Detection of Parkinson's Disease using Machine Learning Algorithms and Handwriting Analysis

Article *in* Journal of Data Mining and Management · December 2022

DOI: 10.46610/JoDMM.2023.v08i01.004

3 authors:

Maya Shelke
Pimpri Chinchwad College Of Engineering

Nihar Madhaw Ranjan
Rajarshi Shahu College of Engineering

Gitanjali Mate
Jayawant Shikshan Prasarak Mandal's Rajarshi Shahu College of Engineering

Journal of
Data Mining and Management

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

www.matjournals.com

https://doi.org/10.46610/JoDMM.2023.v08i01.004

Detection of Parkinson's Disease using Machine Learning Algorithms and Handwriting Analysis

Nihar M. Ranjan[1*], Gitanjali Mate[2], Maya Bembde[3]

[1]Professor, Department of Information Technology, Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

[2]Associate Professor, Department of Information Technology, Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

[3]Assistant Professor, Department of Information Technology, Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

[*]Corresponding Author: nihar.pune@gmail.com

Received Date: March 09, 2023

Published Date: March 29, 2023

ABSTRACT

Parkinson's Disease is a progressive neurodegenerative disorder of movement that affects your ability to control movement. This disease can prove fatal if not detected at an earlier stage. Motor and non-motor symptoms are raised by the loss of dopamine-producing neurons. Currently, there is no test available to detect disease at early stages where the symptoms may be poorly characterised. Handwriting analysis is one of the traditional aspects of studying human personality and also can be used to identify the symptoms of this disease. Identifying such accurate biomarkers provides roots for better clinical diagnosis. In this paper, we proposed a system that makes use of two types of handwriting analysis, spiral and wave drawings of healthy as well as Parkinson's patients as an input to the system. For feature extraction, we are using a histogram of the oriented gradient. The developed system uses a machine learning algorithm and a random forest classifier for the detection of Parkinson's disease among patients. Our model achieved an accuracy of 86.67 % in the case of spiral drawing and 83.30% with wave drawing.

Keywords- Algorithms, Handwriting analysis, Machine learning, Parkinson's disease, Random forest classifier

INTRODUCTION

Parkinson's is a central nervous system disorder that majorly affects movement, including tremors. It mainly destructs the nerve cells which produce dopamine that helps to manage the activities of muscles and their motions [1]. Generally, the initial symptoms of Parkinson's start with a tremor in one hand. The disease also includes other symptoms such as slow movement, stiffness and loss of balance. Generally, among the 1000 people in the universe, 1% of the community which is above 60 years old suffer from PD. There is a higher risk for older people [2]. So, to help patients improve the quality of their lives, there is a need for a practical and definitive detection method. As the symptoms worsen gradually, we can differentiate them into 5 stages. In stage 1 the symptoms are very mild. They are so mild that you cannot guess their severity. But your near and dear ones such as family or friends may observe the difference in your posture, walk, or facial expressions. Stage 2 is an ordinary form of Parkinson's and shows symptoms that are much more distinguishable than the previous stage [3, 4]. Changes such as stiffness, tremors, and trembling may be more detectable. Stage 3 is considered to be a critical movement in the development of the disease. Most of the symptoms are identical to Stage 1. However, PD patients would experience loss of balance and decreased reflexes. The overall rate of movement decreases. Independence is a distinguishable characteristic between stage 3 and stage 4 of Parkinson's disease. The patient does not require any assistance to stand during stage 4. However, the patient may require a walker or other type of assistance. Stage 5 is supercilious to all other stages of Parkinson’s disease. There is advanced stiffness in the legs which causes difficulty in standing or walking [5].

In the human body all actions including

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

https://doi.org/10.46610/JoDMM.2023.v08i01.004

handwriting starts from the brain. Writing or handwriting is a neural activity. For writing purposes, the brain sends signals to the hand by the nervous system. So, the activity of drawing or handwriting involves balancing both hand muscles as well as the brain [6]. Handwriting involves more senses and motor neurons than speaking. As Parkinson's majorly affects the movement of the body, it results in producing a "shake" in the drawings drawn by Parkinson's patients. So, drawing or handwriting can be considered a major characteristic for the detection of Parkinson's Disease [7, 8].

