Karaelmas Science and Engineering Journal
Journal home page: https://dergipark.org.tr/tr/pub/karaelmasfen
DOI: 10.7212/karaelmasfen.1484222

Research Article
Received / Geliş tarihi : 15.05.2024
Accepted / Kabul tarihi : 31.05.2024

Parkinson's Disease Detection Via Machine Learning Using Data Splitting and Validation Methods

Veri Ayırma ve Doğrulama Yöntemleri Kullanılarak Makine Öğrenmesi Aracılığı ile Parkinson Hastalığı Tespiti

Mustafa Alptekin Engin*

Bayburt University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Bayburt, Türkiye

Abstract

Parkinson's disease (PD), a neurological disorder, negatively affects the lives of patients and their caregivers. PD, which is very difficult to diagnose early by examining the clinical characteristics of the person, can be diagnosed using voice recordings. However, the inconsistent performance results of the models obtained from the evaluation of voice recordings through machine learning techniques limit the usability of these models to aid in diagnosing physicians. This study used a database of 195 voice data obtained from 31 individuals, 23 of whom have PD. The classification of the voices as healthy or patient was based on the 22 features in the database. The split ratios 90/10, 80/20, 70/30, 50/50 and 30/70 were used to select the training and test phase data, respectively. In addition, each split ratio was evaluated using 10-fold cross-validation, 5-fold cross-validation, holdout validation and resubstitution validation methods in the training phase, which is the initial process that will directly affect the other classification procedures. In addition, the classification process was performed using quadratic discriminant analysis, support vector machine, ensemble bagged tree, k-nearest neighbours and neural network classifiers. All procedures were repeated 10 times to ensure consistency of results and randomisation of split ratios. As a result, the k-nearest neighbours classifier with 80/20 splitting ratio and 10-fold cross-validation was determined to be the most successful among the compared methods with 95.64±3.21% accuracy. Therefore, it can be seen that much more successful results can be obtained by analysing only the effects of the existing parameters of the classifiers.

Keywords: Classification, cross-validation, machine learning, repeated train/test splitting

Öz

Nörolojik bir bozukluk olan Parkinson hastalığı (PH), hastaların ve bakımlarından sorumlu kişilerin hayatlarını olumsuz olarak etkilemektedir. Kişinin klinik özelliklerinin incelenmesi ile erken tanısı oldukça zor olan PH, konuşma ses kayıtları kullanılarak teşhis edilebilmektedir. Fakat ses kayıtlarının makine öğrenmesi teknikleri aracılığı ile değerlendirilmesinden elde edilen modellerin tutarsız performans sonuçları, bu modellerin hekimlerin teşhis koymasında yardımcı olarak kullanılabilirliğini sınırlamaktadır. Yapılan çalışmada 23'ü Parkinson hastası olan toplam 31 kişiden elde edilen ve 195 ses verisinden oluşan bir veri tabanı kullanılmıştır. Veri tabanındaki her bir konuşma sesinden elde edilen 22 adet öznitelik ile bu seslerin makine öğrenmesi aracılığıyla hasta ve sağlıklı olarak sınıflandırılması gerçekleştirilmiştir. Bu sınıflandırma işleminde eğitim ve test aşamasında kullanılacak verilerin rastgele olarak sırası ile 90/10, 80/20, 70/30, 50/50 ve 30/70 olmak üzere farklı oranlarda bölünmesi sağlanmıştır. Ayrıca her bir ayırma oranı, eğitim aşamasında 10 katmanlı çapraz doğrulama, 5 katmanlı çapraz doğrulama, ayırarak doğrulama ve yeniden ikame doğrulaması yöntemleri kullanılarak değerlendirilmiştir. Bununla beraber kuadratik diskriminant, destek vektör makineleri, toplu torbalı ağaç, k-en yakın komşuluk ve sinir ağları sınıflandırıcıları kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Veri ayırmadaki rastgeleliğin ve tutarlı sonuçların elde edilmesi için tüm işlemler 10 defa tekrar edilmiştir. Yöntemlerin başarımlarının karşılaştırılmasında doğruluk, duyarlılık, özgüllük, kesinlik ve F1 skoru metrikleri aracılığı ile sonuçların ortalama ve standart sapma değerleri hesaplanmıştır. Sonuç olarak 80/20 ayırma oranı ve 10 katmanlı çapraz doğrulama kullanan k-en yakın komşuluk sınıflandırıcısına ait %95.64±3.21 test doğruluğu değeri, karşılaştırılan yöntemler içerisinde en başarılı yöntem olarak tespit edilmiştir. Dolayısıyla sadece sınıflandırıcılara ait mevcut parametrelerin etkileri analiz edilerek çok daha başarılı sonuçların elde edilebileceği görülmüştür.

Anahtar Kelimeler: Çapraz doğrulama, makine öğrenmesi, sınıflandırma, tekrarlı eğitim/test ayrımı

