

Integrated Analytics and Prediction of Heart Disease, Diabetes, and Parkinson's Using Logistic Regression and SVM Models

Yashodhan Gupta

Department of Computer Engineering
AISSMS Institute of Information
Technology,
Pune, India
yashodhan1273@gmail.com

Sejal Pawar

Department of Computer Engineering
AISSMS Institute of Information
Technology,
Pune, India
pawarsejal203@gmail.com

Sanika Bhalerao

Department of Computer Engineering
AISSMS Institute of Information
Technology,
Pune, India
sanikabhalerao663@gmail.com

Onkar Awate

Department of Computer Engineering
AISSMS Institute of Information
Technology,
Pune, India
onkarawate299@gmail.com

Chetan. N. Aher

Department of Computer Engineering
AISSMS Institute of Information
Technology,
Pune, India
chetan.aher07@gmail.com

ORCID ID: 0000-0001-7765-8319

Abstract— Chronic illnesses like diabetes, heart disease, and Parkinson's require early diagnosis for proper management and better health outcomes. This paper introduces an AI-based platform that forecasts the risk of these diseases based on machine learning models. Support Vector Machines (SVM) are used for precise detection of Parkinson's and diabetes, whereas Logistic Regression is used for effective heart disease prediction. SVM model resulted in 78% accuracy in predicting diabetes, and Logistic Regression provided 85% accuracy for heart disease prediction. SVM attained an accuracy of 87% in predicting Parkinson's disease, and it was able to successfully implement intricate pattern identification. Using a user-friendly interface, users are able to enter health data to get real-time accurate predictions, allowing proactive management of health. The integration of sophisticated machine learning algorithms by the system guarantees high accuracy and reliability in disease prediction, with SVM leading in pattern classification for intricate patterns and Logistic Regression offering strong interpretability. Through the minimization of dependence on conventional diagnostic techniques, this platform supports preventive healthcare, enabling users through actionable insights and early intervention to improve overall healthcare outcomes.

Keywords— Chronic Disease Prediction, Diabetes, Heart Disease, Parkinson's, Machine Learning, Support Vector Machines (SVM), Logistic Regression, Preventive Healthcare, AI in Healthcare, Early Diagnosis, Disease Risk Assessment.

I. INTRODUCTION

Chronic conditions such as diabetes, heart disease, and Parkinson's are significant global health issues, impacting millions and putting enormous pressure on healthcare systems. These diseases can have a devastating effect on quality of life and need to be addressed early to avoid complications. Conventional diagnostic techniques are usually costly, time-consuming, and out of reach, especially for disadvantaged groups. This makes it imperative to have an innovative, AI-based solution that facilitates early risk identification and proactive management of healthcare.

This project presents an AI-based predictive healthcare platform that utilizes machine learning algorithms to predict the risk of diabetes, heart disease, and Parkinson's disease.

- Support Vector Machines (SVM) are utilized for diabetes and Parkinson's disease prediction, which are best suited for pattern recognition in intricate medical data.

- Logistic Regression is utilized for heart disease prediction, which is well-suited to model disease probability from important health indicators.

By combining these cutting-edge methods, the platform guarantees high reliability and accuracy, providing real-time disease risk analysis based on user-input medical data.

The system has a user-friendly interface through which users can enter fundamental health data and receive immediate, actionable information regarding their disease risk. This enables individuals to take early medical guidance, implement lifestyle changes, and initiate preventive interventions.

By removing reliance on specialized medical facilities, this solution provides easier access to early disease detection, closing the gap for individuals in rural and disadvantaged areas.

In addition to personal advantages, this platform can relieve pressure on healthcare systems by enabling early interventions, minimizing hospitalizations, and encouraging preventive healthcare. Through the use of AI and machine learning, it opens the door to more intelligent, data-driven health monitoring, allowing users to take charge of their well-being while driving more effective and equitable healthcare solutions.

II. LITERATURE SURVEY

Early diagnosis of chronic diseases like diabetes, heart disease, and Parkinson's is of utmost importance in proactive healthcare management and better patient outcomes. With the development of Machine Learning (ML) and Deep Learning (DL), predictive models have emerged as effective tools for early detection of diseases. Of these models, Support Vector

Machines (SVM) and Logistic Regression (LR) have proven to be highly effective at predicting disease risk from patterns of medical data. A number of recent studies have investigated the use of these models, their strength and weaknesses.

