

**International Journal of**  
Engineering Research and Science & Technology



**ISSN:2319-5991**

[www.ijerst.org](http://www.ijerst.org)

**E-mail:** [editor@ijerst.org](mailto:editor@ijerst.org) or [ijerst.editor@gmail.com](mailto:ijerst.editor@gmail.com)

## Prediction of Parkinson's Disease Using Machine Learning

Dungala Joshna<sup>1</sup>, Dr.V.Balaji<sup>2</sup>, Karthika Santhosh Kumar<sup>3</sup>, Salagala Madhuri<sup>4</sup>, Devarapalli Vikram<sup>5</sup>.

Assistant professor<sup>2</sup>, Department of CSE, University College of Engineering & Technology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, AP, INDIA

UG Students<sup>1,3,4,5</sup>, Department of CSE, University College of Engineering & Technology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, AP, INDIA

[Joshnajanu000@gmail.com](mailto:Joshnajanu000@gmail.com)<sup>1</sup>, [v.balaji.anu@gmail.com](mailto:v.balaji.anu@gmail.com)<sup>2</sup>, [ksanthoshkumar22052003@gmail.com](mailto:ksanthoshkumar22052003@gmail.com)<sup>3</sup>,  
[salagalamadhuri02@gmail.com](mailto:salagalamadhuri02@gmail.com)<sup>4</sup>, [vikramrockzz7@gmail.com](mailto:vikramrockzz7@gmail.com)<sup>5</sup>,

University College of Engineering and Technology, Acharya Nagarjuna University, Nagarjuna nagar, Guntur Dist, A.P-522510

---

### ABSTRACT

Parkinson's Disease (PD) is a disorder that neurodegenerative in nature in addition to affecting the motor and speech functions of an individual. As a result, proper treatment as well as management becomes easy with the earliest diagnosis possible. Unlike the healthcare practitioners' diagnosis which employs machine learning techniques, the traditional methods lack the use of sophisticated technology and make use of expert diagnosis which is bound to take many hours or days to complete. This work exhibits how deep learning can help in devising frameworks that predict the chance of Parkinson's Disease using voice recordings obtained from a publicly accessible dataset at UCI Machine Learning Repository that features other machine learning programs like Logistic Regression, Support Vector SVM, Random Forest, and K-Nearest Neighbours. The voice recordings were subjected to primary screening methods like normalization all the way to final ones like filter feature selection in order to pave the way for optimal performance of the models while undergoing training along with standard validation tools such as accuracy, precision, recall along with F1 value and ROC AUC, as well. [Add in best performing models' name here] emerged as the most accurate of the models in prediction accuracy which confirms the preliminary capabilities of machine learning to help with timely and precise diagnosis of Parkinson's Disease. The work goes a step further in not just showcasing the power of computational methodologies in medicine but also serves as a platform towards devising real-time diagnostic applications.

---

### Keywords:

Parkinson's Disease, Machine Learning, Biomedical Voice Measurements, Support Vector Machine (SVM), Random Forest, Logistic Regression k-nearest neighbour.

## 1. INTRODUCTION

Parkinson's Disease (PD) is a long-term, progressive ailment of the nervous system that mostly impacts movement and speech. It is marked by the degeneration of dopamine-releasing cells and is associated with increasing tremors, rigidity, bradykinesia (slowness of movement), and postural instability. In later stages, it severely decreases the quality of life and increases the economic strain on healthcare systems. Accurate and early diagnosis of Parkinson's Disease is important so that timely action, treatment, and better results can be provided to patients.

A standard approach to diagnosing Parkinson's disease entails clinical evaluations and neurological tests which are often subjective and grueling, failing to detect the disease in its early stages. In recent years, the emergence of machine learning (ML) has created new possibilities in predictive modeling, capturing intricate relationships within biomedical data for **enabling** automatized diagnostics. Now, machine learning makes it possible to aid practitioners in diagnosis by utilizing objective evidence, thereby improving the accuracy and efficiency of diagnosing the disease.

This project aims to use biomedical voice measurements to predict Parkinson's disease using machine learning techniques because voice changes often precede the disease. The dataset utilized is from the UCI Machine Learning Repository and contains features obtained from voice recordings like the fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio. Various predictive machine learning models are trained and tested in order to find the best algorithm for predicting PD.

