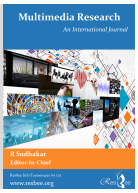
Vol.5 No.2 April 2024
Multimedia Research
Received 15 December, Revised 24 February, Accepted 30 March



An Application of the Filter Feature Selection Method in a Machine Learning Model for the Prognostic of Parkinson's Disease

D. Stalin David
Department of Information Technology
Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Chennai-600062, Tamilnadu, India

Abstract: Parkinson's disease is the most prevalent kind of neurodegenerative sickness that cannot be cured. Neurodegeneration is a word that includes memory loss and other cognitive functions. The conventional methods of medical testing take a lot of time and are not very good at spotting early warning signs. As Parkinson's disease progresses, different treatment approaches are required for patients at different stages of the illness. In this regard, the early identification of Parkinson's disease and the subsequent categorization of its stages may be of great assistance in the process of treating the symptoms of the illness. The objective of the study, on the other hand, is to model a classification approach that could possibly predict the untimely phases of Parkinson's disease by utilizing accurate early-stage gene expression data from the blood that was generated from a clinical Parkinson's dataset. This data is obtained from participants who had the disease. A set of criteria is selected with the use of Information Gain (IG) in order to give sufficient information for differentiating among Normal Control (NC) participants and untimely phases Parkinson's disease (AD) participants. The data is segmented into different sizes, and then 3 unique Machine Learning (ML) methods are used in order to construct the classification approaches: Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (K-NN). The capability of the algorithms to accurately forecast the condition of cognitive impairment is analyzed, compared, and evaluated by utilizing the Weka software tool, as well as a range of metrics to assess the approaches' performance. According to the most recent data, a classification model based on SVM can properly discriminate cognitively impaired Parkinson's patients from normal healthy persons with a success rate of 96.6 percent. As revealed and verified, a gene expression pattern in the blood correctly separates Parkinson's patients from cognitively healthy controls, suggesting that alterations unique to AD may be identified distant from the disease's core location.

Keywords: Gene Expression, Gene Selection, Microarray Technology, Nomadic People Optimizer, Parkinson's disease.

Nomenclature

Abbreviations	Descriptions
IG	Information Gain
NB	Naive Bayes
SVM	Support Vector Machine
HCs	Healthy Controls
RF	Random Forest
NCBI	National Center for Biotechnology Information
MLP-NN	Multi-layer Perceptron Neural Network
AI	Artificial Intelligence
CANFIS	Coac-tive Neuro-Fuzzy Inference System
GA	Genetic Algorithm
LASSO	Least Absolute Shrinkage and Selection Operator
KNN	K-Nearest neighbour
LRR	Logistic Ridge Regression
ML	Machine Learning
MI	Mutual Information
PCC	Pearson correlation coefficient
ANOVA	Analysis of Variance
CNN	Convolutional Neural Network
CN	Cognitively Normal
RBM	Restricted Boltzmann Machine
PCA	Principal Component Analysis
MCI	Mild Cognitive Impairment

© Resbee Info Technologies Pvt Ltd
https://doi.org/10.46253/j.mr.v5i2.a4

1. Introduction

Parkinson's disease Around the globe, 40–50 million people are living with dementia, a figure that has increased by more than 100 percent between the years 1990 and 2016[1]. Parkinson's disease (AD) is now by distant the most prevalent form of dementia, and it is projected that its prevalence will increase as the population as a whole continues to age. The expenditures are going up in line with the increased frequency of the occurrence of it. Parkinson's disease is expected to cost the global economy a total of 604 billion dollars in 2010 [2]. It is estimated that more than 131 million people will be living with Parkinson's disease by the year 2030, which would result in global healthcare costs of $2 trillion. As a consequence of this, Parkinson's disease is rapidly turning out to be a most important economic and global health problem, which has prompted severe scientific research to recognize the fundamental genetic risk factors and regulatory markers and to minimize the evaluated healthcare burden via early detection, mainly at presymptomatic phases, in order to reduce the healthcare expected cost [3]. [4]. Numerous studies have been conducted on the late-onset symptoms of Parkinson's disease. neurofibrillary tangles, neuronal tangles, amyloid plaques, and other tangles are instances[5][6]. Even though these findings are important for diagnostic purposes, it is not yet known how these late-onset Parkinson's disease traits contribute to overall therapy efforts [7]. Furthermore, clinical studies reveal that people with Parkinson's disease have a broad variety of symptoms and react differently to various therapies. This suggests that the illness might have several biological causes. This adds a layer of difficulty to the examination of AD [8][9].

Recently, the data produced by high-throughput gene expression describing has opened new routes for improved consideration of the processes and pathways involved in complex illnesses at the molecular level [10]. However, because of the enormous dimension, the limited sample size, and the noise in the data, detecting embedded patterns in maximum throughput gene expression data may be a challenging task. In the concept of gene expression summary dataset analysis, the strategies to find the majority of explanatory gene subsets via feature selection and data reduction have been split up into two distinct categories[11]: (i) A technique for marginal filtering, and (ii) a technique wrapper that is embedded [12], [13]. There are two types of marginal filtering: univariate marginal filtering and multivariate marginal filtering. For instances univariate filtering processes are the paired t-test (TS), the IG, and the PCC [14], [15].

