Detection of Parkinson's disease using Machine learning

Dr.P.Srinivasu[1],N. Tejaswini[2], M.Likhitha[3], P. Chaithanya Lakshmi [4] and P.Anushka[5]
[1] Associate.Prof..,Dept of ECE, Vignan's Institute of Engineering for Women, Duvvada
[2,3,4,5] Student, Dept of ECE, Vignan's Institute of Engineering for Women, Duvvada
[2] nalamtejaswini188@gmail.com

Abstract:

The most prevalent neurodegenerative disease that significantly impairs elderly people's motor capabilities is Parkinson's disease(PD). The motor functions ofBody, which include speech, handwriting, walking, and coordinated movements, are all affectedby Parkinson's disease. All quantitative indices of motor decline and non-motor biomarkers have been proposed to check the degree of severity asPD is treated as motor disorder. The main objective is classifying both healthyand Parkinson's affected persons in the form of matrix by using feature extraction using machine learning techniques. Results demonstrated nearly 90% accuracy after separation in the confusion matrix form.

Keywords:-Parkinson's diseaseclassification,Machine learning algorithms, Confusion matrix

1. Introduction:

In this present era there are many diseases effecting people. Unlike other diseases, Parkinson's disease is one of the recent affecting disease. It is referred to in the ancient Indian medical system of Ayurveda under the name kampavata (where "kampa" means tremor in Sanskrit). It was in the 1960s that the chemical differences in the brains of Parkinson's patients were identified. The substantia nigra, a region of the brain, has low levels of dopamine and nerve cell degeneration, the researchers discovered. As a result, dopamine agonist therapy for Parkinson's disease became an option. Parkinson's disease(PD) is a neurodegenerative disease affecting significantly the functions of elderly persons. The brain region known as the substantia nigra loses nerve cells, which leads to PD illness. A substance called dopamine is produced by nerve cells in this region of the brain. Between the areas of the brain and nervous system that assist in regulating and coordinating body motions, dopamine serves as a messenger. The amount of dopamine in the brain decreases if these nerve cells are destroyed or die. Parkinson's disease symptoms typically don't appear until the substantia nigra has lost about 50% of its nerve cell activity.

PD symptoms is classified into two types-Motor symptoms and Non-motor symptoms. Motor symptoms are based on movement and Non-motor symptoms are related to non-movement. Motor related examples include tremor, slowness of movements, dizziness, rigidity etc.., and Non-motor related examples include pain cramps, constipation, excessive daytime sleepiness, Low Blood pressure, Sweating problems. Many types of different samples like speech, handwriting, collecting sample of your skin including nerves, Imaging tests can also be considered. Handwriting is one of the motor function mostly affected by PD. Spiral drawing is skilled and complex motor activity. Taking the patient to the clinic for a physical examination and diagnosis based on physical observation requires qualified medical professionals is inconvenient for the patient's carers. Some of the handwriting samples other than spiral drawings such as wavelets, letters, etc., can also be considered.

2. Related work

FouziHarrouet al.[1] have presented their work aiming to compare the mean, standard deviation, median, interquartile range by using techniques such as DEEP, classification tree, boosting,RF, LOGIS,KNN, DIS, SVM. The dataset is primarily collected from PPMI (Parkinson's progression markers initiative). Yipeng Liu et.al [2] presented their work aiming to improve PD detection accuracy, proposed a cascaded system that cascades a chi2 model with adaptive boosting (Adaboost) model. Dataset collection process was asked to perform 6 different tasks. After data collection feature extraction was performed. The proposed system improved the performance of a conventional adaboost by 3.3%. Donato Impedovo et.al [3] suggested the clinical assessment of early signs of Parkinson's disease. This paper aim is to extent the dynamic features of the handwriting process. The good specificity performances have been observed. The experimental results are expressed in terms of some classification performances metrices accuracy, area under the ROC curve, sensitivity and specificity.Giuseppe Pirlo et.al [4] described the analysis of Alzeheimer's disease (AD) and Parkinson's disease (PD) by forming the writing tasks as straight lines, spirals &circles trajectory, tremor and dynamic handwriting (pens down, pen up) & dynamic acquisition are pen-up; black dots :on pad samples)red data on air movement samples. Two types of features can be considered such as function features & Parameter features. The experimental results are Accuracy 79.4%.Musheer Ahmed et.al [5] The study you are referring to explores the potential of using deep transfer learning and optimized feature selection for Parkinson's disease detection. The researchers used a dataset of spiral and wave drawings collected from individuals with Parkinson's disease and healthy individuals.They applied transfer learning by fine-tuning a pre-trained deep convolutional neural network (CNN) to extract features from the handwriting data. They then applied an optimized feature selection method to select the most relevant features for classification.