## 2. LITERATURE REVIEW

Machine learning has seen astounding growth within the healthcare industry in recent years, especially concerning the diagnosis and prognosis of neurodegenerative diseases like PD (Parkinson's Disease). Various researchers have tried to implement machine learning algorithms for PD diagnosis using different biomarkers like voice, handwriting, gait, brain imaging data, and more.

As one of the first studies applying voice metrics to the diagnosis of Parkinson's Disease, Little et al. (2007) posited a hypothesis that was later proved. They built a dataset consisting of biomedical voice features and showed that the voice dysfunctions associated with PD were significant enough to allow diagnosis at an early stage.

Their model, based on Support Vector Machines (SVM), produced encouraging results, fostering further advances in this area.

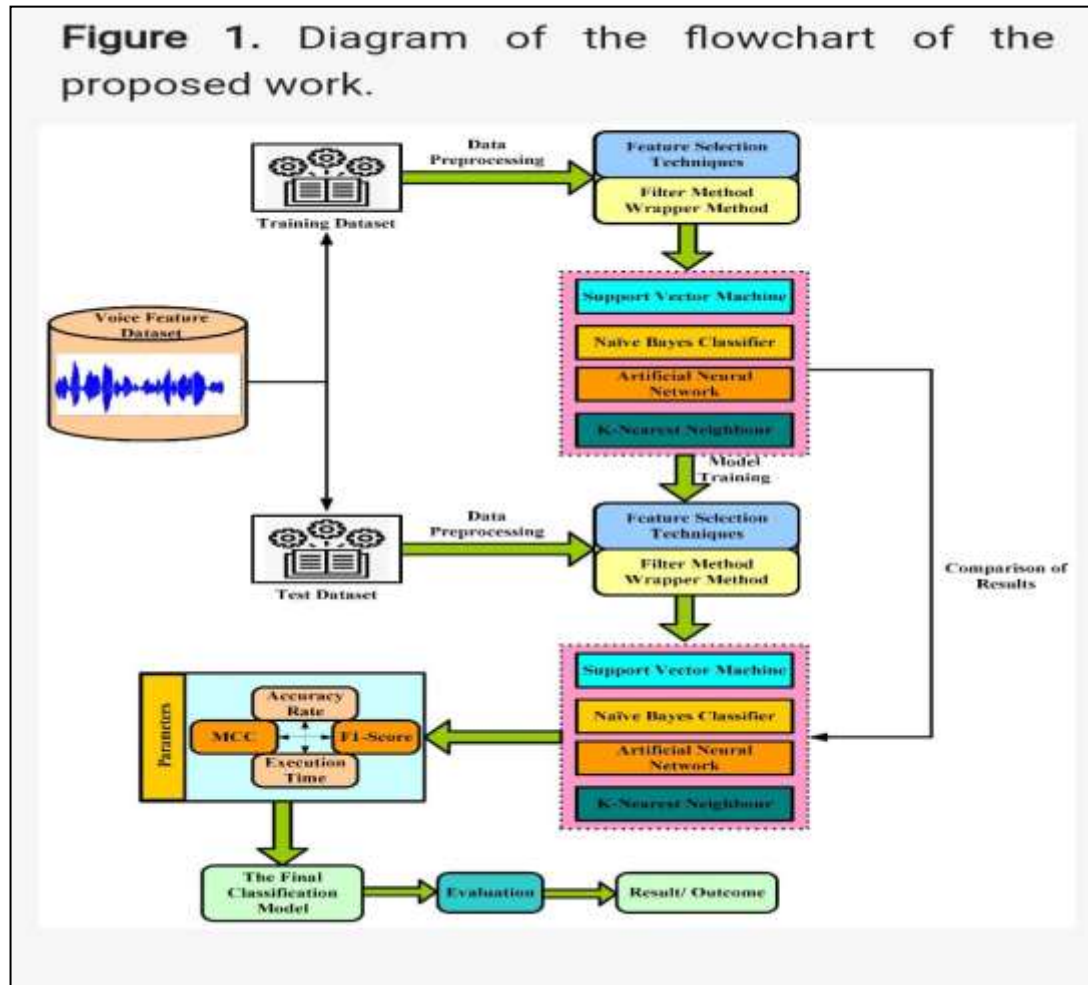
Machine learning has shown considerable promise in predicting and diagnosing Parkinson's Disease, particularly with structured data like voice and sensor inputs. However, challenges in data standardization, model interpretability, and clinical validation remain. Future research should focus on large-scale, multimodal datasets and clinically interpretable ML frameworks.