If there are too many features, there will be problems with overfitting, and if there are not enough features, important features will be left out [16]. As a consequence of this, the selection of features is an essential part of the modeling process [17], [18]. Although it may seem like an excellent situation for constructing a predictive and robust technique, the challenge of picking highly important features included in the gene expression dataset that comprises around 16382 characteristics, makes it difficult to do so. In situations like this, algorithms for feature selection and dimensionality reduction may be of assistance in locating the core feature (or features) that have a major influence on the prediction of the results. This research applied a number of different statistical tests, including the Chi-Squared Test, the IG Test, and the Mean Decrease Gini Test, among others, to locate these gene expressions [19].

As a result, the significance of feature selection in the diagnosis of Parkinson's disease was investigated, as well as an appropriate selection method that is capable of producing more accurate illness forecasts was uncovered as a result of this research. The IG is calculated using three different biomarkers: MRI, PET, and CSF. All three of these biomarkers are approved by the "National Institute of Neurological and Communicative Disorders and Stroke as well as the Parkinson's Disease and Related Disorders Association" [20][21]. In order to make the classifications more accurate, a technique was used to choose the features, and then the feature subsets that rated highest were built out of those characteristics. After that, an SVM algorithm-based classifier will be constructed by making use of the specified characteristics in order to differentiate AD from HCs. A procedure known as 5-cross validation is used in order to assess the efficacy of the predictor. The effectiveness of the suggested technique for the selection of features in the diagnosis of Parkinson's disease was shown by the outcomes of the trials conducted for this research.

2. Related Works

Researchers' Analysis of data from microarrays is attracting the attention of researchers from a diverse variety of disciplines. The subsequent point is a few of the most current suggestions for analyzing microarray data in the areas of AI, ML, and other topics that are linked to these areas. Owing to current developments in both biomedical and information technology, a number of research projects have been freshly performed. These efforts have resulted in the development of a number of algorithms that may be

used for the purpose of AD prediction. Therefore, let's have a look in this part at some of the most current ways that have been created using data mining approaches.

SVM, NB, RF, and MLP-NN are the four kinds of machine learning models that Sekaran and Sudha.[22] employed to analyze the gene expression of AD patients and normal persons. In this investigation, they made use of a dataset known as Gene Expression Omnibus (GEO: GSE1297) that is kept up to date by the NCBI. As a gene selection tool, the statistical t-test approach that requires a significant level of p-value less than 0.05 is used in order to choose the optimal gene subset. The following are the findings that indicated the correctness of the models that were discussed above: (87.10), (90.32), and (97.66) accordingly. The MLP-NN approach carries out superior to the other techniques when it comes to recognizing the difference amid AD genes and normal genes and establishing its effectiveness.

Lena and colleagues [23] developed a classification approach for predicting Parkinson's disease using data from the GSE5281 dataset, which is also known as the AD dataset. A Wrapper of GA/SVM is a feature selection approach that is used to pick a subset of important genes that increases the classification performance. This method is utilized as a part of a genetic algorithm. There have been a total of six distinct classification approaches used, including NB, C4.5 (decision tree), KNN, RF, and SVM using either a linear or a Gaussian kernel. According to the findings, the following models' degrees of accuracy are as follows: (81.4), (78.9), (87.0), (87.0), (85.7), and (91.9) accordingly.

Researchers Yang, L., et al. [34] analyzed the gene expression of Parkinson's disease patients and normal persons using four different machine learning models: SVM, NB, RF, and KNN algorithms. GEO: GSE132903 was the name of the dataset that was utilized in this investigation. It is maintained by the NCBI. The statistical strategy known as MI is used as a gene selection tool in order to choose the most effective gene subset. The findings showed that the aforementioned models were accurate to the following degree: 0.918%. The issues posed by the relatively modest sample sizes are the constraints of this work.

A classification approach for predicting Parkinson's disease using the dataset GSE63060 and GSE63061 that is collectively indicated as the AD dataset was given by Huang, Y, et al. [24]. As a strategy for feature selection, the inclusion of ANOVA and MI is used to pick a subset of important genes, which ultimately leads to an improvement in classification performance. There have been many categorization approaches used, including the k-means algorithm and the CNN. According to the findings, the accuracy of the two models presented above was as follows: 0.886 and 0.929 respectively. Finding the network architecture those results in the highest level of prediction accuracy is one of the most basic challenges posed by deep learning. The selection of network hyperparameters, such as the number of layers, kinds of transformations, and training parameters, is an important part of this process.

A categorization method for Parkinson's disease was provided by Xinzhong et al.[25] utilizing gene expression datasets, such as GSE63061 and GSE63060. These two sets of data have been combined. The approach of feature selection known as LASSO is utilized in order to identify the best possible subset. CN, MCI, and subjects with Parkinson's disease can be distinguished from one another by using the classification models SVM, RF, and LRR, all of which have proven to be predictive. Accuracy was determined to be 0.773 for the SVM classification model, 0.785 for the RF classification model, and 0.765 for the RR classification model.