LITERATURE SURVEY

Donato Impedovo et.al in their research work, Dynamic Analysis for the Assessment of Neurodegenerative Diseases: A pattern recognition perspective use online handwriting analysis for the evaluation of the disease from pattern recognition [9]. The detection is done using data acquisition, feature extraction, data analysis, and classification [5]. Data acquisition includes participant recruiting, acquiring devices and identifying the most applicable writing activities.

Participant Recruiting

The following three features can be considered:
Disease: The seriousness of the disorder must be taken into account with the help of standard clinical test scores. Unified PD Rating Scale (UPDRS) is a rating tool, to measure the seriousness and progression of Parkinson's disease in patients. The UPDRS is made up of many sections and these sections are evaluated by interviews and clinical observations.
Patients: To check whether the patient is going through any aid or not. Such parameters must also be considered.
HCs: A group of healthy people (controls) should also be included. For a fair comparison, elder people and young people must also be considered.

Acquisition Device

There is a wide set of devices, available for data acquisition. An ink pen along with a tablet that has paper fixed to it can be used to acquire parameters such as coordinates of the position of the pen, time stamps, pen orientation and pressure. Electronic (smart) pens can be considered an alternative to tablets. The active sensors are used to record the position, acceleration, and tilt angle of the pen also force and shake.

Writing Tasks

This process involves a complex feedback system and helps to examine several cognitive and motor processes. It includes easy drawing activities such as lines, circles etc. which are utilized for shake, speed and acceleration evaluation.

Feature Extraction

In this, the features considered are function features and parameter features.
Function Features: In function features, the samples are distinguished based on a time function whose values compose the set of features. The common function features are position in terms of coordinates, pressure, altitude, displacement and acceleration.
Parameter Features: In parameter features, the handwritten samples are distinguished as a vector of elements, every element showing the value of the feature. Some of these parameter features are gained using the transformation of the function features.

Data Analysis and Classification

This section mainly focuses on depicting the association between the activities, features and main outcomes which are spotted. Imran Razzak et.al A deep analysis of handwritten notes for early diagnosis of neurological disorders in research work, the system takes handwritten tasks such as meanders, spirals and waves as input to detect Parkinson's disease [10]. Further, the system observes different complexities during handwriting tasks through thorough analysis using four frequently used architectures such as AlexNet, GoogleNet,

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)
https://doi.org/10.46610/JoDMM.2023.v08i01.004

VGGNet, and ResNet for analysing the datasets [1-4, 7]. Compared to other strategies, handwriting because it is less time-consuming and not very expensive. Also, the system uses several deep learning algorithms such as CNN along with an SVM classifier to categorize different handwritten activities performed by Parkinson's patients.

Technologies Used

AlexNet: AlexNet is an 8-layer deep convolutional neural network used for any object detection task.
The input is given to AlexNet from the ImageNet Database.
GoogleNet: GoogleNet architecture is different from AlexNet. To create deeper architecture it uses a different kind of method such as 1 x 1 convolution and global average pooling.
VGG Net: VGG stands for visual geometry group [8]. VGG is a standard multi-layer convolutional neural network. It is used for object recognition [8].
ResNet: ResNet is a residual neural network. It is an artificial neural network that constructs from the pyramidal cell. ResNet was used to create a class for the CNN block.
Elham Dehghanpur Deharab et al. in their research paper**,** Parkinson's Disease Detection Using Dynamic Writing Traces Warping, the system uses Dynamic Writing Traces Warping (DWTW) to evaluate the differences between healthy people and PD-affected patients [11]. Dynamic wrapping is the most important condition for classifying differences and similarities between time series. It can apply time series of different lengths. This system comprises two stages:

Feature Extraction

In feature extraction, dynamic warping is used to evaluate the changes in the two non-linear patterns written by healthy people and PD-affected patients. The feature extraction stage focuses on the separation of kinematic features.

