DE GRUYTER Bio-Algorithms and Med-Systems 2021; 17(3): 181–189

Kamalakannan Kaliyan* and Anandharaj Ganesan

Regularization based discriminative feature pattern selection for the classification of Parkinson cases using machine learning

https://doi.org/10.1515/bams-2021-0064
Received June 2, 2021; accepted July 27, 2021;
published online August 19, 2021

Abstract

Objectives: This paper focuses on developing a regularization-based feature selection approach to select the most effective attributes from the Parkinson's speech dataset. Parkinson's disease is a medical condition that progresses as the dopamine-producing nerve cells are affected. Early diagnosis often reduces the effect on the individuals, minimizes the advancement over time. In recent times, intelligent computational models are used in many complex cases to diagnose a clinical condition with high precision. These models are intended to find meaningful representation from the data to diagnose the disease. Machine learning acts as a tool, gears up the model learning process through a mathematical baseline. But, not in all cases, machine learning will be demanded to perform optimally. It comes with a few constraints, mainly the representation of the data. The learning models expect a clean, noise-free input, which in-turns produces better discriminative patterns over different categories of classes.
Methods: The proposed model identified five candidate features as predictors. This feature subset is trained with different varieties of supervised classifiers to trace out the best-performing model.
Results: The results are validated through accuracy, precision, recall, and receiver's operational characteristic curves. The proposed regularization- based feature selection model outperformed the benchmark algorithms by attaining 100% accuracy on most of the classifiers, other than linear discriminant analysis (99.90%) and naïve Bayes (99.51%).

Conclusions: This paper exhibits the need for intelligent models to analyze complex data patterns to assist medical practitioners in better disease diagnosis. The results exhibit that the regularization methods find the best features based on their importance score, which improved the model performance over other feature selection methods.

Keywords: classification; feature scoring; machine learning; neurodegenerative disorder; Parkinson; regularization.

***Corresponding author: Kamalakannan Kaliyan**, Research Scholar, PG & Research Department of Computer Science, Adhiparasakthi College of Arts & Science, Kalavai, India,
E-mail: kamal.sram@gmail.com
Anandharaj Ganesan, PG & Research Department of Computer Science, Adhiparasakthi College of Arts & Science, Kalavai, India,
E-mail: younganand@gmail.com

Introduction

According to the statistical figures provided by *"Parkinson.org"* [1], nearly a million people in the United States (US) are living with Parkinson's, reported in the year 2020. The people diagnosed with other conditions such as Lou Gehrig's disease (also known as Amyotrophic Lateral Sclerosis), muscular dystrophy, and multiple sclerosis all together combined is less than the number of cases of Parkinson's Disease (PD). Nearly 60,000 people per year from the US are diagnosed with PD and nearly 10 million people across the globe are affected by PD. It is observed that the chance of developing PD increases with age. Conversely, four percent of people are affected before 50 years of age. Men are more likely to develop this condition 1.5 times than women are. In the case of India, a clinical study is conducted by comparing the PD prevalence among different ethnic groups. This study reveals that the prevalence over Anglo-Indians is five times lesser than the general Indian population [2]. A door-to-door survey is conducted in Bangalore of South Karnataka in the year 2004 where the prevalence rate is found to be in the count of 33 per 1,00,000 [3, 4]. In Kashmir, the prevalence rate was 14.1 in every 1,00,000, whereas it is 134 per 1,00,000 when the age factor is adjusted [5].

The statistical data reveals an alarming fact that Parkinson's disease is one of the most dangerous ones that affect people, especially in their elder stages. The major challenges are the lack of proper infrastructures, medical facilities, and poor accessibility, which leads to a steady increase in the number of cases every year. If the condition is left undiagnosed, it leads to severe effects on the individual. In Parkinson's, the cases are categorized into five stages if diagnosed [6]. In the initial stage, the people experience mild signs of PD, which won't affect daily routines. Few physical changes can be observed in the postures and facial expressions. In stage two, symptoms are highly visible, often with tremor and rigidity.

The next stage is marked as mid-level PD. Loss of balance; slowness in movements become more common. It interferes in simple activities like eating, dressing; sometimes push the individual to fall. Stage four requires people with dependency on the walker, needs help from others to fulfill the activities. Stiffness in the legs, hallucinations, delusions are the symptoms observed in the person in Stage 5 of PD. The person should be bedridden, needs a care from medical professional all the time [7].

