphone recording (SP) and the professional microphone (AC). The results are better than those from [26], but not as good as those achieved using the LRNN. The LRNN results show a clear improvement in particular over the smartphone results from [27], which are probably more comparable to ours, than the professional microphone ones.

A noteworthy statistic is that throughout all testing and gathering of results there have been no cases where any of the models have achieved 100 per cent classification. The highest recorded results were within the 100 runs of the LRNN with 47 neurons, which is also the architecture with the highest average. Two results are tied for the highest accuracy. Table II shows the statistic from this result.

Looking at the other benchmark results (all done for the uniform data) the Fast-Large Margin model produced a very interesting result. This model produces a sensitivity (recall)

TABLE II
BEST RESULTS OF 100 RUNS OF THE LRNN

TP	TN	FP	FN	Accuracy	Coherent	Incoherent
120	118	2	0	0.9916667	78	2
118	120	0	8	0.9916667	78	2

TABLE III
FURTHER PERFORMANCE ANALYSIS

Method	F-Score	ρ	π
[26]	0.867	0.842	0.894
FFNN	0.857	0.848	0.867
LRNN	0.960	0.964	0.956
FLM	0.760	0.943	0.637
SVM	0.854	0.852	0.856

value of 0.943, on the face of it, this is a very good result to achieve. However, it does come at the expense of the other performance metrics which are considerably lower than the other results from the benchmark. The four performance metrics were chosen for their relevance within the industry and to create a direct comparison with the authors work. One metric that could have been included is ‘‘F-Score’’. This metric combines recall and precision to provide insight when results improve one statistic at the expense of another and most importantly shows if that trade-off is worth it. F-Score is calculated through the following equation:

$$\text{F-Score} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

Table III shows several of these results. This clearly shows that even the second-highest value for recall the result given by the FLM model does not justify the sacrifice to precision. This is backed up even further by the very low Specificity. Throughout all of the results, it is clear that the layer recurring network surpassed expectations, gives the best result for all metrics, and improved upon the accuracy of [26] by 0.089.

B. Individualised data

To avoid being misled by the uniform data, we also look at coherency for individuals. Not only does the work produced in this study provide higher accuracy than the previous work, but it also produces a higher level of coherency, as shown in Table IV. In [26], there were twenty-two individuals with incoherent results (that is from the three recordings at least one result did not match). Our LRNN improves this to only 8 incoherent results. As before, the FFNN performs roughly similar to existing results.

Looking at the two best runs out of the 100 runs of the LRNN, which were reported in Table II, we note that

TABLE IV
COHERENCY COMPARISON

Method	Coherent	Incoherent
[26]	58	22
FFNN	62	18
LRNN	72	8

all individuals where the results were coherent have correct results. Only if the results were not coherent, there were errors. Since this shows that the coherent results are better than the incoherent ones, we calculated the accuracy of all coherent predictions of the 100 runs. This is 0.99562 for the LRNN and 0.926881 for the FFNN. This result means that we distinguish between trustworthy, coherent results and less trustworthy, incoherent results, and can be very confident in a result if it is coherent. For trustworthy results, we have almost 100% accuracy.

C. Practical considerations

Despite the good results produced, there are several considerations for the practical use of the method. One concern is the use of a voice recording to diagnose Parkinson’s. Although it is clear that the system developed is capable of detecting a difference between a ‘‘healthy’’ individual and those showing Parkinson’s-like symptoms, it is unclear if the system is capable of discerning difference between Parkinson’s and other Voice Disorders. If an individual with a common voice disorder were to be fed through the system, the system may identify them as someone possessing Parkinson’s. The system may be capable of picking up on this, but it is unlikely to be that advanced at this stage. It also would not be able to differentiate that difference between PD and similar tremor-related illnesses. Essential tremor is very similar to Parkinson’s and is regularly present within those suffering from PD but is separate. If someone were to possess essential tremor but not Parkinson’s this may also trigger a positive response from the system and be given a misdiagnosis. To our knowledge, there exist no datasets with recordings of these different types of patients. Alternatively, we suggest the system to be used for advisory purposes. If the person is suspect of holding one of these ailments, then the system can be used to assist a GP’s referral. This would then be followed by the patient attending a consultant diagnosis.

V. FUTURE RESEARCH

In order to deploy the application to potential patients, applications could be created either on a smartphone or through a web page. This would allow users to easily record their voice, if this returned a positive diagnosis it would provide them with material informing them about the disease and how to get professionally diagnosed. Another direction that could be taken is by applying a different methodology. The data for the neural network is purely numerical, this means that the only information the system is getting is numbers. A possible alternative is to replace the numbers with an image of an audio wave. Neural networks have been used within audio recognition in the past.