Classification

And for the classification phase, this system uses an SVM classifier along with a linear kernel.

Classification results are evaluated using different parameters such as accuracy, sensitivity and specificity.
Sensitivity = TP/(TP+FN) Specificity =TN/(TN+FP)
Accuracy=TP+TN/(TP+FN+TN+FP)
Anita R et.al in their research work early detection of Parkinson's disease using machine learning the detection of Parkinson's disease is done with the help of speech as well as image analysis [12]. But we have considered only the image analysis part. The system takes spiral drawings as input to the system. It includes the following three parts:

Pre-Processing

This part involves image acquisition, pre-processing and segmentation. Pre-processing the image helps in improving the quality of the input image so that the resulting output images are better. Image acquisition aims to improve the quality of images which results in better quality, low noise and distortion. Whereas segmentation helps to improve the image representation so that it becomes easier to analyse it.

Feature Extraction

This section uses a mean and median filter for processing of selecting the images. Mean filtering is mainly used to reduce the changes in the intensity between one pixel and the next. Whereas to reduce the salt-pepper noise from an image median filter is used. The main aim of both the mean and median filter is to reduce noise.

The classification is done using a random forest classifier which constructs decision trees on data samples and also predicts the output with the help of voting.

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

https://doi.org/10.46610/JoDMM.2023.v08i01.004

Data Set

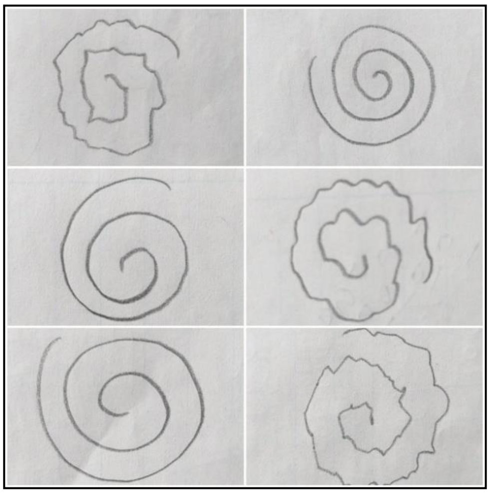

Figure 1: *Spiral drawing.*

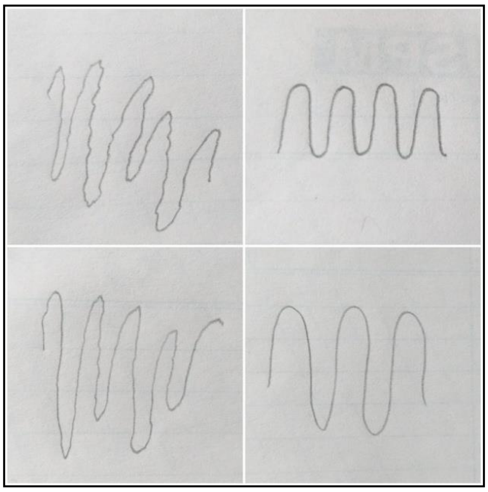

Figure 2: *Wave drawing.*

Fig. 1 and 2 show some of the spiral and wave images that we have taken as input for the proposed system. The system includes a dataset that has a total of 108 images. Out of which 84 are a part of the training dataset and the remaining 24 are a part of the testing dataset. The 84 images for training are further divided into 42 images for each healthy and Parkinson's. And the remaining 24 images for testing are parted as 12 images for each of healthy and Parkinson's.

PROPOSED SYSTEM

The dataset of the proposed system includes drawings of spirals and waves. This dataset is taken as input, where the drawings are drawn by both healthy people as well as Parkinson's patients. For the detection of Parkinson's disorder, the proposed model uses a Histogram of Oriented Gradients (HOG) and a machine learning algorithm such as a Random Forest Classifier as shown in (Fig. 3).