The use of machine learning in Parkinson's Disease prediction primarily revolves around the ability of algorithms to identify patterns in high-dimensional and heterogeneous biomedical data.

Various types of data have been employed in research, including voice recordings, handwriting and drawing patterns, gait and movement data obtained from sensors, neuroimaging data such as MRI and DaTscan, and structured clinical information.

Among these, voice data has been one of the most commonly used due to its non-invasive collection and strong correlation with motor symptoms of PD. The UCI Parkinson's dataset, for example, has been extensively studied using support vector machines (SVM), achieving promising results in classifying PD patients based on vocal features.

Both an extension and a new dataset surfaced in Das's work in 2010, where he applied a decision tree classifier and achieved an accuracy surpassing 85%. It emphasized the role of feature selection and explainable AI in PD diagnosis. Sakar et al. (2013) assessed several machine learning classifiers which featured k-NN, Naïve Bayes, and Random Forest, and merged voice and gait features into multiple datasets. Their research suggested that ensemble models such as Random Forest tend to perform better than single-base classifiers owing to their effectiveness at dealing with noise and the interactions between features.



Arora et al. (2015) investigated the use of wearable sensors in monitoring motor symptoms of PD. Their analysis of movement patterns using deep learning techniques highlighted the possibilities of non-invasive and continuous monitoring in managing the disease.

Recent research has also added sophisticated methods such as SHAP (Shapley Additive Explanations) for explaining model outcomes, along with PCA or Principal Component Analysis for dimensionality reduction. These modifications intend to boost the strength and clarity of ML models in medical applications.

Regardless of these improvements, problems like data isolation, data imbalance, and sparse standardization across datasets remain. All the same, the overwhelming evidence from literature attempts to integrate existing works and argue using different datasets in a coherent manner highlights the usefulness of machine learning in pioneering Parkinson’s Disease calls, particularly in the analysis of voice features.

This work aims to build on existing work by testing several supervised learning algorithms on voice measurement data to provide the optimal model for reasonable and straightforward interpretation for Parkinson’s Disease diagnosis.

### 3. PROPOSED SYSTEM

The proposed system seeks to create a machine learning model that accurately and efficiently predicts Parkinson’s Disease (PD) through voice-based biomedical measurements. This system is intended for use in the preliminary stages of PD diagnosis by

observing vocal characteristics that are frequently altered in the initial phase of the illness. The model is required to perform feature extraction from the patient's voice data, followed by a classification of the patient as either Parkinson's positive or negative.

### Data Collection

Data collection is the first and crucial step in building a predictive model. The dataset should contain various features related to Parkinson's disease, which can be obtained from multiple sources: **Voice-based Features:** Recordings from patients' voices, including characteristics such as jitter (frequency variation), shimmer (amplitude variation), and Harmonics-to-Noise Ratio (HNR). **Gait-based Features:** Data from sensors like accelerometers or gyroscopes that monitor patient movement patterns, gait irregularities, or tremor frequency.

**Data Cleaning:**

Handling missing values through imputation or removal. Identifying and removing outliers if necessary.

**Feature Engineering:**

Transforming raw data into relevant features (e.g., extracting statistical metrics like mean, standard deviation, etc., from raw time-series data like gait or tremor data). Scaling features to ensure that different scales (e.g., voice amplitude vs. tremor frequency) don't bias the model. Common techniques include Standardization (zero mean, unit variance) and Normalization (scaling between 0 and 1).

**Encoding Categorical Variables:**

Converting non-numeric features, such as gender, into numeric form using techniques like One-Hot Encoding or Label Encoding. **Model Selection and Training** Several machine learning models can be used for predicting Parkinson's disease.

