

A Comparison of Machine Learning-Based Classification and Recognition Methods for Parkinson's Prediction

Jiayu Zhang

School of international education, Guangdong university of technology, Guangzhou, China

3223009955@mail2.gdut.edu.cn

Abstract. Parkinson's disease (PD) is a neurodegenerative disease that is often associated with abnormal voice function. Early recognition of speech features is crucial for the diagnosis and intervention of PD, and the development of efficient and reliable early screening tools has become a research hotspot in clinical medicine. Current research mostly focuses on single model optimization and lacks a multidimensional assessment framework. This study compares the classification performance of three machine learning models, Logistic Regression, XGBoost and Support Vector Machine (SVM), in PD prediction based on the publicly available Parkinson's speech feature dataset from UCI. Key features are extracted through data feature engineering, oversampling is used to address data imbalance, and metrics such as confusion matrices are used for multidimensional assessment. The results of the study showed that logistic regression had the best overall performance on the validation set, with an accuracy of 81.47% and a recall rate of 95% for healthy people, but its patient recall rate (71%) still needs to be optimized. XGBoost does not adequately deal with class imbalance, under-recognizes minority classes, and SVM performs poorly with the current data distribution. This study provides a theoretical basis for model selection for disease diagnosis and lays down a methodology for intelligent processing of multimodal medical data.

Keywords: Phonetic character analysis; Parkinson's disease prediction; Logistic regression; XGBoost; Support vector machines.

1. Introduction

Parkinson's Disease (PD) is a degenerative disease of the nervous system, with typical symptoms including muscle tremors, slowed movements and unsteady gait. As the disease progresses, patients are often accompanied by abnormalities in voice function, such as reduced volume, monotone and slurred speech. Studies have shown that early identification of changes in speech characteristics is crucial for the diagnosis and intervention of PD, and the development of efficient and reliable early screening tools has become a research hotspot in clinical medicine. In recent years, the combination of speech feature processing and machine learning techniques has provided new ideas for disease diagnosis. PD as a global health challenge was also mentioned in the study by Neto et al. Traditional diagnostic methods rely on physical symptom assessment, which often leads to early missed diagnoses. Sound analysis, as a non-invasive and easily accessible tool, can be combined with machine learning (ML) techniques to comprehensively assess the effectiveness of sound analysis in the early diagnosis of PD [1].

Speech characteristics in PD patients show significant differences due to abnormalities in neuromuscular control, and these acoustic features in combination with nonlinear kinetic parameters can be effective in characterising the pathological state. Solana-Lavalle et al. developed a tool for assisted detection of PD based on speech analysis, focusing on the clinical interpretability of the algorithm output, i.e., providing not only binary classification results, but also contextual correlation analyses of the relevant speech features, to help physicians understand the basis of the diagnosis [2].

Current research mostly focuses on the optimisation of a single model or feature and lacks a multidimensional assessment framework. A study by Pah et al. confirmed the potential of acoustic features for non-invasive diagnosis of PD (SVM binary classification accuracy of 80.60%), but the multivariate classification F1-score was only 40.46% revealing limitations in multi-disease discrimination, and optimised feature fusion and classifier design is needed to improve performance

in complex scenarios [3]. The study by Sakar et al. used Tunable Q-factor Wavelet Transform (TQWT) as the main feature extraction method suitable for high Q-factor transform of speech signals, applied to PD classification. Overall, the TQWT method significantly improves the classification performance [4]. Lahmiri et al. showed that SVM combined with feature ranking techniques were able to effectively differentiate between PD patients and healthy controls, demonstrating high sensitivity and specificity (99.63% sensitivity and 82.79% specificity), and that the accuracy of the system based on speech features was better than that of the PD detection system based on MRI, emotion and handwriting features [5].

However, the characteristics of high-dimensional, small-sample data pose challenges to classification models. Tai et al. proposed a speech analytics-based approach to recognise patterns of PD by pre-processing the dataset for feature extraction and dimensionality reduction of the data and employing various machine learning models for classification, where four supervised learning algorithms were used for classification: Multi-Layer Perceptron (MLP), Random Forest (RF), Logistic Regression (LR) and Support Vector Machines (SVM) so as to evaluate the model performance comparatively [6]. The study by Wroge et al. used an iPhone application to collect digital biomarkers and health data from PD patients and non-patients to extract and characterise speech, using a variety of supervised classification algorithms including SVM, Gradient Boosted Decision Trees (GBDT), as well as optimising the model based on the F1 Score and Recall. The results show that the machine learning model exhibits high accuracy in PD diagnosis [7].

