



Prediction of Parkinson's Diseases using SVM and Logistic Regression Algorithm

Gokila S¹ | Joy Princy J¹ | Monicka G¹ | Durkka Devi S¹ | Pavithra M²

¹UG Students, Department of C.S.E, Jansons Institute of Technology, Coimbatore, India

²Assistant Professor, Department of C.S.E, Jansons Institute of Technology, Coimbatore, India

To Cite this Article

Gokila S, Joy Princy J, Monicka G, Durkka Devi S and Pavithra M, "Prediction of Parkinson's Diseases using SVM and Logistic Regression Algorithm", *International Journal for Modern Trends in Science and Technology*, Vol. 07, Issue 04, April 2021, pp.:100-106.

Article Info

Received on 15-March-2021, Revised on 29-March-2021, Accepted on 06-April-2021, Published on 13-April-2021.

ABSTRACT

Predicting Parkinson's Disease by using Data mining in systematic way. This project aims at detecting Parkinson's disease through data mining. Since there is no standard test to detect parkinsonism, we propose a statistical approach using the most common symptoms of PD which are gait, tremors and micro-graphia. This includes analyzing the co-relation between the symptoms and classifying the achieved data using different classification algorithms in order to find the algorithm which gives the highest accuracy in diagnosing PD patients.

INTRODUCTION

Machine learning is the computational learning using algorithm to learn from and make predictions on data. Machine Learning is the powerful new technology for analyst to focus on the most important information in their data warehouse. Support vector machine is an open source it helps to extract the featured attribute from the dataset. Sequential Minimal Optimization use Poly Kernels to predict the rate of disease in graphical manner. A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data.

Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological

and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector- Machine (SVM). This idea leads to improved quality of patient service and good patient retention and satisfaction. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM).

A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition

prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data.

Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data. Standard univariate analysis of neuroimaging data has revealed a host of neuroanatomical and functional differences between healthy individuals and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data.

and patients suffering a wide range of neurological and psychiatric disorders. Significant only at group level however these findings have had limited clinical translation, and recent attention has turned toward alternative forms of analysis, including Support-Vector-Machine (SVM). A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data.

Parkinson's disease (PD) is a neurodegenerative disease which often affects patients' movement. Currently, PD is diagnosed via various neurological examinations by specialists. The most common symptoms of PD are tremor, gait disturbance, stiffness, and slowness. Through this project we are trying to co-relate different symptoms in order to increase the accuracy in diagnosing Parkinson's. The dataset will include features such as jitters and stride. This data will be analyzed using different classification techniques thus providing a reliable and accurate approach to diagnose Parkinson's at an early stage.

LITERATURE SURVEY

Rahul R Zaveri, Prof. Pramila M. Chawan "The International Research Journal of Engineering and Technology" vol. 07, issue: 10, 2020.

Dr. Anupam Bhatia and RaunakSulekh [1]
"Predictive Model for Parkinson's Disease through Naive Bayes Classification" In this study, Naive Bayes was applied to predict the performance of the dataset. Rapid miner 7.6.001 is a tool, which was used to explore, statistically analyze, and mine the data. The Naive Bayes model performs with 98.5 % accuracy, and 99.75% of precision.

Carlo Ricciardi, et al [2]
"Using gait analysis' parameters to classify Parkinsonism: A data mining approach" In this system, Random Forest is used for classification along with comparing it with Gradient Boosted Trees. These results are being categorized into 3 different categories namely PSP, De Novo Parkinson's Disease and Stable Parkinson's Disease with their accuracy being as high as 86.4% as compared to Gradient Boosted Trees which were accurate to a meagre 70%. Also the precision rate of Random Forest was maximum of 90 % against

Gradient Boosted Trees which were around maximum of 85%.

MehrbakhshNilashi et al [3]

"A hybrid intelligent system for the prediction of Parkinson's Disease progression using Machine Learning techniques" In this system a method was proposed for the UPDRS (Total-UPDRS and Motor-UPDRS) prediction using machine learning. ISVR was used to predict the Total-UPDRS and Motor-UPDRS. SOM and NIPALS were used for clustering and data dimensionality reduction. The results show that the method combining SOM, NIPALS, and ISVR techniques was effective in predicting the Total-UPDRS and Motor-UPDRS.

