HHS Public Access

Author manuscript

Procedia Comput Sci. Author manuscript; available in PMC 2023 July 27.

Published in final edited form as:

Procedia Comput Sci. 2023 ; 218: 1529–1541. doi:10.1016/j.procs.2023.01.131.

A non-invasive method for prediction of neurodegenerative diseases using gait signal features

Vipin Syam[a], **Shivesh Safal**[a], **Ongmu Bhutia**[a], **Amit Kumar Singh**[a], **Diksha Giri**[b], **Samrat Singh Bhandari**[c], **Ranjit Panigrahi**[b,*]

[a]Department of Electrical & Electronics Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, India

[b]Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, India

[c]Department of Psychiatry, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok, Sikkim

Abstract

The steady degeneration of neurons is the hallmark of neurodegenerative illnesses, which are, by definition, incurable. Corticobasal Syndrome (CS), Huntington's Disease (HD), Dementia, Amyotrophic Lateral Sclerosis (ALS), Progressive supranuclear palsy (PSP) and Parkinson's Disease (PD) are some of the common neurodegenerative diseases which has impacted millions of people, predominantly among the older population. Various computational techniques, including but not limited to machine learning, are emerging as discrimination and detection of neuro-related diseases. This research proposed a machine learning-based framework to correctly detect PD, HD, and ALS from the gait signals of subjects both in binary and multi-class detection environment. The detection approach proposed here combines the classification power of Naïve Bayes and Logistic Regression jointly in a modern UltraBoost ensemble framework. The proposed method is unique in its ability to detect neuro diseases with a small number of gait features. The proposed approach ascertains most essential gait features through three state-of-the-art feature selection schemes, infinite feature selection, infinite latent feature selection and Sigmis feature selection. It has been observed that the gait signal features of the subjects are identified through Infinite Feature Selection manifests better detection results than the features obtained through Infinite Latent Feature and Sigmis feature selection while detecting Parkinson's and Huntington's Disease in a multi-class environment. So far as the binary detection environment is concern, the Amyotrophic lateral sclerosis is detected with 99.1% detection accuracy using 18 Sigmis gait features, with 99.1% sensitivity and 98.9% specificity, respectively. Similarly, Huntington's disease was detected with 94.2% detection accuracy, 94.2% sensitivity, and 94.5% specificity using 5 Sigmis gait features. Finally, Parkinson's disease was detected with 98.4% sensitivity, specificity, and detection accuracy.

*Corresponding author. Tel.: +91-70014-33347. ranjit.p@smit.smu.edu.in.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Keywords

Parkinson's Detection; Neurodegenerative disease detection; UltraBoost ensemble; Infinite feature selection; Sigmis feature selection; Infinite latent feature selection

1. Introduction

Neurodegenerative diseases are related to the nervous system, where the progressive deterioration of neurons occurs. The primary cause of neurodegenerative diseases is the continuous activation of neuron death. The risk factor for various neurodegenerative diseases rises with age [1]. It has been estimated that nearly 6.8 million people expired every year due to neurological disorders [2]. The population keeps on mounting because of state-of-the-art medical advancements and hygiene. As a result, the ageing population is increasing exponentially, and thus, increased number of people suffer from neurodegenerative maladies. Therefore, it is crucial to diagnose the neuro related diseases at an early age to curtail the damages that the diseases impart on the human brain. Early detection of the diseases at a prior stage would warrant accurate diagnoses which will enable imparting correct treatment at an early stage. However, detecting neuro diseases at an early stage is challenging, not only for the affected individuals or their caregiver who may not recognize the initial symptoms but also for the clinicians who may not be able to diagnose the condition confidently. The symptoms of neurodegenerative diseases are generally voice impairment, loss of memory, difficulty in gait movement etc. Neurodegenerative diseases, mainly Parkinson's Disease (PD) and Huntington's Disease (HD) are basal ganglia disorders that impact motor and cognitive functions [3]. Amyotrophic Later Sclerosis (ALS) belongs to motor neuron diseases (MND), which are caused by slow degeneration and death of motor neurons [4]. As a result, degeneration of the muscles takes place. Motor neurons are part of the Central Nervous System (CNS) which controls all our muscle movements. The pathophysiology of these diseases is varied as well also the presentations, as in which some cause memory loss, cognitive impairments and others affect human gait. Cognitive impairment and gait have received recent attention. Gait is how someone walks. Medical professionals can tell if someone has neurologic, muscular, or skeletal problems from their gait. Swing and stance are gait phases. Swing phase begins when the foot leaves the ground and ends when the heel strikes. The stance phase includes all foot-on-ground activity, from heel strike to toe off. The stance phase makes up 60% of the gait cycle [5].

Stride, stance, and swing intervals helps to identify neurodegenerative disorders [6–9]. At the beginning of the disease, stride, stance, and swing intervals vary where variable time intervals of stride, stance and swing can predict neurodegeneration [10]. The Hausdorff et al. [11,12] gait signal dataset was investigated to understand gait rhythm variation. The dataset includes gait dynamics of neurodegenerative and healthy individuals. The dataset contains gait signals of 13 ALS, 15 PD, and 20 HD patients. The patients were asked to walk normally for 5 minutes along a 77-m hallway. Force-sensitive insoles and an ankle-worn recorder were used to evaluate gait rhythm and cycle timing. The gait signals were sampled at 300Hz using an internal 12-bit A/D converter. A UNIX workstation was used to analyze digitized data to determine the duration of each stride's gait cycle. Force-sensitive insoles

recorded left swing interval, left stride interval, left stance interval, right swing interval, right stride interval, right stance interval, double support interval, left swing interval (% of stride), left stance interval (% of stride), right swing interval (% of stride), right stance interval (% of stride), and double support interval (percent of stride).

The swing, stance and stride intervals of both right and left legs of a random control and random subject suffering ALS, HD and PD has been plotted in Figure 2 to understand the signal variation and the possible scope for developing an automatic neuro disease detection method thereof. The stride-to-stride fluctuations in distance and time due to malfunction of the nervous system are closely associated with neurodegenerative diseases. According to Figure 1, it can be seen that the control stride, stance and swing intervals are consistent both for left and right leg and within the normal range. On the other hand, the ALS, PD and HD patient's stride, stance and swing intervals are highly modulated. Such a huge variation in the time intervals of stride, stance and swing intervals can be sent to a machine learning models to develop a prediction engine.