Parkinson's disease affects the ability to control movement which causes a "shake" in the handwriting or drawings of Parkinson's patients. The HOG has mainly been used to quantify the changes i.e. the shake in the drawings of spirals and waves. It helps to determine how the direction of both spirals and waves change. Another important part of the proposed system is the Random Forest Classifier. The Random Forest classifier is mainly formed from the idea of ensemble learning. The random forest consists of numerous decision trees where prediction from each tree is taken and based on the majority of prediction, the final result is calculated.

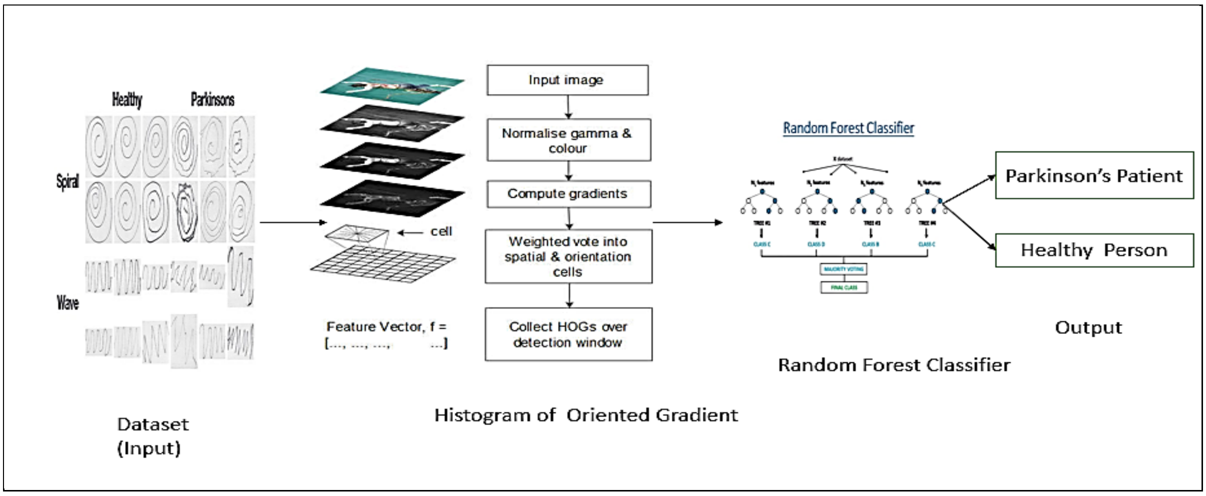

Figure 3: *Proposed system.*

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

www.matjournals.com
https://doi.org/10.46610/JoDMM.2023.v08i01.004

Histogram of Oriented Gradients (HOG)

HOG is a feature descriptor that extracts features from the image. It helps to display a simpler depiction of the image and contains only the prominent information about the image. The features are extracted using the gradient and orientation i.e. the magnitude and directions of the edges.

The orientations in the image are calculated in 'limited' sections which implies that the entire image is divided into small-scale sections. For every section, the magnitude and direction are evaluated. In the end, HOG would create a Histogram for every section individually. For the creation of histograms, the magnitude and direction of the pixel values are used.

Algorithm (Random Forest Classifier)

Random Forest, being one of the supervised machine learning algorithms, uses the concept of ensemble learning. So, the "Forest" resembles an ensemble of decision trees. Ensemble learning helps to improve predictive performance by combining the prediction of multiple models, as shown in Fig. 4.

The working process of the Random Forest is as follows:

Step 1: From the training dataset, select randomly K data points.

Step 2: Related to the data points (Subsets) which are chosen, construct the decision trees.

Step 3: To construct the decision trees, choose the value of N.

Step 4: Perform Step 1 and Step 2 again.

Step 5: Discover the predictions of each decision tree, for the new data points, and assign the new data points to the group which wins the majority votes.