Many modern medical procedures are introduced in the field of neurological, neurodegenerative medicine to effectively treat PD-affected individuals. The procedures often vary based on dependent factors such as age, ethnicity, the field of employment, accidents, past medical history, genetic factors, and familial background. The diagnosis methods often include medical imaging techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI) for accurate diagnosis, and Diffusion MRI [8]. Additionally, the vocal test is included in the preliminary assessment, where the voice pattern of the normal and PD cases is analyzed to find the changes. Medications for PD suggested to be effective are Levodopa (L-Dopa), dopamine agonists, and MAO-B inhibitors [9]. In advanced stages of PD, deep brain stimulation is suggested to improve the condition when the drugs are ineffective [10].

The images captured from the devices are visually analyzed by the medical experts to diagnose and categorize the stages of PD if affected. Recent advancements in the field of computer vision open up new pathways to analyze the images automatically, finding morphological structures, and spatial information to make accurate predictions over the condition [11–13]. Deep learning models often outperform the traditional methods in every application, sometimes even beats human experts. Diabetic retinopathy, breast cancer, lung cancer, brain tumor, bone dislocation, and even Covid-19 from the lung images can also be diagnosed with high precision by deep learning models, developed by the team of medical and computer experts [14, 15]. The main pitfall is the lack of access to data, as the medical information is very much sensitive. Not only from the images but based on the data availability, an intelligent diagnosis model can be built. In this paper, the vocal recording dataset of Parkinson's and healthy cases are analyzed to find the important feature point that determines the discrimination between two cases. Regularization- based feature selection is adapted to rank the features upon their importance score. Then, the selected features are inputted into different learning algorithms, trained with 5 fold cross-validation method. The test results are evaluated under performance evaluation metrics.

The rest of the paper is organized as follows. In Section 2, the literature work is conducted from various databases related to the current study. The materials and methods section details the process of the experiment, explains the dataset properties, feature selection methods, significance of regularization approaches, model training, and validation. The results obtained in this study are presented in Section 4 with visual graphs and charts. The findings and path for future enhancements of this study are discussed in Section 5, which concludes the work.

Background study

Many studies based on hypothetical assumptions are conducted on heterogeneous medical data such as electronic health records, electromagnetic signals, digital images, and genetic data by applying intelligent learning models. The outcome of the study often reveals more insights from the data, with logical implications and associations between several factors.

Irrespective of any disease condition, the concept of machine learning is widely adopted in various sectors with no bound. This section discusses some of the literature exhibiting their findings, which are related to the present study.

In recent work, a fusion model is proposed to find the significant parameters from the Parkinson's speech dataset. Discrete Wavelet Transform fused with Mel Frequency Cepstral Coefficients identifies best features from the data. This finding is statistically validated through descriptive and statistical methods. The model attained a 0.86 ROC value when tested with logistic regression and has reached 76.78% accuracy with descriptive analysis, which is comparatively less than the accuracy score obtained during factor analysis as 78.23% [16]. Another study employs model-based logic strategies to find the most significant indicators of PD-affected individuals. It uses logistic regression, support vector classifier, gradient boosting model, *k*-nearest neighbor, random forest classifier. The model is trained and validated using 5 fold cross-validation and the scores are calculated using *f*1 score, precision, confusion matrix, and recall. SVC among other classifiers attained 93.83% accuracy where bagging reached only 73.28% [17]. A clinical approach is proposed by combining two effective algorithms, chaos-mapped bat algorithm (CMBA) and support vector machines to find a feature subset for the Parkinson's speech data. This coordination is said to be minimized the problem of parameter tuning in SVM. The CMBA intends to find the features and SVM validates the subset [18].

The biomedical experts prefer the decision forests over decision trees in case of massive acoustic signal data for PD detection [19]. In this study, a similar model is examined with two recent decision forest techniques such as Systematically Developed Forest (SysFor) and Penalizing Attributes o1f decision forest (ForestPA). This approach achieved maximum accuracy with the fewer number of decision trees thereby optimized the model complexity. The highest detection accuracy is observed in the range of 94.12–95.00% [19]. A similar study employed a mixture of *k*-means and decision trees to find the pattern from PD and normal samples of the dataset. The spiral drawing inputs are added up with the existing data, which improved the model performance significantly. The principal vectors are extracted from the spiral drawing and inputted into a support vector machine for classification [20]. Another study uses the recursive feature elimination (RFE) and features importance calculation techniques for feature selection.

For the classification, artificial neural networks, support vector machine, and classification regression tree algorithms. The combination of SVM and RFE attained 93.84% accuracy with the least number of features [21].