3. Implementation

For Parkinson's disease any type of datasets can be considered like Spiral images, waves, letters, voice samples can be taken. We considered both spiral and wave images for this project.

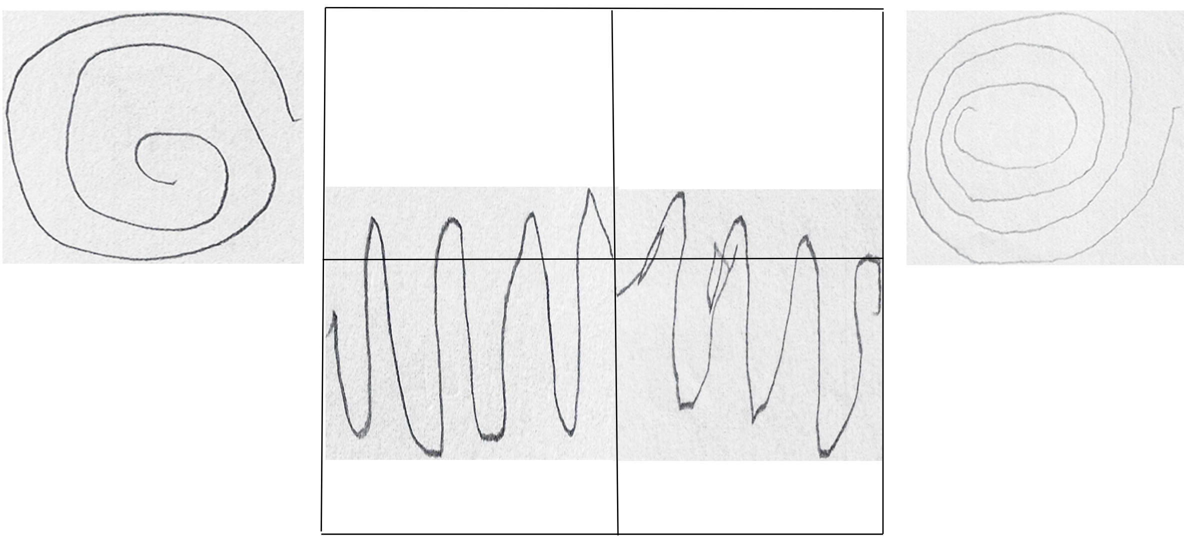


Fig.1(a). Spiral drawing (healthy sample), 1(b). Spiral drawing (patient sample)
Fig.1(c). Wave drawing(healthy sample), 1(d). Wave drawing (patient sample)

3.1 *Block diagram:*

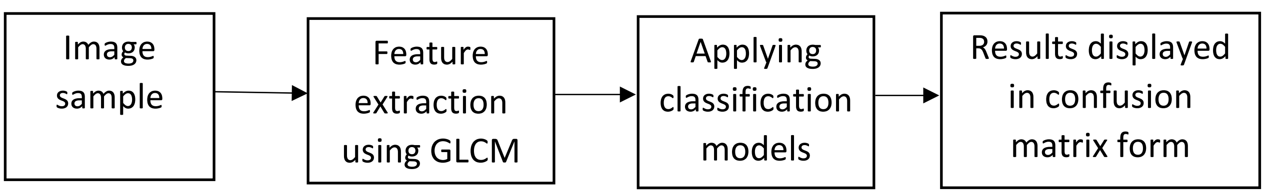
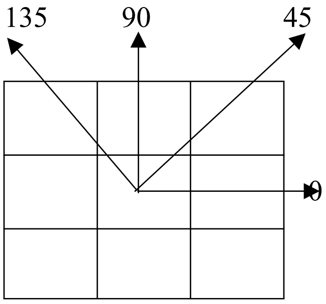