Arvind Kumar Tiwari [4]

"Machine Learning based Approaches for Prediction of Parkinson's Disease," In this system, minimum redundancy maximum relevance feature selection algorithms were used to select the most important feature among all the features to predict Parkinson's disease. This system of feature selection along with Random Forests provided an accuracy of 90.3% and precision of 90.2%.

M. Abdar and M. Zomorodi-Moghadam [5]

"Impact of Patients' Gender on Parkinson's disease is using Classification Algorithms" In this system, SVM and Bayesian Networks were used for classification of data based on the gender of the patient. The accuracy for SVM was 90.98% and Bayesian network was 88.62%. This test proved that the SVM algorithm had a great ability to identify a patient's gender suffering from PD.

DraganaMiljkovic et al [6]

"Machine Learning and Data Mining Methods for Managing Parkinson's Disease" In this system, based on the initial patients examination and medications taken, the Predictor part was able to predict each Parkinson's Disease symptom separately covering 15 different Parkinson's Disease symptoms in total. The accuracy of prediction ranges from 57.1% to 77.4% depending on the symptom where the highest accuracy is achieved from tremor detection. 7) Md. Redone Hassan et al [7], "A Knowledge Base Data Mining based on Parkinson's Disease" In this system, the results and output of the vector support machine (SVM), K nearest neighbor and the output figures for the decision tree algorithms were shown in the output section of the train data. The decision tree offered the highest precision of 78.2%.

SatishSrinivasan, Michael Martin &AbhishekTripathi [8]

“ANN based Data Mining Analysis of Parkinson’s Disease” In this study, it was intended to understand how the different types of preprocessing steps could affect the prediction accuracy of the classifier. In the process of classifying the Parkinson’s Disease dataset using the ANN based MLP classifier a significantly high prediction accuracy was observed when the dataset was pre-processed using both the Discretization and Resample technique, both in the case of 10-fold cross validation and 80:20 split. Whereas in the 70:30 split it was found that the combination of the preprocessing steps namely resampling and SMOTE on the dataset resulted towards the higher prediction accuracy using the MLP classifier. On an 80:20 split of the pre-processed (Discretized and Resampled) dataset the ANN based MLP classifier achieved a 100% classification accuracy with F1-score and MCC being 100%.

Ramzi M. Sadek et al [9]

“Parkinson’s Disease Prediction is using Artificial Neural Network” In this system, 195 samples in the dataset were divided into 170 training samples and 25 validating samples. Then importing the dataset in the Just Neural Network (JNN) environment, we trained, validated the Artificial Neural Network model. The most important attributes contributing to the ANN model were made known of. The ANN model was 100% accurate. Table -1: Summary of Literat

2.2 Saykin, A. 1., Shen, L., Foroud, T. M., Potkin, S. G., Swaminathan, S., Kim, S., et al. "Alzheimer's disease Neuroimaging Initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans". *Alzheimers Dement.* Vol. 6, No.3, pp. 265- 273, 2010.

The role of the Alzheimer’s Disease Neuroimaging Initiative Genetics Core is to facilitate the investigation of genetic influences on disease onset and trajectory as reflected in structural, functional, and molecular imaging changes; fluid biomarkers; and cognitive status. Major goals include (1) blood sample processing, genotyping, and dissemination, (2) genome-wide association studies (GWAS) of longitudinal phenotypic data, and (3) providing a central resource, point of contact and planning group for genetics within the Alzheimer’s Disease Neuroimaging Initiative.

Genome-wide array data have been publicly released and updated, and several neuroimaging

GWAS have recently been reported examining baseline magnetic resonance imaging measures as quantitative phenotypes. Other preliminary investigations include copy number variation in mild cognitive impairment and Alzheimer’s disease and GWAS of baseline cerebrospinal fluid biomarkers and longitudinal changes on magnetic resonance imaging.

Blood collection for RNA studies is a new direction. Genetic studies of longitudinal phenotypes hold promise for elucidating disease mechanisms and risk, development of therapeutic strategies, and refining selection criteria for clinical trials. A type of machine learning, SVM allows categorization of an individual's previously unseen data into a predefined group using a classification algorithm, developed on a training data set. In recent years, SVM has been successfully applied in the context of disease diagnosis, transition prediction and treatment prognosis, using both structural and functional neuroimaging data

EXISTING SYSTEM:

Clinical decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patient.