Detecting and diagnosing the disease at early stage requires specialized expertise and years of practice. Therefore, medical practitioners have now a days started employing artificial intelligence techniques on available gait and motor signal data of patients. There are available datasets that use gait biomarkers of patients which contain conditions of these diseases [11]. In the past few years, many intellectual techniques have come up to classify neurodegenerative diseases from patients and controls' gait signals [10]. Recent machine learning algorithms make use of decision trees and decision forests for classification challenges. Decision forests are also used over conventional decision trees as decision forests aggregate many decision trees and give more precise, accurate results and can limit overfitting and underfitting conditions [14]. Decision trees along with many other supervised classification techniques have been used for neurodegenerative disease prediction, where k-NN, Linear Discriminant Analysis (LDA) and SVM plays a crucial role in the detection process [15].

Many of the existing Neurodegenerative Disease Detection (NDD) techniques employs different approaches of feature selection and classification during detection of the disease. For an instance, a non-invasive approach of Parkinson's detection scheme has been proposed using wavelet decomposition of the gait signals [16]. It has been observed that the left stance interval shows 90.32% detection accuracy when features are extracted using wavelet transformation. Similarly, gait fluctuations are analyzed to detect ALS, PD and HD using multiple gait signal along with vertical ground reaction force, stride, stance and swing duration [17]. A multi-class detection engine has been prepared using various neuro disease subjects where the AdaBoost shows promising detection accuracy of 99.17%. A couple of years ago Long Short-Term Memory (LSTM) based multi-feature extraction model has been proposed for ultimate detection of neurodegenerative diseases [18]. The LSTM method proved to be superior to the manual observation of gait rhythm. Similarly, a state-of-the-art multimodal method of neurodegenerative disease gait recognition method has been proposed recently using correlative memory neural network on representative gait features generated by a spatial feature extractor [19]. The proposed neural network model shows 99.31% detection accuracy in a multi-class detection environment. Hidden Markov

Model (HMM) proved to be an excellent detection method for NDD [20]. The gait stride interval with HMM magnificently reveals a detection rate of 90.3%. An automatic diagnosis model attracts the medical practitioners that employs Support Vector Machine (SVM) using Radial Basis Function (RBF) kernel. Genetic algorithm has been used to select relevant gait features where the RBF based SVM witnessed 90.63% detection accuracy combining PD, HD, ALS and control subjects.

It has been observed that the gait signal features are prominent in detecting neuro diseases. Many state-of-the-art neuro disease detection methods have been explored. Considering the strengths and weaknesses of existing approaches, a non-invasive neuro disease detection method has been proposed in section 2. The proposed approach has been analyzed in detail in section 3, followed by the conclusion in section 4.

2. Materials & Methods

The proposed work employs naïve bayes and logistic regression in an ensemble framework known as ultaboost, empowered by the adaptively boosting power of Adaboost principle. The ensemble framework is tested on the most prominent gait signal features ascertained through three leading feature selection techniques Infinite Feature Selection (IFS), Infinite Latent Feature Selection (ILFS) and Sigmis Feature Selection (SFS). The most frequent features of ILFS, IFS and SFS are considered to design the proposed neurodegenerative detection engine. In addition, the Synthetic Minority Over-sampling Technique (SMOTE) technique has been employed to address the class imbalance issue. The strength of UltraBoost is that it combines heterogeneous classifiers and adaptively boosts the detection process.

The proposed neuro disease detection model consists of 5 distinct steps. At first, the data was aggregated. As the data is time series in nature, the traditional classifiers cannot be applied directly to the data. Therefore, the time series data of each sensor has been aggregated through the mean and standard deviation of sensor readings. The detailed procedure for realizing 27 features with distinct instances per subject has been explained in Section 3.1. In the second step, three binary and one multi-class dataset are prepared. In this process, instances of HD and Controls, PD and Controls, ALS and Controls are combined separately to form three binary class datasets. Similarly, the HD, PD, ALS and Control subjects' aggregated data are combined together to form a multi-class dataset. Subsequently, in the third step, the most prominent features of three binary datasets and one multi-class dataset are selected using Infinite Latent Feature Selection (ILFS), Infinite Feature Selection (INFS) and Sigmis feature selection (SFS); thus, creating nine binary and three multi-class datasets containing selected features. The twelve datasets contain little number of instances for each subject, and all these disease and control instances are prone to class imbalance issues. Therefore, more artificial instances are needed to counter the class imbalance. In this regard, the Synthetic Minority Over-sampling Technique (SMOTE) [21] has been employed at the fourth step to generate significant artificial samples to counter the class imbalance issue. In the final step, the twelve datasets formed are passed to an Adaboost ensemble framework known as UltraBoost where the Naïve Bayes and Logistic Regression acts as

the ensemble. The preprocessing of the dataset, various methods used and the ensemble detection techniques have been explained in the following sections.

2.1. The Dataset

It is mentioned earlier that the dataset used here has been derived by Hausdorff et al.[11,12]. The force-sensitive insole sensors of the subjects under study successfully recorded various time series parameters such as time intervals of left and right leg swing, stride, stance and double support. Apart from these time series information other parameters such as age, height, weight, gender and speed of gait have also been recorded. Since the stride, stance intervals are timeseries in nature and is recoded with elapsed Time (sec), therefore the data cannot be used directly for automated detection of the diseases. Therefore, as a preprocessing step the mean and standard deviation of these time feature attributes are calculated to obtain a unique record per subject. In a nutshell a total of 27 gait features pertaining to mean and standard deviation of gait signals are ascertained for each subject. A correlation matrix of all the gait features is generated and presented in Figure 3.

The correlation matrix presented in Figure 3 indicates that many features having correlation almost equal to 1. The high feature correlation necessitates to employ feature selection techniques to select uncorrelated features for neuro disease detection.

2.2. Feature Ranking to Feature Selection

As it is mentioned earlier, the proposed neuro disease detection approach ranks and selects relevant features at the preprocessing stage. The IFS [22], ILFS [23] and SFS [24] feature selection schemes are used as the rankers and feature selectors. As a ranker these schemes ranks the features from where relevant features are selected.