K. Muhammed Niyas and colleagues [26] published a paper on the detection of Parkinson's disease via machine learning methods. [27] In order to categorize photos, Roman Filipovych and colleagues employed a supervised approach using the SVM model. D. Jha et al.[28] presented a cluster analysis method for diagnosing Parkinson's disease using MRIs of the brain in light of the selection tree. The random neural system group was suggested by Xia-an Bi et al. [29] as a means of improving the order of execution. The Parkinson's disease Neuroimaging Initiative provided the researchers with the dataset that they used in their investigation. It was determined that the Elman Neural Network is an appropriate basis classifier since it makes use of the arbitrary neural system cluster and is reliant on the consequences of highlighted choices. The accuracy of the Elman Neural Network was proven to be 92.31 percent.

Xia-an Bi et al. [29] employed the random SVM clustering technique to identify Parkinson's disease and to categorize the disease areas into superior frontal gyrus, inferior frontal gyrus, precentral gyrus, and cingulate cortex. This approach was used to classify the disease regions. Using a variety of machine learning approaches, including multilayer perceptron, bagging, decision tree, CANFIS , and GA, Sandhya Joshi et al. [30] constructed numerous models for the categorization of Parkinson's disease. The CANIFIS approach achieved an accuracy of 99.55 percent in terms of categorization. A hybrid model consisting of SVM and Bayesian Classifier was created by Claudia Plant and her colleagues [31] to identify the patterns of brain atrophy as well as to forecast Parkinson's disease. The pattern matching index for their approach came in at 92 percent after being calculated. A novel strategy for classifying MR

brain pictures between those that are normal and those that are afflicted by Parkinson's disease was developed by Luis et al. [32]. Their approach begins with the collection of MRI features in the wavelet domain, then moves on to the dimensionality reduction and SVM classification stages. A modern technology that is based on the PCA was developed by Li et al. [33]. It employs techniques of continuous selection and dropout in addition to the model of RBM that is an advanced instructional strategy. Ye et al. [34] developed a method to assess AD from medical pictures with ML-based multimodal data fusion as the primary data source.

3. Materials and Methods

In this part of the article, the dataset, several methods for getting the data ready, and the machine learning algorithm that powers the classification model are all discussed. This section also describes the statistical model performance evaluator that is available in Weka. This evaluator may be utilized to evaluate and evaluate the robustness and techniques dependability that have been built.

3.1 Dataset

It is possible to have access to a big number of different biological datasets. The Gene Expression Omnibus that is a data repository that is open to the general public, was used in order to obtain the data that were used for this specific paper (GEO). It was first made accessible to the general public on August 5, 2015, by the NCBI. The information that was presented by the AddNeuroMed Cohort indicated that the accession numbers for the dataset are as follows: (GSE63061 and GSE63060). These two separate datasets have been combined into a single one that is indicated as the AD dataset so that the total number of samples may be increased. In order to conduct a study on the number of gene expression that was seen in the AD dataset, the microarray technique was used. It is structured with suitable columns for 16382 genes and rows for each gene (569 samples). It is possible to divide it into ("245 patients with AD, 142 MCIs, and 182 CTLs"). Fig. 1 illustrates an instance of the steps included in creating a classification approach to predict Parkinson's disease in its early phases.