Deep learning models have shown their prominence in medical image classification, segmentation, and object tracking. Some studies adopted deep learning models to improve the prediction performance on the numerical dataset. However, sometimes this leads to model overfitting as the deep learning models are highly sophisticated and complex. This study uses a nine-layered Convolution Neural Network (CNN) for the training and validation process. The proposed model is tested on the same dataset used in this current study. The performance validation is done with Leave-One-Person-Out (LOPO) cross-validation. Since the data is suffering from class imbalance, it uses f-measure and Matthew's correlation coefficient alongside accuracy metrics. The findings of the study exhibit the importance of deep learning algorithms, not only successful but effective in building up the discrimination between subclasses [22].

Materials and methods

In this section, the concepts behind the techniques are detailed with mathematical algorithms. The pipeline of the work is discussed with the process under every phase such as selecting the features, training the models, validating its performance, and benchmarking with algorithms. The workflow of the current system is given in Figure 1.

Dataset details

The experimental study is conducted with the Parkinson's speech dataset. The data is collected from the population of two different subgroups, Parkinson's affected and healthy cases. The vocal recordings of the individuals are collected for analysis to find the pattern that shows the discrimination between the subgroups. This dataset is accessed from the University of California (UCI) Irvine Machine Learning repository [23]. It consists of a total of 40 records, with 20 Parkinson's and 20 healthy cases. The test is carried out with 26 multiple varieties of sound recordings from all the cases (26 voice samples including sustained vowels, numbers, words and short sentences) which in turn produces 1,040 samples. The properties of the dataset are represented in Table 1.

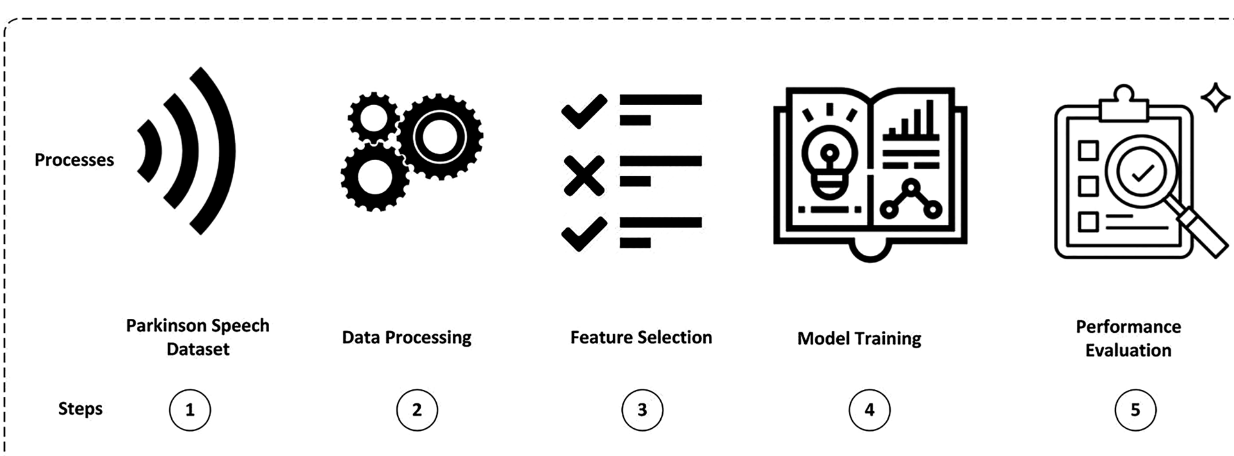

Figure 1: Systematic model workflow.

Table 1: Data set details.

Data set	Details
Source	https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings
Characteristics	Multivariate, real valued attributes
Number of instances	1,040
Number of attributes	26
Class	Binary target

Feature selection

In any machine learning application pipeline, the important phase that determines the model performance is feature selection. The learning models have a strong dependency on the data. If the data is noisy, inconsistent, and vague, then the model will learn the same during training that significantly affects the results on testing it. Technically, this can be denoted as "overfitting" of the model. The concept of feature selection is limited only to finite cases, which can't be useful when the data is complex and dynamic [24]. For instance, in images, there aren't any effective procedures available to select the features as the information is represented in terms of pixels. The spatial data can only be observed, extracted, and transformed into another form, where the result will be the vectors. Similarly in feature selection, the main advantage is, the outcome of every feature selection algorithm will be a subset of all available features, where each feature can be identified individually.