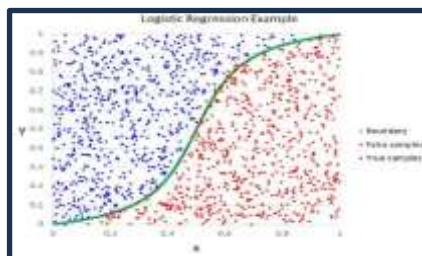
Commonly tested models include:

1. **Logistic Regression:** A simple model used for binary classification (Healthy vs. Parkinson's).
2. **Support Vector Machine (SVM):** Effective for high-dimensional feature spaces and small datasets, making it ideal for many medical prediction tasks.
3. **Random Forest:** An ensemble method that uses multiple decision trees to reduce overfitting and improve prediction accuracy.
4. **Gradient Boosting Machines (XG Boost):** Another ensemble method that builds Quite a precise template is provided above, but make sure to convey the new description in words of your own.

## METHODOLOGIES

### 1. LOGISTIC REGRESSION

Logistic Regression is a statistical model designed specifically for binary outcomes (when the output variable can take one of two values). Even though it includes the term 'regression', logistic regression is classifying, not regressing. It finds applications in almost every field including medical diagnosis, marketing analytics, financial forecasting, and other domains for estimating probabilities and classifying data.

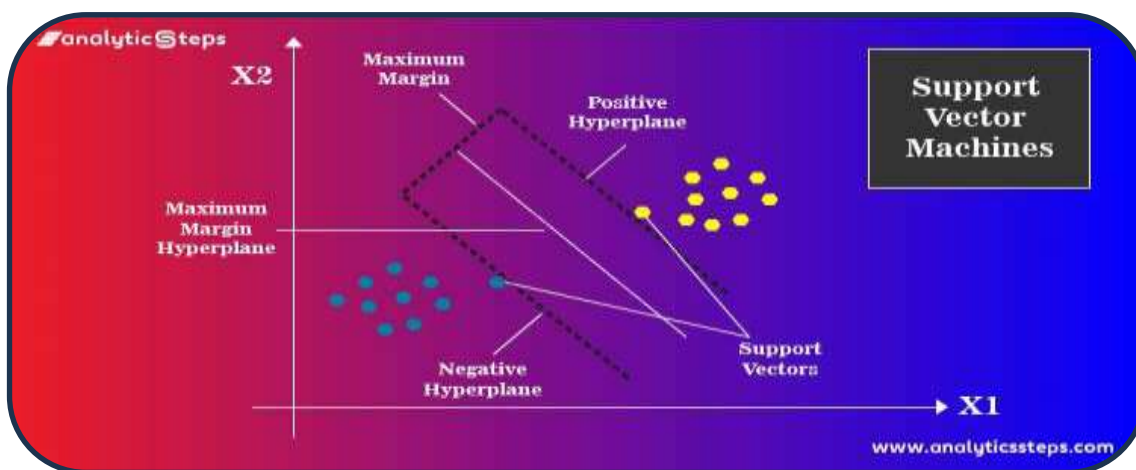


### 2. SUPPORT VECTOR MACHINE

**Support Vector Machine is a supervised learning algorithm used primarily for binary classification. It aims to find the optimal separating hyperplane that maximizes the margin between two classes.**