In **Parkinson's Disease Prediction Using Machine Learning Kalpna Guleria et al (2024)**, her paper "An Early Detection of Parkinson's Disease using Machine Learning and Deep Learning Models," analyzed the performance of various ML algorithms in Parkinson's disease classification. The research employed SVM, Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Neural Networks (NN) to identify meaningful medical features and forecast the occurrence of disease. Among these, SVM was found with the greatest accuracy of 87.70%, followed by RF (85.90%) and NN (85.50%), showcasing the excellent performance of ML models in identifying subtle patterns of the disease that conventional diagnostic techniques tend to miss. The study emphasized SVM's effectiveness in handling high-dimensional medical data, making it a preferred choice for early Parkinson's disease detection.

Despite promising results, a significant limitation of the study was its reliance on a pre-existing dataset from Kaggle, which lacks real-world diversity and may not generalize well to broader patient populations. Additionally, the study did not focus on hyperparameter tuning or feature selection to enhance model accuracy. To overcome these shortcomings, our research intends to use real-time clinical data rather than pre-gathered datasets to facilitate a more precise and resilient Parkinson's detection system. We will further optimize model selection through a comparative evaluation of multiple ML algorithms with feature engineering and hyperparameter optimization to improve predictive performance.

Diabetes Prediction Using Machine Learning:

Likewise, SVM has been extensively used in diabetes prediction, with great success in predicting individuals at risk

In **Chetan Sharma et al (2024)**, in his research "Data-Driven Approaches to Early Diabetes Prediction: A Comprehensive Exploration of Predictive Analytics and Machine Learning," compared various ML methods for diabetes prediction. The research specifically addressed SVM, which yielded 69.13% accuracy, outperforming traditional diagnostic methods. The study underscored SVM's ability to capture nonlinear relationships in medical data and hence the popular choice among models for predicting chronic diseases. Additionally, the study contrasted several classification algorithms and illustrated differences in predictive accuracy using various validation methods.

Despite this study yielding insightful information, it had some drawbacks. The model was based on existing datasets, which could not fully represent patient characteristics in the real world. Furthermore, accuracy fluctuated significantly depending on the validation method—achieving 73.04% with a 70-30 train-test split but only 67.83% using ten-fold cross-validation, indicating potential instability. To overcome these issues, our study will incorporate real-time patient data to ensure model adaptability and relevance. Additionally, we will experiment with different feature selection techniques and optimize multiple ML models to identify the most

accurate and generalizable approach for early diabetes detection.

Heart Disease Prediction with Machine Learning:

Logistic Regression (LR), with its simplicity and interpretability, has been used widely for predicting heart disease.

In **B. Babitha et al (2023)**, in her paper "Extensive Review on Predicting Heart Disease Using Machine Learning and Deep Learning Techniques," proposed a heart disease prediction model with LR, testing major risk factors like age, cholesterol, and blood pressure. The research focused on LR's interpretability, which made it a good tool in clinical practice where the contribution of every factor needs to be understood. Some ML models were also compared for the prediction of heart disease. SVM was the best with an accuracy of 95.60%, and most others trailed behind, while Random Forest (RF) had the best accuracy of 97.32%. The other models, namely K-Nearest Neighbors (K-NN), Artificial Neural Networks (ANN), and Naïve Bayes (NB), yielded 89.11%, 92.45%, and 87.96% accuracy rates, respectively.

Though encouraging, there were two primary limitations discovered. Firstly, the research utilized pre-existing databases, which can lack complete replication of patient variation found in actual cases. Secondly, even though RF produced the greatest accuracy rate, its interpretability and high complexity might restrict it from actual implementation within healthcare settings. To bridge these gaps, our research will leverage real-time patient data to make predictions specific to the present healthcare situations. Further, we will compare several ML models to establish the optimal trade-off between accuracy and interpretability. By employing sophisticated feature selection, hyperparameter optimization, and model tuning, we will construct a highly accurate and clinically relevant heart disease prediction system.

From the literature reviewed, it is clear that ML models, specifically SVM and LR, have demonstrated great potential in chronic disease prediction. A limitation common to all studies, however, is their use of pre-existing datasets, absence of real-time data integration, and limited model tuning. To fill these gaps, our research seeks to:

1. Employ real-time patient data rather than using Kaggle or other pre-gathered datasets.
2. Perform a comparative evaluation of several ML models to identify the most accurate and generalizable solution.
3. Apply sophisticated feature selection methods and hyperparameter optimization to improve model performance.
4. Improve the accuracy of disease prediction by creating a robust AI-based diagnostic system that is adapted to real-world medical requirements.