Thereby, the transparency will be more; also the factors affecting the results can be traced, but in the extraction process, the entire procedure turns out to be a black box. There exist several feature selection algorithms, each is categorized into different sections depends upon the mathematical constructs and background evidence [25].

The majority of algorithms are categorized into subtypes such as information theory-based feature selection, correlation feature selection, regularization methods, wrapper techniques, metaheuristic methods, and some variants of neural networks. This paper focuses on employing a regularization-based feature selection technique to find the features with higher significance based on the scores of individual features [26].

Regularization based feature selection

In general mathematical problems, regularization techniques are applied by adding additional information to control the model overfitting, thereby improving model stability and performance. The term often a kind of penalty added to the minimization function to lead the model towards an optimal solution. Regularization techniques intend to reduce the generalization error.

The generalization error is measured on testing data evaluated on the trained model. There are two major types of regularization such as L1 (also known as LASSO – Least Absolute Shrinkage and Selection Operator) [27] and L2 (Ridge or Tikhnov) [28, 29]. The L1 type employs a regression technique, which calculates the linear dependency among the features by mapping the relationship between input and output vectors. L2 regularization is also built with regression, alongside adds a penalty to the model. It penalizes the feature coefficients having larger values that prevent the problem of overfitting. The penalty term is often represented as 'λ', added to the loss function of the model. The method "shrinking" in L2 regularization focuses on eliminating redundant features from the feature set. The major difference between these techniques lies in navigating the feature coefficients near to zero. In this case, L1 often produces sparse results, which forces weak features of the dataset to have zero coefficients. These features are considered to have no importance in discriminating different classes. While L2 is not expected to turn the features to zero coefficients, rather a non-zero value is often produced [30].

The general equation of LASSO is given below as Eq. (1).

$$\min \in \frac{Rd}{a} \frac{1}{2} \parallel y - X^T \ a \parallel \frac{2}{2} + \lambda \|a\| 1 \tag{1}$$

In the above equation, the term denotes the vector coefficient, and is the regularization term. The major drawback of the L1 technique is its inability to find the non-linear representation between the features and target.

L1 regularization is often used as a feature selection technique as it precisely identified important features with zero and non-zero coefficients, where L2 never makes a feature coefficient to zero. But, based on the application, the type can be decided. In the case of larger dimensional datasets such as gene expressions, L1 will perform better, whereas the dataset with very limited features can be applied L2. The latter method gives each feature a coefficient; therefore an effective threshold technique can be used to fix the limit to select feature subset. The general cost function (Mean Squared Error) of the model is given in Eq. (2).

$$J(\varnothing) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_0 (x)^{(i)} - (y)^{(i)}\right) \tag{2}$$

The penalty term of L2 is given as,

$$R = \lambda \sum_{j=1}^{n} \theta j^2 \tag{3}$$

The penalized cost function is represented in Eq. (4).

$$J(\varnothing) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_0 (x)^{(i)} - (y)^{(i)}\right) 2 + \lambda \sum_{j=1}^{n} \theta j^2 \tag{4}$$

Threshold criteria

The threshold to choose the best feature subset from all the scores is done by applying a sigmoid function to the ranking vector. This criterion finds five features from the entire feature set as optimal sets. The sigmoid function is given in Eq. (5).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

The 'x's' in the equation denotes the steepness of the curve. Then after the process of applying sigmoid to the coefficient vectors, five features are selected as best among all the features. These features are inputted into the next phase to train the models. The labels of the five features are Jitter (Local), Shimmer (dda), Number of Pulses, Number of Periods, and UPDRS. The coefficient value of each feature is given in Table 2.

Supervised learning models

Machine learning gaining momentum from its well-performing, heterogeneous varieties of algorithms. These wide classes of algorithms fall under three major categories like supervised, unsupervised, and reinforcement models. Based on the kind of data to be analyzed, the algorithms can be chosen. In this system, the data is labeled; hence supervised learning algorithms are the best fit for the problem identified in this study.

Linear and non-linear algorithms, regression techniques, tree-based models, ensemble methods, neural network models, probabilistic approaches are the most common types of algorithms categorized under supervised learning strategies [31]. This paper employs five classifiers such as Support Vector Machines, Random Forest, Naïve Bayes, Back Propagation Neural Network, and Linear Discriminant Analysis for training and validation.