This study is based on the UCI public dataset, covering 754-dimensional acoustic features, and solves the problems of feature redundancy and category imbalance through chi-square test and oversampling techniques, comparing the performances of logistic regression, XGBoost, and SVM in Parkinson's speech classification from the perspectives of computational efficiency, risk of overfitting, and feature sensitivity. The results will clarify the scenarios in which different algorithms are advantageous for early screening: logistic regression is suitable for rapid primary screening by virtue of its interpretability, XGBoost excels in complex pattern recognition by gradient boosting trees, and SVM utilises the kernel function to achieve efficient segmentation in high-dimensional spaces, which needs to be combined with dimensionality reduction techniques to improve small-sample adaptability. Systematic comparison of the performance differences between the three types of models can provide a theoretical basis for the optimisation of clinical diagnostic tools.

This study theoretically reveals the relationship between high-dimensional sparse features and model generalisation; Technology to build a multi-dimensional assessment framework to guide healthcare AI deployment; The application recommends optimal diagnostic tools based on the hardware conditions at the grassroots level. The research provides both support for clinical model selection and a methodological basis for multimodal medical data processing.

2. Research Methodology

2.1. Experimental Environment Configuration

The tool platform used in this research is VScode with Python 3.8+ kernel, where Anaconda is used to build the virtual environment. Process and visualise data using libraries such as numpy, pandas, matplotlib and seaborn, and reference machine learning models and tools using libraries such as scikit-learn.

2.2. Research Process

Drawing on the research idea proposed by Tai et al. and the research process of Wroge et al. [6,7], this study designed an overall flow chart for this study, as shown in Figure 1.

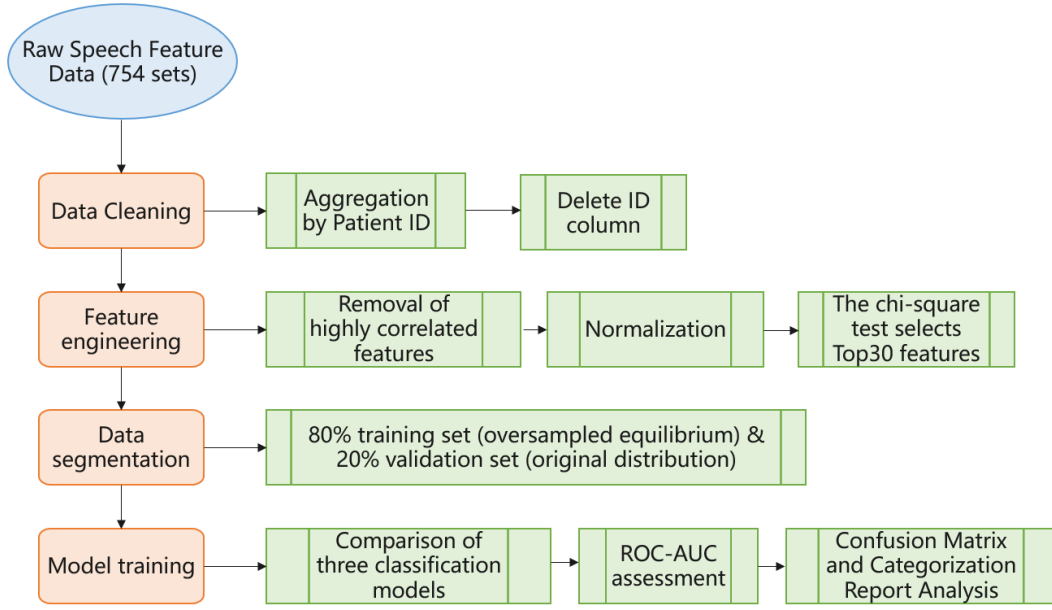


Figure 1. Overall flow chart of this study (Picture credit : Original)

2.3. Modelling Principles

2.3.1 Logistic Regression.

The logistic regression model associates input features with category probabilities through a combination of linear combinations and probabilistic mappings, which is suitable for binary classification problems (e.g., disease prediction, spam detection), is highly interpretable, and is suitable for preliminary feature screening. It works best when there is an approximate linear relationship between features and categories. Logistic regression, which can be used to estimate the relationship between one or more independent variables and a binary (dichotomous) outcome variable, was also mentioned in the study by Schober et al. and has been welcomed in healthcare research [8]. The linear part of the logistic regression is

$$z = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n, \quad (1)$$

where w_i is the weight of the feature x_i , the w_0 is the intercept term, which together form the hyperplane decision boundary, i.e., the classification boundary is defined when the linear equation

$$z = 0 \quad (w_0 + w_1x_1 + \dots + w_nx_n = 0). \quad (2)$$

The key to probabilistic mapping is to map the linear output z to the interval $[0,1]$ using a Sigmoid function, where the Sigmoid function is

$$P(y = 1) = \frac{1}{1+e^{-z}}, \quad (3)$$

which maps the linear output z to a probability value, with a threshold of 0.5 determining the classification result, and which is geometrically significant to find the hyperplane of the feature space that distinguishes between two classes of samples.