INFS considers a selection of features as a path in a graph, leading to infinite number of permits dealing with relevance and redundancy tenets [19]. This ensures the ranking of the features in a sophisticated way by inhibiting the computational complexity of the selection procedure. Truly interesting trait of this method is that it assesses the significance of a given feature whilst considering all the possible subsections of features. On the other hand, Infinite Latent Feature Selection shape a feature relevance, using pre-processing, graph weighting, and ranking [23]. This approach considers all sub-features by using power series of matrices. In contrast, SFS is a feature selection algorithm based on correlation method for processing continuous features [20]. Unlike INFS and ILFS, the SFS selects features based on feature correlation. Correlation is employed to find links between features from the sample data. In this algorithm, *t*-test is used on correlation coefficient to obtain the significant of features. The *t*-value can be estimated as –

$$t = s\sqrt{\frac{m-2}{1-s^2}} \tag{1}$$

where, m is the number of instances and s is the correlation coefficient of the input data. The *t*-test values and the correlation values are used to calculate significant value. The significant values are further compared with the critical values. If the significant value is more than the critical value, the feature is then selected, otherwise dropped.

2.3. UltraBoost – Combining heterogeneous classifiers

After getting requisite features through INFS, ILFS and SFS, an ensemble framework using Naïve Bayes and Logistic Regression has been prepared through a modern UltraBoost ensemble framework [25,26]. The UltraBoost framework facilitate to combine heterogeneous classifiers and adaptively boosts (AdaBoosts) them. The name UltraBoost is coined since it was formed at the Ultrasound Research Lab at the University of Pennsylvania. In the configuration of UltraBoost, Naive Bayes is the first classifier that works on nominal attributes and Logistic Regression that takes numeric attributes.

3. Results & Discussion

The result of the proposed model has been deduced in two broad ways. At first, the gait data pertaining to all the ALS, HD, PD and control subjects are combined together and relevant features are selected through ILFS, IFS and SFS. Both the ILFS and SFS; thus, creating a multi-class neurodegenerative disease detection model. Secondly, the ALS, HD and PD subjects are combined separately in order to create three diverse binary neurodegenerative disease detection environments. Similar to multi-class detection, revenant features were also selected using ILFS, IFS and SFS for binary neuro disease detection engines.

3.1. Multi-class Detection

In a multi-class detection system, as it is mentioned earlier, all the gait signal instances of ALS, HD, PD and control subjects are combined together; thus, having four class detection scenarios. The combined data when sent to ILFS, IFS and SFS, the ILFS selects all the features in the order of their feature importance. Similarly, SFS also selects all the features with respect to their t-test and correlation coefficient. Unlike ILFS and SFS, the IFS selects only 22 prominent features for neuro disease detection. When the ILFS, IFS and SFS induced selected features were sent to the UltraBoost ensemble framework, a good number of satisfactory results were derived. The detail results of multi-class neuro disease detection in terms of Accuracy, Average Absolute Error (AAE), Root Average Squared Error (RASE), False Negative Rate (FNR), False Positive Rate (FPR), Precision, Sensitivity and Specificity have been tabulated in Table 1.

From Table 1 it is evident that the IFS provides most discriminative features for neurodegenerative disease detection. It is because the IFS extracts only 22 features with the highest classification results. On the other hand, the ILFS states that all the features are equally discriminative by extracting all 27 features. With all the 27 features in hand, the ILFS exhibits 3.7% lower detection accuracy compared to the original set of features. The sensitivity and specificity were also degraded in the case of ILFS. In the meantime, the SFS claims all the features as the most discriminative by extracting all the features. Therefore, the classification results for both SFS and original features appears to be exactly the same. Since the IFS shows the highest classification results in the UltraBoost framework, the combination of UltraBoost and IFS has been proposed as a suitable neurodegenerative disease detection mechanism. The Receiver Operating Curve (ROC) and Precision Recall Curve (PRC) analysis had also been conducted to visualize the efficiency of the UltraBoost

and IFS combination for individual neuro diseases. Figure 4 represents the ROC and PRC curves of the proposed model for multi-class neurodegenerative disease detection.

The ROC curve shows that the proposed multi-class detection model designed through UltraBoost and IFS detects the ALS subjects brilliantly, followed by the controls. However, the controls are highly discriminated as compared to HD and PD subjects. In a nutshell, the proposed model demonstrates superior true positive rates with increase in false positives. A similar interpretation has been received through the precision recall curve.

3.2. Binary-class Detection

In a binary-class detection system, each non-control class instances are combined with control class instances, which are then sent for feature selection and detection. Therefore, we received three binary model datasets, viz., ALS+CONTROLS, PD+CONTROLS and HD+CONTROLS. Initially, the original data with 27 features of the binary mixtures were classified through the UltraBoost framework, followed by binary mixtures having selected features through ILFS, IFS and SFS.

For ALS+CONTROLS mixtures, the ILFS, IFS and SFS select 27, 5 and 18 features respectively. The ILFS selects all the 27 features with its own ranks leading to a detection accuracy of 99.1%. A similar amount of sensitivity (99.1%) has also been derived. The specificity received by the UltraBoost+ILFS is 98.9%. The UltraBoost+ILFS model also exhibits a very small number of false positives and false negatives. So far as IFS is concerned, the feature selection derived only 5 prominent features with a detection accuracy of 96.2%. This leads to a compromise of 2.9% detection accuracy as compared to UltraBoost+ILFS model. But here the winner is the UltraBoost+SFS model. The UltraBoost+SFS model shows the optimum detection accuracy of 99.1% with just 18 features. In other words, the UltraBoost+SFS model reveals the exact same results of as UltraBoost+ILFS but with only 18 features. Not only this, the UltraBoost+SFS model also shows little amount of average absolute error and root average squared error compared UltraBoost+ILFS model. Therefore, overall, for ALS prediction the UltraBoost+SFS model is the suitable predictor of gait signal data.

The proposed UltraBoost+SFS model has also been explored through concentration analysis. The concentration analysis of the UltraBoost+SFS model has been presented in Figure 5. It can be seen from Figure 5 is that the SFS features on the UltraBoost framework reveals inflated detection results. The false positives are also significantly reduced as compared to the multi-class detection model discussed earlier. The true negatives and true positives are satisfactorily improved. In a nutshell, the UltraBoost+SFS predictor is suitable for ALS prediction on gait signal data.

Similarly, the combination of controls and Huntington’s Disease subjects was also explored through the same UltraBoost+ILFS, UltraBoost+IFS and UltraBoost+SFS predictors. Here also the UltraBoost+SFS combination evolved as the most prominent predictor. The UltraBoost+SFS model reveals the highest detection accuracy only with 5 prominent features. The SFS also leads to a leading specificity of 94.5%. The average absolute error and root average squared errors were also observed to be far lower than their

counterparts. Like ALS, a concentration graph has also been prepared for UltraBoost+SFS model for predicting Huntington's Disease. Figure 6 shows the concentration graph of the proposed UltraBoost+SFS model both for HD and controls subjects. It can be seen from the concentration graph that the true negatives and true positives are significantly improved. At the same time, a deviation about the predictive model has also been observed. The deviation is about the ability of the model dealing with HD and control subjects. The proposed UltraBoost+SFS model generates a smaller number of false positives while predicting controls. In contrast, the same model also generates a significant number of false alarms while predicting HD subjects. Since, both the True positives and True negatives are in the acceptable range, the proposed model is also acceptable for discriminating HD and controls.