L2 Regularization Based Feature Selection on Parkinson's Speech Dataset

Input: *m (samples), n (features), y (target class), λ (regularization term of coefficients).*

Initialize the data processing steps:
Split the data into independent and target separately
Perform grid search to find the optimal C value for the L2 model of logistic regression model grid_params = [0.001, 0.01, 0.1, 1.0, 10.0]
for i in grid_params:
fit the data into LR model with grid_params[i] calculate the performance at the chosen param
if (performance(grid[$i + 1$]) > performance(grid[i])) max = grid[$i + 1$]
else
max = grid[i]
end for
apply L2 feature selection with identified params
train the classifiers on the selected features with 5-fold cv calculate the performance of the model with performance metrics

Output: Optimal feature subset S_o, Scores of the classifiers S_c

Table 2: Coefficients of the feature set identified by L2 regularization.

Feature	Coefficient value
Jitter (local)	7.58
Shimmer, dda	7.49
Number of pulses	7.28
Number of periods	7.05
UPDRS	6.92

Model validation and performance evaluation

The learning models are first trained and then validated on unseen data. The dataset should be divided into two parts, one for training and another set is for testing the performance of the trained model on new data. Many techniques exist such as hold-out validation, k-fold cross-validation, and stratified k-fold [32]. In the case of hold-out, under a fixed ratio/percentage, the data is divided into two sets but comes with few drawbacks. The data may sometimes be biased, or too many samples are considered for training, which may perform poorly on unseen data. But in k-fold cross-validation, these problems can be avoided. The number of folds should be decided before the partition of data. In this pipeline, 5 fold cross-validation is adapted as the number of samples is less. So, based on the amount of data passed during every epoch, 4 fold data undergone training, and 1 fold is reserved for testing. The average result on all epochs is calculated and the final performance will be decided upon the score.

The training time of the proposed model is much less compared to the training time of the conventional or baseline NN (trained on full features). The processing time of the proposed method is 9.0653 which was 5,194.5 in the baseline method. Hence, it is identified the proposed method exhibits outstanding performance in terms of classification accuracy. The performance of any learning model is identified by evaluating through the validation metrics. For supervised classification models, many metrics are available, notably accuracy, precision, recall (sensitivity), f-score, true positive rate, false-positive rate, and specificity [33–35]. The aforementioned classifiers employed in this framework are validated through these metrics.

Implementation details

The configuration of the system in the view of both hardware and software is detailed as follows. Anaconda distribution for installing and managing python machine learning libraries and PyCharm Integrated Development Environment for creating and building the project is used. The packages NumPy, pandas, sklearn, and matplotlib [36, 37] are accessed to manipulate the data, implementing a feature selection module, training and validating models, and visualize the results. The hardware consists of 8 GB RAM, 4 GB graphics memory, and 1 TB hard drive with Intel i7 core processor (fifth generation).

Table 3: Performance scores of the classifiers on L2 identified feature subset in (%).

Classifiers	Accuracy	F-measure	True positive rate	False positive rate
NB	0.9951	0.995	0.995	0.005
SVM	1.0	1.0	1.0	0.0
RF	1.0	1.0	1.0	0.0
LDA	0.999	0.999	0.999	0.001
BPNN	1.0	1.0	1.0	0.0

Table 4: Comparison of performance from previous literature.

Model	Accuracy	Number of features
Correlation-based FS – ant colony – SVM	95.00%	16
mRMR – stacked autoencoder	97.00%	8
L2 – (RF/SVM/BPNN)	100%	5

Results and discussion

The regularization-based feature selection technique finds five best features as predictor attributes from the feature set of 28. The identified features are inputted into classifiers for the model training and validation process. The classifiers attained the best results, reached 100% accuracy on three algorithms such as SVM, RF, and BPNN, whereas the LDA