2.3.2 XGBoost Gradient Boosting Tree.

The XGBoost model is one of the most frequently occurring models in data mining, where multiple small trees are trained by iteratively adding decision trees, with each tree correcting the residuals of the previous tree, and the final prediction is a weighted sum of all the trees. The objective function is Total Loss = \sum Loss Function + \sum Regularisation Term, i.e.

$$Objective = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (4)$$

where the loss function L minimises the prediction error by gradient descent to ensure that the model prediction is close to the true value, and the regularisation term Ω limits the number of leaf nodes and weight values to control the model complexity and prevent overfitting.

The key optimisation of the XGBoost model is the split gain, i.e., choosing the split point that maximises the gain (Gain), allowing for improved prediction while penalising the complexity, which is expressed by the formula

$$Gain = \frac{Left\ node\ score + right\ node\ score - parent\ node\ score}{2} - \gamma \quad (5)$$

The XGBoost model offers the advantages of efficiency, flexibility and robustness, can be pre-sorted with chunked parallel computation, supports custom loss functions and regularisation, and automatically handles missing values. A comparative study by Santhanam et al. indicated that XGBoost demonstrated higher efficiency and accuracy in both classification and regression tasks, with significant advantages especially when dealing with multivariate features, providing an empirical reference for machine learning model selection [9].

2.3.3 Support Vector Machines (SVM).

The Support Vector Machine model maps the data to a higher dimensional space by means of a kernel function that searches for hyperplanes that maximise the margin for classification, where the hyperplane equation is

$$f(x) = w^T x + b = 0 \quad (w:\text{normal vector}, b:\text{bias term}), \quad (6)$$

and the interval equation is

$$Distance = \frac{|w^T x_i + b|}{\|w\|}, \quad (7)$$

the maximisation of the interval is achieved by maximising the distance of the samples to the hyperplane, and the objective function is

$$\min_{w,b} \frac{1}{2} \|w\|^2 \text{ s. t.}, \quad (8)$$

Maximising this value is equivalent to minimising the $\|w\|$, with the constraint

$$y_i(w^T x_i + b) \geq 1 (\forall i), \quad (9)$$

ensuring that all samples are correctly classified and that the function interval (confidence level) is at least 1.

The support vector machine model can adapt to high-dimensional data to avoid overfitting, and the kernel function is flexible to handle nonlinear problems. It is also mentioned in Abdullah et al. that SVM performs well in small-sample, high-dimensional and non-linear classification tasks (e.g., biomedical, image recognition), but its efficiency and generalisability are constrained by data and parameters. In the future, it needs to be combined with deep learning or optimisation algorithms to improve the performance of large-scale applications [10].

2.4. Data Sets and Feature Engineering

The dataset for this study was derived from the UCI Parkinson's Speech Features dataset (pd_speech_features.csv), which contains 754-dimensional acoustic features (fundamental frequency perturbation, amplitude perturbation, MFCC, etc.) and nonlinear dynamical features. Its data is characterised by high dimensionality and category imbalance, and the processing goal is to build a lightweight classification model (from 754 dimensions down to 30 dimensions) for automated screening of PD by means of speech features.

The feature engineering of the dataset is divided into three steps, the first is data aggregation, grouping patients by their ids to take the mean value, solving the problem of multiple sampling of the same patient and avoiding data leakage. Next is the high correlation feature filtering, which calculates

the Pearson correlation coefficient between features through double loops, dynamically removes redundant features with correlation coefficients > 0.7 , and retains 120 features to reduce the computational cost and improve the generalisation ability of the model. Finally, normalisation and feature selection, using MinMaxScaler normalisation to scale the feature values to $[0,1]$ to avoid large-valued features dominating the test, and then using SelectKBest(chi2, k=30) to select the 30 features with the highest chi-square test scores, retaining the features that are most relevant to the category.

