Literature Survey

The study [2] used the voice dataset from UCI Machine Learning repository, Orange v2.0b, and Weka v3.4.10 for analysis. SVM achieved 88.9% accuracy, Random Forest achieved 90.26% accuracy, and Naïve Bayes obtained an accuracy of 69.23%.

Speech signals from 252 subjects for PD diagnosis were used in [3] where significant differences were found in voice recordings of PD patients compared to healthy individuals. Deep neural networks were employed for voice data analysis. Ensemble of Classifiers achieved an accuracy of 95%, F1-score of 0.97 and Mathews Correlation Coefficient (MCC) of 0.87.

The study [4] highlights the importance of feature selection and algorithm choice in Parkinson's disease diagnosis. Performance evaluation of 20 ML algorithms for Parkinson's disease using acoustic datasets was compared. Different cross validation techniques like LOOCV and Stratified K-Fold cross validation and hyperparameter tuning techniques like Grid Search were used for evaluation. The dataset from UC Irvine ML Repository, consisting of 195 voice recordings and a dataset collected at Istanbul University, including 188 records from Parkinson's patients and 64 records from healthy individuals were used and an accuracy of 96.41% and 97.35% were obtained using ensemble voting classifiers.

EEG, MRI, handwriting, and speech data were used for classification in [5] and it was concluded that deep learning showed promising results. Future directions include integration of different data modalities for improved accuracy, longitudinal data analysis for disease progression forecasting, emphasis on explainable AI and federated learning for collaborative model training.

The study [6] looks at how different fats (lipids) in the blood might help us understand and predict how Parkinson's disease (PD) progresses in patients. Researchers found over 900 types of fats in the blood of PD patients and used machine learning to see if these fats could predict how patients' motor skills and mood would change after two years. The models demonstrated optimal performance using baseline serum lipids, achieving a normalized root mean square error of 0.7 for both UPDRS III and Geriatric Depression Scale scores. Thus it was concluded that changes in blood fats are strongly connected to PD and can help in predicting the disease.

2600 disease data points (MDS – UPRDS) and protein data from 248 patients were used in [7]. Principal Component Analysis (PCA) was applied to reduce the data dimension to 500. Random forest achieved 27.54 Mean Squared Error (MSE) and Symmetric Mean Absolute Percentage (SMAPE) of 69.50, outperforming the multiple regression model, which had an MSE of 55.81 and an SMAPE of 93.73.

The study [8] introduces the use of three predictive models for Parkinson's disease: Fully Complex-Valued Radial Basis Function network (FC RBF), Meta-Cognitive Fully Complex-Valued Radial Basis Function network (Mc-FCRBF), and Extreme Learning Machine (ELM). These models predicted the severity of the disease using Unified Parkinson's Disease Rating Scale (UPDRS). For patients who have not received treatment, the UPDRS scale ranges from 0 to 176.

Feature selection was done through Feature Importance and Recursive Feature Elimination (RFE) methods. Classification, regression trees, Artificial Neural Networks, SVM were used where SVM with RFE achieved an accuracy of 0.9384 with the least number of voice features.

The dataset used in [10] comprised of 240 records with 46 acoustic features extracted from three voice recordings of 80 patients. Different machine learning models along with feature selection were applied and KNN classifier with wrapper – based feature selection achieved an accuracy of 88.33%.

Automated diagnostic models were developed in paper [11] using Multilayer Perceptron, , Random Forest, BayesNet and Boosted Logistic Regression. The study found that Boosted Logistic Regression provided the best performance, with an accuracy of 97.159% and an area under the ROC curve of 98.9%.

The study [12] proposes the use of various machine learning models and the processing speed of these models are also analysed, to understand their suitability for light weight mobile applications in the ubiquitous computing environment. The authors conclude that if the number of voice samples increase, the performance of the predictive models could further increase.

The proposed ensemble learning technique that uses stacking in [13] surpasses existing methods such as SVM, KNN, RF, DT, MLP, Stacking Classifier, and Logistic Regression with 94.87% accuracy, 81.99% Matthews Correlation Co-efficient (MCC), and 94.52% F1 score.

The study [14] used the UCI’s Parkinson’s Telemonitoring Vocal Data Set of patients and a neural network was created to predict the severity of the disease. Further machine learning models were applied to detect PD.

Research was carried out on the MDVP audio data consisting of 30 PWP and healthy people for training of machine learning models. Random Forest obtained an accuracy of 91.83% and 0.95 sensitivity [15].

The authors of [16] used Waikato Environment for Knowledge Analysis (WEKA) for data mining and algorithms like SVM, ANN, KNN were applied. Ensemble methods like AdaBoost, Bagging and MLP were used. Combined approach of Adaboost.M1 and MLP with KNN achieved an accuracy of 91.28%.

Voice data were collected through mPower containing data of both healthy and PD patients. Using an app from I-Phone, a clinical observational study was conducted by Sage Bionetworks [17]. Here raw audio was cleaned with Voice Activation Detection (VAD) algorithm before applying machine learning algorithms.