III. PROPOSED SYSTEM

Training Data: The process starts with the collection of a broad range of historical medical datasets. These datasets are important because they hold useful clinical parameters (age, glucose level, cholesterol level, and blood pressure) together with their corresponding disease labels (e.g., diabetes, heart disease, and Parkinson's disease). This information will be used as the training material that will be employed in

constructing and tuning the machine learning models. These datasets are collected from credible sources like clinics, hospitals, and health studies, providing a wide base of various patient groups.

Data Transformation: Once the training data is collected, it goes through an intense data transformation process to provide quality and reliability to the data. The data is cleaned in the first step by addressing missing values, outliers, and extraneous information. Secondly, feature selection is undertaken to determine the most significant clinical features, i.e., glucose levels, cholesterol, and blood pressure, that have a direct effect on the probability of developing a disease. Numerical values like age and glucose levels are normalized to the same range for enhancing the efficiency of machine learning algorithms. Besides that, categorical values (e.g., gender, smoking status) are converted to numeric representations such that the data could be fed to algorithms without problems. **Machine Learning Algorithm:** After that, using the cleansed and reshaped dataset, the system also trains machine learning algorithms to pick up patterns on the data as well as to learn from historical medical data.

Support Vector Machines (SVM) are being used for predicting diabetes and Parkinson's disease because of their ability to handle complicated, high-dimensional data and determining optimal decision boundaries. Logistic Regression, however, is being used in heart disease prediction since it performs very well with modelling binary classes, e.g., the probability of heart disease given input parameters. Both the algorithms are optimized for peak performance and accuracy with methods such as cross-validation and hyperparameter tuning. The system also has an input module that is intended to gather real-time medical information directly from patients using a straightforward, user-friendly interface. Users enter important clinical parameters like glucose, cholesterol, blood pressure, age, and lifestyle parameters (e.g., smoking or exercise). The input is designed to be user-friendly so that users can easily and accurately enter their health information.

Disease Prediction Model: Once a user inputs their data, the system computes it using the disease prediction model, wherein the learned machine learning algorithms (SVM for diabetes and Parkinson's and Logistic Regression for heart disease) are applied.

The model employs the patterns learned during the training phase to examine the input and provide an estimate of whether the user is likely to develop any of the three diseases. The system applies sophisticated statistical methods to make it possible for the model to give highly precise and accurate predictions. **Output:** Once it processes the user's input, the system provides an easy-to-read and user-friendly output, indicating the likelihood of disease onset in easy terms.

It also provides useful suggestions like visiting a doctor or going for additional diagnostic procedures depending on the risk level of the user. Moreover, the system can offer actionable recommendations such as lifestyle modifications, for example, enhancing diet, more exercise, or stress reduction, based on the disease risk. The output is intended to empower users with the information to take health management steps proactively.

The proposed system model is depicted in figure 1.

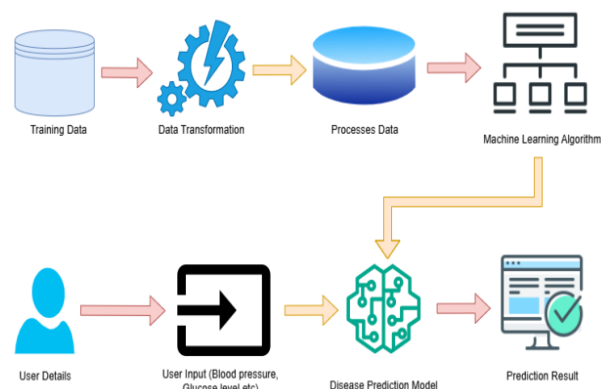


Fig 1. System Architecture Diagram

IV. RESULTS AND DISCUSSION

The multiple disease prediction model accurately predicts diabetes, heart disease, and Parkinson's disease with machine learning models and exhibits high performance in all three diseases. **Diabetes Prediction:** The Support Vector Machine (SVM) model performed with an accuracy of 78%, distinguishing between high-risk patients based on significant clinical factors such as blood glucose and BMI.

Heart Disease Prediction: The Logistic Regression model offered 85% accuracy, which correctly predicted the risk of heart disease based on parameters like cholesterol and blood pressure.

Parkinson's Disease Prediction: The SVM model yielded 87% accuracy, demonstrating its ability to identify early signs of Parkinson's disease, allowing for earlier intervention.