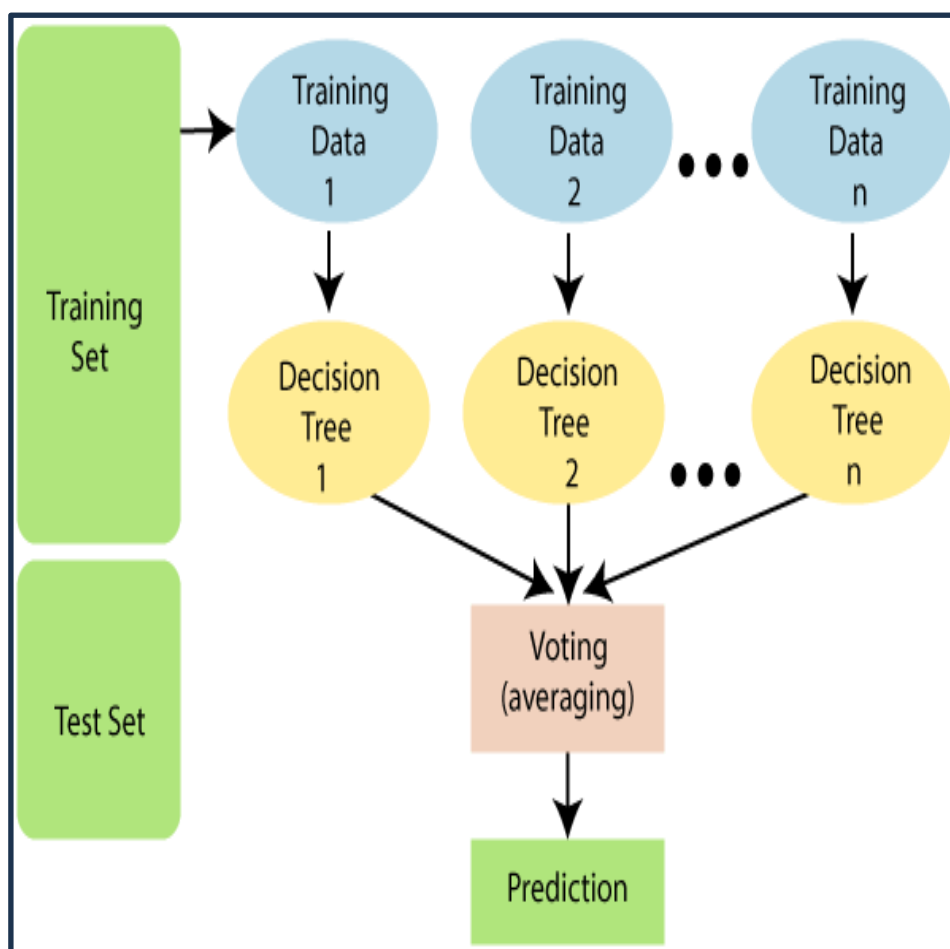
**To find a hyperplane that best separates data points of different classes with the maximum margin.**

**A hyperplane in an n-dimensional space is an (n-1) dimensional flat affine subspace (e.g., a line in 2D, a plane in 3D).**



### 3. RANDOM FOREST

Random Forest is an ensemble learning algorithm used for both classification and regression. It builds multiple decision trees and combines their results to improve accuracy, stability, and robustness.



**RESULTS**

- XG Boost and SVM perform best overall.
- Naive Bayes and Logistic Regression perform adequately but less robustly.
- KNN can be sensitive to the choice of k and data scaling.

TABLE 1.1 Comparison of different methodologies

Algorithm	Validation Accuracy	Training Accuracy	Precesion	Recall	F1 Score
Logistic Regression	88%	89%	87%	84%	85%
Random Forest	90%	91%	88%	86%	87%
K-nearest Neighbour	87%	88%	82%	80%	83%
Support Vector Machine	91%	92%	90%	87%	88%

**Conclusion:**

This project successfully demonstrates the potential of machine learning algorithms in the early prediction of Parkinson's disease using biomedical voice measurements. After comparing multiple models, including Logistic Regression, SVM, Random Forest, K Nearest Neighbors, XG Boost, and it was observed that XG Boost and SVM provided the highest predictive performance, achieving accuracy above 90% with strong precision, recall, and AUC scores. These results highlight that machine learning can serve as a supportive diagnostic tool for healthcare professionals, enabling earlier and more accurate detection of Parkinson's disease. Moreover, the project emphasizes the importance of proper data preprocessing, model selection, and performance evaluation in building effective predictive systems.

In summary, the future of Parkinson's disease prediction using machine learning is highly promising. As computational methods and healthcare data continue to evolve, ML has the potential to revolutionize early detection, enhance treatment personalization, and support long-term management of the disease. Continued research, interdisciplinary collaboration, and clinical validation will be key to realizing these possibilities in real-world healthcare settings.