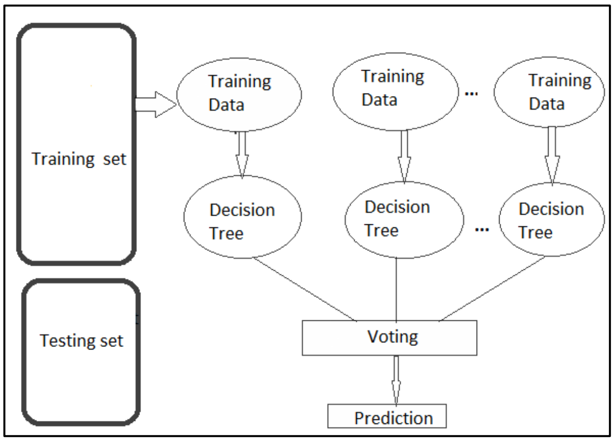

Figure 4: *Random forest classifier.*

Confusion Matrix

In machine learning, a confusion matrix is used to measure the performance of classification problems where output can be divided into two or more two classes. It is a matrix with 2 rows and 2 columns that contain predicted and actual values, as shown in Fig. 4.

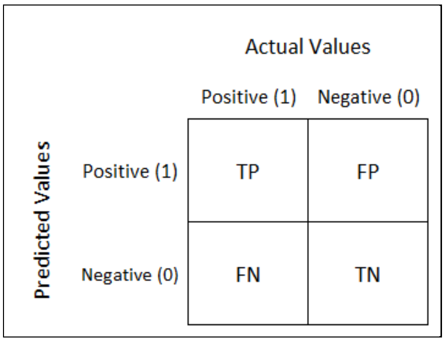

Figure 5: *Confusion matrix.*

Journal of
Data Mining and Management

www.matjournals.com

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

https://doi.org/10.46610/JoDMM.2023.v08i01.004

TP (True Positive)- Interpretation: You forecasted positive and it's true.
FP (False Positive)- Interpretation: You forecasted positive and it's false.
FN (False Negative)- Interpretation: You forecasted negatively and it's false.
TN (True Negative)- Interpretation: You forecasted negatively and it's true.

Performance Parameters

Three performance parameters Accuracy, Precision and Recall is used to find the efficiency and effectiveness of the system

$$\text{Accuracy}=\frac{(\text{True Positive}+\text{True Negative})}{(True\ Positive+True\ Negative+False\ Positive+False\ Negative)}$$

$$\text{Precision}=\frac{\text{True Positive}}{(True\ Positive+False\ Positive)}$$

$$\text{Recall}=\frac{\text{True Positive}}{(True\ Positive+False\ Negative)}$$

RESULT ANALYSIS

The confusion matrix for the spiral drawings is shown in Fig. 6 which gives us the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values of the given input spiral images. So, the TP = 14, FP = 1, FN = 3, TN = 12. Therefore, the proposed system gives an accuracy of 86.67% for the spiral drawings as input to the system. Fig. 7 shows the output for the spiral drawings.