*Corresponding author: maengin@bayburt.edu.tr

Mustafa Alptekin Engin orcid.org/0000-0003-3399-9343

1. Introduction

Parkinson's disease (PD), a neurological disease named after James Parkinson, who first described its typical symptoms in the 1800s, is increasingly diagnosed, especially in individuals over a certain age (Huang et al. 2024). Although experts with 91% accuracy detect this disease in the first five years with the full application of detailed and challenging clinical diagnostic criteria, this correct detection rate decreases to 76% when examined by non-experts (Virameteekul et al. 2023). The fact that the disease occurs at an advanced age causes patients to face certain difficulties in the application of clinical diagnostic criteria and analyses. Therefore, there is a significant demand for low-cost methods that will help the patient overcome the difficulties in PD detection more easily (Esmer et al. 2020). It has been observed in the literature that a loss of control in motor activity is a common cause of a variety of voice and speech disorders in patients diagnosed with PD (Orozco-Arroyave et al. 2016). Speech disorders, which is one of the cognitive-motor skills, is a common diagnostic criterion seen in varying degrees in 90% of people with Parkinson's disease (Smith and Caplan 2018). Acoustic sound analyses and measurement methods may be useful biomarkers for diagnosing PD at an early stage of the disease, potentially enabling remote monitoring of patients. Furthermore, they may provide valuable feedback in sound therapy for clinicians or patients (Rusz et al. 2011). Especially in the last two decades, thanks to the development of software and hardware technologies, the idea of improving patients' quality of life by detecting PD based on machine learning at an early stage using speech sounds has come to the forefront of studies. There are many studies in the literature on the evaluation of dysphonic symptoms of PD with machine learning (Bang et al. 2023, Islam et al. 2024). In one of the pioneering studies, a kernel support vector machine and 50 replicates of bootstrap resampling methods were used to classify 195 speech data from 31 people, 23 of whom had Parkinson's disease, with 91.4% success (Little et al. 2009). In another classification study in which Parkinson's disease was evaluated with speech sounds using the probabilistic neural network method and the database was divided into 70% training and 30% testing phases, training accuracy was 81.74%, and testing accuracy was 81.28% (Ene 2008). In a study on the Neural Network Classification method, 65% of the database was used for training, and the remaining 35% was used for testing; 100% training and 92.9% testing accuracy values were achieved (Das 2010). In another study on the Adaptive Neuro-Fuzzy Classifier with the linguistic hedges method, the database was divided into training and testing by 50%, and the classification accuracy values of the model were calculated as 95.38% in the training phase and 94.72% in the testing phase (Çağlar et al. 2010). In another study, 85.03% classification accuracy was achieved using a similarity classifier and feature selection using fuzzy entropy criteria after splitting the data into training and testing by 50% (Luukka 2011). The Classification and Regression (C&R) Tree, Bayes Net and C5.0 are used to generate an ensemble method in a classification study on detecting Parkinson's patients and healthy subjects. An accuracy of 95.31% was achieved in this study, where training and test data were split by 70% and 30%, respectively (Inzamam-Ul-Hossain et al. 2015). Classification accuracy of 92.19% was achieved in a study using The Optimised Cuttlefish algorithm for early diagnosis of Parkinson's disease, where the dataset was divided into 70% training and 30% test data (Gupta et al. 2018). In the study using the Modified Grey Wolf Optimization method and Random Forest classifier, the data set was divided by 70% and 30% to be used in the training and testing phases, respectively, and a classification accuracy of 93.87% was calculated (Sharma et al. 2019). In a study, the classification accuracy of 93.84% was calculated with the support vector machine classifier, and the features were evaluated using the recursive feature elimination method"(Senturk 2020). Another classification process was performed using entropy-based feature selection using the k-nearest neighbour algorithm, feed-forward Extreme Learning Machine, and Fast Learning Machine methods for Parkinson's detection. The study obtained 80% classification accuracy using half of the features via the Fast Learning Machine method (Abdulateef et al. 2023). In a recent study, the % classification accuracy of 88.5% was obtained with the cascade forest (casForest) algorithm using deep ensemble transformers, a fast, scalable approach for dimensionality reduction problems (Nareklishvili and Geitle 2024). All these studies, which focus on different feature selection techniques or classification methods, have in common the use of the Oxford Parkinson's Disease Detection Dataset for PD detection (Little 2007). In these studies, a single and different ratio was utilised to divide the database into distinct segments designated for use in both the training and testing phases. The impact of this ratio on performance was not explicitly elucidated. Moreover, as the first step in a classification process, this separation can directly affect the following steps. In addition, many of the studies discussed above do not mention the validation technique used during the training phase. Conventionally, when developing a machine learning classification model, a large amount of training data is used in the training phase. The training validation with this large amount of data ensures that the quality and quantity of the data are adjusted, i.e. that the classification model performs well and, in particular, achieves reliable and consistent results. In addition to these deficiencies, conflicting performance results obtained using different classification models limit the clinical applicability of machine learning-based methods developed for PD detection (Iyer et al. 2023). The availability of a low-cost method with reliable results will both alleviate the clinical workload and help patients over a certain age group overcome the many physical challenges they will face during the diagnostic phase. Generally, there is no established method for finding the most appropriate model for the database to be used in a classification problem. In this respect, different classification algorithms should be evaluated with data separation and validation methods to find the optimal model. In the proposed study, the Oxford Parkinson's Disease Detection Dataset was used to classify patients and healthy subjects. In this classification process, training and test data were separated by dividing the entire database into 90/10, 80/20, 70/30, 50/50 and 30/70 ratios, respectively. In addition, in the training phase, the performance of 10-fold cross-validation, 5-fold cross-validation, holdout validation and resubstitution validation methods were analysed at each split ratio. The classification model was then built using Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), Ensemble Bagged Tree (EBT), k-nearest neighbours (KNN) and Neural Network (NN) classifiers. The whole process was repeated 10 times for each data split ratio to ensure randomness and reliable results. The mean and standard deviation values of the results were calculated using the metrics of accuracy, sensitivity, specificity, precision, and F1 score to compare the performance of the methods in all processes.

2. Material and Methods

The block diagram of the study is shown in Figure 1.

2.1. Dataset

The database used in this study is the Oxford Parkinson's Disease Detection Dataset (Little 2007), which consists of biomedical voice measurements from 31 individuals. Of the 31 people from whom the voice recordings in the database were taken, 23 had PD. The database consists of a total of 195 voice data, 48 of which are healthy and 147 of which are PD labelled. The ages of all subjects ranged between 46 and 85; 19 of them were male, and 12 of them were female. For individuals with PD, the disease duration ranges from 0 to 28 years since diagnosis. The voice recordings were captured utilising a microphone positioned at a distance of 8 cm from the lips, with a sampling frequency of 44100 Hz. After the voice recordings were digitized, a total of 22 features (Average vocal fundamental frequency (F0), Maximum vocal fundamental frequency (Fhi), Minimum vocal fundamental frequency (Flo), jitter as a percentage (Jitter(%)), absolute jitter in microseconds (Jitter(Abs)), Relative Amplitude Perturbation (RAP), five-point Period Perturbation Quotient (PPQ), Average absolute difference of differences between cycles, divided by the average period (DDP), local shimmer (Shimmer), local shimmer in decibels (Shimmer(dB)), Three-point Amplitude Perturbation Quotient (APQ3), Five-point Amplitude Perturbation

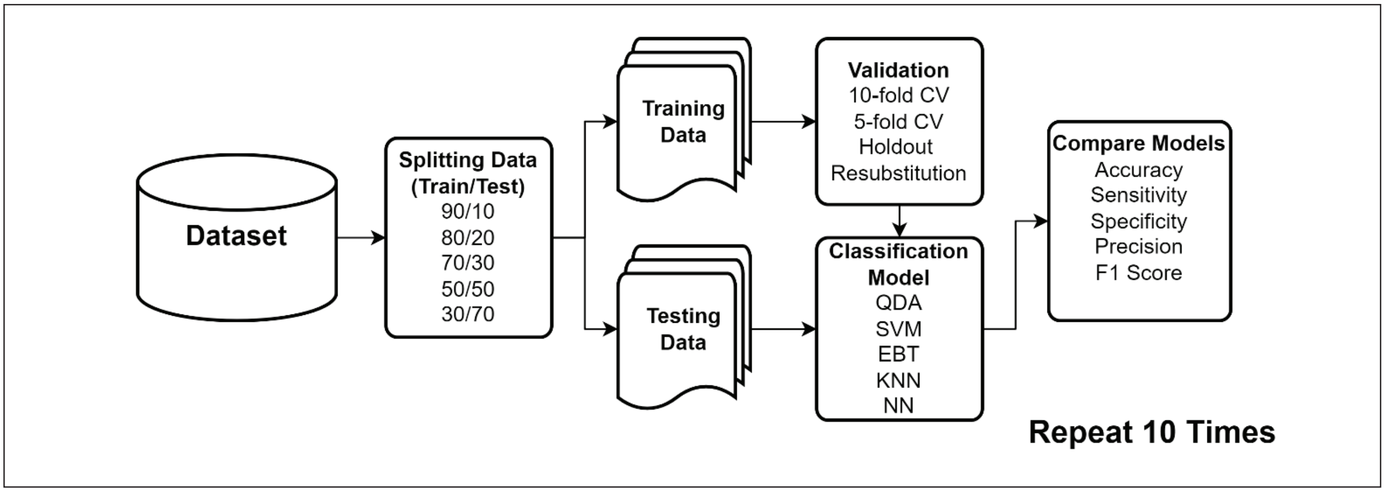

Figure 1. Block diagram of the study.