Finally, the binary mixture of Parkinson's and controls are analyzed through ILFS, IFS and SFS. A total of 124 instances have been explored separately through UltraBoost+ILFS, UltraBoost+IFS and UltraBoost+SFS. The ILFS returned 16 features, IFS and SFS returned 6 features respectively. Although the UltraBoost+ILFS model takes a bit higher number of features but the detection process reveals similar detection result. Nevertheless, the UltraBoost+IFS is the winner as it shows highest specificity, sensitivity and detection accuracy with just six features in hand. Although, the UltraBoost+SFS took a similar number of features as that of UltraBoost+IFS, but the former falls behind the latter due to lower accuracy and error rate.

Similar to ALS and HD, the proposed UltraBoost+IFS model has been explored through concentration graphs to understand the real detection capability of the model. Figure 7 shows the concentration plot for the UltraBoost+IFS model. The plot shows that the proposed model successfully generates significant number of true negatives and true positives. The best encouraging aspect is that the model generates the least number of false positives while predicting Parkinson's and control subjects.

4. Conclusion

Mostly the elderly population is afflicted by neurodegenerative disorders like Huntington's disease (HD), amyotrophic lateral sclerosis (ALS), and Parkinson's disease (PD). There is a proliferation of computational methods for discriminating and diagnosing neuro-related disorders. This article provided a machine learning-based system for accurately detecting PD, HD, ALS and Controls in a binary and multi-class detection environment. Together, Naive Bayes and Logistic Regression provided an impressive classification of neuro diseases in the state-of-the-art UltraBoost ensemble framework. The suggested method is novel since it uses a limited number of gait parameters to detect neuro disorders. The suggested method uses the three most cutting-edge feature selection algorithms infinite feature selection, infinite latent feature selection, and Sigmis feature selection to determine the most crucial gait features. It has been found that, in a multi-class setting, Infinite Feature Selection produces more accurate detection findings for Parkinson's and Huntington's disease. However, the proposed system proved more effective in binary classification environments. Although the proposed solutions perform relatively well in binary-class environments, they suffer slightly in multi-class settings due to the class imbalance issue. Even though

additional artificial samples have been generated using a synthetic minority oversampling technique to address the issue of class imbalance, it still remains a challenge.

Data Availability

The dataset used in this paper is publicly available via the PhysioNet Repository with the labels and link as: Gait in Neurodegenerative Disease Database (https://physionet.org/content/gaitndd/1.0.0/)

References

[1]. Hou Y, Dan X, Babbar M, Wei Y, Hasselbalch SG, Croteau DL et al. (2019) Ageing as a risk factor for neurodegenerative disease. Nature Reviews Neurology, 15, 565–81. 10.1038/s41582-019-0244-7 [PubMed: 31501588]

[2]. Organization, W.H. and others. (2007) Neurological disorders affect millions globally. International Journal of Health Care Quality Assurance, 20.

[3]. Nichols-Larsen DS, Kegelmeyer DA, Buford JA, Kloos AD, Heathcock JC and Basso DM (2016) Basal Ganglia Disorders: Parkinson's and Huntington's Diseases. Neurologic Rehabilitation: Neuroscience and Neuroplasticity in Physical Therapy Practice, McGraw-Hill Education, New York, NY.

[4]. National Institute of Neurological Disorders and Stroke. (2013) Amyotrophic Lateral Sclerosis (ALS) Fact Sheet | National Institute of Neurological Disorders and Stroke [Internet]. 16–916.

[5]. Burnfield M (2010) Gait analysis: normal and pathological function. Journal of Sports Science and Medicine, 9, 353.

[6]. Zhao A, Li J, Dong J, Qi L, Zhang Q, Li N et al. (2021) Multimodal gait recognition for neurodegenerative diseases. IEEE Transactions on Cybernetics, IEEE.

[7]. Beyrami SMG and Ghaderyan P (2020) A robust, cost-effective and non-invasive computer-aided method for diagnosis three types of neurodegenerative diseases with gait signal analysis. Measurement, Elsevier. 156, 107579.

[8]. Alaskar H, Hussain AJ, Khan W, Tawfik H, Trevorrow P, Liatsis P et al. (2020) A data science approach for reliable classification of neuro-degenerative diseases using gait patterns. Journal of Reliable Intelligent Environments, Springer. 6, 233–47.

[9]. Yan Y, Omisore OM, Xue Y-C, Li H-H, Liu Q-H, Nie Z-D et al. (2020) Classification of neurodegenerative diseases via topological motion analysis—A comparison study for multiple gait fluctuations. IEEE Access, IEEE. 8, 96363–77.

[10]. Ye Q, Xia Y and Yao Z (2018) Classification of gait patterns in patients with neurodegenerative disease using adaptive neuro-fuzzy inference system. Computational and Mathematical Methods in Medicine, Hindawi. 2018.

[11]. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG et al. (2000) Components of a new research resource for complex physiologic signals. PhysioBank, PhysioToolkit, and Physionet,.

[12]. Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D and Goldberger AL (2000) Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. Journal of Applied Physiology, American Physiological SocietyBethesda, MD.

[13]. Nordin NHD, Muthalif AGA and M Razali MK (2018) Control of transtibial prosthetic limb with magnetorheological fluid damper by using a fuzzy PID controller. Journal of Low Frequency Noise, Vibration and Active Control, SAGE Publications Sage UK: London, England. 37, 1067–78.

[14]. Pramanik M, Pradhan R, Nandy P, Bhoi AK and Barsocchi P (2021) Machine Learning Methods with Decision Forests for Parkinson's Detection. Applied Sciences, Multidisciplinary Digital Publishing Institute. 11, 581.