Fig.2: Block diagram of obtaining confusion matrix using GLCM method with machine learning

3.2 Methodology:

Feature extraction may achieve better accuracy when classification is performed. Feature extraction using grey-level co-occurrence matrix(GLCM), in this step the conversion of image pixel matrix is converted into co-occurrence matrix. GLCM characterize the texture of image by specifying the spatial relationship between adjacent pixels.



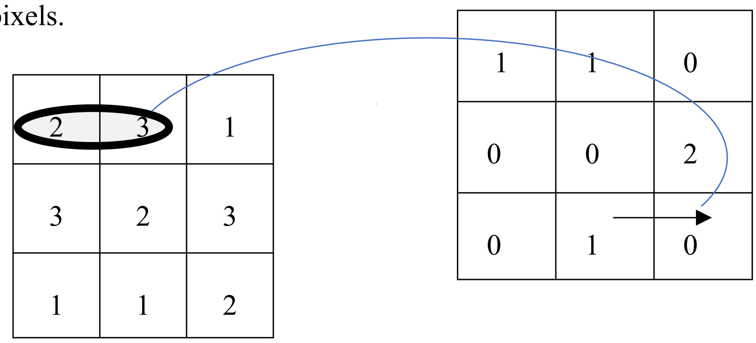


Fig.3:Representation of Θ

Fig.4: Conversion of image pixel matrix to GLCM matrix

When the image pixel matrix is converted into co-occurrence matrix,the matrix form is taken as data to the next step. As in machine learning the dataset is divided into training and testing dataset. To obtain perfect results giving 70-80% for training and remaining 20-30% for testing data set is better way. Now machine learning algorithms are applied on the data in individual format.

3.3 Classifiers & algorithms:

Support vector Machine(SVM):It is one of the most prominent supervised learning techniques, which is used for classification issues. In n-dimensional space, there may be several lines or decision boundaries used to separate the classes, but we must identify the optimum decision boundary that best aids in classifying the data points. The hyperplane of SVM is a name for this optimal boundary. The dataset's features determine the hyperplane's dimensions, therefore if there are only two characteristics, the hyperplane will be a straight line. Additionally, hyperplane will be a two-dimensional plane if there are three features. We always construct a hyperplane with a maximum margin, or maximum.

Random Forest:It can be applied to classification and regression issues in machine learning. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance. A classifier called random forest uses a number of decision trees on different subsets of the input dataset and averages the results to increase the dataset's predicted accuracy. Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions. Higher accuracy and overfitting are prevented by the larger number of trees in the forest.

K-Nearest neighbour: K-Nearest Neighbour assigns the new case to the category that is most similar to the existing categories based on the assumption that the new case/data and the existing cases are comparable. When new data is generated, it can be quickly and readily sorted into a well-suited category by employing the K-NN algorithm because it maintains all of the already accessible data. Although the K-NN approach is most frequently employed for classification problems, it can also be utilised for regression. It is referred to as a lazy learner since it keeps the dataset instead of learning from the training set right away and then uses that dataset when classifying data.

Adaboost algorithm: One of the well-known learning ensemble modelling strategies, boosting is used to create strong classifiers from a variety of weak classifiers. Adaboost (Adaptive boosting) was the first boosting method to merge many weak classifiers into a single strong classifier. It focuses mostly on categorization problems, such binary classification.