Quotient (APQ5), 11-point Amplitude Perturbation Quotient (APQ), Average absolute difference between consecutive differences between the amplitudes of consecutive periods (Shimmer: DDA), Noise-to-Harmonics Ratio (NHR), Harmonics-to-Noise Ratio (HNR), Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), Correlation dimension (D2), Three nonlinear measures of fundamental frequency variation spread1,spread2 and Pitch period entropy (PPE)) were calculated using Kay Pentax Multi-Dimensional Voice Program (MDVP) (Little et al. 2009).

2.2. Data Splitting

The first step for a machine learning-based classification process is to split the data to be used in the training and testing phases. For a blind and fair classification process, the data used in the training phase should not be used in the testing phase. There is no optimum split ratio in the classification process. The most crucial point here is that parameter estimates have high variance with less training data, while less test data leads to high variance in performance measurements. Hence, random and repeated database examination with different split ratios is vital in classification. Therefore, in this study, the data collected from patient and non-patient samples in the whole database were split into 90/10, 80/20, 70/30, 50/50 and 30/70 ratios to be used in the training and testing phases, respectively, and the performance of different ratios was examined. In addition, the results were computed by repeatedly re-dividing all classification operations randomly with the same split ratio.

2.3. Validation

Machine learning-based classification research focuses on building more accurate models that can automatically learn from the real world. However, the issue of validating the accuracy of machine learning methods is less popular than implementing new methods (Xie et al. 2011). Regardless of which model is used, it is necessary to prevent overfitting in the training phase, which causes the model to underperform against new data not seen when well-trained on the training data. Therefore, there is a need for a control in the training phase, i.e. validation algorithms. During the training phase, validation will show whether the features represent the classes well enough and measures can be taken to improve the performance of the classification model. In this respect, the k-fold cross-validation method, one of the most preferred validation techniques in the literature, was also used in this study. The steps of this method involve randomly dividing the training set into *k* equal number of groups, considering the first fold as a validation set and applying the same method to the remaining *k* - 1 folds respectively (James et al. 2013). The validation result of the model is calculated as the average of the results obtained in all iterations. Figure 2 shows a schematic representation of 5-fold cross-validation.

Cross-validation allows the model to be tested more frequently in the training phase, except for test data. This reduces overfitting and improves the generalizability of the model. Another method used in the study is 10-fold cross-validation. In this method, the training set is divided into 10 equal parts and the same process is repeated. As illustrated in

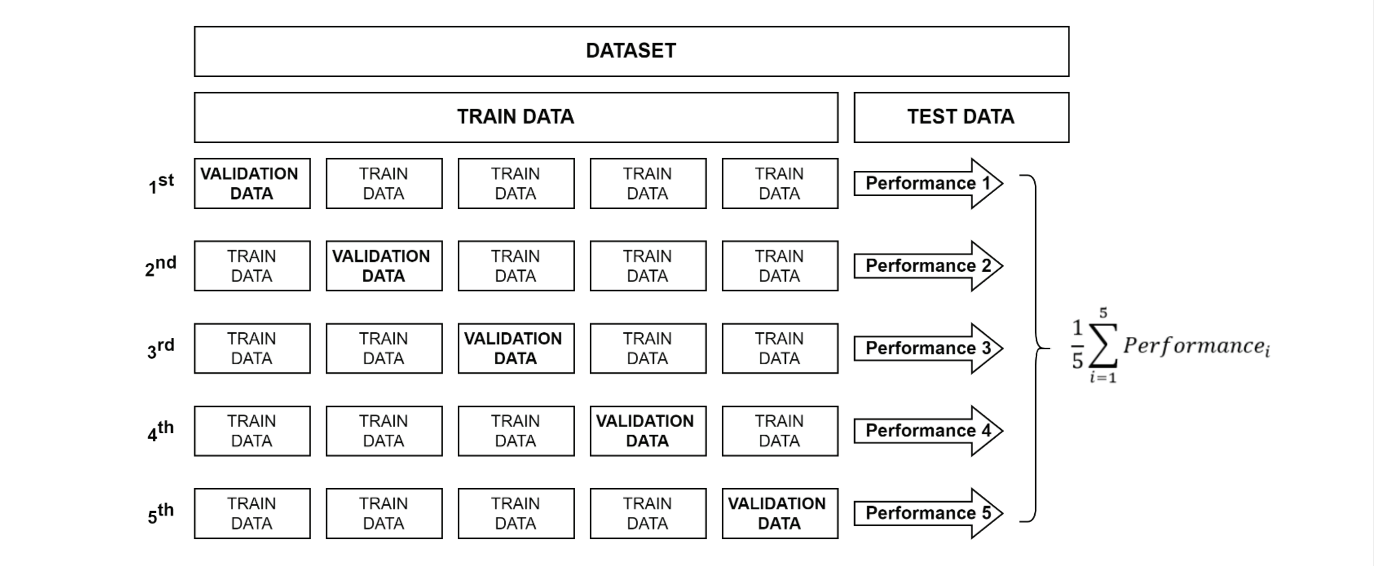

Figure 2: 5-fold cross-validation.

Figure 3, the holdout validation method involves randomly dividing the training set into two subsets for training and validation at a specific ratio. In this method, the model is constructed and validated only once.

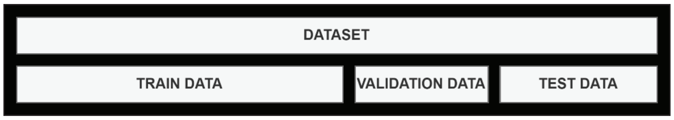

Figure 3. Holdout validation.

The entire training set is used to build the model without splitting it in the resubstitution validation method. Despite generating overly optimistic predictions of performance during the training phase, the actual classification performance of the model is evaluated by comparing it with test data. As a result, the test dataset is used to test the accuracy of a particular model; the training dataset trains different algorithms to build the classification model. In contrast, the validation dataset compares the performance of different algorithms (with different hyperparameters) and decides which is suitable (Molera 2024). This study divided the training and test datasets into the same specific ratios, and the four validation methods mentioned above were applied sequentially in the training phase. Thus, it is determined which method's training validation performance is closest to the test accuracy performance for the database.

2.4. Classification Models

This study used Quadratic Discriminant Analysis, Support Vector Machine, Ensemble Bagged Tree, k-nearest neighbours and Neural Network classifiers, which are widely preferred in classification studies (Ekpezu et al. 2022). Among these popular methods, discriminant analysis, which is fast, accurate and easy to interpret, assumes that different classes generate data based on different Gaussian distributions. It is more suitable for large datasets as it tends to have a lower bias. In training, the fitting function estimates the parameters of the Gaussian distribution for each class. The trained classifier finds the class with the smallest misclassification cost to predict the classes of new data. Quadratic discriminant analysis is more flexible because it can learn quadratic boundaries (Duda et al. 2022). The full covariance matrix structure is utilised in this study for the quadratic discriminant analysis classifier. In the SVM method, to separate the points belonging to the classes on a plane, a line is drawn at the maximum distance to both classes, and the closest points to be drawn to the decision points are called supports (Cortes and Vapnik 1995). However, since it is mostly not enough to separate the classes by drawing a line, different kernel functions are used to multiply the axes and the classes are separated by a non-linear line. In this study, the cubic kernel function is employed via the parameters of standardised data, the automatic kernel scale and box constraint level one. Among the ensemble classification methods created to provide a more reliable and higher performance than a single decision tree classifier, the bagged tree process is preferred because it reduces overfitting and minimises the variance in the decision tree classifier (Bhavsar et al. 2022). The method firstly divides the training set into sub-sets of the same size so as not to overlap, providing a unique data sample and creating a separate decision tree for each sub-set, combining the results from multiple decision trees. In the study, the number of learners was set at 30. In contrast to most classifiers, the KNN algorithm can achieve high performance in difficult situations where data are intertwined. While predicting the class, it looks at the data to be tested and the class of the *k* closest points in the training set. If the *k* closest training set data has the most data belonging to which class, it is decided that the test data belongs to that class (Fix and Hodges 1951). For this reason, the coefficient *k* is often chosen as one in order to make a precise decision. In the present study, the Euclidean distance metric is employed in conjunction with a value of k, equalling one. A neural network, a collection of algorithms that simulate the workings of the human brain, is used in classification to determine the connectivity between features in a dataset. Although neural network models, which usually have predictive solid accuracy, can be complex to understand compared to other classification methods, the size and amount of fully connected layers in the neural network increases model flexibility (Jana et al. 2023). In this study, the Trilayered Neural Network (TNN) classifier was used in combination with rectified linear unit activation.