- ✓ Use of wearable technologies through the implementation of Internet of things
- ✓ Handwriting as a marker for the diagnosis of PD using support vector machine achieving the accuracy of 88.13
- ✓ Using 3D visualization techniques to provide an intuitive tool for assessment of Parkinson’s
- ✓ Visually guided tracking performance of PD patients using data mining technique.
- ✓ Using Voice and speech data to detect Parkinson’s.

DISADVANTAGES

- ✓ Speech samples require speech segmentation and noise removal.
- ✓ Breath samples require dedicated sensors.
- ✓ Handwriting samples can be influenced by other factors
- ✓ Considering single symptom requires less calculation.
- ✓ Results and accuracy are based on a single symptom.

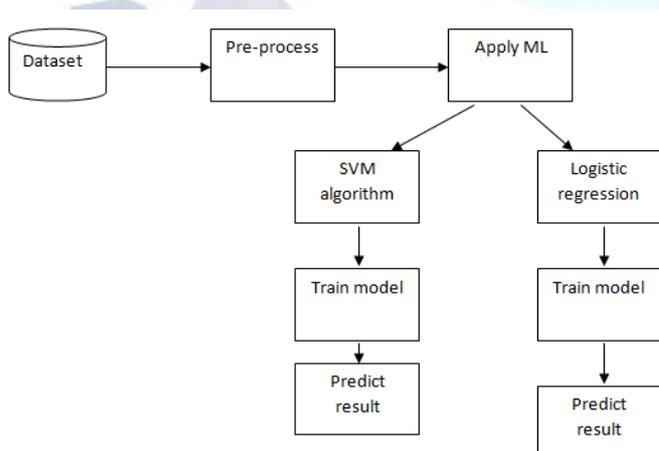
PROPOSED SYSTEM

- ✓ Parkinson's disease detection using gait, tremors and handwriting samples as the dataset, in order to increase the accuracy by finding the co-relation between these symptoms.
- ✓ Since individual analysis of every symptom has some drawback attached to it, for example handwriting is a complex activity where other factors can influence motor movement, in speech recognition additional steps such as noise removal and speech segmentation are required, using breath samples has been proved to fail to meet clinically relevant results.
- ✓ Thus in order to avoid the above problems, we have included multiple symptoms rather than relying on one of them.

ADVANTAGES

- ✓ No such additional steps needed
- ✓ No requirement for any special sensors and no need to solve the typical problems of acoustic signal acquisition and processing.
- ✓ Including additional symptoms into account.
- ✓ Analysis of multiple symptoms may require additional calculations.
- ✓ Results and accuracy are based on multiple co-related symptoms hence making it more reliable.

ARCHITECTURAL DESIGN



MODULES

DATA PRE-PROCESSING

We have taken multiple symptoms in our case study, in which we combined the patient's dataset with speech and keystroke dataset. Pre-processing of dataset is done for converting the string attributes to numerals and missing data records

are dropped. The pre-processed data is stored in "newdata.csv" file, which is given as input for machine learning models.

SUPPORT VECTOR MACHINES ALGORITHM

A Support Vector Machine is a supervised learning algorithm. An SVM models the data into k categories, performing classification and forming an N-dimensional hyper plane. These models are very similar to neural networks. Consider a dataset of N dimensions. The SVM plots the training data into an N dimensioned space. The training data points are then divided into k different regions depending on their labels by hyper-planes of n different dimensions. After the testing phase is complete, the test points are plotted in the same N dimensioned plane. Depending on which region the points are located in, they are appropriately classified in that region.

Split our dataset to train and test set and fit the dataset to SVM model, as given below.

```

X_train, X_test, Y_train, Y_test =
train_test_split(X_all, y_all)
clf = svm.LinearSVC()
clf.fit(X_train, Y_train)
pred = clf.predict(X_test)
Result is stored in csv file by using below code
result2=open("Output/resultSVM.csv","w")
result2.write("ID,Predicted Value" + "\n")
for j in range(len(pred)):
result2.write(str(j+1) + "," + str(pred[j]) + "\n")
result2.close()
  
```

Output is stored in csv as below

	A	B	C
ID	Predicted Value		
1		0	
2		0	
3		0	
4		0	
5		0	
6		0	
7		0	
8		0	
9		0	
10		1	
11		0	
12		0	
13		0	
14		0	
15		0	
16		0	

Logistic Regression

Logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. When selecting the model for the logistic regression analysis, another important consideration is the model fit. Adding independent variables to a logistic regression model will always increase the amount of variance. A pseudo R2 value is also available to indicate the adequacy of the regression model.