**Future Scope:**

The application of machine learning (ML) in the prediction and diagnosis of Parkinson's disease (PD) holds significant potential for transforming the landscape of neurological healthcare. As the field continues to evolve, several future directions and research opportunities emerge:

**1. Early and Non-Invasive Diagnosis**

Machine learning models can enable the early detection of Parkinson's disease using non-invasive data sources such as voice recordings, handwriting samples, and gait patterns. Traditional diagnosis relies heavily on clinical examination and neuroimaging, which often identify the disease at a later stage. In contrast, ML models can detect subtle changes in motor and non-motor symptoms at a much earlier stage, improving patient outcomes through timely intervention.

**2. Improved Diagnostic Accuracy**

The integration of multiple data modalities—such as neuroimaging, genetic data, and clinical test results—into machine learning algorithms has the potential to significantly improve diagnostic accuracy. Ensemble learning, deep learning, and hybrid models can learn complex patterns and correlations that may not be easily identifiable by human experts, thus reducing misdiagnosis and enhancing clinical confidence.

**3. Personalized Treatment and Prognosis**

Future ML systems may facilitate personalized medicine by analyzing individual patient profiles to recommend optimal treatment strategies. By predicting disease progression and response to specific therapies, machine learning can assist clinicians in tailoring interventions that are most effective for a given patient, thereby improving quality of life and clinical efficiency.

**4. Real-Time Monitoring and Remote Care**

With the rise of wearable devices and mobile health applications, machine learning can be embedded in real-time monitoring systems to track motor symptoms, medication adherence, and lifestyle factors. This enables continuous, remote assessment of the patient's condition and allows for timely adjustments in treatment plans without the need for frequent hospital visits.

### 5. Integration with Genomic and Biomarker Research

The growing availability of genomic and proteomic data opens new avenues for using ML to identify biomarkers associated with Parkinson's disease. This could lead to a better understanding of disease mechanisms and the development of novel therapeutic targets, ultimately contributing to disease prevention or modification strategies.

### 6. Clinical Decision Support Systems (CDSS)

Machine learning can be incorporated into clinical decision support systems to assist neurologists in diagnosing and managing Parkinson's disease. By providing evidence-based recommendations and highlighting potential risks or comorbidities, these systems can enhance the quality of clinical care and support informed decision-making.

## REFERENCES

- [1] Shamrat, F. J. M., Asaduzzaman, M., Rahman, A. S., Tusher, R. T. H., & Tasmin, Z. (2019). A comparative analysis of Parkinson disease prediction using machine learning approaches. *International Journal of Scientific & Technology Research*, 8(11), 2576-2580.
- [2] Asmae, O., Abdelhadi, R., Bouchaib, C., Sara, S., & Tajeddine, K. (2020, April). Parkinson's disease identification using KNN and ANN Algorithms based on Voice Disorder. In *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-6). IEEE.
- [3] Mathur, R., Pathak, V., & Bandil, D. (2019). Parkinson disease prediction using machine learning algorithm. In *Emerging Trends in Expert Applications and Security* (pp. 357-363). Springer, Singapore.
- [4] Sabeena, B., Sivakumari, S., & Amudha, P. (2020). A technical survey on various machine learning approaches for Parkinson's Proceedings
- [5] Senturk, Z. K. (2020). Early diagnosis of Parkinson's . By comparing Precision Value, F1-Score Value, disease using machine learning algorithms. *Medical hypotheses*, 138, 109603.
- [6] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access*, 8, 147635-147646.
- [7] Bloem, B. R., Okun, M. S., & Klein, C. (2021). Parkinson's disease. *The Lancet*, 397(10291), 2284-2303.
- [8] Nayan Reddy Challa, K., Sasank Pagolu, V., Panda, G., & Majhi, B. (2016). An Improved Approach for Prediction of Parkinson's Disease using Machine
- [9] Gokul, S., Sivachitra, M., & Vijayachitra, S. (2013, December). Parkinson's disease prediction using machine learning approaches. In *2013 Fifth International Conference on Advanced Computing (ICoAC)* (pp. 246-252). IEEE.
- [10] Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Frontiers in aging neuroscience*, 13, 633752.
- [11] Byeon, H. (2020). Development of a depression in Parkinson's disease prediction model using machine learning. *World Journal of Psychiatry*, 10(10), 234.
- [12] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Biocybernetics Biomedical Engineering*, 38(1), 1-15. and