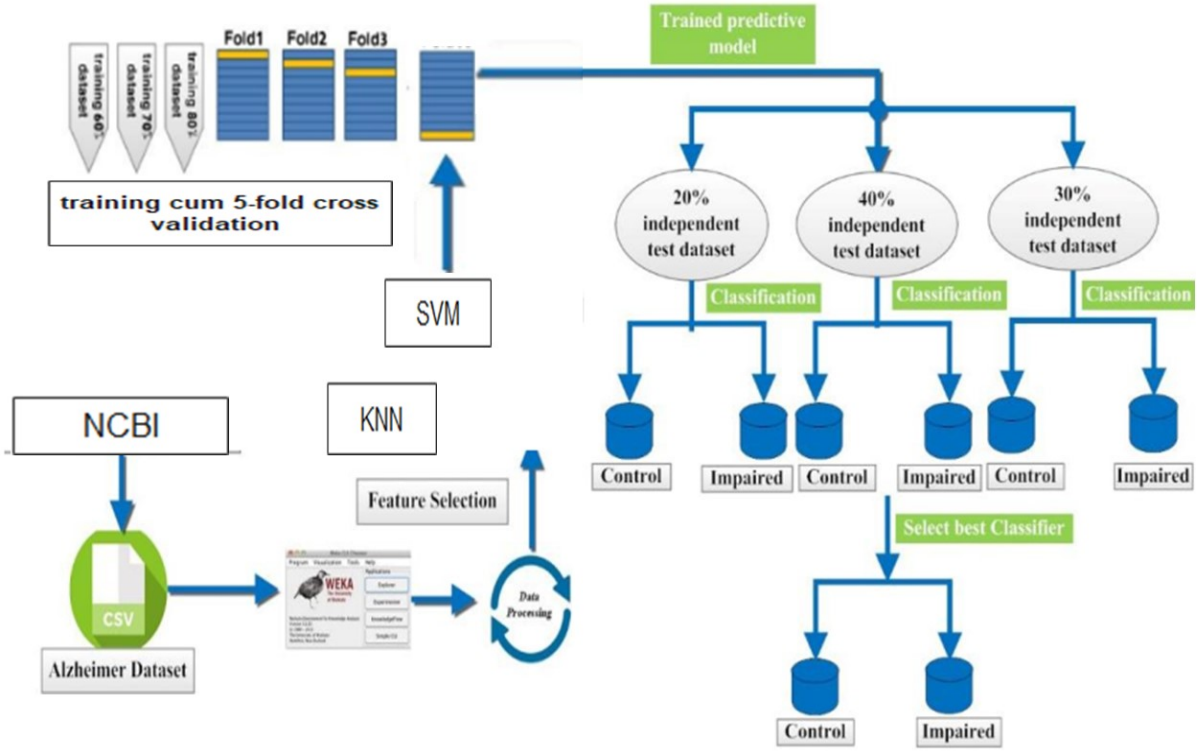


Fig.1.*Flow chart for predicting Parkinson's disease in its early stages*

3.2 Feature Cleaning/ Feature Selections to Reduce Redundant Data

When dealing with very large datasets, one common strategy for improving model correctness or boosting processing time is to decrease the dimensionality of the data by selecting fewer characteristics to analyze. This is a common approach. The following is a list of some of the reasons why the strategy of feature selection has to be used in this scenario:

1. The training process for some machine learning algorithms may become unmanageably drawn out if more than 16382 features are input into the system.
2. The models have been made more straightforward so that analysts and customers will have an easier time understanding them.
3. A reduction in instances of overfitting leads to an increase in generalization as a consequence (formally, reduction of variance).
4. It's conceivable that a lot of the features in the data are redundant (strongly connected, linearly dependent), or else they're not important at all. This is a possibility, but it's also possible that none of the characteristics are meaningful. It is feasible to disable these features without causing any noticeable loss of data in the process.

3.2.1 Information Gain

The degree of impurity in a random collection of samples may be calculated using their entropy. Information gain is the term used to describe the anticipated reduction in entropy that occurs as a result of partitioning the samples prior to dividing the feature node. It is a technique for figuring out how the inputs and outputs are connected to one another. Therefore, the higher the amount of new information obtained the better [35]. Fig. 2 demonstrates the information gain bar chart.

3.3 Select Features and Integrate

The number of relevant features that may be acquired from a total of 44 characteristics via the use of IG is determined by applying a variety of feature cleaning and selection processes. This results in a considerable decrease in dimensionality, which is a significant accomplishment considering there were over 16382 characteristics before. A model for making predictions is constructed with the help of these 44 attributes.

3.4 Machine Learning Algorithms for Model Building

A method that uses machine learning to analyze relationships between attributes is used to classify the subjects. In this study, three well-known machine learning techniques, known as SVM, NB, and K-NN, were utilized to classify individuals into impaired and healthy control groups on the basis of specified characteristics. Several different statistical approaches were used to estimate and contrast the accuracy of each model's predictions. The following is a brief summary of the ML methodologies that were used in the creation of a classification model for the present research that can differentiate between those who are affected by MCI and healthy individuals:

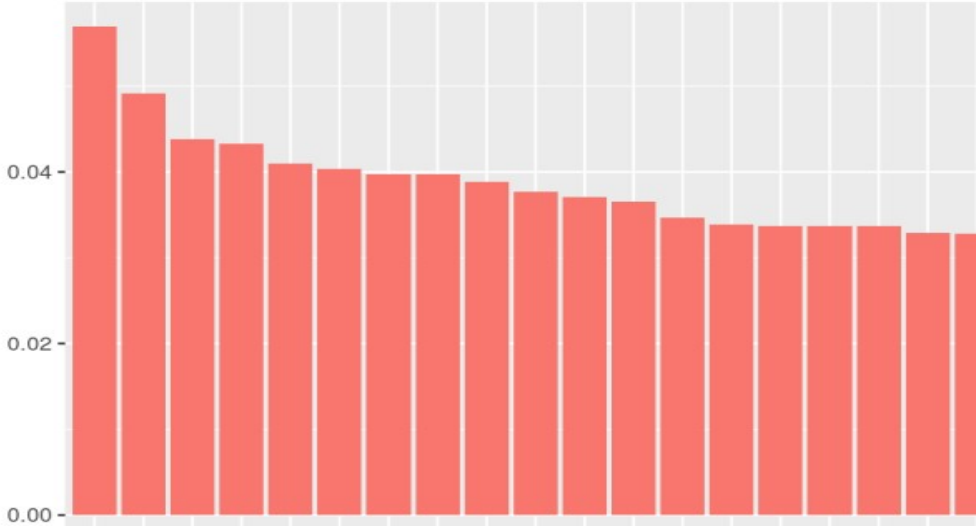


Fig.2*.Information GAIN Bar Chart*

a) NB: The NB method is based on the assumption that all of the predictive characteristics included in the training dataset (X1, X2, etc.) are conditionally independent of one another. The Bayes theorem is used in order for the NB technique to categorize the qualities that are present in the test dataset. It determines the likelihood of a characteristic being placed in one of the specified classes at some point in the past and calculates that likelihood. According to the Bayes rule, previous experience is what defines the prior probability of an attribute. As a consequence of this, the participants in the test case are separated into groups based on the conditional probabilities associated with numerous different attributes. Second, the possibility of a subject being categorized into one of the classes is determined by the proportion of other topics in any of the classes that have similarities with the subject at hand. In terms of NB analysis, the final classification of a subject in a data set is determined by multiplying prior and probability information on an attribute to produce a posterior possibility. This is done in order to create a posterior possibility. If a person has a greater likelihood of possessing characteristics that are associated with a certain class, then that person will be placed in that group [36]. The explanation of the NB technique, is as follows:

Presume that the person's likelihood " X "with definite characteristics is $Z = < z1..., zn >$ belonging to the impaired class that is indicated as the letter " I " and is exhibited as below:

$P(I1)$ refers to previously related probability with class $I1$, and $P(I1 | zi)$ refers to posterior probability. Hence, have " n " different hypotheses.

b) SVM: SVM is a supervised ML approach that can undertake problems in regression and classification. Support Vector (frontier) is just the coordinates of individual observation in the hyper-plane that optimally differentiates or segregates the two classes [37]. The decision function of SVM is indicated as, subsequent to solving a convex optimization issue,

$$f(x) = \sin(wT x + b) \quad (1)$$

Wherein, b indicated biased as well as w indicates weight vector.

Advantages:

- The main effectual and efficient supervised ML approaches with high dimensional spatial data.
- Clearer margin of separation amid the support vector outcomes in superior prediction.
- It's particularly helpful while the number of dimensions goes beyond the number of samples [38] that is the perfect ML approach for the gene expression dataset (16382 dimensions, 569samples). Additionally, it is a memory-effectual technique [39].

Disadvantages:

Basically, Computation time is higher than a conventional ML approach. It carries out with noise data like data with large numbers of highly correlated features [40], which is shown in Fig. 3.

c) K-NN: The K-NN model is a straightforward data mining tool that might be employed to resolve both regression and classification issues. The K-NN classification approach presents an object to a certain class based on the majority classes of its K neighbors. The number of neighbors to be considered to poll is explained by the value of K that is a positive integer. In this analysis, the value of K is 11, that is selected by exploiting the trial and error m[41][4] and it is exhibited in Fig. 4.

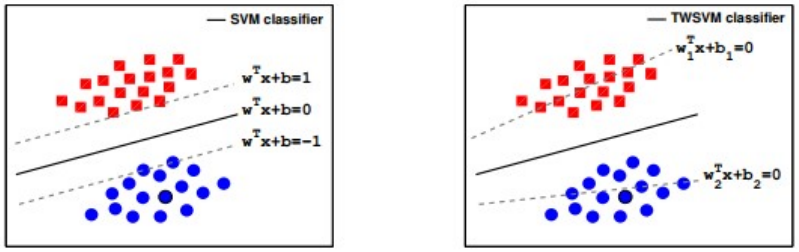


Fig.3. *Graphical representation of SVM and TWSVM classifiers [40]*

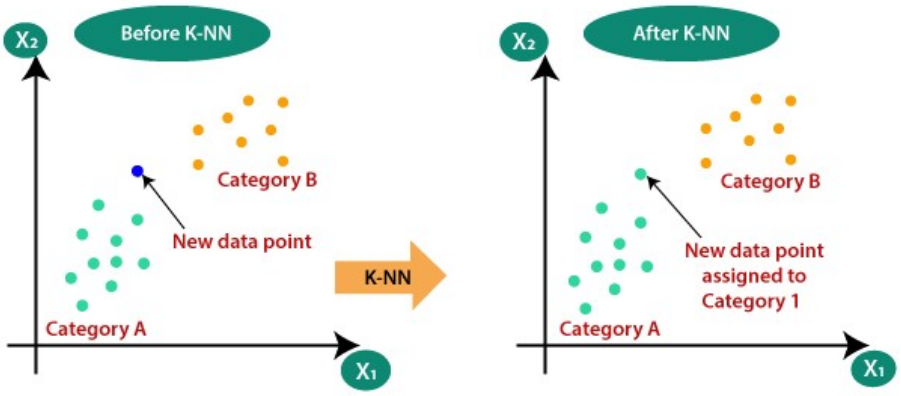


Fig.4. *KNN classifier [4]*

4. Result and analysis

It is vital to do data preparation in order to correctly feed data to each algorithm used in ML. In addition, the performance of the ML approach is significantly influenced by the quality of the data [42]. Data transformation has been carried out since the format of the dataset including gene expression information is not ideal for incorporation into the algorithms. This dataset on gene expression is now laid out in the format of x column (individual) and row (gene expression), however, it must be laid out in the format of row (individual) x column (individual) (gene expression). Table 1 displays the results of the classifiers' performance in terms of their ability to classify data throughout the testing and validation phase.