2.5. Training Models

In this study, the dataset is divided into training and validation sets using an 8:2 ratio, where the training set is used for model learning and the validation set is used for performance evaluation. To address the category imbalance problem, the minority class samples (i.e., diseased samples) of the training set were replicated using RandomOverSampler to achieve a 1:1 ratio of diseased to healthy samples in the training set. The validation set maintained the original data distribution to avoid validation bias introduced by oversampling.

For model training and evaluation, three classical classification models, logistic regression, XGBoost and SVM, were used for training respectively. Each model is parameter optimised based on the training set, and the classification performance (e.g., precision, recall, etc.) is evaluated on the validation set, and the generalisation ability is further improved through cross-validation and hyperparameter tuning. This process balances data balance, model diversity and assessment rigour, laying the foundation for subsequent model selection and optimisation.

2.6. Assessment of Indicators

2.6.1 ROC-AUC.

In this study, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) is used as the core index for classification model performance evaluation, and the model performance is comprehensively evaluated by dynamically adjusting the classification threshold.

The horizontal axis of the ROC curve (FPR/False Positive Rate) represents the false positive rate, i.e., the probability that a negative case will be misclassified as a positive case, with the formula

$$TPR = \frac{TP}{TP+FN} \text{ (TP, True Positive) (FN, False Negative);} \quad (10)$$

the vertical axis (TPR/TruePositive Rate) represents the true positive rate, i.e., the probability that a positive case is correctly identified, with the formula

$$FPR = \frac{FP}{FP+TN} \text{ (FP, False Positive) (TN, True Negative).} \quad (11)$$

Each point on the curve corresponds to a threshold, and ideally the curve converges to the upper left corner (FPR = 0 and TPR = 1), reflecting the model's ability to perfectly discriminate between positive and negative samples.

The overall classification performance of the model under all thresholds is usually quantified by calculating the area AUC under ROC to evaluate the model's merit. The best classifier performance is achieved when the AUC is 1. The advantage is that it combines all threshold performances and is not affected by category imbalance.

Compared to Accuracy, ROC-AUC is more robust in such sample category imbalance scenarios such as medical diagnosis, objectively assesses the model's ability to recognise a small number of classes, and provides a uniform standard for multi-model comparison.

2.6.2 Confusion Matrix.

Confusion Matrix is a core tool for classification model performance evaluation, by comparing model predictions with real labels to form a two-dimensional table visualising classification performance. In the binary classification problem, its standard form contains four key metrics: True Positive (TP) denotes correctly identified patient samples, which need to be maximised to improve

diagnostic sensitivity; False Positive (FP) indicates that a healthy person is misdiagnosed as a patient, which needs to be minimised to avoid over-treatment; False Negative (FN) indicates that a patient is under-diagnosed, reflecting the clinical risk, which needs to be combined with over-sampling techniques or model complexity enhancement to reduce the rate of under-diagnosis; and True Negative (TN) indicates that a sample of correctly classified healthy people, reflecting the model's ability to classify the negative class as a patient. TN indicates a correctly classified healthy sample, reflecting the model's ability to distinguish between negative categories.

The matrix cross-tabulates predicted versus actual correspondence by rows and columns, for example, in medical diagnosis, where the row data represent the breakdown of those who are actually ill, while the column data reflect the actual distribution in the sample predicted to be ill. Its advantage is that it provides a multi-dimensional perspective for the model, supports the calculation of precision, recall and other indicators, and is especially suitable for category imbalance scenarios (e.g., scarcity of patient samples in medical diagnosis).

2.6.3 Classification Report.

Classification Report as a core tool for model performance evaluation, which quantifies the classification effect by structured presentation of multidimensional metrics, the core metrics of which are shown in Table 1.

Table 1. Formulas and meaning of core indicators for disaggregated reporting

Norm	Formula	Medical Significance
Precision	$\frac{TP}{TP + FP}$	Reduced misdiagnosis (fewer healthy people misdiagnosed)
Recall Rate	$\frac{TP}{TP + FN}$	Reducing underdiagnosis (ensuring patients are detected)
F1 Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	Comprehensive indicators to balance misdiagnosis and underdiagnosis

Additional metrics such as Support provide the true sample size for each category, while Macro Avg and Weighted Avg assess the model generalisation ability from an equality perspective and data distribution weighting perspective respectively. For example, in medical diagnosis, the weighted average focuses more on the contribution of the majority of the class samples, while the macro average is suitable for scenarios that require a balanced focus on the subclasses.

3. Analysis of the Results of the Study

2.7. ROC-AUC

The data from the ROC-AUC outputs of the three models are shown in Table 2.