[15]. Biagetti G, Crippa P, Falaschetti L, Tanoni G and Turchetti C (2018) A comparative study of machine learning algorithms for physiological signal classification. Procedia Computer Science, Elsevier B.V. p. 1977–84. 10.1016/j.procS.2018.07.255

[16]. Joshi D, Khajuria A and Joshi P (2017) An automatic non-invasive method for Parkinson's disease classification. Computer Methods and Programs in Biomedicine, Elsevier. 145, 135–45. [PubMed: 28552119]

[17]. Fraiwan L and Hassanin O (2021) Computer-aided identification of degenerative neuromuscular diseases based on gait dynamics and ensemble decision tree classifiers. Gadekallu TR, editor. PLOS ONE, 16, e0252380. 10.1371/journal.pone.0252380

[18]. Zhao A, Qi L, Dong J and Yu H (2018) Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases. Knowledge-Based Systems, 145, 91–7. 10.1016/j.knosys.2018.01.004

[19]. Wu Y and Krishnan S (2010) Statistical Analysis of Gait Rhythm in Patients With Parkinson's Disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18, 150–8. 10.1109/TNSRE.2009.2033062 [PubMed: 20650700]

[20]. Khorasani A and Daliri MR (2014) HMM for Classification of Parkinson's Disease Based on the Raw Gait Data. Journal of Medical Systems, 38, 147. 10.1007/s10916-014-0147-5 [PubMed: 25354664]

[21]. Chawla N. v, Bowyer KW, Hall LO and Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research.

[22]. Roffo G, Castellani U, Vinciarelli A, Cristani M and others. (2020) Infinite feature selection: a graph-based feature filtering approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE.

[23]. Roffo G, Melzi S, Castellani U and Vinciarelli A (2017) Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. Proceedings of the IEEE International Conference on Computer Vision, Institute of Electrical and Electronics Engineers Inc. p. 1407–15. 10.1109/ICCV.2017.156

[24]. Blessie EC and Karthikeyan E (2012) Sigmis: A feature selection algorithm using correlation based method. Journal of Algorithms \& Computational Technology, SAGE Publications Sage UK: London, England. 6, 385–94.

[25]. Moustafa AF, Cary TW, Sultan LR, Schultz SM, Conant EF, Venkatesh SS et al. (2020) Color Doppler Ultrasound Improves Machine Learning Diagnosis of Breast Cancer. Diagnostics, Multidisciplinary Digital Publishing Institute. 10, 631. [PubMed: 32854253]

[26]. Venkatesh SS, Levenback BJ, Sultan LR, Bouzghar G and Sehgal CM (2015) Going beyond a first reader: a machine learning methodology for optimizing cost and performance in breast ultrasound diagnosis. Ultrasound in Medicine\& Biology, Elsevier. 41, 3148–62. [PubMed: 26354997]

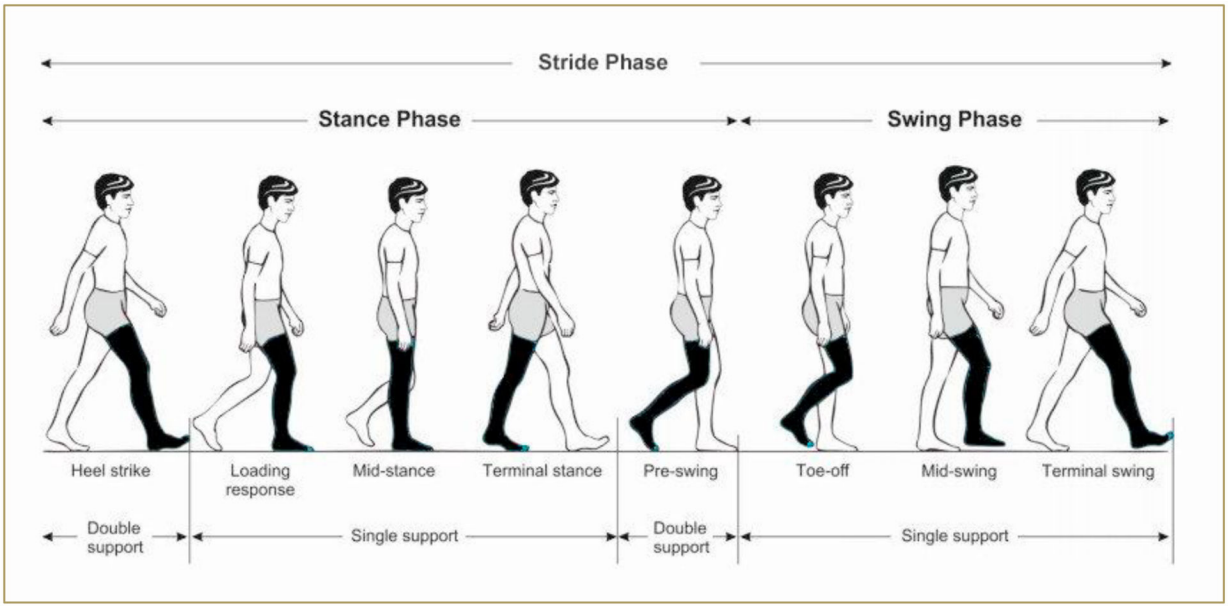
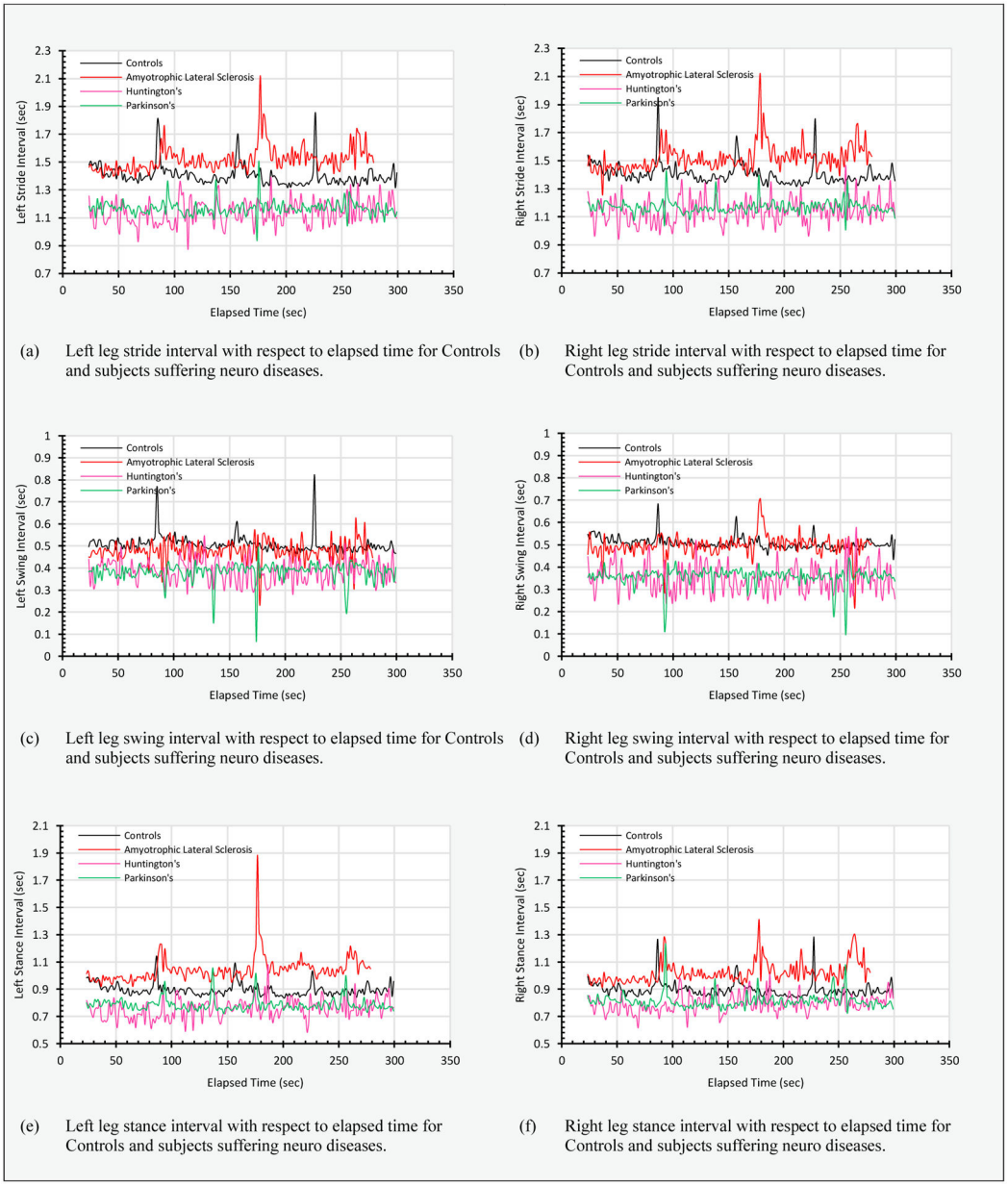