2.5. Comparison Metrics

When evaluating a classification model's performance, the model's prediction results should be compared with the actual values. In this comparison, the number of positive predictions for positively labelled data (TP), negative predictions for positively labelled data (FN), negative predictions for negatively labelled data (TN) and positive predictions for negatively labelled data (FN) are calculated. Accuracy, Sensitivity, Specificity, Precision and F1 Score metrics obtained through these values were used in the study. Among these metrics, accuracy is the percentage at which positive and negative data can be detected in total, and it is calculated as shown in Equation 1.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \qquad (1)$$

Sensitivity is the percentage of samples with a positive value that can be positively detected and is calculated as shown in Equation 2.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \qquad (2)$$

Specificity refers to the percentage of a sample with a negative value that can be detected as negative and is calculated as shown in Equation 3.

$$Specificity = \frac{TP}{TN + FP} \times 100\% \qquad (3)$$

Precision indicates how much of the data predicted as positive is actually positive and is calculated as shown in Equation 4.

$$Precision = \frac{TP}{TP + FP} \times 100\% \qquad (4)$$

The F1 Score corresponds to the harmonic mean of the Sensitivity and Precision values and is calculated as shown in Equation 5.

$$F1\ Score = 2 \times \frac{Precision \times Sensitivity}{Precision \times Sensitivity} \times 100\% \qquad (5)$$

3. Results and Discussion

In order to ensure randomness in data selection and to obtain valid and stable results, the procedures of all methods used for classification were repeated 10 times. The mean and standard deviation values of the results obtained through Accuracy, Sensitivity, Specificity, Precision and F1 Score metrics were calculated to compare the performance of the methods. Table 1 shows the calculated results of different validation and classification methods based on using 90% of all data in the training phase and 10% in the testing phase. Tables 1-5 show the comparison metric values calculated after separating the training and test data in the 80/20, 70/30, 50/50, and 30/70 ratios, respectively. The KNN model exhibited notable performance with the highest test accuracy of 95.64% using an 80/20 split and 10-fold Cross-Validation, showcasing balanced sensitivity (92.00%) and specificity (96.90%). The QDA model achieved the highest test accuracy of 94.5±6.43% with a 90/10 data-splitting ratio and 10-fold Cross-Validation, demonstrating superior performance across multiple metrics, including sensitivity (82.0%), specificity (98.66%), precision (95.50%), and F1 score (86.77%). The SVM model maintained consistent performance with test accuracies mean around 90%, demonstrating robustness across different splits. The EBT and TNN models displayed varied results, with the EBT model achieving a test accuracy of around 91.76% with both 80/20 and 70/30 splits and the TNN model achieving a test accuracy of 91.55% under the 90/10 split. These results highlight the effectiveness of KNN and QDA in classification tasks, particularly with smaller test sets, while SVM, EBT, and TNN offer reliable alternatives depending on specific metric priorities, showcasing the adaptability and performance of each model under different validation conditions. Generally, the test performance results obtained by applying the 10-fold Cross-Validation method among the validation methods used in the training phase are higher. Notably, the performance results for the 50/50 and 30/70 ratios, where the amount of data allocated for training is lower, are lower than the other cases.

Table 1. Results for the data-splitting ratio taken as 90/10.

Validation	QDA	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	88.86±1.38	61.40±3.50	97.80±1.04	90.20±4.72	73.01±3.44
	Test	94.50±6.43	82.00±20.41	98.66±2.81	95.50±9.56	86.77±16.14
5-fold CV	Train	87.94±0.87	55.35±3.43	98.56±0.90	92.84±4.13	69.24±2.75
	Test	92.10±5.36	80.00±22.37	96.45±2.71	93.36±10.48	84.52±16.18
Holdout 25%	Train	88.60±2.99	54.00±14.30	99.09±1.46	95.83±6.80	67.75±11.74
	Test	93.40±5.42	78.00±15.01	94.12±2.54	94.67±12.25	84.32±15.20
Resubstitution	Train	98.51±0.30	100.00±0.00	98.03±0.39	94.31±1.07	97.07±0.57
	Test	92.30±6.64	81.0±18.43	95.26±2.38	93.48±11.26	83.13±16.42

Table 1. Cont.

Validation	SVM	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	91.09±2.52	82.56±6.32	93.86±1.62	81.42±4.85	81.94±5.25
	Test	93.62±3.50	94.00±7.66	92.67±2.11	81.20±5.22	86.91±6.74
5-fold CV	Train	89.66±2.34	76.98±7.79	93.79±1.12	80.07±3.64	78.39±5.42
	Test	93.00±3.42	96.00±8.43	92.24±2.81	80.14±5.38	87.24±6.43
Holdout 25%	Train	91.63±3.14	86.00±10.75	93.33±1.92	79.64±5.88	82.49±7.10
	Test	93.12±3.57	94.00±8.86	92.67±2.57	81.10±5.24	84.91±6.92
Resubstitution	Train	99.89±0.24	100.00±0.00	99.85±0.32	99.55±0.96	99.77±0.48
	Test	93.50±3.33	96.00±8.34	92.66±2.24	81.33±5.16	88.00±6.37
Validation	**EBT**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	90.11±1.55	73.02±4.93	95.68±1.47	84.81±4.47	78.37±3.59
	Test	91.00±6.58	88.00±6.87	92.00±6.89	81.10±6.09	83.03±2.16
5-fold CV	Train	89.77±0.74	73.26±1.98	95.15±0.73	83.15±2.11	77.87±1.59
	Test	87.50±5.89	78.00±4.76	90.66±7.83	77.21±7.45	75.94±10.40
Holdout 25%	Train	93.26±4.02	84.00±6.99	96.06±3.51	87.23±11.01	85.45±8.41
	Test	91.50±5.30	80.00±6.33	95.33±6.32	88.29±15.85	82.33±10.83
Resubstitution	Train	99.94±0.18	100.00±0.00	99.92±0.24	99.77±0.72	99.89±0.36
	Test	92.00±5.87	84.00±18.38	94.66±6.13	86.62±15.35	83.65±12.19
Validation	**KNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	94.85±0.97	93.26±2.31	95.39±0.97	86.85±2.49	89.92±1.90
	Test	91.00±5.68	98.00±6.32	88.67±7.73	76.33±13.06	85.13±8.50
5-fold CV	Train	93.94±1.33	91.16±3.25	94.85±1.12	85.27±2.90	88.09±2.60
	Test	90.86±6.08	97.20±6.51	86.13±7.45	77.42±14.56	84.12±7.35
Holdout 25%	Train	93.24±3.37	89.00±9.94	94.53±4.24	84.64±10.85	86.08±6.44
	Test	90.08±5.34	94.01±7.73	87.43±7.42	76.15±13.64	82.15±11.87
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	89.12±6.49	91.00±8.31	86.27±7.93	76.26±14.86	85.16±12.92
Validation	**TNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	91.54±2.05	83.02±5.15	94.32±1.68	82.75±4.59	82.82±4.18
	Test	91.55±4.10	84.33±15.72	94.00±4.92	84.45±11.53	83.15±9.04
5-fold CV	Train	90.06±1.35	82.56±3.51	92.50±1.49	78.31±3.37	80.32±2.56
	Test	90.00±6.24	88.00±10.33	90.66±9.53	79.44±14.92	82.27±8.78
Holdout 25%	Train	91.58±4.82	80.50±16.06	94.85±4.75	83.65±12.39	80.86±12.32
	Test	89.00±5.68	86.00±21.19	90.00±4.71	74.20±8.99	78.63±13.62
Resubstitution	Train	99.83±0.28	99.77±0.74	99.85±0.32	99.55±0.96	99.65±0.56
	Test	90.5±4.38	92.00±10.33	90.00±6.48	77.49±12.98	83.19±7.34