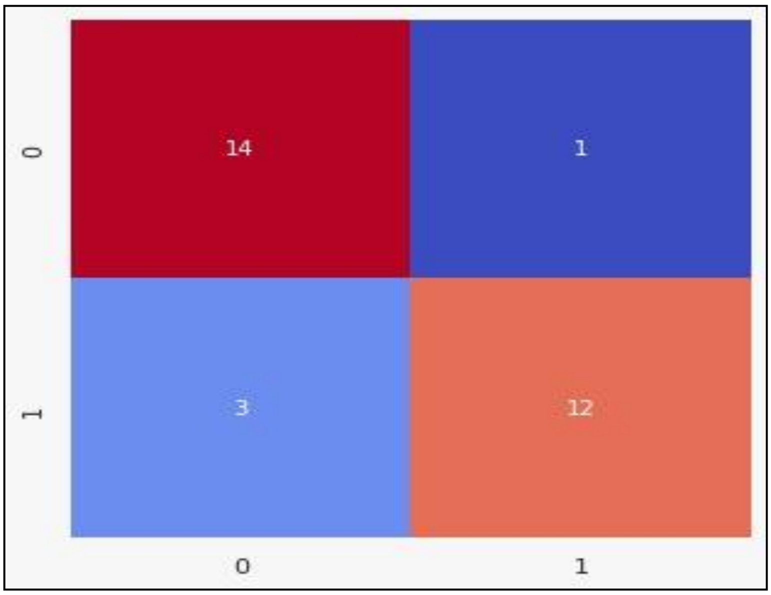

Figure 6: *Confusion matrix of spiral drawing.*

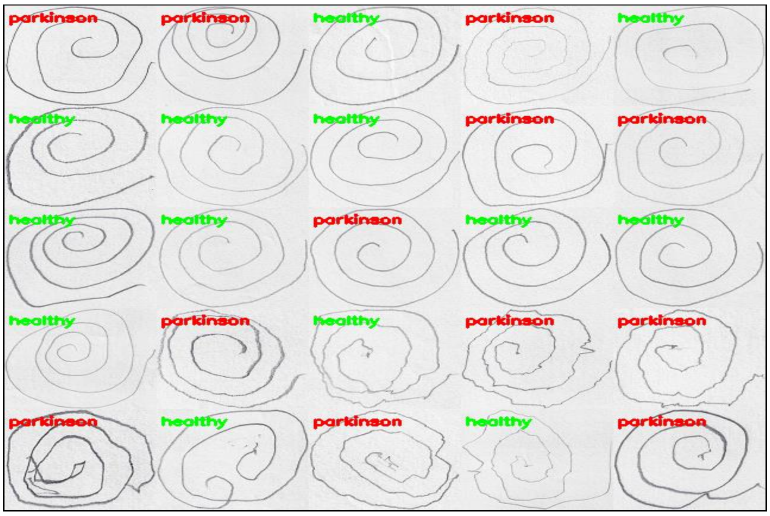

Figure 7: *Output for spiral drawings.*

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)
https://doi.org/10.46610/JoDMM.2023.v08i01.004

The confusion matrix for the wave drawings is shown in Fig. 8 which gives us the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values of the given input wave images. So, the TP = 12, FP = 13, FN = 2, TN = 13. Therefore, the proposed system gives an accuracy of 83.3% for the wave drawings as input to the system. Fig. 9 shows the output for the wave drawings.