Figure 1.
Human Gait cycle demonstrating stride, stance and swing [13]



(a) Left leg stride interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

(b) Right leg stride interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

(c) Left leg swing interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

(d) Right leg swing interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

(e) Left leg stance interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

(f) Right leg stance interval with respect to elapsed time for Controls and subjects suffering neuro diseases.

Figure 2.
Stance, Swing and Stride intervals of controls and subjects suffering neuro diseases.

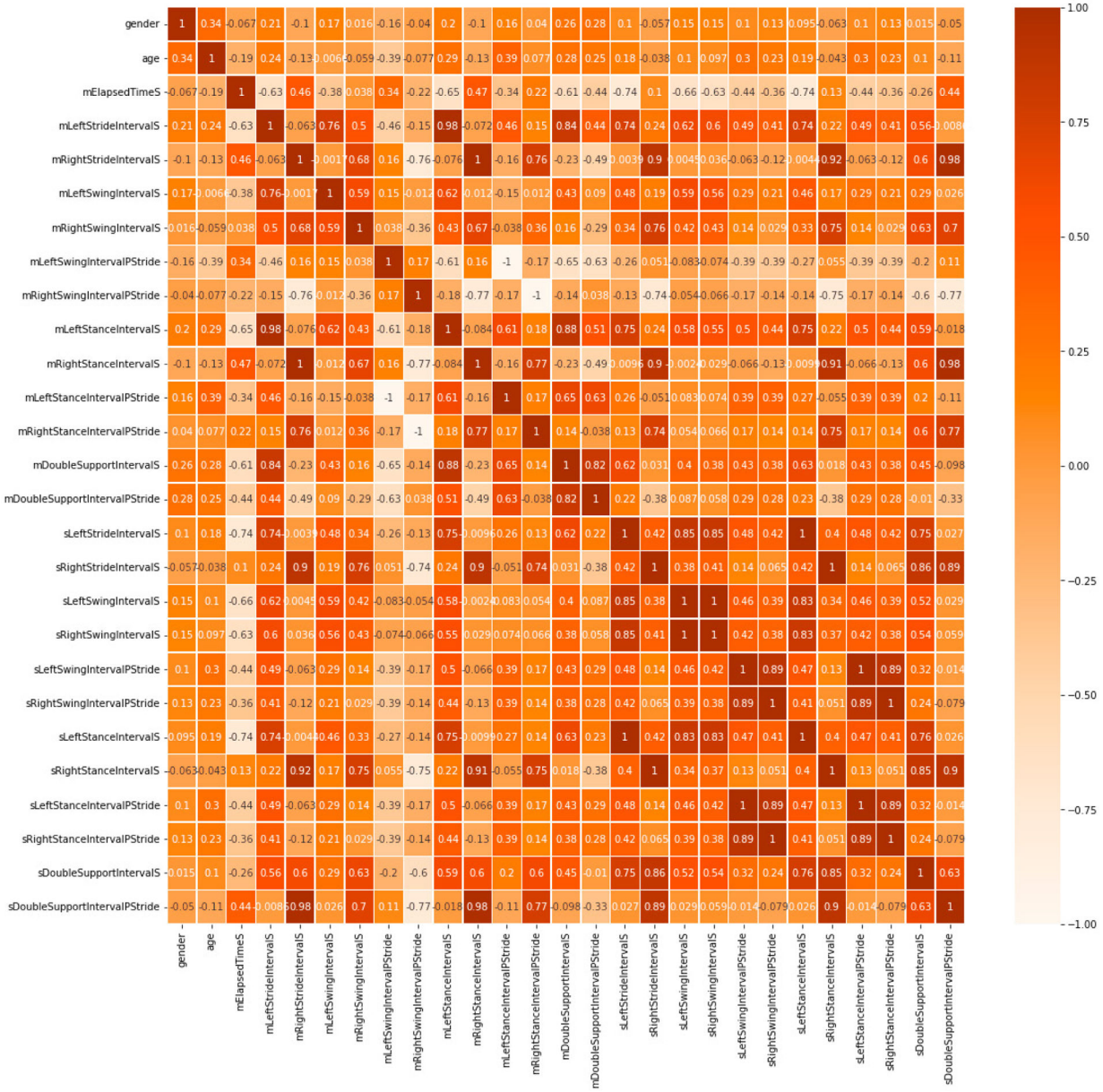


Figure 3.
The feature correlation matrix of the 27 features of gait dataset

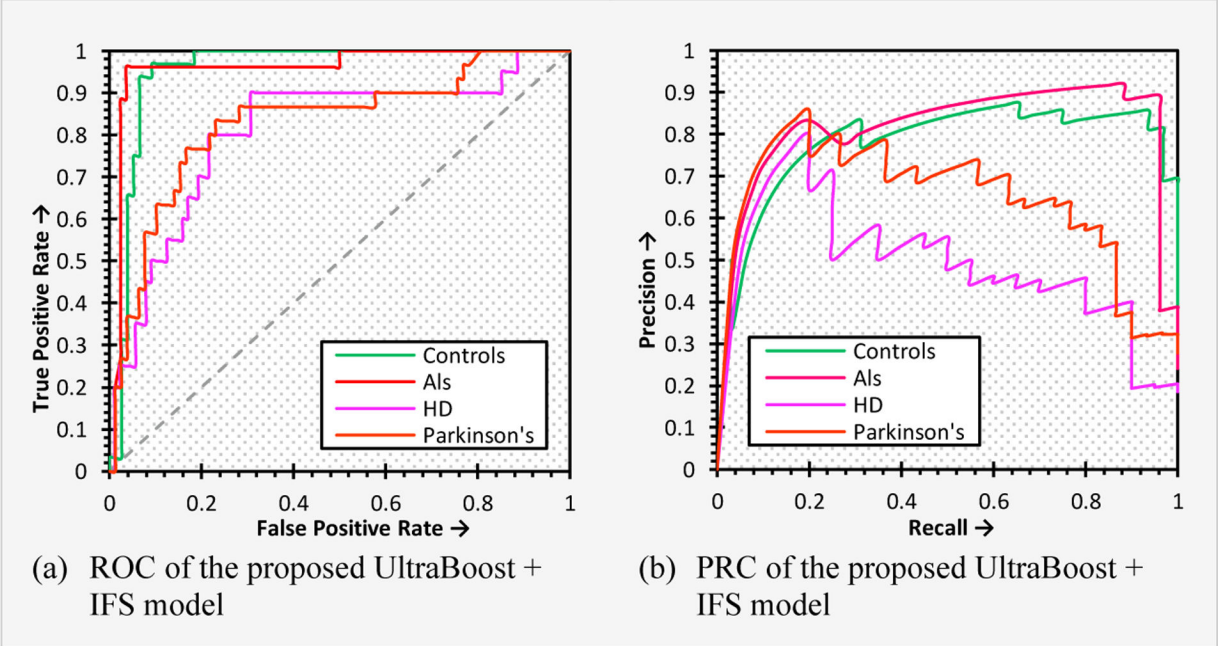


Figure 4.
ROC and PRC of the proposed multi-class UltraBoost and IFS neuro disease detection model

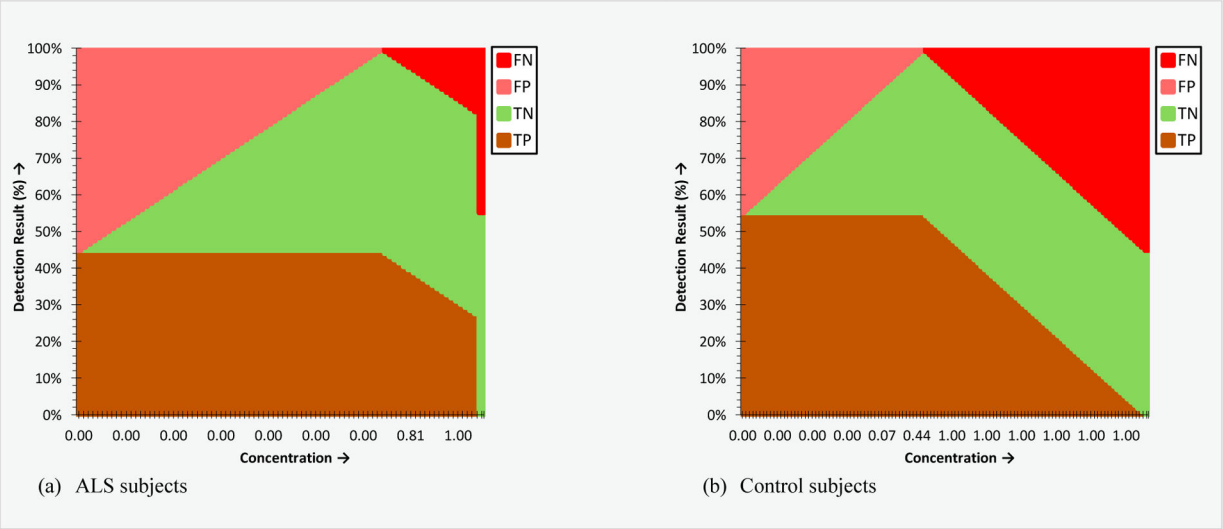


Figure 5.
Detection Result with concentration of the proposed UltraBoost+SFS binary-class detection model for ALS and Control subjects

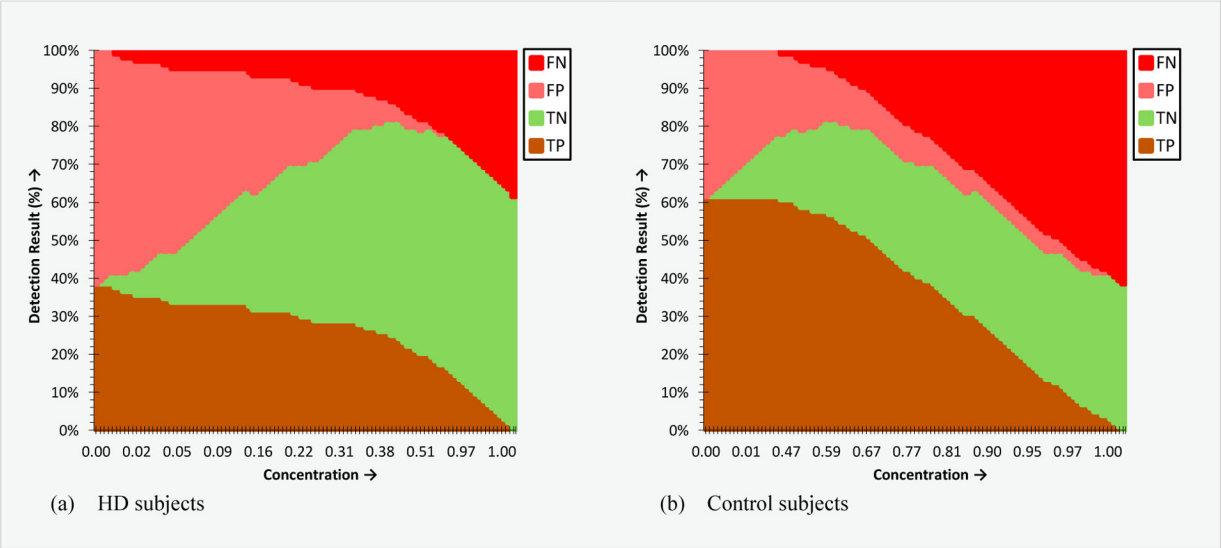


Figure 6.
Detection Result with concentration of the proposed UltraBoost+SFS binary-class detection model for HD and Control subjects