Decision tree: While random forest algorithms combine several decisions, the decision tree combine some decisions only. Decision tree algorithm mainly used for operating large datasets which can operate easily and in a fast manner. It is capable of working for both classification and regression techniques. Decision tree thinks like a human and take decisions which is easy.Decision trees work by recursively partitioning the data based on input field values. The data partitions are called branches . The initial branch (sometimes called the root) encompasses all data records. The root is split into subsets, or child branches , based on the value of a particular input field.

CONFUSION MATRIX :

confusion matrix is used to describe how well a classification system performs. A confusion matrix shows a classification algorithm's performance.

		Actual	
		True	False
Predicted	Positive	True positive(tp)	False positive(fp)
	Negative	True Negative(tn)	False Negative(fn)

Fig.5: Representation of confusion matrix

True Positive: The number of times our actual positive values are equal to the predicted positive. You predicted a positive value, and it is correct.

False Positive: The number of times our model wrongly predicts negative values as positives. You predicted a negative value, and it is actually positive.

True Negative: The number of times our actual negative values are equal to predicted negative values. You predicted a negative value, and it is actually negative.

False Negative: The number of times our model wrongly predicts negative values as positives. You predicted a negative value, and it is actually positive

3.4 Evaluation metrics:

Accuracy: The closeness of the measurements in relation to a given value is referred to as accuracy. Accuracy is also a term used to describe systematic error. Moreover, accuracy provides us with a measurement of statistical bias.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

Precision: Also called as positive value is the bit of applicable cases. The perfection is intimately the capability of the classifier not to label a sample as positive if it's negative. High the value, better the classifier.

$$precision = \frac{tp}{tp + tn + fp + fn} \qquad (2)$$

Sensitivity: Represents the percentage of all relevant instances that were actually retrieved. The classifier's capacity to locate all the positive samples is known as recall. The accuracy increases as the value rises.

$$recall = \frac{tp}{tp + fn} \qquad (3)$$

Specificity: Specificity relates to the classifier’s ability to identify negative results.

$$specificity = \frac{tn}{tn + fp} \qquad (4)$$

F1 Score: The F1 score is a machine learning evaluation metric that measures a model’s accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$f1 = \frac{\left(2*\left(\frac{tp}{tp+fp+tn+fn}\right)\right)*\left(\left(\frac{tp}{tp+fn}\right)\right)}{\left(\left(\frac{tp}{tp+fp+tn+fn}\right)\right)+\left(\frac{tp}{tp+fn}\right)} \qquad (5)$$