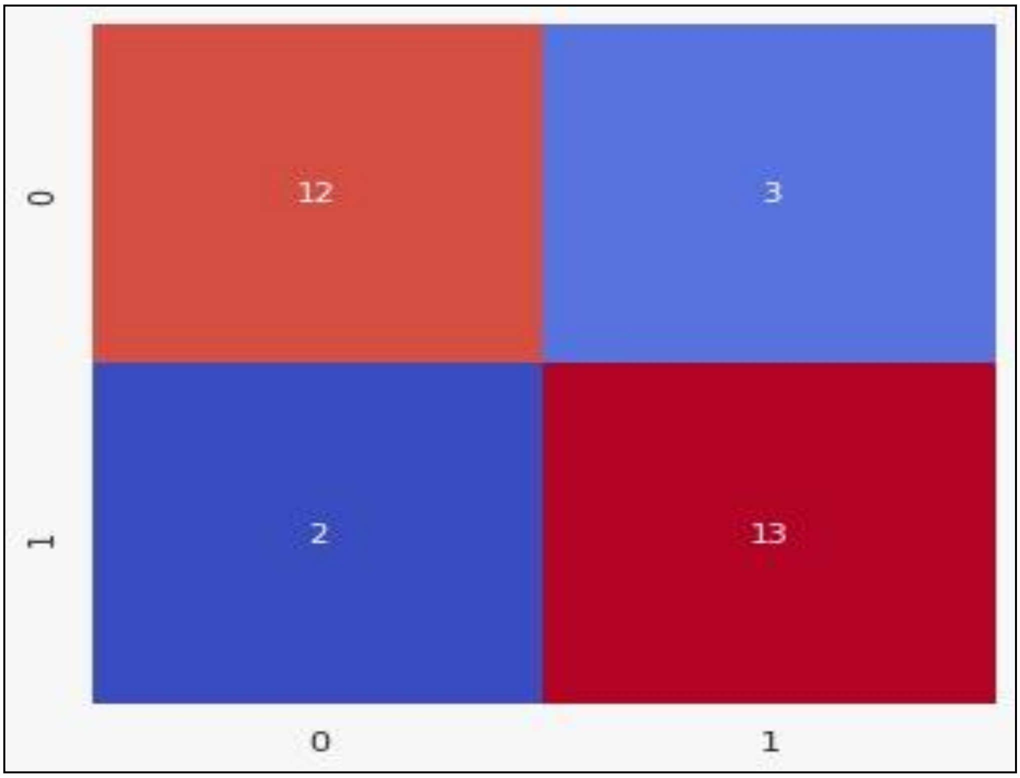

Figure 8: *Confusion matrix for wave drawing.*

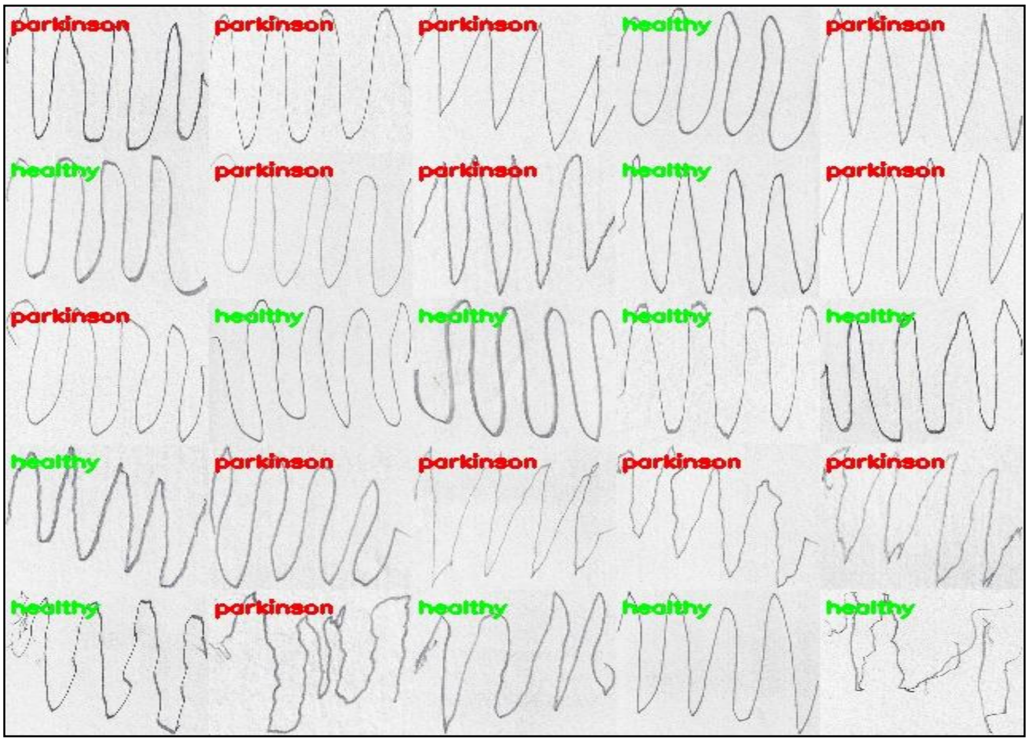

Figure 9: *Output for wave drawings.*

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)
www.matjournals.com
https://doi.org/10.46610/JoDMM.2023.v08i01.004

CONCLUSION

In this paper, we have designed a developed a system that will help to detect Parkinson's patients by taking the drawings of spirals and waves as input to the system. We make use of a Histogram of Oriented Gradient (HOG) and Random Forest Classifier. The Histogram of Oriented Gradient (HOG), which are extremely powerful descriptors help in describing the structure, shape and appearance of the input image. The drawings of spirals and waves are given as input to the HOG which is used to depict the prominent features of these images such as finding the "shake" in these drawings which are expected from a PD patient. So, the output of HOG is a feature vector and this is given as input to the Random Forest Classifier. Based on the input, the Random Forest Classifier performs the classification and predicts the output. However the accuracy of this model can be increased further by training the model by using a larger dataset, identification of some critical and discriminatory features during feature selection can also enhance the efficiency of the model significantly. There must be a system that could detect the disease at the earlier stage, mobile application development can play a crucial role in this regard as it will help every individual to use it anytime everywhere.