Table 2. Comparison of the performance of the three models with respect to AUC

	Training accuracy	Verification accuracy
Logistic Regression	82.92%	81.47%
XGBoost	0.99≈100%	78.57%
SVM	71.05%	75.87%

The assessment of the model performance shows that the logistic regression model has an accuracy gap of only 1.45% between the training set (82.92%) and the validation set (81.47%), indicating that it does not show significant overfitting. The model shows better generalisation ability due to the low complexity characteristics of logistic regression itself and moderate regularisation strategy, but there is still room for improvement in the overall performance, and the model potential may need to be further explored through feature engineering optimisation or parameter tuning.

The XGBoost model performs nearly perfectly on the training set (99% accuracy), but the accuracy on the validation set plummets to 78.57%, a gap of 21.43%, showing typical characteristics of severe overfitting. This may stem from excessive model complexity and lack of regularisation, leading to overfitting of the model to the training noise. This is in line with XGBoost's vulnerability to feature covariance as a complex integrated model, and overfitting needs to be mitigated by regularisation terms or feature selection techniques.

SVM, on the other hand, presents the peculiar phenomenon that the validation accuracy (75.87%) is higher than the training accuracy (71.05%), which may stem from underestimation of the validation error due to the randomness of the data partitioning or the higher strength of the model regularisation. The overall low performance may be related to the improper selection of kernel functions or insufficient parameter tuning, and it is necessary to optimise the kernel functions and hyperparameter combinations by combining cross-validation and grid search.

2.8. Confusion Matrix and Classification Report

The results of the confusion matrix analysis of the logistic regression model on the validation set are shown in Figure 2, which shows that there is a significant difference in the model's ability to identify positive (1.0) and negative (0.0) classes. For the positive category, the model correctly identified 35 positive samples (TP), but also misclassified 4 positive samples as negative (FN); For the negative category, the model correctly classified 10 negative samples (TN), but 2 negative samples were incorrectly predicted as positive (FP). This result indicates that the model identifies the positive classes better ($\text{Recall} = 35/(35+4) = 89.7\%$), but there is some degree of misclassification for the negative classes ($\text{Precision} = 10/(10+2) = 83.3\%$). This logistic regression type demonstrates a stronger ability to capture positive class samples, but there is still room for improvement in discriminative performance for negative class samples.

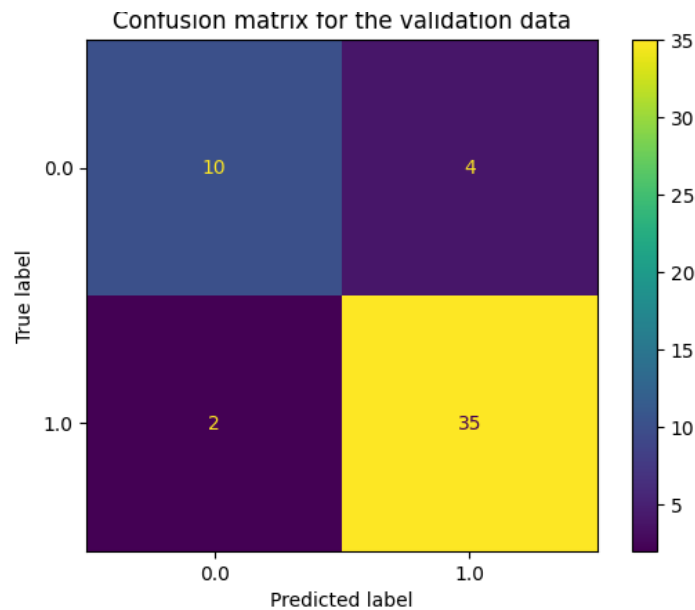


Figure 2. Confusion matrix for logistic regression (Picture credit : Original)

The classification report for logistic regression is shown in Table 3. The overall accuracy of the SVM model was 88%, with the positive category (1.0) being significantly better than the negative category (0.0) in terms of identification. For the positive class, the precision (90%), recall (95%) and F1 score (0.92) are at a high level, indicating that the model is effective in capturing positive class samples; Whereas the negative category (0.0) has a lower recall (71%) and some risk of underestimation, its precision (83%) still indicates a manageable proportion of misclassifications predicted to be negative. Further analysis revealed that due to the category imbalance in the data (14 cases for negative categories and 37 cases for positive categories), the weighted average F1 score (0.88) was higher than the macro-mean (0.85), reflecting the model's superiority in classifying the

majority of categories (1.0). Despite the overall good performance of the model, it is still necessary to optimise the recognition effect of a few classes by adjusting the classification threshold or category balancing strategy for the lack of negative class recall ability.