Table 2. Results for the data-splitting ratio taken as 80/20.

Validation	QDA	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	88.27±1.00	57.11±4.31	98.31±0.89	91.79±4.03	70.27±3.11
	Test	90.00±3.91	66.00±14.30	98.28±1.82	93.25±7.29	76.44±11.07
5-fold CV	Train	85.71±1.48	43.68±5.14	99.24±0.84	95.03±5.34	59.69±5.07
	Test	89.40±4.02	65.00±14.78	97.28±2.23	92.31±7.92	74.12±11.52
Holdout 25%	Train	83.08±4.22	35.00±13.79	99.31±1.45	95.00±10.54	49.73±16.26
	Test	85.35±4.27	63.00±12.20	95.28±3.18	91.27±7.14	71.44±12.14
Resubstitution	Train	98.65±0.82	100.00±0.00	98.22±1.09	94.85±3.06	97.34±1.61
	Test	88.03±4.92	64.00±15.27	96.28±1.49	92.40±8.49	73.44±13.21
Validation	SVM	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	89.55±1.96	78.42±5.92	93.14±1.01	78.58±3.26	78.45±4.39
	Test	90.51±4.37	78.50±14.83	94.83±4.38	85.24±11.12	80.38±9.89
5-fold CV	Train	89.29±1.96	78.68±5.33	92.71±1.34	77.67±3.89	78.13±4.25
	Test	90.00±4.75	78.00±16.19	94.14±4.61	83.55±12.35	79.52±10.62
Holdout 25%	Train	87.44±5.73	68.78±19.43	93.89±6.12	81.68±15.90	72.45±13.89
	Test	89.74±4.68	77.00±16.36	94.09±4.32	83.16±11.88	78.85±10.60
Resubstitution	Train	99.36±0.43	98.16±2.17	99.75±0.41	99.23±1.24	98.67±0.90
	Test	90.26±4.80	79.00±15.24	93.14±4.98	83.55±12.35	80.26±10.53
Validation	EBT	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	90.26±1.76	74.47±3.93	95.34±1.56	83.87±4.76	78.84±3.69
	Test	91.79±5.10	76.00±15.78	97.24±2.72	90.56±9.73	81.95±12.21
5-fold CV	Train	89.10±1.57	70.79±4.72	95.00±1.52	82.19±4.37	75.96±3.56
	Test	91.79±3.59	74.00±15.06	97.93±2.41	93.29±7.49	81.51±9.79
Holdout 25%	Train	89.74±5.27	72.67±16.15	95.56±4.27	85.43±13.13	77.52±12.80
	Test	91.79±3.97	77.00±13.37	96.90±2.54	89.73±7.72	82.33±9.84
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	91.79±3.97	75.00±13.54	97.59±3.27	92.40±10.66	82.01±9.60
Validation	KNN	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	93.46±1.12	90.00±2.99	94.58±1.34	84.35±3.17	87.03±2.14
	Test	95.64±3.21	92.00±9.19	96.90±3.43	91.83±8.22	91.52±6.24
5-fold CV	Train	92.44±1.20	87.63±4.12	93.98±1.16	82.50±2.87	84.93±2.53
	Test	94.44±4.53	91.12±9.32	95.85±3.21	90.63±7.48	91.38±6.03
Holdout 25%	Train	93.33±3.46	87.78±8.17	95.22±3.30	86.42±8.71	86.85±6.87
	Test	93.56±4.71	91.17±9.94	95.64±3.67	90.43±8.11	90.04±7.22
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	94.34±4.27	91.24±9.08	96.90±4.02	89.88±9.52	90.72±7.54
Validation	TNN	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	88.97±2.63	79.73±4.48	91.95±2.41	76.35±6.06	77.95±4.93
	Test	88.97±4.53	76.00±17.13	93.45±5.74	82.87±13.73	77.35±10.32
5-fold CV	Train	88.78±1.26	79.21±5.75	91.86±1.70	76.00±3.44	77.42±3.00
	Test	90.29±4.93	80.27±16.17	93.79±4.24	82.60±10.25	80.54±10.96
Holdout 25%	Train	90.00±5.60	80.89±13.63	93.18±5.78	81.44±14.67	80.28±11.01
	Test	87.95±4.69	78.00±15.49	91.38±5.69	77.32±11.36	76.48±10.04
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	91.79±3.97	83.00±13.37	94.83±3.35	85.21±7.64	83.47±8.90

Table 3. Results for the data-splitting ratio taken as 70/30.