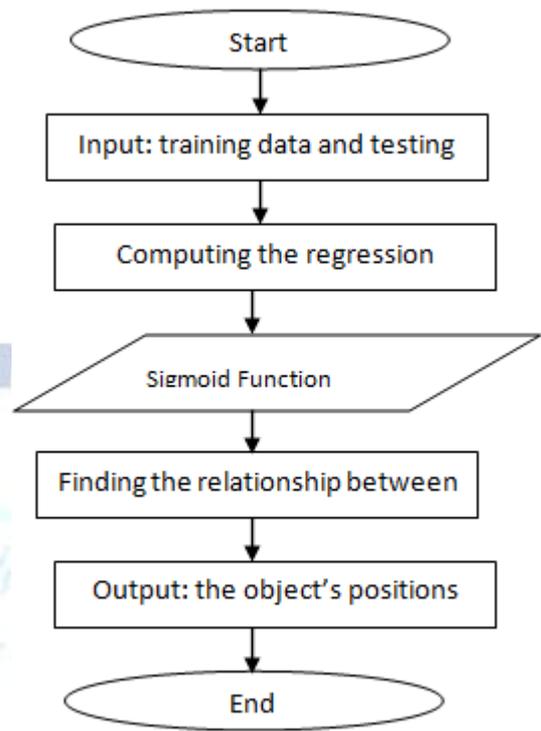
Split our dataset to train and test set and fit the dataset to Logistic regression model, as given below.

```
X_train, X_test, Y_train, Y_test =
train_test_split(X_all, y_all)
scf = linear_model.LogisticRegression(C=1e5)
scf.fit(X_train,Y_train)
pred = scf.predict(X_test)
```

```
result2=open("Output/resultLogisticRegression.csv", "w")
result2.write("ID,Predicted Value" + "\n")
for j in range(len(pred)):
result2.write(str(j+1) + "," + str(pred[j]) + "\n")
result2.close()
```

Output is stored in csv as below

A	B	C
ID	Predicted Value	
1	0	
2	0	
3	0	
4	0	
5	0	
6	0	
7	0	
8	0	
9	0	
10	0	
11	0	
12	0	



**CONCLUSION
RESULT AND DISCUSSION**

In our implementation, Parkinson disease prediction is done using two machine learning algorithms such as Logistics regression and Support vector machine. The result shows that Regression model achieves high accuracy than support vector machine. The following table shows the implemented algorithm and its accuracy arrived.

Algorithm	Accuracy
SVM	81.81
Logistic Regression	88.63

The following chart represents the performance of two algorithm over considered dataset.

The following charts shows the error values such as mean squared error (MSE), mean absolute error (MAE), Root Mean Square Error (RMSE) and R-squared value.

CONCLUSION

The use of multiple instance learning for detecting Parkinson disease symptoms is studied. Proposed work addressed the formulation of PD symptom detection from weakly labeled data as a semi-supervised multiple instance learning problem. The features were carefully chosen to address the subject and symptom specific nature of the problem. We show promising preliminary

results on four days of monitoring performed with two PD subjects.

FUTURE ENHANCEMENTS

In future work, we plan to increase our subject pool and utilize optimal feature selection strategies under MIL frameworks for developing robust person-specific models. These techniques can potentially be adapted to various other physiological sensing and monitoring applications as well.

REFERENCES

- [1] DraganaMiljkovic et al, "Machine Learning and Data Mining Methods for Managing Parkinson's Disease" LNAI 9605, pp 209-220, 2016.
- [2] Arvind Kumar Tiwari, "Machine Learning based Approaches for Prediction of Parkinson's Disease," Machine Learning and Applications- An International Journal (MLAU) vol. 3, June 2016.
- [3] Dr. Anupam Bhatia and RaunakSulekh, "Predictive Model for Parkinson's Disease through Naive Bayes Classification" International Journal of Computer Science & Communication vol. 9, March 2018..
- [4] M. Abdar and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" Journal of AI and Data Mining, vol. 6, 2018.
- [5] Md. Redone Hassan et al, "A Knowledge Base Data Mining based on Parkinson's Disease" International Conference on System Modelling& Advancement in Research Trends, 2019.