A neural network called WaveNet [34] has been developed specifically to analyse and create audio waves. It has been altered and used to create speech to text programs and to create realistic language programs. It has even been used to identify where a language comes from (geographically). If this technology was combined with the methodology outlined in

this could allow for the diagnosis of Parkinson's by analysing the audio waves themselves. Despite the system already being very quick, this could potentially speed up the process even further, since there is no numerical data to process. The current technique requires the audio data to be recorded. Then several different data points need to be extracted and processed all of which takes time. Once the data has been converted from audio to numerical it can then be passed through the neural network which takes seconds. If this proposal of using waves works, it would simply be a case of recording the audio then sending it through WaveNet to generate the wave and diagnosis.

A further area of research is to evaluate the method using a larger and more diverse dataset. Whilst the current dataset is large enough to indicate the benefits of our method, a larger dataset can establish them more firmly. Increased diversity of the dataset (e. g. different mother tongues of speakers) can also help to determine the use of the method.

If two proposals outlined were successful, it would not only improve the accuracy of traditional diagnosis but would also definitely provide the solution with a considerable lifespan. It would also increase the speed at which diagnosis takes place, as well as making the diagnosis a much simpler process. Along with these improvements, it would also provide a novel use of a pre-existing system that hopefully could promote creativity and innovation within the industry.

VI. CONCLUSION

Numerical results highlight the effectiveness of the methodology proposed in this study, which displays outstanding classification performances, outperforming the state-of-the-art in multiple aspects. In terms of accuracy, the 87.1% achievement is significantly outperformed thanks to the use of our neural system with a very satisfactory 96%. Moreover, it is worth mentioning that our method has also a higher coherency than all the other comparison methods. This is a significant achievement which means that the inconsistency within an individual's diagnosis is minimised. In this light, the outcome of this study is promising and leads to:

- a faster and more practical diagnosis approach: the algorithm is very fast, since executing a trained neural network is a matter of seconds in the worst case. This also means the diagnosis is much more practical than the patient travelling to hospitals several times.
- High accuracy and coherence: we have achieved an accuracy of 96%, by using an optimised LRNN over 100 runs of the uniform dataset, ignoring the individuals. Looking at these, we achieve coherent results for 80% of the patients. Furthermore, if we restrict ourselves to patients with coherent predictions, we can state that almost 100% of these are correctly classified. This result is better than all comparable studies, and it is also better than typical medical diagnoses.
- A better accessibility: the algorithm, once the network is trained, is inexpensive to run. It can be offered via mobile devices as well as via the internet. This makes frequent

diagnoses possible. This is of particular importance, since the disease may take longer to show itself.

In conclusion, we suggest that an individual being diagnosed should submit three audio recordings and if these recordings produce coherent, then this can then be considered a valid diagnosis. If this indicates the presence of Parkinson's disease, the individual should be required to attend a professional diagnosis. Those with a strong classification should be given priority. If a patient receives an incoherent diagnosis, this should be taken through traditional channels.

REFERENCES

- [1] V. Santucci, A. Milani, and F. Caraffini, "An optimisation-driven prediction method for automated diagnosis and prognosis," *Mathematics*, vol. 7, no. 11, p. 1051, Nov 2019.
- [2] N. Summerton, "The medical history as a diagnostic technology," *British Journal of General Practice*, vol. 58, no. 549, pp. 273–276, 2008.
- [3] N. C. Franklin, C. J. Lavie, and R. A. Arena, "Personal health technology: A new era in cardiovascular disease prevention," *Postgraduate Medicine*, vol. 127, no. 2, pp. 150–158, 2015.
- [4] S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr, and H. Adeli, "Parkinson's disease: Cause factors, measurable indicators, and early diagnosis," *Computers in Biology and Medicine*, vol. 102, pp. 234 – 241, 2018.
- [5] A. Ahmed, N. Ehsan, E. Mirza, S. A. Awan, and A. Ishaque, "Information technology: A means of quality in healthcare," in *2010 3rd International Conference on Computer Science and Information Technology*, vol. 9. IEEE, 2010, pp. 26–30.
- [6] M. Pavel, H. B. Jimison, H. D. Wactlar, T. L. Hayes, W. Barkis, J. Skapik, and J. Kaye, "The role of technology and engineering models in transforming healthcare," *IEEE reviews in biomedical engineering*, vol. 6, pp. 156–177, 2013.
- [7] V. Jahmunah, S. L. Oh, J. K. E. Wei, E. J. Ciaccio, K. Chua, T. R. San, and U. R. Acharya, "Computer-aided diagnosis of congestive heart failure using ecg signals – a review," *Physica Medica*, vol. 62, pp. 95–104, 2019. [Online]. Available: www.scopus.com
- [8] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British journal of radiology*, vol. 78, no. suppl_1, pp. s3–s19, 2005.
- [9] I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. Jiménez-Hoyuela, and S. J. Ortega Lozano, "Automatic assistance to parkinson's disease diagnosis in datscan spect imaging," *Medical Physics*, vol. 39, no. 10, pp. 5971–5980, 2012.
- [10] I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. Jiménez-Hoyuela, and S. J. Ortega Lozano, "Automatic assistance to parkinson's disease diagnosis in datscan spect imaging," *Medical Physics*, vol. 39, no. 10, pp. 5971–5980, 2012.
- [11] F. J. Martínez-Murcia, A. Ortiz, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, and I. A. Illán, "A 3d convolutional neural network approach for the diagnosis of parkinson's disease," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer, 2017, pp. 324–333.
- [12] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of Medical Imaging*, vol. 5, no. 3, p. 036501, 2018.
- [13] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [14] N. Jothi, N. A. Rashid, and W. Husain, "Data mining in healthcare – a review," *Procedia Computer Science*, vol. 72, pp. 306 – 313, 2015, the Third Information Systems International Conference 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050915036066>
- [15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