Validation	QDA	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	86.06±1.70	47.06±5.37	98.93±0.85	93.56±5.03	62.50±5.41
	Test	88.45±2.82	58.57±11.07	97.95±1.68	90.68±8.08	70.49±9.09
5-fold CV	Train	82.77±0.99	32.35±3.92	99.42±0.50	95.02±4.35	48.13±4.35
	Test	86.36±3.81	56.30±10.94	95.58±2.23	89.57±8.10	70.26±8.99
Holdout 25%	Train	80.59±3.45	20.56±13.22	100.00±0.00	100.00±0.00	36.38±13.49
	Test	82.42±2.97	53.57±11.07	94.22±2.72	86.46±9.58	69.23±9.17
Resubstitution	Train	99.27±0.91	100.00±0.00	99.03±1.21	97.25±3.31	98.58±1.74
	Test	84.28±4.26	54.57±11.07	95.35±2.41	88.81±11.03	68.41±9.22
Validation	**SVM**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	90.37±1.68	80.59±2.48	93.59±1.95	80.82±4.86	80.64±2.99
	Test	91.72±1.09	82.14±13.57	94.77±4.42	85.53±9.22	82.47±3.27
5-fold CV	Train	89.85±1.59	77.94±3.18	93.79±1.90	80.77±4.46	79.25±2.92
	Test	91.52±1.78	81.53±14.63	94.57±4.68	85.70±10.41	82.52±4.29
Holdout 25%	Train	87.65±4.76	71.25±12.80	92.98±2.47	76.35±9.20	73.58±10.74
	Test	89.25±7.13	80.14±14.59	85.00±13.21	83.70±11.31	81.43±4.07
Resubstitution	Train	99.34±0.64	97.65±2.70	99.90±0.31	99.71±0.90	98.65±1.33
	Test	90.37±1.99	81.14±14.95	94.32±5.38	84.96±11.42	82.02±4.12
Validation	**EBT**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	88.83±1.09	72.35±5.41	94.27±0.72	80.68±1.56	76.19±3.10
	Test	90.52±2.33	72.14±7.86	96.36±1.59	86.46±5.75	78.46±5.81
5-fold CV	Train	88.10±2.26	70.59±5.55	93.88±1.65	79.28±5.19	74.62±4.91
	Test	91.72±2.54	75.00±9.07	97.05±2.16	89.42±6.43	81.22±6.11
Holdout 25%	Train	87.94±3.52	71.53±11.42	93.46±5.75	81.03±16.78	74.30±6.97
	Test	92.41±1.67	77.86±7.86	97.05±2.64	90.39±7.93	83.13±3.65
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	91.55±2.87	77.14±8.11	96.14±2.84	87.10±9.21	81.47±6.68
Validation	**KNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	93.87±1.69	92.06±5.20	94.47±1.38	84.66±3.21	88.14±3.33
	Test	93.28±3.09	90.00±10.75	94.32±4.58	84.98±10.58	86.63±6.50
5-fold CV	Train	92.70±2.64	88.82±5.51	93.98±2.09	83.07±5.44	85.81±5.10
	Test	92.12±3.29	89.50±11.45	93.45±4.64	83.65±11.75	86.52±6.11
Holdout 25%	Train	92.65±2.50	86.81±12.22	94.55±2.00	83.96±4.25	84.87±6.23
	Test	92.22±3.58	89.15±10.87	92.82±4.49	82.90±10.63	85.03±6.63
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	91.67±4.01	88.90±11.98	92.95±5.57	83.18±11.78	85.83±7.50
Validation	**TNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	89.49±2.34	80.29±3.12	92.52±2.63	78.36±6.21	79.22±4.09
	Test	89.14±5.08	74.29±21.87	93.86±3.72	79.18±9.88	74.76±18.86
5-fold CV	Train	89.27±2.53	78.53±3.93	92.82±2.34	78.51±5.93	78.48±4.68
	Test	89.31±3.02	80.71±10.67	92.05±3.59	77.12±7.58	78.32±6.60
Holdout 25%	Train	87.94±3.52	67.64±13.72	94.52±4.19	82.70±12.69	72.94±7.80
	Test	89.14±3.36	85.00±11.88	90.45±5.12	75.55±10.30	79.01±6.63
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	88.52±5.20	82.14±10.24	90.53±5.92	75.21±14.33	77.77±9.60

Table 4. Results for the data-splitting ratio taken as 50/50.

Validation	QDA	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	70.31±3.92	87.92±3.07	64.59±4.93	44.83±3.64	59.32±3.58
	Test	67.84±5.41	93.75±8.39	59.32±8.04	43.50±4.49	59.21±4.52
5-fold CV	Train	71.12±3.73	87.92±3.07	65.68±4.78	45.61±3.67	59.98±3.44
	Test	66.32±5.22	91.43±10.12	57.32±8.31	42.67±6.96	57.13±4.81
Holdout 25%	Train	66.67±10.21	80.00±13.15	62.22±12.51	42.68±9.99	55.18±10.16
	Test	68.27±4.41	92.61±12.27	58.32±8.82	43.21±9.67	58.61±5.98
Resubstitution	Train	71.84±2.97	92.08±1.32	65.27±3.82	46.38±2.80	61.65±2.60
	Test	69.54±6.43	92.45±11.43	59.27±9.64	43.42±12.16	58.91±6.42
Validation	**SVM**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	89.18±237	76.25±8.11	93.38±1.34	78.85±4.16	77.39±5.65
	Test	85.46±3.81	67.92±11.63	91.23±4.20	72.84±9.42	69.62±7.70
5-fold CV	Train	86.94±3.78	72.50±9.04	91.62±2.97	73.99±8.85	73.06±8.08
	Test	83.69±5.21	66.63±12.11	90.18±5.20	72.67±11.12	69.23±7.13
Holdout 25%	Train	85.85±8.40	63.33±24.60	93.33±5.74	76.15±21.35	67.74±20.77
	Test	84.31±4.67	65.82±12.72	90.26±3.12	72.84±10.14	66.62±9.20
Resubstitution	Train	99.18±0.94	97.08±3.95	99.86±0.43	99.60±1.26	98.28±2.03
	Test	85.20±5.47	67.54±12.97	91.11±6.23	72.84±12.57	67.72±8.71
Validation	**EBT**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	88.88±1.12	75.00±6.51	93.38±1.85	78.95±4.23	76.67±2.92
	Test	90.01±1.88	76.25±7.87	94.52±2.51	82.72±6.68	78.94±4.34
5-fold CV	Train	88.06±1.74	68.33±3.51	94.46±1.74	80.19±5.04	73.73±3.59
	Test	85.88±4.98	74.17±9.58	89.73±7.65	73.55±14.02	72.60±7.27
Holdout 25%	Train	84.58±6.53	68.33±12.30	90.00±5.11	70.05±13.83	69.03±12.62
	Test	87.63±3.67	75.83±12.55	91.51±6.55	77.26±11.43	75.19±6.30
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	88.66±2.57	75.42±9.71	93.01±3.32	78.79±7.38	76.54±6.01
Validation	**KNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	92.35±2.77	86.25±6.23	94.32±2.53	83.40±6.46	84.68±5.59
	Test	90.72±2.00	86.67±5.83	92.35±3.58	79.00±7.03	82.30±3.25
5-fold CV	Train	91.12±2.55	83.33±5.20	93.65±2.18	81.11±5.81	82.16±5.10
	Test	88.64±3.05	83.67±5.27	90.07±3.64	77.00±7.29	79.94±5.20
Holdout 25%	Train	90.83±6.46	81.67±14.59	93.89±6.11	83.14±16.17	81.75±13.27
	Test	86.17±3.57	82.67±5.68	91.68±5.15	79.00±7.53	81.37±4.12
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	87.22±4.01	84.67±7.84	92.05±5.52	79.00±8.13	82.12±6.15
Validation	**TNN**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**
10-fold CV	Train	89.29±2.90	76.25±7.87	93.51±1.89	79.34±6.22	77.61±6.53
	Test	87.01±4.01	82.08±10.77	88.63±6.87	72.58±11.09	75.96±5.70
5-fold CV	Train	83.88±5.54	70.00±12.39	88.38±3.67	66.12±10.68	67.93±11.21
	Test	83.81±6.07	70.83±9.42	88.08±7.45	68.80±14.20	68.88±8.83
Holdout 25%	Train	89.58±7.92	86.67±20.49	90.56±5.89	75.73±17.35	80.18±17.15
	Test	86.80±3.07	82.21±16.02	88.77±5.66	72.90±8.89	75.63±6.19
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	85.67±6.49	77.08±11.99	88.49±7.42	71.14±15.95	73.08±11.25

Table 5. Results for the data-splitting ratio taken as 30/70.