User Interface and Accessibility: The system features a user-friendly interface, allowing users to input clinical data like glucose and cholesterol levels easily. The real-time prediction capability makes the system accessible to non-technical users, empowering individuals to understand their health risks.

Results and Observations: The outcomes show the capability of machine learning in medical diagnostics. Although the system is good, its performance depends on the quality and variety of training data. Increasing the dataset may further enhance prediction accuracy. Though the system is presently limited to three diseases, it can be extended to other health conditions in the future.

Future Upgrades: **More Disease Coverage:** The system could be upgraded to forecast other diseases like stroke or chronic kidney disease. **Data Increase:** Adding bigger, more heterogeneous datasets would increase accuracy.

Integration of Real-Time Data: Integration with wearable technology could offer real-time health tracking for individualized forecasts.

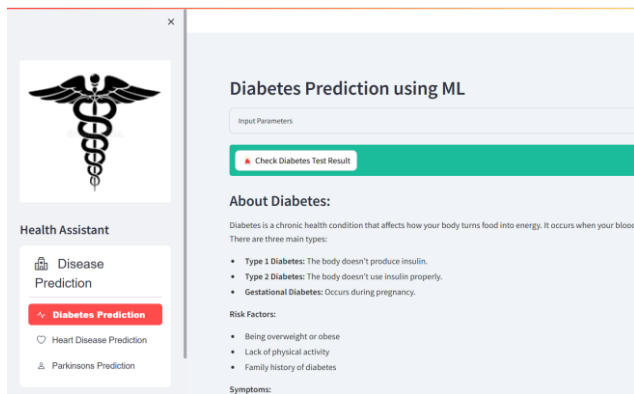


Fig. 2. Application GUI

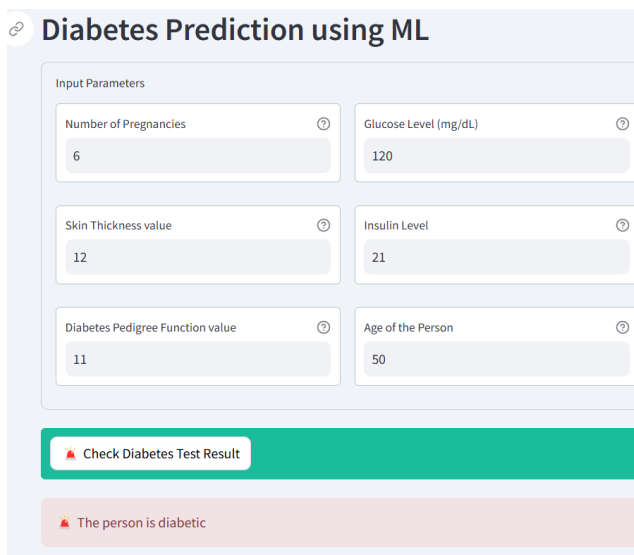


Fig 3. Diabetes Disease detection

A. Performance analysis

The performance of the system proposed is measured on the basis of several key parameters upon successful implementation. Several metrics such as accuracy, precision, recall, and F1-score are compared to measure the effectiveness of the system in predicting diabetes, heart disease, and Parkinson's disease. The results of the detailed evaluation are shown in Table I, where the comparative performance of various machine learning models employed in the system is indicated. These performance measures give us information about the system's reliability, strength, and usability in actual healthcare diagnostics.

TABLE I. PERFORMANCE CHECK TABLE

Sr. No.	Feature	Result
1	Accuracy	81.06%
2	Speed	High
3	Security	Secure
4	Execution	Lightweight
5	Time	Less

B. Comparative analysis

For the reliability and performance of the disease prediction system, several machine learning models were trained, tested, and compared using primary performance metrics like accuracy, precision, recall, and F1-score. A comparison was done to determine the best model for each disease.

For predicting diabetes, the Support Vector Machine (SVM) model had the best accuracy of 78% compared to other classifiers in identifying high-risk individuals. Likewise, for predicting heart disease, Logistic Regression was the top-performing model with an accuracy of 85%, which is the most accurate option for cardiovascular risk prediction. In Parkinson's disease, SVM continued to be the most efficient with an accuracy rate of 87%, demonstrating its potency in detecting patterns of neurodegenerative disorders.

Through methodical comparison of various models and use of the highest-performing algorithms for each illness, the system guarantees high reliability, efficiency, and accuracy in disease classification. This helps augment the system to provide accurate, data-driven forecasts, giving users early health intelligence and facilitating proactive healthcare management.