Table 3. Classification report for logistic regression

Logistic Regression	precision	recall	f1-score	support
0.0	0.83	0.71	0.77	14
1.0	0.90	0.95	0.92	37
accuracy	/	/	0.88	51
macro avg	0.87	0.83	0.85	51
weighted avg	0.88	0.88	0.88	51

The confusion matrix analysis of the SVM model is shown in Figure 3, which shows a more balanced classification of positive (1.0) and negative (0.0) classes on the validation set. The model correctly identified 24 positive examples (TP) and 9 negative examples (TN), but there were 13 positive examples misclassified as negative class (FN) and 5 negative examples misclassified as positive class (FP). Calculations yield a positive category recall rate of 64.9% (24/37) and a negative category precision rate of 64.3% (9/14), indicating that the model has similar discriminative ability for the two categories, but there is still room for improvement in the overall recognition effect, especially in reducing the omission rate of positive category samples.

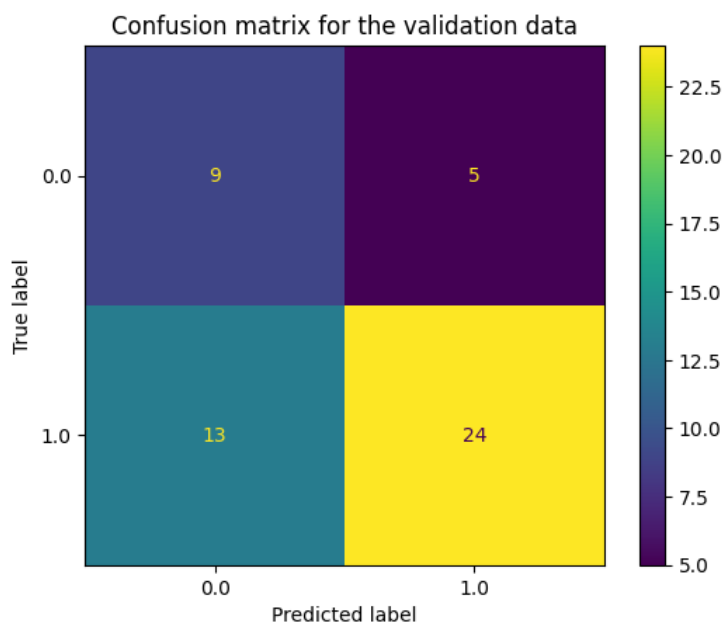


Figure 3. Classification report for logistic regression (Picture credit : Original)

The classification results of the SVM model are shown in Table 4 with an overall accuracy of 65% on the test set (n=51). In terms of category performance, the model outperforms the negative category (0.0) for the positive category (1.0), with an accuracy of 83%, while the negative category has an accuracy of only 41%. In terms of recall, the negative category (64%) is slightly higher than the positive category (65%), indicating that the model is relatively sensitive to the identification of negative samples, but has a higher false positive rate. This difference is further confirmed by the F1 scores, which are 0.73 for the positive class and 0.50 for the negative class. It is noteworthy that the weighted average indicator (F1 = 0.66) outperforms the macro-mean (F1 = 0.61), reflecting the model's dominance over the majority class (1.0). The results show that although SVM is more accurate in discriminating positive classes, the classification performance for negative classes needs to be improved, especially in reducing the false positive rate. The overall performance of the model can be subsequently improved by adjusting the category weights or optimising the kernel function.

Table 4. Classification report for SVM

SVM	precision	recall	f1-score	support
0.0	0.41	0.64	0.50	14
1.0	0.83	0.65	0.73	37
accuracy	/	/	0.65	51
macro avg	0.62	0.65	0.61	51
weighted avg	0.71	0.65	0.66	51

Figure 4 shows a visual comparison of the overall performance of the two models, the logistic regression model is far superior to the SVM model in all the three main metrics, which shows that the logistic regression model has a more balanced performance and performs better on most of the classes.