Validation	QDA	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	68.45±6.35	82.14±6.94	64.09±6.67	42.56±6.47	55.97±6.96
	Test	71.02±4.73	87.65±5.85	65.53±5.48	45.89±4.50	60.16±4.83
5-fold CV	Train	69.66±6.15	82.86±6.02	65.45±6.84	43.79±6.39	57.17±6.44
	Test	70.56±4.57	85.37±5.85	63.13±5.37	44.27±4.69	58.96±4.63
Holdout 25%	Train	68.57±15.13	83.33±6.57	64.55±6.72	42.56±6.51	54.94±6.99
	Test	71.27±4.51	87.82±5.85	65.21±5.87	45.98±4.55	60.56±4.23
Resubstitution	Train	70.34±5.68	88.57±3.69	64.55±6.79	44.78±5.94	59.35±5.64
	Test	70.42±4.82	86.25±5.85	65.57±6.63	45.96±5.80	59.18±7.83
Validation	SVM	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	83.45±5.02	62.86±14.98	90.00±4.44	66.91±12.17	64.21±12.66
	Test	85.77±3.77	60.88±13.23	93.98±2.81	77.03±9.42	67.40±10.93
5-fold CV	Train	81.90±8.10	60.00±20.54	88.86±4.96	62.51±19.16	60.93±19.28
	Test	86.92±3.85	61.58±10.12	94.16±3.57	80.23±10.21	68.67±10.23
Holdout 25%	Train	83.57±8.28	63.33±33.15	89.09±7.17	61.30±19.07	64.63±20.95
	Test	86.79±3.86	61.62±12.13	94.07±2.67	78.61±9.12	68.42±10.15
Resubstitution	Train	98.97±1.45	95.71±6.02	100.00±0.00	100.00±0.00	97.72±3.23
	Test	85.56±2.34	60.18±9.21	92.99±3.82	77.08±12.54	67.29±9.91
Validation	EBT	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	83.28±6.61	55.71±14.98	92.05±5.05	69.89±19.04	61.54±15.47
	Test	85.62±3.95	63.24±10.76	93.01±4.60	77.01±13.26	68.41±5.64
5-fold CV	Train	83.62±4.16	58.57±14.60	91.59±3.40	69.04±11.53	62.69±11.85
	Test	86.64±3.89	63.24±8.69	94.37±4.41	80.52±13.26	70.20±8.06
Holdout 25%	Train	82.86±11.76	53.33±39.13	90.91±6.87	66.67±26.35	65.60±26.56
	Test	86.79±2.74	65.88±7.87	93.69±2.97	78.22±8.49	71.15±6.03
Resubstitution	Train	99.66±0.73	98.57±3.01	100.00±0.00	100.00±0.00	99.26±1.56
	Test	85.91±2.62	61.76±6.65	93.88±3.86	78.48±10.54	68.55±4.92
Validation	KNN	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	85.34±3.92	75.71±7.68	88.41±3.93	68.04±8.46	71.47±7.17
	Test	88.61±3.27	78.24±10.30	92.04±3.13	76.88±7.80	77.18±7.10
5-fold CV	Train	85.52±5.09	67.86±12.26	91.14±4.60	71.93±13.87	69.30±10.89
	Test	88.57±4.13	78.11±11.12	91.98±4.12	75.96±7.56	77.09±8.23
Holdout 25%	Train	85.00±10.88	66.67±27.22	90.00±10.00	70.17±27.76	65.45±23.75
	Test	88.43±2.21	77.97±10.56	91.94±3.53	76.58±6.97	77.11±7.52
Resubstitution	Train	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Test	87.61±6.21	77.61±12.37	91.08±3.57	75.68±10.81	76.03±7.12
Validation	TNN	Accuracy	Sensitivity	Specificity	Precision	F1 Score
10-fold CV	Train	81.90±5.41	65.00±12.35	87.27±3.89	61.98±11.26	63.33±11.39
	Test	84.67±4.09	67.06±15.43	90.49±4.41	70.50±8.25	67.75±11.14
5-fold CV	Train	80.69±6.84	60.71±17.25	87.05±4.16	59.38±14.41	59.87±15.56
	Test	84.96±5.04	66.47±19.32	91.07±3.39	70.46±10.15	67.46±14.86
Holdout 25%	Train	80.00±12.51	56.67±35.31	86.36±11.54	52.67±36.69	60.88±14.74
	Test	84.89±2.71	70.88±15.09	89.51±3.57	69.40±5.09	69.28±8.38
Resubstitution	Train	99.66±0.73	98.57±3.01	100.00±0.00	100.00±0.00	99.26±1.56
	Test	84.59±6.75	64.71±18.91	91.17±4.79	70.52±14.87	66.86±16.05

4. Conclusion and Suggestions

Early diagnosis of PD, a neurodegenerative disorder characterised by a range of motor and non-motor symptoms, especially speech abnormalities, is crucial for timely intervention and disease management. In this study, we investigate the effectiveness of machine learning methods applied to speech data for Parkinson's disease detection, focusing on different data separation and validation methods. As a result of the detailed investigations, the KNN classifier model, which separates the data into training and test sets using an 80/20 ratio and uses 10-fold cross-validation in the training phase, was the most successful method. In the training phase, the accuracy of our model was 93.46 ± 1.12%, demonstrating its ability to classify individuals as PD or healthy with high precision accurately. Sensitivity, specificity, precision and F1 score were also computed and obtained 90.00 ± 2.99%, 94.58 ± 1.34%, 84.35 ± 3.17% and 87.03 ± 2.14%, respectively. In evaluating the test set, our model maintained its strong performance, achieving an accuracy of 95.64 ± 3.21%. More importantly, the sensitivity, specificity, precision and F1 score remained consistently high, with 92.00 ± 9.19%, 96.90% ± 3.43%, 91.83 ± 8.22% and 91.52 ± 6.24%, respectively. These findings emphasise the generalizability of our model to unseen data and its potential as a diagnostic tool for Parkinson's disease. A comparison of our model with other studies in the literature is presented in Table 6.

The results of the proposed method, which is the most successful among the compared studies, emphasise the robustness and generalizability of the KNN classifier in accurately identifying individuals with Parkinson's disease based on speech data. The success of the KNN classifier can be attributed to its ability to classify data points based on their proximity to neighbouring examples in the feature space. Utilising the natural structure of the data, the K-NN algorithm effectively captured the subtle differences in speech features between individuals with and without PD, thus facilitating accurate classification. In addition, although the 70/30 ratio is commonly used in the literature, the experimental studies found that the 80/20 ratio is more appropriate for data separation when using the KNN classifier. In contrast to most of the compared studies, high mean and low standard deviation values of 95.64 ± 3.21%

Table 6. Comparative analysis of studies on PD.

Study	Split Rate Train/Test	Classification Method	Accuracy %
Ene 2008	70/30	Probabilistic Neural Network	81.28
Little et al. 2009	Bootstrap resampling 50 replicates	Gaussian Radial Basis Kernel Support Vector Machine	91.4
Das 2010	65/35	Artificial Neural Network	92.9
Çağlar et al. 2010	50/50	Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges	94.72
Luukka 2011	50/50	Fuzzy Entropy Measures + Similarity	85.03
Inzamam et al. 2015	70/30	Ensemble Method Generated with C&R Tree, Bayesian Network and C5.0	95.31
Gupta et al. 2018	70/30	The Optimized Cuttlefish algorithm	92.19
Sharma et al. 2019	70/30	Modified Grey Wolf Optimization and Random Forest	93.87
Senturk 2020	N/A	Recursive Feature Elimination Support Vector Machines	93.84
Abdulateef et al. 2023	N/A	Fast Learning Machine	80.00
Vu et al. 2023	70/30	Random Forest	95.42
Nareklishvili and Geitle 2024	N/A	Deep Ensemble Transformers (casForest) algorithm	88.5
Proposed	80/20	K-Nearest Neighbours	95.64

were achieved by performing all operations repetitively in order to minimise the effect of randomness in data separation on the generalised results and thus achieve more consistent results. In conclusion, the promising results on progressive PD pave the way for the development of robust and reliable diagnostic tools for the early detection and management of this debilitating condition.