Table 1: Accuracy of data mining classification

Classifier	Accuracy
K-NN	0.91%
Naive Bayes	0.88%
SVM	**0.98%**

Table 1 summarizes the accuracy of the data mining classifiers' classification was evaluated both during the phase of validation and the phase of testing. 3. At the conclusion of each iteration, the mean value of the performance measure will be provided using k-fold cross-validation.

a) Evaluation of performance: Overfitting happens while parameters of models are identified and examined on a similar dataset, ensuing in flawless accuracy while training with observed data but considerably erroneous outcomes while training with unobserved data. Cross-validation is exploited in this paper to avoid over-fitting when accuracy and AUC are exploited to quantify performance[43].

(i) Accuracy: Each forecast consists of two values: the first is the likelihood of having an illness, which is represented by the number "1," and the second is the "probability of not having a disease, which is represented by the number "0". Another value is the probability of not having a disease, which is represented by the number 0. The accuracy is determined by using a threshold of 0.1, which indicates that the value is a "1" if it contains 0.1, and else it is a "0." Because it is believed that the cost of a false positive and a false negative would be the same, the threshold has been set at 0.1.

(ii) The AUC is a measurement of accuracy that is derived from the ROC curve and illustrates the balance that must be struck between sensitivity and specificity[44]. The connection between sensitivity and specificity is one of inversion (maximizing sensitivity outcomes in minimizing specificity). The AUC curve that is optimal and accurate best covers the whole of the ROC space, stretching from the bottom right to the top left. The ROC space curve that has a 45-degree diagonal has a prediction power of 50 percent that is a categorization that is generated at random.

(iii) Cross Validation : In order to test the model, k-fold cross-validation separates the training set into a number of smaller sets referred to as validation sets. The remaining part of the data set is then used to train the model. At each iteration, the model is put through its paces using progressively more minimal subsets, starting with the smallest one. The outcome is the simple average of all of the scores.

1. The k value used to train each model in this investigation is 5.
2. A more limited data set is used to validate the performance metric, which in this instance is accuracy.

5. Discussions

The training dataset was used for all of the classification models in this work, and they were all trained using five folds of cross-validation. Using the training dataset for cross-validation helps prevent models from being too tailored to their data. The performance of the classifiers is evaluated with the help of an unidentified test dataset. By exploiting the IG approach, we were able to choose the characteristics that are the most discriminative and would contribute the most to the early-stage diagnosis of Parkinson's disease.

Applying the SVM classifier to the information in Fig. 5 will help you determine which traits are the most important.

The quantity of characteristics used has a direct bearing on the degree of accuracy produced by the analysis. It had the highest accuracy of 96.6 percent and had 44 characteristics. In this method of feature selection, the accuracy of the mean cross-validation is improved with an increase in the number of features when the threshold term is lowered (that is when it is set to a value that is less than 125 mean).

6. Conclusions

We recommended using a set of characteristics to enhance AD diagnosis prediction. Feature selection identifies the best SVM training features. This research performed a 5-fold cross-validation to show the resilience of our technique. ML and data mining are used in medical and healthcare to identify and diagnose illnesses early. Our technique has an advantage over prior feature selection techniques since the training system mechanically replenishes essential features for improved prediction. Several of our technique's properties have been connected to Parkinson's disease or other psychiatric illnesses, proving its usefulness. The research showed that ML as well as data mining can reliably identify, forecast, and diagnose illnesses. Increase the quantity of AD and NC class instances so the approach may be trained with adequate and balanced data to improve AD stage classification.

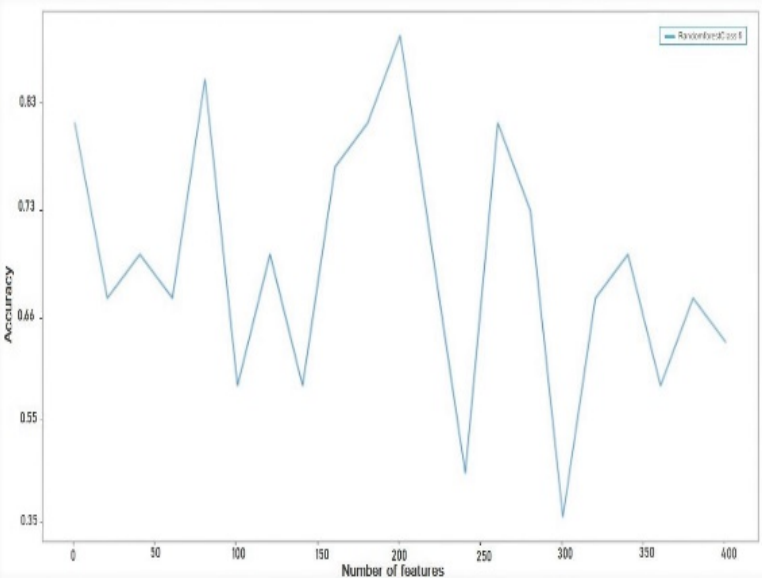


Fig.5. *Graphical representation of SVM classifier to choose features*

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] David, D. S., Arun, S., Sivaprakash, S., Raja, P. V., Sharma, D. K. et al. Enhanced Detection of Glaucoma on Ensemble Convolutional Neural Network for Clinical Informatics. CMC-Computers, Materials & Continua, vol. 70, no.2, pp. 2563–2579. 2022.

