Chapter 3
Prediction of Parkinson's Disease Severity Based on Feature Optimization

Tao Zhang
https://orcid.org/0000-0003-0450-724X
Yanshan University, China

Zaifa Xue
Yanshan University, China

Zesheng Huo
Yanshan University, China

ABSTRACT

Voice disorders are one of the incipient symptoms of Parkinson's disease (PD). Most of the existing PD severity prediction methods are based on baseline features, and generally select features with high relevance for dimensionality reduction. To further improve the prediction performance, the LDFSF and GWT-RF-Att methods are proposed respectively from the perspective of feature selection and feature transformation. The LDFSF method utilizes dynamic feature selection strategy based on SOM clustering to select feature subsets with high correlation, low redundancy, and high complementarity from voice features. The GWT-RF-Att method uses graph wavelet transform to extract the more effective feature set based on the baseline features, and uses random forest improved by attention mechanism to improve the prediction performance of the model. The results on the Parkinson's telemonitoring dataset show that the performance of the two methods is better than that of the existing comparison methods, thus verifying their effectiveness.

DOI: 10.4018/979-8-3693-1582-8.ch003

INTRODUCTION

Parkinson's disease (PD) is a common degenerative nervous system disorder characterized by trembling, slow movement, postural balance disorder, and other motor symptoms. And PD patients will also be accompanied by symptoms such as voice disorders (Xue et al., 2022; Zhang et al., 2023a). Fortunately, voice disorders are one of the earliest symptoms of PD, which can generally be observed five years before clinical diagnosis (Zhang et al., 2023b). By analyzing the voice features of subjects, it is possible to use machine learning methods to analyze the severity of the diagnosed patients' conditions, which is of great significance for the early diagnosis and treatment of PD patients.

For the regression tasks for Parkinson's disease, using voice data for severity assessment of PD has shown promising results. The Unified Parkinson's Disease Rating Scale (UPDRS) is an internationally recognized measure for assessing the severity of Parkinson's disease symptoms (Yoon & Gaw, 2021). It consists of two components: motor-UPDRS and total-UPDRS, which assess motor symptoms and overall assessment of PD symptoms. In the past few decades, simple linear regression (LR), nonlinear classification regression tree, k-nearest neighbors (KNN) regression, and support vector regression (SVR) are commonly used traditional prediction methods for Parkinson's disease symptom severity due to their computational efficiency. However, these models may not achieve the desired prediction accuracy (Tsanas et al., 2010). In recent years, machine learning and neural network methods have gained popularity in UPDRS prediction. Mayo & Frank (2017) applied Bayesian regression and its modified version to improve prediction accuracy and generalization ability. Neural networks, such as feed-forward neural networks (Xiang et al., 2016) and adaptive network-based fuzzy inference systems (ANFIS) (Nilashi et al., 2017), have also been employed due to their ability to handle nonlinear relationships. However, neural networks require large datasets and careful parameter tuning, and may be prone to overfitting and lack of generalization ability (Hlavica et al., 2016).

To enhance the generalization ability of the PD severity prediction model, Tunc et al. (2020) proposed a combined approach of feature selection and extreme gradient boosting (XGBoost) based on decision trees (DT) to predict UPDRS scores in PD patients. Despotovic et al. (2020) used the automatic relevance determination (ARD) method to rank PD voice feature weights. In the study by Tsanas et al. (2021), random forest (RF), SVR, and XGBoost were used to evaluate PD severity, and RF exhibits superior prediction performance compared to SVR and XGBoost. RF has been found to possess higher effectiveness, stability, and robustness compared to artificial neural networks (Dong et al., 2021), making it a suitable choice for predicting UPDRS scores in PD patients. Furthermore, the selection of different input features for RF can impact the model's prediction performance (Liu et al., 2018).

However, the above method has the following two problems, which may increase the prediction error of the model. On the one hand, while these methods consider the correlation of the feature to the target value, relying solely on the property may be insufficient. The redundancy and complementarity between features are conducive to the selection of feature subsets that can improve the performance of the model. On the other hand, most of the existing methods for predicting the severity of Parkinson's disease are based on baseline features and the RF regressor, but the information of baseline features is limited, and some decision trees in RF have poor prediction results, which may affect the performance of the model.

To address the above two issues, this chapter proposes two methods to address the aforementioned issues in PD research. Firstly, the local dynamic feature selection fusion (LDFSF) method is introduced for feature selection based on relevance, redundancy, and complementarity in PD symptom severity prediction tasks. This method introduces the maximal information coefficient (MIC) as an effective