5. References

Abdulateef, SK., Ismael, AN., Salman, MD. 2023. Feature weighting for Parkinson's identification using single hidden layer neural network. Computing, 225–230. https://doi.org/10.47839/ijc.22.2.3092

Bang, C., Bogdanovic, N., Deutsch, G., Marques, O. 2023. Machine learning for the diagnosis of Parkinson's disease using speech analysis: a systematic review. International Journal of Speech Technology, 26(4), 991–998. https://doi.org/10.1007/s10772-023-10070-9

Bhavsar, K., Vakharia, V., Chaudhari, R., Vora, J., Pimenov, DY., Giasin, K. 2022. A comparative study to predict bearing degradation using discrete wavelet transform (DWT), tabular generative adversarial networks (TGAN) and machine learning models. Machines, 10(3), 176. https://doi.org/10.3390/machines10030176

Çağlar, MF., Çetişli, B., Toprak, İB. 2010. Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier. Mühendislik Bilimleri Ve Tasarım Dergisi, 1(2), 59–64.

Cortes, C., Vapnik, V. 1995. Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/bf00994018

Das, R. 2010. A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568–1572. https://doi.org/10.1016/j.eswa.2009.06.040

Duda, RO., Hart, PE., Stork, DG. 2022. Pattern Classification (3rd ed.). Standards Information Network.

Ekpezu, AO., Katsriku, F., Yaokumah, W., Wiafe, I. 2022. The use of machine learning algorithms in the classification of sound: A systematic review. International Journal of Service Science Management Engineering and Technology, 13(1), 1–28. https://doi.org/10.4018/ijssmet.298667

Ene, M. 2008. Neural network-based approach to discriminate healthy people from those with Parkinson's disease. Annals of the University of Craiova-Mathematics and Computer Science Series, 35, 112–116.

Esmer, S., Uçar, MK., Çil, İ., Bozkurt, MR. 2020. Parkinson Hastalığı Teşhisi İçin Makine Öğrenmesi Tabanlı Yeni Bir Yöntem. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 8(3), 1877–1893. https://doi.org/10.29130/dubited.688223

Fix, E., Hodges, JL. 1951. Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties USAF School of Aviation Medicine.

Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., de Albuquerque, VHC. 2018. Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cognitive Systems Research, 52, 36–48. https://doi.org/10.1016/j.cogsys.2018.06.006

Huang, Y., Chen, Q., Wang, Z., Wang, Y., Lian, A., Zhou, Q., Zhao, G., Xia, K., Tang, B., Li, B., Li, J. 2024. Risk factors associated with age at onset of Parkinson's disease in the UK Biobank. NPJ Parkinson's Disease, 10(1), 3. https://doi.org/10.1038/s41531-023-00623-9

Inzamam-Ul-Hossain, M., MacKinnon, L., Islam, MR. 2015. Parkinson disease detection using ensemble method in PASW benchmark. 2015 IEEE International Advance Computing Conference (IACC).

Islam, MA., Hasan Majumder, MZ., Hussein, MA., Hossain, KM., Miah, MS. 2024. A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. Heliyon, 10(3), e25469. https://doi.org/10.1016/j.heliyon.2024.e25469

Iyer, A., Kemp, A., Rahmatallah, Y., Pillai, L., Glover, A., Prior, F., Larson-Prior, L., Virmani, T. 2023. A machine learning method to process voice samples for identification of Parkinson's disease. Scientific Reports, 13(1), 20615. https://doi.org/10.1038/s41598-023-47568-w

James, G., Witten, D., Hastie, T., Tibshirani, R. 2013. An introduction to statistical learning: With applications in R (1st ed.). Springer.

Jana, DK., Bhunia, P., Adhikary, SD., Mishra, A. 2023. Analyzing of salient features and classification of wine type based on quality through various neural network and support vector machine classifiers. Results in Control and Optimization, 11(100219), 100219. https://doi.org/10.1016/j.rico.2023.100219

Little, M. 2007. Parkinsons [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C59C74

Little, MA., McSharry, PE., Hunter, EJ., Spielman, J., Ramig, LO. 2009. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Bio-Medical Engineering, 56(4), 1015–1022. https://doi.org/10.1109/tbme.2008.2005954

Luukka, P. 2011. Feature selection using fuzzy entropy measures with similarity classifier. Expert Systems with Applications, 38(4), 4600–4607. https://doi.org/10.1016/j.eswa.2010.09.133

Molera, LM. 2024. Machine learning Q&A: All about model validation. Mathworks.com. https://ch.mathworks.com/campaigns/offers/next/all-about-model-validation.html

Nareklishvili, M., Geitle, M. 2024. Deep ensemble transformers for dimensionality reduction. IEEE Transactions on Neural Networks and Learning Systems, 1–12. https://doi.org/10.1109/tnnls.2024.3357621

Orozco-Arroyave, JR., Vdsquez-Correa, JC., Honig, F., Arias-Londono, JD., Vargas-Bonilla, JF., Skodda, S., Rusz, J., Noth, E. 2016. Towards an automatic monitoring of the neurological state of Parkinson's patients from speech. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E. 2011. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. The Journal of the Acoustical Society of America, 129(1), 350–367. https://doi.org/10.1121/1.3514381

Senturk, ZK. 2020. Early diagnosis of Parkinson's disease using machine learning algorithms. Medical Hypotheses, 138(109603), 109603. https://doi.org/10.1016/j.mehy.2020.109603

Sharma, P., Sundaram, S., Sharma, M., Sharma, A., Gupta, D. 2019. Diagnosis of Parkinson's disease using modified grey wolf optimization. Cognitive Systems Research, 54, 100–115. https://doi.org/10.1016/j.cogsys.2018.12.002

Smith, KM., Caplan, DN. 2018. Communication impairment in Parkinson's disease: Impact of motor and cognitive symptoms on speech and language. Brain and Language, 185, 38–46. https://doi.org/10.1016/j.bandl.2018.08.002

Virameteekul, S., Revesz, T., Jaunmuktane, Z., Warner, TT., De Pablo-Fernández, E. 2023. Clinical diagnostic accuracy of Parkinson's disease: Where do we stand? Movement Disorders: Official Journal of the Movement Disorder Society, 38(4), 558–566. https://doi.org/10.1002/mds.29317

Vu, TA., Ha, NTT., Duc, LM., Huy, H. Q., Dung, NV., Huong, PTV., Thanh, NT. 2023. A comparison of machine learning algorithms for Parkinson's disease detection. 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS).

Xie, X., Ho, JWK., Murphy, C., Kaiser, G., Xu, B., Chen, TY. 2011. Testing and validating machine learning classifiers by metamorphic testing. The Journal of Systems and Software, 84(4), 544–558. https://doi.org/10.1016/j.jss.2010.11.920