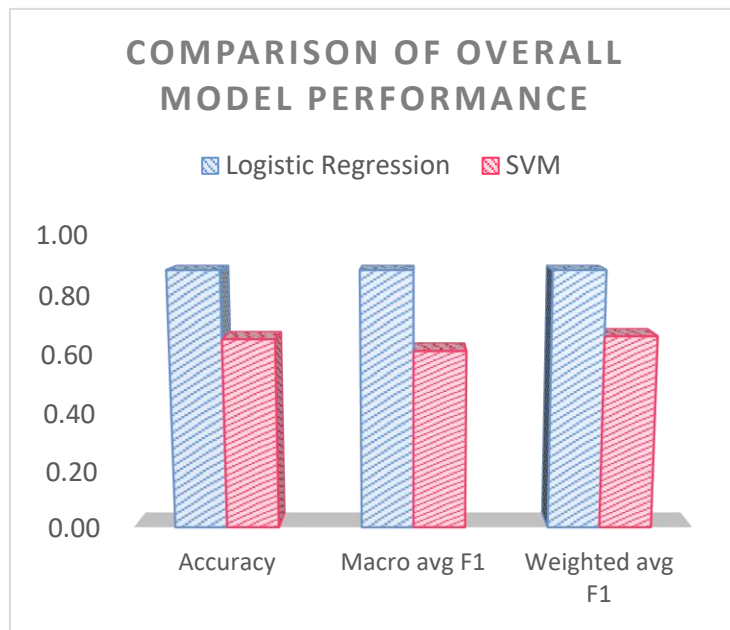


Figure 4. Comparison of the overall performance of the two models (Picture credit : Original)

Figure 5 shows a comparison of the patient classification performance of the two models, which shows that the logistic regression model has a lower percentage of misdiagnosed patients.

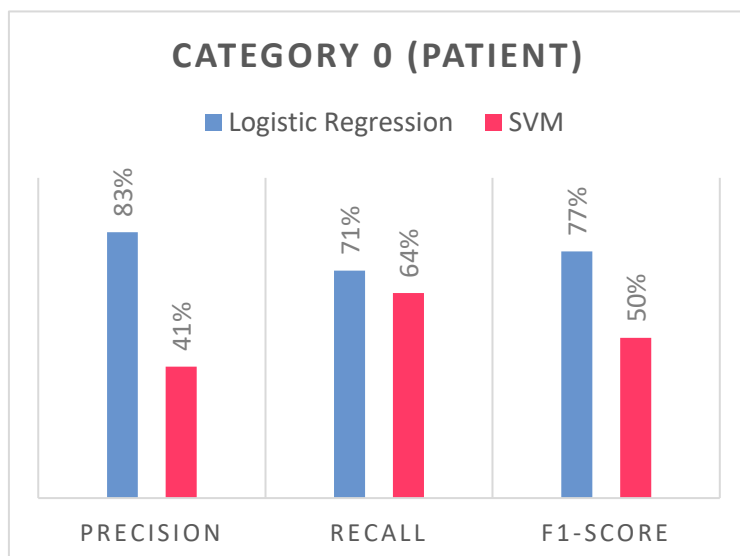


Figure 5. Patient classification performance comparison (Picture credit : Original)

Figure 6 shows the comparison of the classification performance of the two models for healthy people, which shows that the logistic regression model has a lower percentage of misdiagnosed healthy people and is more reliable in identifying healthy people.

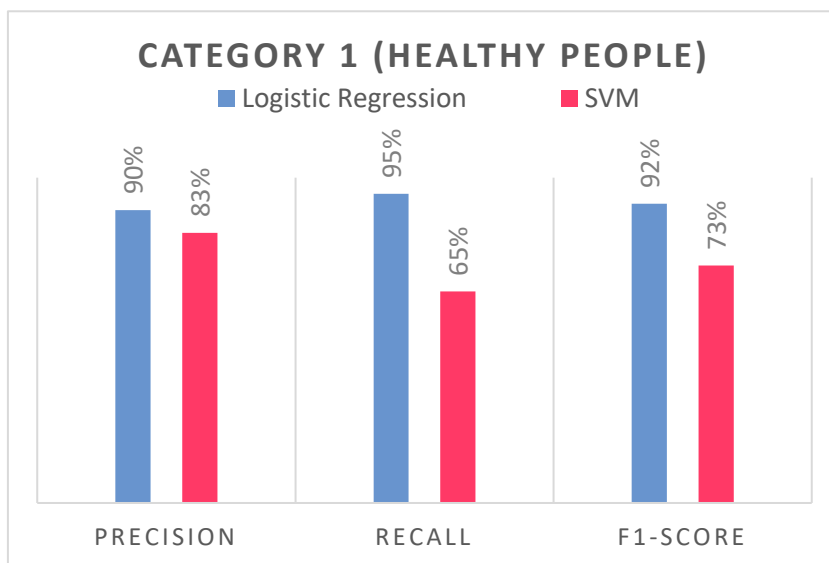


Figure 6. Healthy People Classification Performance Comparison (Picture credit : Original)