The study [18] aimed to identify individuals who have high risk for PD by combining plasma proteins with clinical-demographic variables. Using data from the UK Biobank, researchers applied machine learning to select 22 plasma proteins and various clinical-demographic factors for predicting PD risk. The Light Gradient Boosting Machine model achieved an area under the curve (AUC) of 0.832, with key predictors including education, age, serum creatinine and past traumatic brain injury.

References:

[1] Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', IEEE Transactions on Biomedical Engineering.

[2] Sriram, T. V., M. Venkateswara Rao, G. S. Narayana, D. S. V. G. K. Kaladhar, and T. Pandu Ranga Vital. "Intelligent Parkinson disease prediction using machine learning algorithms." *Int. J. Eng. Innov. Technol* 3, no. 3 (2013): 1568-1572. 1

[3] Yuan, Linlin, Yao Liu, and Hsuan-Ming Feng. "Parkinson disease prediction using machine learning-based features from speech signal." *Service Oriented Computing and Applications* 18, no. 1 (2024): 101-107. 2

[4] Saleh, Shawki, Bouchaib Cherradi, Oussama El Gannour, Soufiane Hamida, and Omar Bouattane. "Predicting patients with Parkinson's disease using Machine Learning and ensemble voting technique." *Multimedia Tools and Applications* 83, no. 11 (2024): 33207-33234. 3 4

[5] Ibrahim, Aiesha Mahmoud, and Mazin Abed Mohammed. "A comprehensive review on advancements in artificial intelligence approaches and future perspectives for early diagnosis of Parkinson's disease." *International Journal of Mathematics, Statistics, and Computer Science* 2 (2024): 173-182.

[6] Galper, Jasmin, Giorgia Mori, Gordon McDonald, Diba Ahmadi Rastegar, Russell Pickford, Simon JG Lewis, Glenda M. Halliday, Woojin S. Kim, and Nicolas Dzamko. "Prediction of motor and non-motor Parkinson’s disease symptoms using serum lipidomics and machine learning: a 2-year study." *npj Parkinson's Disease* 10, no. 1 (2024): 123.

[7] Qian, Shuxin Jeremy. "Predicting Parkinson's Disease Progression with Random Forests." In *Proceedings of the 2024 4th International Conference on Bioinformatics and Intelligent Computing*, pp. 191-199. 2024.

[8] Gokul, S., M. Sivachitra, and S. Vijayachitra. "Parkinson's disease prediction using machine learning approaches." In *2013 fifth international conference on advanced computing (ICoAC)*, pp. 246-252. IEEE, 2013.

[9] Senturk, Zehra Karapinar. "Early diagnosis of Parkinson’s disease using machine learning algorithms." *Medical hypotheses* 138 (2020): 109603.

[10] Saeed, Faisal, Mohammed Al-Sarem, Muhannad Al-Mohaimeed, Abdelhamid Emara, Wadii Boulila, Mohammed Alasli, and Fahad Ghabban. "Enhancing Parkinson’s disease prediction using machine learning and feature selection methods." *Computers, Materials and Continua* 71, no. 3 (2022): 5639-5658.

[11] Challa, Kamal Nayan Reddy, Venkata Sasank Pagolu, Ganapati Panda, and Babita Majhi. "An improved approach for prediction of Parkinson's disease using machine learning techniques." In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pp. 1446-1451. IEEE, 2016.

[12] Kumar, Tapan, Pradyumn Sharma, and Nupur Prakash. "Comparison of Machine learning models for Parkinson’s Disease prediction." In *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0195-0199. IEEE, 2020.

[13] Mall, Pawan Kumar, Rajesh Kumar Yadav, Arun Kumar Rai, Vipul Narayan, and Swapnita Srivastava. "Early warning signs of Parkinson’s disease prediction using machine learning technique." *Journal of Pharmaceutical Negative Results* (2022): 4784-4792.

[14] Raundale, Pooja, Chetan Thosar, and Shardul Rane. "Prediction of Parkinson’s disease and severity of the disease using Machine Learning and Deep Learning algorithm." In *2021 2nd International Conference for Emerging Technology (INCET)*, pp. 1-5. IEEE, 2021.

[15] Govindu, Aditi, and Sushila Palwe. "Early detection of Parkinson's disease using machine learning." *Procedia Computer Science* 218 (2023): 249-261.

[16] Mathur, Richa, Vibhakar Pathak, and Devesh Bandil. "Parkinson disease prediction using machine learning algorithm." In *Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018*, pp. 357-363. Springer Singapore, 2019.

[17] Wroge, Timothy J., Yasin Özkanca, Cenk Demiroglu, Dong Si, David C. Atkins, and Reza Hosseini Ghomi. "Parkinson’s disease diagnosis using machine learning and voice." In *2018 IEEE signal processing in medicine and biology symposium (SPMB)*, pp. 1-7. IEEE, 2018.

[18] You, Jia, Linbo Wang, Yujia Wang, Jujiao Kang, Jintai Yu, Wei Cheng, and Jianfeng Feng. "Prediction of Future Parkinson Disease Using Plasma Proteins Combined With Clinical-Demographic Measures." *Neurology* 103, no. 3 (2024): e209531.