REFERENCES

1. N Ranjan, et al. (2022). Diagnosis of Parkinson disease using handwriting analysis, *International Journal of Computer Applications*, 184(1), 13-16, Available at: https://www.researchgate.net/publication/359494208_Diagnosis_of_Parkinson_Disease_using_Handwriting_Analysis.
2. S Singh, et al. (2022). Parkinson’s disease detection using machine learning, *International Journal of Development Research*, 12(4), 55117-55119, Available at: https://www.researchgate.net/publication/360139302_PARKINSON%27S_DISEASE_DETECTION_USING_MACHINE_LEARNING.
3. K Simonyan and A Zisserman (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations*, (pp. 1-14). ICLR, Available at: https://arxiv.org/pdf/1409.1556.pdf.
4. S Ghodake and N Ranjan (2019). AI algorithm for health risk prediction, *Journal of Emerging Technologies and Innovative Research*, 6(6), 600-607, Available at: http://www.jetir.org/papers/JETIR1906D01.pdf.
5. R.V. Darekar, et al. (2023). A hybrid meta-heuristic ensemble based classification technique speech emotion recognition, *Advances in Engineering Software*, 180, Available at: https://doi.org/10.1016/j.advengsoft.2023.103412.
6. N M. Ranjan and M Chakkaravarthy (2021). Evolutionary and incremental text document classifier using deep learning, *International Journal of Grid and Distributed Computing*, 14(1), 587-595, Available at: https://www.lincolnedu.education/pdf/research/Evolutionary%20and%20Incremental%20Tex.pdf.
7. A Krizhevsky, I Sutskever and G E. Hinton (2014). ImageNet classification with deep convolutional neural networks, *Communications of the ACM*, 60(6), 84-90, Available at: https://doi.org/10.1145/3065386.
8. S Singh, et al. (2022). Detection of Parkinson’s disease using machine learning algorithm, *International Journal of Computer Applications*, 184(6), 24-29, Available at: https://www.ijcaonline.org/archives/volume184/number6/singh-2022-ijca-922016.pdf.
9. D Impedovo and G Pirlo (2018). Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective, *IEEE Reviews in Biomedical Engineering*, 12, 209-220, Available at: https://doi.org/10.1109/RBME.2018.2840679.
10. I Razzak, I Kamran and S Naz (2020). Deep analysis of handwritten notes for early diagnosis of neurological disorders. *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Available at: https://doi.org/10.1109/IJCNN48605.2020.9207087.
11. E Dehghanpur Deharab and P Ghaderyan (2020). Parkinson's disease detection using dynamic writing traces warping. *2020 28th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, Available at: https://doi.org/10.1109/ICEE50131.2020.9260903.
12. R Anitha, et al. (2020). Early detection of Parkinson’s disease using machine learning, *International Journal of Advance Research and Innovative Ideas in Education*, 6(2),

e-ISSN: 2456-9437
Volume-8, Issue-1 (January-April, 2023)

https://doi.org/10.46610/JoDMM.2023.v08i01.004

505-511, Available at: https://ijariie.com/AdminUploadPdf/EARLY_DETECTION_OF_PARKINSON_S_DISEASE_USING_MACHINE_LEARNING_ijariie11591.pdf.

CITE THIS ARTICLE

Nihar Ranjan, et al. (2023). Detection of Parkinson's Disease using Machine Learning Algorithms and Handwriting Analysis, *Journal of Data Mining and Management*, 8(1), 21-29.