- [16] U. Raghavendra, U. R. Acharya, and H. Adeli, "Artificial intelligence techniques for automated diagnosis of neurological disorders," *European neurology*, pp. 1–24, 2019.
- [17] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, Feb 2007.
- [18] G. Mirzaei and H. Adeli, "Segmentation and clustering in brain mri imaging," *Reviews in the Neurosciences*, vol. 30, no. 1, pp. 31–44, 2018.
- [19] J. Garre-Olmo, M. Faúndez-Zanuy, K. López-de Ipiña, L. Calvó-Perxas, and O. Turró-Garriga, "Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, alzheimer disease and healthy controls," *Current Alzheimer research*, vol. 14, no. 9, pp. 960–968, 2017.
- [20] Nhs.uk, Parkinson's disease, (2019), [Online] Available at: <https://www.nhs.uk/conditions/parkinsons-disease/> [Accessed 27 January 2020].
- [21] Neuroscientifically Challenged, Know your brain: Substantia nigra, (2019), [Online] Available at: <https://neuroscientificallychallenged.com/blog/know-your-brain-substantia-nigra> [Accessed 27 January 2020].
- [22] D. W. Dickson, H. Braak, J. E. Duda, C. Duyckaerts, T. Gasser, G. M. Halliday, J. Hardy, J. B. Leverenz, K. Del Tredici, Z. K. Wszolek *et al.*, "Neuropathological assessment of parkinson's disease: refining the diagnostic criteria," *The Lancet Neurology*, vol. 8, no. 12, pp. 1150–1157, 2009.
- [23] D. Calne, B. Snow, and C. Lee, "Criteria for diagnosing parkinson's disease," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 32, no. S1, pp. S125–S127, 1992.
- [24] A. Schrag, Y. Ben-Shlomo, and N. Quinn, "How valid is the clinical diagnosis of parkinson's disease in the community?" *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 5, pp. 529–534, 2002.
- [25] 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, p. e0188226, 2017.
- [26] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and J. Martín, "Addressing voice recording replications for parkinson's disease detection," *Expert Systems with Applications*, vol. 46, pp. 286–292, 2016.
- [27] J. S. Almeida, P. P. Rebouças Filho, T. Carneiro, W. Wei, R. Damaševičius, R. Maskeliūnas, and V. H. C. de Albuquerque, "Detecting parkinson's disease with sustained phonation and speech signals using machine learning techniques," *Pattern Recognition Letters*, vol. 125, pp. 55–62, 2019.
- [28] P. Gómez-Vilda, Z. Galaz, J. Mekyska, J. M. F. Vicente, A. Gómez-Rodellar, D. Palacios-Alonso, Z. Smekal, I. Eliasova, M. Kostalova, and I. Rektorova, "Vowel articulation dynamic stability related to parkinson's disease rating features: Male dataset," *International journal of neural systems*, vol. 29, no. 02, p. 1850037, 2019.
- [29] F. J. Martínez-Murcia, J. M. Górriz, J. Ramirez, and A. Ortiz, "Convolutional neural networks for neuroimaging in parkinson's disease: is preprocessing needed?" *International journal of neural systems*, vol. 28, no. 10, p. 1850035, 2018.
- [30] L. Naranjo, uCI Machine Learning Repository: Parkinson Dataset with replicated acoustic features Data Set, (2019), [Online] Available at: <http://archive.ics.uci.edu/ml/datasets/Parkinson+Dataset+with+replicated+acoustic+features+> [Accessed 26 July 2019].
- [31] "Image by chrislb, wikimedia," <https://commons.wikimedia.org/wiki/File:RecurrentLayerNeuralNetwork.png>, [Accessed 25 January 2020].
- [32] M. Hofmann and R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. Chapman & Hall/CRC, 2013.
- [33] "Gartner magic quadrant for data science platforms," <https://rapidminer.com/resource/gartner-magic-quadrant-data-science-platforms/>, [Accessed 25 January 2020].
- [34] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.