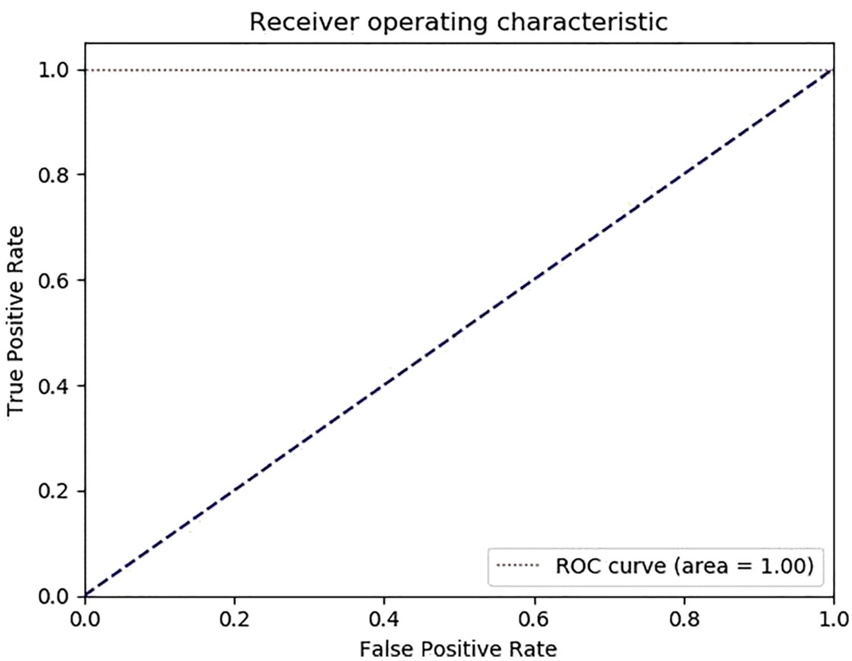

Figure 2: ROC curve of L2-SVM model.

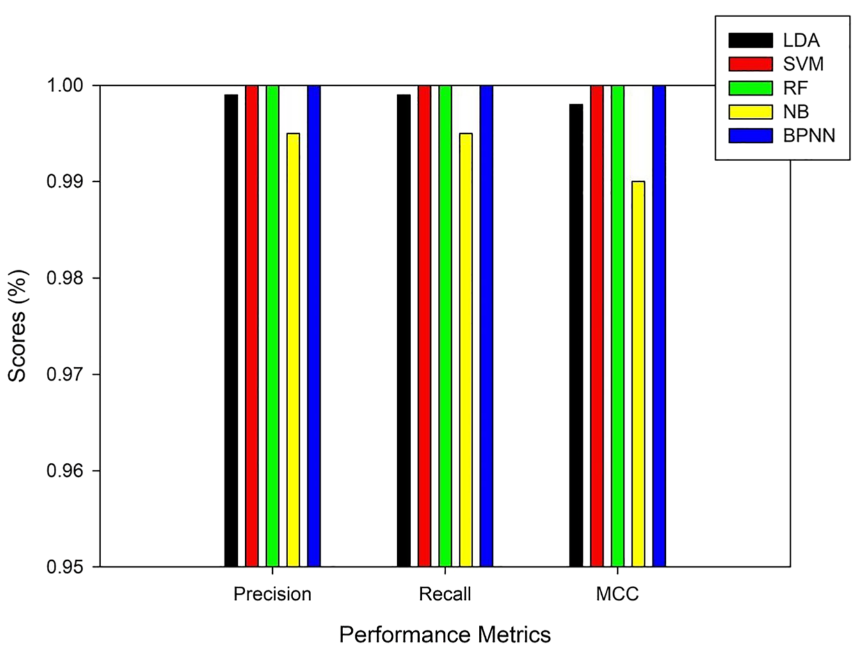

Figure 3: Performance of classifiers under different evaluation metrics.

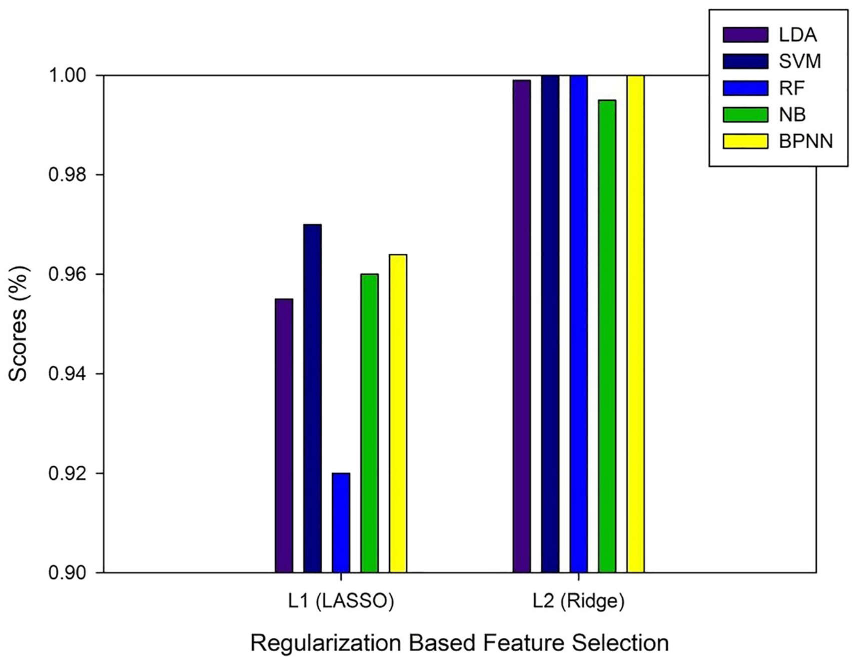

Figure 4: Comparison of performance between L1 and L2 regularization methods.

and NB scored 99.90 and 99.51% accuracy with one and five misclassifications respectively. The results of the algorithms attained under the subset are given in Table 3.

