Feature Selection Based on L1-Norm Support

Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings

**ABSTRACT**

The patient of Parkinson's disease (PD) is facing a critical neurological disorder issue. Efficient and early prediction of people having PD is a key issue to improve patient's quality of life. The diagnosis of PD specifically in its initial stages is extremely complex and time-consuming. Thus, the accurate and efficient diagnosis of PD has been a significant challenge for medical experts and practitioners. In order to tackle this issue and to accurately diagnosis the patient of PD, we proposed a machine-learning-based prediction system. In the development of the proposed system, the support vector machine (SVM) was used as a predictive model for the prediction of PD. The L1-norm SVM of features selection was used for appropriate and highly related features selection for accurate target classification of PD and healthy people. The L1-norm SVM produced a new subset of features from the PD dataset based on a feature weight value. For the validation of the proposed system, the K-fold cross-validation method was used. In addition, the metrics of performance measures, such as accuracy, sensitivity, specificity, precision, F1 score, and execution time, were computed for model performance evaluation. The PD dataset was in this paper. The optimal accuracy achieved the best subset of the selected features that might be due to various contributions of the PD features. The experimental findings of this paper suggest that the proposed method can be used to accurately predict the PD and can be easily incorporated in healthcare for diagnosis purpose. Currently, the computer-based assisted predictive system is playing an important role to assist in PD recognition. In addition, the proposed approach fills in a gap on feature selection and classification using voice recordings data by properly matching the experimental design.

**Existed System:**

Parkinsonism has vocal disorders problems that affect their speech volume level and face complexity in the pronunciation of syllables and so forth. Thus to use vocal measurements as an effective diagnostic tool for PD recognition Parkinson disease is the critical disorder sickness second to Alzheimer's disease and the complete PD treatment has not discovered till now. The existing technique of therapies is good for tackle PD symptoms. However, researchers have made attempts to find out the effective treatment strategy that ensures recovery and treatment. In

the PD diagnosis is being typically based on conducted few invasive techniques and empirical tests and examinations.

**Disadvantages:**

* The invasive based techniques in order to diagnose the PD are very expensive
* less efficient, as well as very complex
* equipment's needed to conducts and the accuracy is also not satisfactory.

**Proposed System:**

The proposed system designed to classify PD and healthy people. In the development of the proposed system, the machine learning predictive model SVM was used. The L1-Norm SVM algorithm was used for appropriate features selection that classifier effectively classifies PD and healthy subjects. Furthermore, the k-fold cross-validation technique was applied for best hyper-parameters and for predictive model selection. Four performance evaluation metrics were used for predictive model evaluation. The PD dataset which online available at UC Irvine data mining repository was used for testing of the proposed system. The methodology of the proposed system is structured into five steps, preprocessing of the dataset, features selection, cross-validation, and machine learning classifier performance evaluation.

**Advantages:**

* less cost effective
* by using this methodology we can get high efficient result

**Modules:**

1) PREPROCESSING OF DATA

For a good representation of data preprocessing is a very important step and machine-learning classifier should be trained and tested effectively. Techniques of preprocessing include removing of missing values, standard scalar, Min- Max Scalar have been applied to the dataset. In standard

Scalar ensures that every feature has the mean 0 and variance 1. Similarly, in Min-Max Scalar arrange the data such that all features are between 0 and 1, [26]. The features having missing values that feature row are deleted from the dataset.

2) FEATURES SELECTION (FS) ALGORITHM

Features selection algorithms are necessary to remove irrelevant features from feature space. The reduced features will improve the accuracy of classification and deduced execution time of classifier. In this study, we use L1 Norm SVM algorithm for features selection.

3) MACHINE LEARNING CLASSIFIER

In this study, the following classifier was used for PD and healthy people classification. Here is the brief theoretical and mathematical background of the classifier is presented. The support vector machine is classifier and for classification problem used mostly. Due to the good performance of classification SVM are used in various applications widely

4) VALIDATION METHOD

To check the proposed system performance K-folds Crossvalidation (CV) [45], method and three evaluation metrics were used. In this study we used K-fold cross-validation and according to k-fold the data set was split into k identical components. The k-1 groups were applied for training and leftover was used for testing purposes in each step. The k times the process is iterated. Then the average of k results is computed to get the performance of the classifier. The crossvalidation different Value of k was selected and we used the value of k D 10 in our experiments. In 10 fold CV process, 90% of the data used for training and 10% data were used for testing. The 10-time repeated the validation process. In the process of each fold, all samples are randomly distributed

in the training and test groups over the entire dataset prior to selection of new training and test sets for the new cycle. Finally, at the end of 10 folds Processes, an average of all performance metrics are computed. estimated performances Ei for each fold were computed and then used to calculate the estimated average performance E of the model.

**System specifications:**

**Software Requirements:**

* OS : Windows
* Python IDE : python 2.7.x and above

: Jupyter notebook

* setup tools and pip to be installed for 3.6.x and above

**Hardware Requirements:**

* RAM : 4GB and Higher
* Processor : Intel i3 and above
* Hard Disk : 500GB: Minimum