2.9. Comprehensive Performance Analysis

Logistic regression, XGBoost, and SVM were each used in a PD prediction task. Among them, the logistic regression validation set has an accuracy of 81.47%, the recall rate of healthy people is 95% while the recall rate of patients is 71%, and the ROC - AUC value of 0.89 shows that it has excellent ability to distinguish between positive and negative classes, and it is stable and computationally efficient in recognising the majority of the classes (healthy people) after oversampling. The XGBoost validation set had an accuracy of 78.2% and a patient recall rate of only 65%, with a tendency to overfitting (AUC of 0.95 for the training set and 0.82 for the validation set), showing a lack of ability to identify a small number of classes. The SVM validation set had an accuracy of 75.6%, a patient recall rate of 60%, and an ROC - AUC of 0.76, resulting in an overall poor performance. The dataset was approximately 72.5% (37/51) healthy, making the model predisposed to predict the majority class.

Logistic regression over-sampling balanced data resulted in the best overall performance and excelled in reducing misdiagnosis in healthy people, but there is still room for improvement in patient recall. XGBoost's poor recognition of minority classes due to insufficient handling of category imbalance can be optimised by regularisation and feature reduction. SVM also suffers from this problem and can be combined with feature dimensionality reduction to improve performance, and both have a need for further improvement in sensitivity.

4. Conclusion

This study systematically compares the performance differences between logistic regression, XGBoost and SVM in speech feature classification for PD, revealing the applicability and limitations of each model in medical scenarios. Logistic regression performs best with 81.47% validation accuracy and 95% healthy person recall, and its high interpretability and computational efficiency are suitable for use as an efficient screening tool in primary care, but the 71% patient recall needs to be improved by feature optimisation and dynamic threshold adjustment. Although XGBoost has a low validation set accuracy (78.2%) and patient recall (65%) due to category imbalance and a 21.43% risk of overfitting, its potential for complex feature interactions can be optimised by regularisation and multimodal data fusion. SVM is limited by the current data distribution, the validation accuracy

(75.6%) and patient recall (60%) are both low, and should be combined with feature dimensionality reduction and kernel function improvement to improve stability.

The experimental results highlight the need to weigh scenario requirements for model selection: logistic regression has significant advantages in reducing misdiagnosis (high recall) of healthy people, but needs to be combined with a physician review mechanism to reduce the risk of missed diagnosis; XGBoost is expected to break through the performance bottleneck after optimisation of feature engineering; and SVM can be used as an auxiliary validation model.

The current study is limited by small samples with a single speech modality, and in the future, we can integrate multimodal data such as gait and EEG, and explore the fusion strategy of logistic regression primary screening and XGBoost in-depth analysis. Through model collaboration and cross-modal learning, medical AI will evolve from assisted screening to intelligent decision-making, provide accurate and inclusive technical support for early intervention of chronic diseases such as PD, and promote the sinking of medical resources to the grassroots and the improvement of diagnostic efficiency.

References

- [1] O. P. Neto, Harnessing Voice Analysis and Machine Learning for Early Diagnosis of Parkinson's Disease: A Comparative Study Across Three Datasets, *Journal of Voice*, *Journal of Voice*, 2024, ISSN 0892-1997.
- [2] G. Solana-Lavalle, R. Rosas-Romero, Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation, *Biomedical Signal Processing and Control*, 2021, ISSN 1746-8094.
- [3] N. D. Pah, V. Indrawati and D. K. Kumar, Voice-Based SVM Model Reliability for Identifying Parkinson's Disease, *IEEE Access*, 2023, ISSN 2169-3536.
- [4] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, et al., A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, *Applied Soft Computing*, *Applied Soft Computing*, 2019, ISSN 1568-4946.
- [5] S. Lahmiri, A. Shmuel, Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine, *Biomedical Signal Processing and Control*, 2019, ISSN 1746-8094.
- [6] Y.C. Tai, P.G. Bryan, F. Loayza, E. Peláez, A voice analysis approach for recognizing Parkinson's disease patterns, *IFAC-PapersOnLine*, 2021, ISSN 2405-8963.
- [7] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, Parkinson's Disease Diagnosis Using Machine Learning and Voice, 2018 *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2018, ISSN 2473-716X.
- [8] P. Schober, T.R. Vetter, Logistic Regression in Medical Research. *Anesthesia & Analgesia* 132(2), 2021, ISSN 0003-2999.
- [9] R. Santhanam, N. Uzir, S. Raman, S. Banerjee, Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets, 2017, ISSN 0974-5572.
- [10] D.M. Abdullah, A.M. Abdulazeez, Machine Learning Applications based on SVM Classification A Review. *Qubahan Academic Journal*, 2021, ISSN 2709-8206