In Table 4, the performance of the models under different feature selection methods is benchmarked from the previous studies. Above all, the current method shows better significance in its performance by attaining 100% discrimination between two subclasses. In Figure 2, the ROC curve of the L2 – SVM model is given. Although other models such as RF and BPNN scored the same, to avoid repetition in the graphs, only SVM classifier is considered. In Figure 3, the precision, recall, and Matthew's correlation coefficient scores are plotted as a bar graph is given.

In Figure 4, the comparison between L1 and L2 regularization-based feature selection techniques is represented. Since L1 has too many sparse coefficients, only three features are selected out from 28, which also eliminates the natural bias from the data thereby reduced the model performance when compared with L2.

The regularization-based feature selection is widely used in bioinformatics for the selection of biomarkers [38, 39], which can also be effective in low-dimensional datasets [40, 41].

Conclusion

In this paper, an effective regularization-based feature selection technique is used to find the best predictor features of Parkinson's disease. The voice recordings fetched from the healthy and Parkinson's affected individuals are collected for analyzing the patterns. The final subset identified from all the 28 features is five. The reduction in the number of features further minimizes the complexity of the learning model to understand the hidden insight from the data. Furthermore, it simplifies the task of discriminating the category of classes as the data is projected with comparatively lesser dimensions with the selective features. These features are then trained with supervised learning models. To benchmark the scores among different classifiers, validation metrics were used. The model is evaluated through 5-fold cross-validation. Every four among 5-folds were undergone training; the remaining 1 fold is employed for testing the model classification performance. Additionally, a few more feature selection techniques are applied on the same pipeline to find the efficacy of the regularization-based technique. The results exhibit that the regularization methods find the best features based on their importance score, which improved the model performance over other feature selection methods. The main advantage of the current model is the minimization of features, thereby decreases model complexity and improves performance. In the future, this study will be further enhanced by collecting data in various aspects such as clinical trials, brain activity reports, gait information, and genetic data to provide more precise reports on the conditions of the individuals. Deep learning frameworks will be beneficial for modeling and fitting complex data to find the patterns hidden deep inside the data.

Research funding: No funding support.
Author contributions: Kamalakannan Kaliyan and Anandharaj Ganesan contributed equally.
Competing interests: Authors state no conflict of interest.
Informed consent: Informed consent was obtained from all individuals included in this study.
Ethical approval: Not applicable.