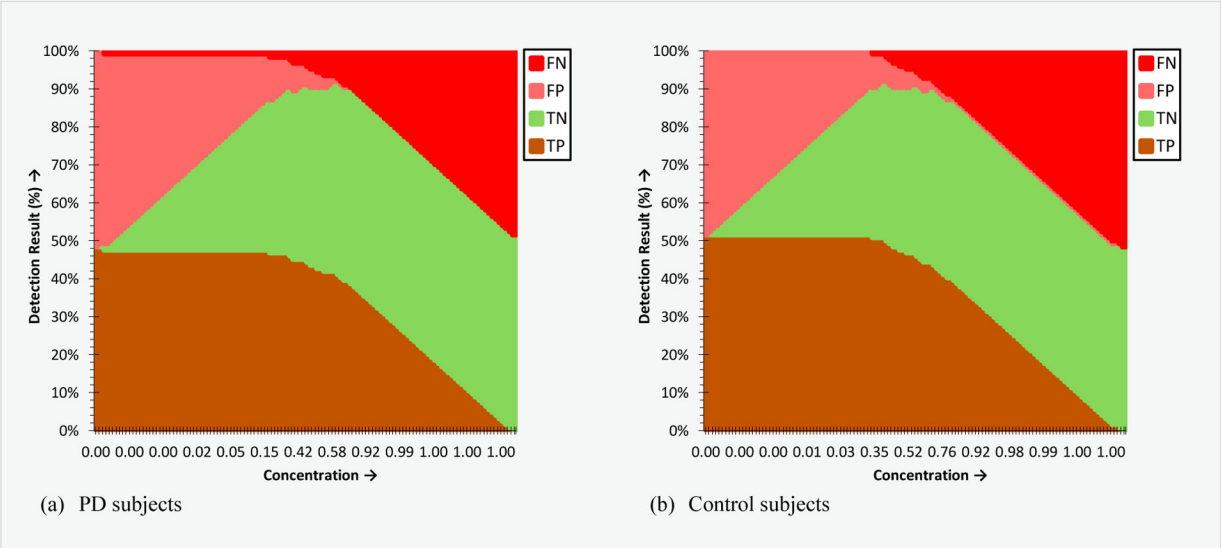


Figure 7.
Detection Result with concentration of the proposed UltraBoost+IFS binary-class detection model for PD and Control subjects

Table 1.

All Neurodegenerative disease detection (n = 108) in a multi-class detection environment of Amyotrophic Lateral Sclerosis, Huntington's, Parkinson's, and Controls using various feature selection schemes through the Naïve Bayes and Logistic Regression ensemble in the UltraBoost ensemble framework.

PARAMETERS	ORIGINAL	ILFS	IFS	SFS
Features (f)	27	27	22	27
Accuracy	0.713	0.676	0.769	0.713
AAE	0.158	0.172	0.146	0.158
RASE	0.329	0.334	0.306	0.329
Precision	0.713	0.682	0.772	0.713
Sensitivity	0.713	0.676	0.769	0.713
FNR	0.287	0.324	0.231	0.287
FPR	0.101	0.106	0.074	0.101
Specificity	0.899	0.894	0.926	0.899
Kappa	0.613	0.566	0.689	0.613

Table 2.

Amyotrophic Lateral Sclerosis detection (n = 116) in a binary detection environment of Amyotrophic Lateral Sclerosis and Controls using various feature selection schemes through the Naïve Bayes and Logistic Regression ensemble in the UltraBoost ensemble framework.

PARAMETERS	ORIGINAL	ILFS	IFS	SFS
Features (f)	27	27	5	18
Accuracy	0.983	0.991	0.962	0.991
AAE	0.072	0.069	0.086	0.062
RASE	0.190	0.171	0.208	0.163
Precision	0.983	0.992	0.962	0.992
Sensitivity	0.983	0.991	0.962	0.991
FNR	0.017	0.009	0.038	0.009
FPR	0.021	0.011	0.052	0.011
Specificity	0.979	0.989	0.948	0.989
Kappa	0.965	0.983	0.918	0.983

Table 3.

Huntington's Disease detection ($n = 104$) in a binary detection environment of Huntington's and Controls using various feature selection schemes through the Naïve Bayes and Logistic Regression ensemble in the UltraBoost ensemble framework

PARAMETERS	ORIGINAL	ILFS	IFS	SFS
Features (f)	14	6	14	5
Accuracy	0.942	0.942	0.942	0.942
AAE	0.098	0.104	0.070	0.076
RASE	0.232	0.232	0.209	0.222
Precision	0.942	0.942	0.942	0.944
Sensitivity	0.942	0.942	0.942	0.942
FNR	0.058	0.058	0.058	0.058
FPR	0.064	0.074	0.074	0.055
Specificity	0.936	0.926	0.926	0.945
Kappa	0.878	0.877	0.877	0.879

Table 4.

Parkinson's Disease detection (n = 124) in a binary class detection environment of Parkinson's and Controls using various feature selection schemes through the Naïve Bayes and Logistic Regression ensemble in the UltraBoost ensemble framework.

PARAMETERS	ORIGINAL	ILFS	IFS	SFS
Features (f)	27	16	6	6
Accuracy	0.976	0.984	0.984	0.960
AAE	0.040	0.024	0.038	0.064
RASE	0.159	0.121	0.140	0.195
Precision	0.976	0.984	0.984	0.960
Sensitivity	0.976	0.984	0.984	0.960
FNR	0.024	0.016	0.016	0.040
FPR	0.024	0.016	0.016	0.041
Specificity	0.976	0.984	0.984	0.959
Kappa	0.952	0.968	0.968	0.919