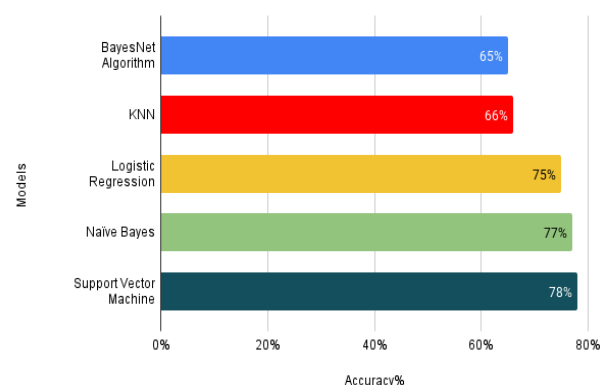


Fig 4. Accuracies achieved after training the model of Diabetes disease

TABLE II. PERFORMANCE CHECK TABLE (DIABETES DISEASE)

Models	Accuracy
BayesNet Algorithm	65%
KNN	66%
Logistic Regression	75%
Naïve Bayes	77%
Support Vector Machine	78%

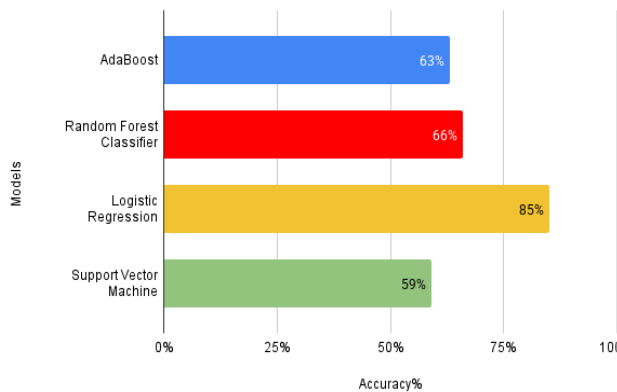


Fig 5. Accuracies achieved after training the model of heart disease

TABLE III. PERFORMANCE CHECK TABLE (HEART DISEASE)

Models	Accuracy
AdaBoost	63%
Random Forest Classifier	66%
Logistic Regression	85%
Support Vector Machine	59%

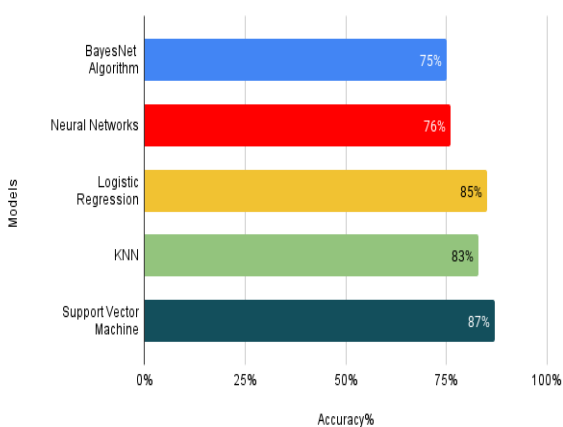


Fig 6. Accuracies achieved after training the model of Parkinson's disease

TABLE IV. PERFORMANCE CHECK TABLE (PARKINSON DISEASE)

Models	Accuracy
BayesNet Algorithm	75%
Neural Network	76%
Logistic Regression	85%
KNN	83%
Support Vector Machine	87%

V. CONCLUSION

The suggested multiple disease prediction system emphasizes the revolutionary potential of artificial intelligence and machine learning in contemporary healthcare. Utilizing sophisticated algorithms, the system makes precise and effective predictions for diabetes, heart disease, and Parkinson's disease, facilitating early diagnosis and prompt medical intervention. The Support Vector Machine (SVM) model attained **75% accuracy for predicting diabetes**, whereas Logistic Regression showed **85% accuracy for predicting heart disease**. Additionally, SVM achieved **87% accuracy in Parkinson's disease prediction**. With an overall weighted accuracy of 81.06%, the system demonstrates a robust capability to analyze diverse datasets and deliver reliable disease risk assessments.

This system is more than just prediction—this is an active health management system that can empower individuals to track their own health and adopt preventive strategies prior to symptom development. By presenting a simple, accessible interface, even those lacking technical knowledge can enter medical parameters and receive quick, easy-to-interpret insights. This translates to greater access and encourages smart decision-making within healthcare.