3.5 Flow diagram:

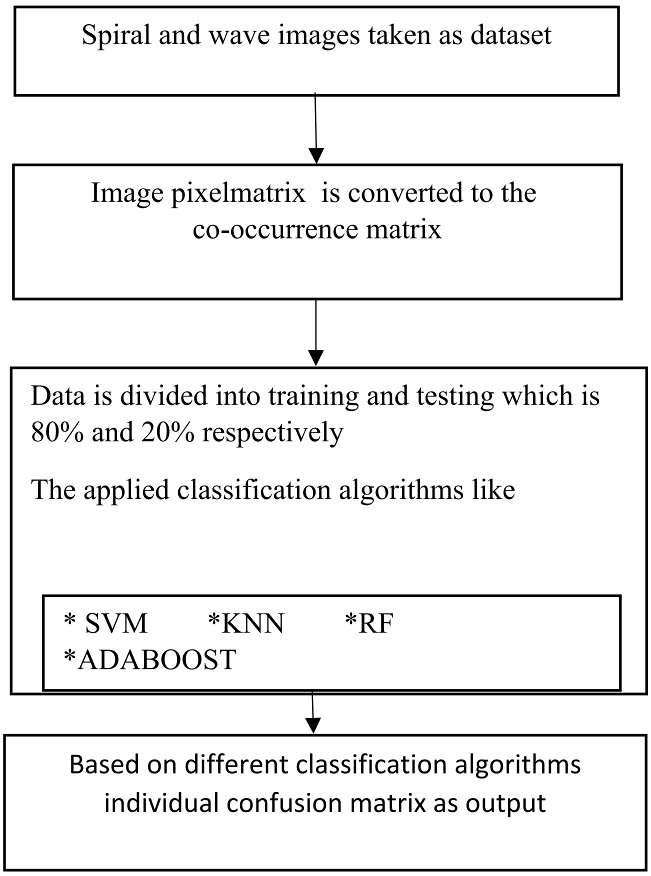


Fig 6: Flow chart of classification in confusion matrix

4. Result and Conclusion:

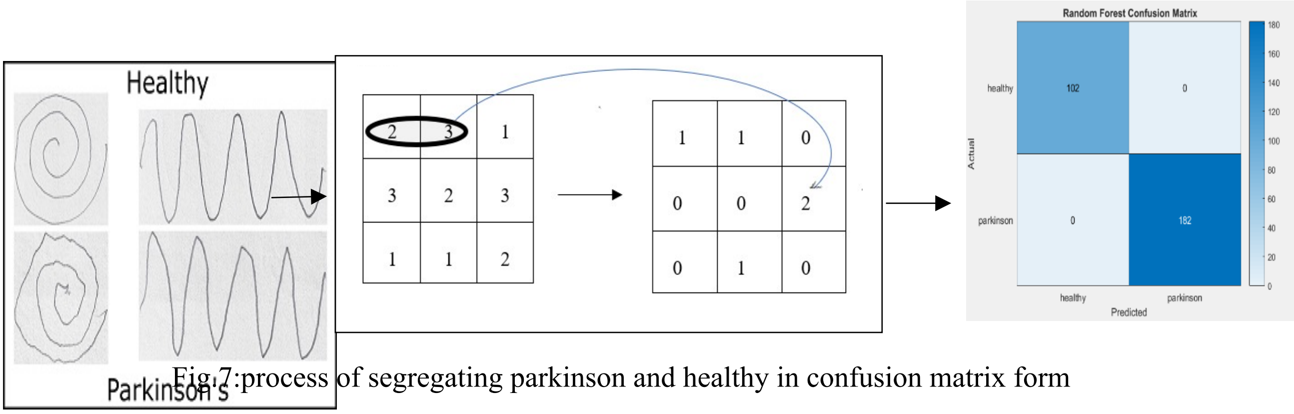


Fig.7:process of segregating parkinson and healthy in confusion matrix form

We applied feature extraction technique for the samples apart using machine learning. By applying the feature extraction using GLCM form makes easy which is matrix form to segregate healthy and parkinson's individual. In the confusion matrix the classification is observed with better acccuracy. We can observe the variations from individual confusion matrix of applied machine learning algorithm. When the two cases like actual value is parkinson and predicted one is false , actual is false and predicted is true are "0" then that individual algorithm is having higher accuracy than other algorithm.Here in our project we observed Random forest (RF),K-Nearest heighbour (KNN), ,adaboost shows better accuracy as those shows values "0" in mentioned above cases than Support vector classification (SVM), decision tree algorithm. For decision algorithm actual is parkinson and predicted is healthy for 1 sample other than that the other case is "0". Overall, the GLCM conversion makes easier for applying algorithms. This type of technique with applying machine learning algorithm is more useful for classification or segregation type tasks.

References

[1] FouziHarrou,et al., "*Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning* " ,2020.

[2] Yipeng Liu, et al., "*Reliable Parkinson's Disease Detection by Analyzing Handwritten Drawings: Construction of an Unbiased Cascaded Learning System Based on Feature Selection and Adaptive Boosting Model*".

[3] D. Impedovo, G. Pirlo, "*Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective"*, in: IEEE Reviews in Biomedical Engineering, vol. 12, 2019, pp. 209–220, https://doi.org/10.1109/ RBME.2018.2840679.

[4] Donato Impedovo, et al., "*Dynamic Handwriting Analysis for Supporting Earlier Parkinson's Disease Information*" 9, 2018, p. 247, https://doi.org/10.3390/ info9100247, 10.

[5] Musheer Ahmed, et.al.," *Deep Transfer Learning Based Parkinson's Disease Detection Using Optimized Feature Selection* ".