References

1. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, et al. Prevalence of Parkinson's disease across North America. Npj Parkinson's Dis 2018;4:1–7.
2. Ragothaman M, Murgod UA, Gururaj G, Kumaraswamy SD, Muthane U. Lower risk of Parkinson's disease in an admixed population of European and Indian origins. Mov Disord 2003;18: 912–4.
3. Gourie-Devi M, Gururaj G, Satishchandra P, Subbakrishna DK. Prevalence of neurological disorders in Bangalore, India: a community-based study with a comparison between urban and rural areas. Neuroepidemiology 2003;23:261–8.
4. Surathi P, Jhunjhunwala K, Yadav R, Pal PK. Research in Parkinson's disease in India: a review. Ann Indian Acad Neurol 2016;19:9–20.
5. Razdan S, Kaul RL, Motta A, Kaul S, Bhatt RK. Prevalence and pattern of major neurological disorders in rural Kashmir (India) in 1986. Neuroepidemiology 1994;13:113–9.
6. Braak H, Ghebremedhin E, Rüb U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res 2004;318:121–34.
7. Lewis SJG, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. J Neurol Neurosurg Psychiatr 2005;76:343–8.
8. Brooks DJ. Imaging approaches to Parkinson disease. J Nucl Med 2010;51:596–609.
9. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. JAMA 2020;323:548–60.
10. Pedrosa DJ, Timmermann L. Management of Parkinson's disease. Neuropsychiatric Dis Treat 2013;9:321–40.
11. Oh SL, Hagiwara Y, Raghavendra U, Yuvaraj R, Arunkumar N, Murugappan M, et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Comput Appl 2020;32: 10927–33.
12. Pereira CR, Weber SA, Hook C, Rosa GH, Papa JP. Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In: 29th SIBGRAPI conference on graphics, patterns and images. Sao Paulo, Brazil; 2016.
13. Vásquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, Eskofier B, Klucken J, Nöth E. Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE J Biomed Health Inf 2018;23:1618–30.
14. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221–48.
15. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
16. Kuresan H, Samiappan D, Ghosh S, Gupta AS. Early diagnosis of Parkinson's disease based on non-motor symptoms: a descriptive and factor analysis. J Ambient Intell Humaniz Comput 2021 Mar 1. https://doi.org/10.1007/s12652-021-02944-0 [Epub ahead of print].
17. Yadav S, Singh MK. Hybrid machine learning classifier and ensemble techniques to detect Parkinson's disease patients. SN Comput Sci 2021;2:1–10.
18. Sahu B, Mohanty SN. CMBA-SVM: a clinical approach for Parkinson disease diagnosis. Int J Inf Technol 2021;13:647–55.
19. Pramanik M, Pradhan R, Nandy P, Bhoi AK, Barsocchi P. Machine learning methods with decision forests for Parkinson's detection. Appl Sci 2021;11:581.
20. Anudeep P, Mourya P, Anandhi T. Parkinson's disease detection using machine learning techniques. In: Advances in electronics, communication and computing. Singapore: Springer; 2021.
21. Senturk ZK. Early diagnosis of Parkinson's disease using machine learning algorithms. Med Hypotheses 2020;138: 109603.
22. Gunduz H. Deep learning-based Parkinson's disease classification using vocal feature sets. IEEE Access 2019;7:115540–51.
23. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inf 2013; 17:828–34.
24. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40:16–28.
25. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
26. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: 2015 38th international convention on information and communication technology, electronics and microelectronics. Opatija, Croatia; 2015.
27. Fonti V, Belitser E. Feature selection using lasso. VU Amst Res Pap Bus Anal 2017;30:1–25.
28. Paul S, Drineas P. Feature selection for ridge regression with provable guarantees. Neural Comput 2016;28:716–42.
29. Karthik S, Sudha M. A regularization-based feature scoring criterion on candidate genetic marker selection of sporadic motor neuron disease. In: Intelligent data engineering and analytics. Singapore: Springer; 2021.
30. Sekaran K, Sudha M. Predicting autism spectrum disorder from associative genetic markers of phenotypic groups using machine learning. J Ambient Intell Humaniz Comput 2020;12: 3257–70.
31. Cunningham P, Cord M, Delany SJ. Supervised learning. In: Machine learning techniques for multimedia. Berlin, Heidelberg: Springer; 2008.
32. Karthik S, Sudha M. A survey on machine learning approaches in gene expression classification in modelling computational diagnostic system for complex diseases. Int J Eng Adv Technol 2018;8:182–91.

33. Karthik S, Sudha M. Diagnostic gene biomarker selection for Alzheimer's classification using machine learning. Int J Innovative Technol Explor Eng 2019;8:2348–52.
34. Karthik S, Perumal RS, Mouli PC. Breast cancer classification using deep neural networks. In: Knowledge computing and its applications. Singapore: Springer; 2018.
35. Sekaran K, Sudha M. Prediction of lipopolysaccharides simulation responsiveness on gene expression profiles of major depression disorder affected cases using machine learning. Int J Sci Technol Res 2019;8:21–4.
36. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
37. Raschka S. Python machine learning. Birmingham: Packt Publishing Ltd; 2015.
38. Sekaran K, Sudha M. Predicting drug responsiveness with deep learning from the effects on gene expression of obsessive–compulsive disorder affected cases. Comput Commun 2020;151: 386–94.
39. Karthik S, Sudha M. Predicting bipolar disorder based non-overlapping genetic phenotypes using deep neural network. Evol Intell 2021;14:619–34.
40. Kamalakannan K, Anandharaj G. Stacked autoencoder based feature compression for optimal classification of Parkinson disease from vocal feature vectors using immune algorithms. Int J Adv Comput Sci Appl 2021;12: 470–6.
41. Kamalakannan K, Anandharaj DG. Deep feature selection from the vocal features for effective classification of Parkinson's disease. Int J Adv Sci Technol 2020;29:1661–72.