[2] David, D. S., Anam, M., Kaliappan, C., Arun, S., Sharma, D. K. et al. Cloud Security Service for Identifying Unauthorized User Behaviour. CMC-Computers, Materials & Continua, vol. 70,no. 2, pp. 2581–2600. 2022.

[3] Jayachandran, A., and D. Stalin David. "Textures and Intensity Histogram Based Retinal Image Classification System Using Hybrid Colour Structure Descriptor." Biomedical and Pharmacology Journal, vol. 11, no. 1, 2018, p. 577+. Accessed 12 Feb. 2021.

[4] D. Stalin David, "Parasagittal Meningiomia Brain Tumor Classification System based on MRI Images and Multi Phase level set Formulation", Biomedical and Pharmacology Journal, Vol.12, issue 2, pp.939-946. 2019.

[5] Thendral R., David D.S. An Enhanced Computer Vision Algorithm for Apple Fruit Yield Estimation in an Orchard. In: Raje R.R., Hussain F., Kannan R.J. (eds) Artificial Intelligence and Technologies. Lecture Notes in Electrical Engineering, vol 806. Springer, Singapore. 2022. https://doi.org/10.1007/978-981-16-6448-9_27

[6] D. S. David and A. Jeyachandran, "A comprehensive survey of security mechanisms in healthcare applications," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, pp. 1-6, 2016, doi: 10.1109/CESYS.2016.7889823.

[7] Stalin David, D., Jayachandran, A. A new expert system based on hybrid colour and structure descriptor and machine learning algorithms for early glaucoma diagnosis. Multimed Tools Appl 79, 5213–5224 (2020). https://doi.org/10.1007/s11042-018-6265-1.

[8] D Stalin David, A Jayachandran, 2018,Robust Classification of Brain Tumor in MRI Images using Salient Structure Descriptor and RBF Kernel-SVM, TAGA Journal of Graphic Technology, Volume 14, Issue 64, pp.718-737.

[9] D Stalin David, Robust Middleware based Framework for the Classification of Cardiac Arrhythmia Diseases by Analyzing Big Data, International Journal on Recent Researches In Science, Engineering & Technology, Volume 4, Issue 9, pp.118-127. 2018.

[10] M. Rajdhev, D. Stalin David, "Internet of Things for Health Care", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 800-805, March-April 2017.

[11] P. Prasanth, D. Stalin David, "Defensing Online Key detection using Tick Points", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 758-765, March-April 2017.

[12] A. Sudalaimani, D. Stalin David, "Efficient Multicast Delivery for Data Redundancy Minimization over Wireless Data Centres", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 751-757, March-April 2017.

[13] R. Abish, D. Stalin David, "Detecting Packet Drop Attacks in Wireless Sensor Networks using Bloom Filter", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 730-735, March-April 2017.

[14] A. Vignesh, D. Stalin David, "Novel based Intelligent Parking System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 724-729, March-April 2017.

[15] D Stalin David, 'Diagnosis of Alzheimer's Disease Using Principal Component Analysis and Support Vector Machine, International Journal of Pharmaceutical Research, Volume 12, Issue 2, PP.713-724. 2020.

[16] Jaswanth K S, Dr. D. Stalin David, "A Novel Based 3d Facial Expression Detection Using Recurrent Neural Network", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 48-53, March-April 2020.

[17] D Stalin David, 'An Intellectual Individual Performance Abnormality Discovery System in Civic Surroundings' International Journal of Innovative Technology and Exploring Engineering, Volume 9, Issue 5, PP.2196-2206. 2020.

[18] D Stalin David, 'Machine learning for the prelude diagnosis of dementia', International Journal of Pharmaceutical Research, Volume 13, Issue 3, PP.2329-2335. 2020.

[19] David, D.S. and Y. Justin, A Comprehensive Review on Partition of the Blood Vessel and Optic Disc in RetinalImages.Artech J. Eff. Res. Eng. Technol., vol. 1, pp. 110-117. 2020.

[20] D. Stalin David and A.A. Jose, Retinal image classification system for diagnosis of diabetic retinopathy using SDCMethods.Artech J. Eff. Res. Eng. Technol., vol. 1: pp. 87-93. 2020.

[21] D. Stalin David and T. Joseph George, Identity-based Sybil attack detection and localization.ArtechJ. Eff. Res. Eng. Technol., vol. 1, pp. 94-98, 2020.

[22] David, D.S. and L. Arun, Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme.Artech J. Eff. Res. Eng. Technol., vol. 1, pp. 57-63. 2020.
[23] David, D.S., Retinal image classification system for diagnosis of diabetic retinopathy using morphological edgedetection and feature extraction techniques.Artech J. Eff. Res.Eng. Technol., vol.1, pp. 28-33, 2020.
[24] David, D.S., A novel specialist system based on hybrid colour and structure descriptor and machine learningalgorithms for early diabetic retinopathy diagnosis.Artech J. Eff. Res. Eng. Technol., vol. 1, pp50-56. 2020.
[25] David, D.S. and M. Samraj, A comprehensive survey of emotion recognition system in facial expression.Artech J. Eff. Res. Eng. Technol., vol. 1, pp. 76-81. 2020.
[26] David, D.S. and L. Arun, Multi-view 3D face renovation with deep recurrent neural networks. ArtechJ. Eff. Res. Eng. Technol., vol. 1, pp. 64-68, 2020.
[27] David, D.S. and S. Namboodiri, Improvement of framework for the grouping of CA diseases by investigating bigdata.Artech J. Eng. Appl. Technol., vol. 1,pp. 7-14. 2020.
[28] Stalin David D , Saravanan M, 'Multi-perspective DOS Attack Detection Framework for Reliable Data Transmission in Wireless Sensor Networks based on Trust', International Journal of Future Generation Communication and Networking , Volume 13, Issue 4, PP.1522–1539.2020.
[29] J. K. S and D. S. David, "A Novel Based 3D Facial Expression Detection Using Recurrent Neural Network," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262287.
[30] Stalin David D, Saravanan M, "Enhanced Glaucoma Detection Using Ensemble based CNN and Spatially Based Ellipse Fitting Curve Model", Solid State Technology, Volume 63, Issue 6, PP.3581-3598.
[31] Stalin David D, Saravanan M, Jayachandran A, "Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification", Solid State Technology, Volume 63, Issue 6, PP.3599-3623.
[32] D. Jayakumar; Dr.U. Palani; D. Raghuraman; Dr.D. StalinDavid; D. Saravanan; R. Parthiban; S. Usharani. "Certain Investigation On Monitoring The Load Of Short Distance Orienteering Sports On Campus Based On Embedded System Acceleration Sensor". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2477-2494.
[33] R. Parthiban; S. Usharani; D. Saravanan; D. Jayakumar; Dr.U. Palani; Dr.D. StalinDavid; D. Raghuraman. "Prognosis Of Chronic Kidney Disease (Ckd) Using Hybrid Filter Wrapper Embedded Feature Selection Method". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2511-2530.
[34] Dr.U. Palani; D. Raghuraman; Dr.D. StalinDavid; R. Parthiban; S. Usharani; D. Jayakumar; D. Saravanan. "An Energy-Efficient Trust Based Secure Data Scheme In Wireless Sensor Networks". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2495-2510.
[35] Dr. D. Stalin David; R. Parthiban; D. Jayakumar; S. Usharani; D. RaghuRaman; D. Saravanan; Dr.U. Palani."Medical Wireless Sensor Network Coverage And Clinical Application Of Mri Liver Disease Diagnosis". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2559-2571.
[36] D.Raghu Raman; D. Saravanan; R. Parthiban; Dr.U.Palani; Dr.D.Stalin David; S. Usharani; D. Jayakumar."A Study On Application Of Various Artificial Intelligence Techniques On Internet Of Things". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2531-2557.
[37] D.Saravanan; Dr.D.Stalin David; S.Usharani;D.Raghuraman; D.Jayakumar; Dr.U.Palani; R.Parthiban. "An Energyefficient Traffic-Less Channel Scheduling Based Data Transmission Inwireless Networks". European Journal of Molecular & Clinical Medicine, 2020, Volume 7, Issue 11, Pages 5704-5722.
[38] S. Usharani; D.Jayakumar; Dr.U.Palani; D.Raghuraman; R.Parthiban; D.Saravanan; Dr.D.Stalin David. "Industrialized Service Innovationplatform Based On 5g Network And Machinelearning".European Journal of Molecular & Clinical Medicine, 2020, Volume 7, Issue 11, Pages 5684-5703.
[39] P Gopala Krishna, D StalinDavid, "An Effective Parkinson's Disease Prediction Using Logistic Decision Regression And Machine Learning With Big Data", Turkish Journal of Physiotherapy and Rehabilitation; 32(3), Pages 778-786.
[40] Jaswanth K S, Dr. D. Stalin David, "A Novel Based 3d Facial Expression Detection Using Recurrent Neural Network", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 48-53, March-April 2020.
[41] T. Babu, H. Roopa, Arvind Kumar Shukla, D. Stalin David, S. Jayadatta, A.S. Rajesh, Internet of things-based automation design and organizational innovation of manufacturing enterprises,MaterialsToday:Proceedings,2021,ISSN:2214-7853, https://doi.org/10.1016/j.matpr.2021.10.459.
[42] M. Chandragowda, C. Gnanavel, D. Saravanan, D. Stalin David, R. Parthiban, A.S. Rajesh,Consequence of silanee combination representative on the mechanical possessions of sugarcane bagasse and polypropylene amalgams,Materials Today: Proceedings,2021, ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2021.10.455.
[43] T.V.V. Pavan Kumar, Shafqat Nabi Mughal, Radhika Gautamkumar Deshmukh, S. Gopa Kumar, Yogendra Kumar, D. Stalin David,A highly consistent and proficient class of multiport dc-dc converter based sustainable energy sources,Materials Today: Proceedings, 2021,ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2021.10.458.
[44] David, D.S. Enhanced glaucoma detection using ensemble based CNN and spatially based ellipse fitting curve model. J Ambient Intell Human Comput (2021). https://doi.org/10.1007/s12652-021-03467-4.