In addition, this AI-based process can potentially light the load for healthcare systems in terms of providing early detection, reducing hospital appointments, and enhancing resource utilization. It also confronts some serious challenges related to conventional diagnostic tools, including expense, time pressures, and absence of accessibility for remote locations.

In the future, possible upgrades may include extending the scope of the diseases being addressed, incorporating deep learning methods to enhance precision, and including real-time information from wearable health monitors. AI-driven lifestyle and medical history-based individualized health recommendations could also augment preventive care measures further.

With ongoing innovations and growing datasets, this system has the potential to revolutionize AI-assisted diagnostics and make healthcare more predictive, personalized, and accessible.

By connecting technology and medical science, it opens up avenues for a wiser, more efficient, and patient-centric approach to disease prevention and management.

ACKNOWLEDGMENTS

We place special thanks to AISSMS Institute of Information Technology, Pune for their necessary resources and infrastructure to conduct this project efficiently.

The functioning version of the disease prediction system can be found online at <https://multi-disease-predict.streamlit.app/>, and we encourage interested readers to try the application.

REFERENCES

- [1] Kalpna Guleria; Shagun Sharma; Rajni Sobti (2024), "An Early Detection of Parkinson's Disease using Machine Learning and Deep Learning Models," 2024 International Conference on Computational Intelligence and Computing Applications (ICCICA), IEEE, pp. 154–159, doi:10.1109/ICCICA.2024.154.
- [2] Chetan Sharma; Shamneesh Sharma; Sakshi (2024), "Data-Driven Approaches to Early Diabetes Prediction: A Comprehensive Exploration of Predictive Analytics and Machine Learning," 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS), IEEE, pp. 872, ©2024.
- [3] B. Babitha; S. Hemalatha; T. Kavitha; K. Chitra (2023), "Extensive Review on Predicting Heart Disease Using Machine Learning and Deep Learning Techniques," 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE, doi:10.1109/ICCCI56745.2023.10128617.
- [4] G. Pahuja; T. N. Nagabhushan (2021), "A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection," IETE Journal of Research, vol. 67, no. 1, pp. 4–14, Jan. 2021, doi:10.1080/03772063.2018.1531730.
- [5] O. Cigdem; H. Demirel (2018), "Performance analysis of different classification algorithms using different feature selection methods on Parkinson's disease detection," Journal of Neuroscience Methods, vol. 309, pp. 81–90, Nov. 2018, doi:10.1016/j.jneumeth.2018.08.017.
- [6] L. Moro-Velazquez; J. A. Gómez-García; J. D. Arias-Londoño; N. Dehak; J. I. Godino-Llorente (2021), "Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," Biomedical Signal Processing and Control, vol. 66, p. 102418, Apr. 2021, doi:10.1016/j.bspc.2021.102418.
- [7] M. Gil-Martín; J. M. Montero; R. San-Segundo (2019), "Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks," Electronics (Basel), vol. 8, no. 8, p. 907, Aug. 2019, doi:10.3390/electronics8080907.
- [8] O. Karaman; H. Çakın; A. Alhudhaif; K. Polat (2021), "Robust automated Parkinson disease detection based on voice signals with transfer learning," Expert Systems with Applications, vol. 178, p. 115013, Sep. 2021, doi:10.1016/j.eswa.2021.115013.
- [9] G. Solana-Lavalle; J.-C. Galán-Hernández; R. Rosas-Romero (2020), "Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features," Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 505–516, Jan. 2020.
- [10] C. Quan; K. Ren; Z. Luo (2021), "A deep learning based method for Parkinson's disease detection using dynamic features of speech," IEEE Access, vol. 9, pp. 10239–10252, Jan. 13 2021, doi:10.1109/ACCESS.2021.3051432.
- [11] M. Aljalal; S. A. Aldosari; K. AlSharabi; A. M. Abdurraqeab; F. A. Alturki (2022), "Parkinson's Disease Detection from Resting-State EEG Signals Using Common Spatial Pattern, Entropy, and Machine Learning Techniques," Diagnostics (Basel), vol. 12, no. 5, art. 1033, Apr. 2022, doi:10.3390/diagnostics12051033.
- [12] A. Kurmi; S. Biswas; S. Sen; A. Sinitca; D. Kaplun; R. Sarkar (2022), "An Ensemble of CNN Models for Parkinson's Disease Detection Using DaTscan Images," Diagnostics (Basel), vol. 12, no. 5, art. 1173, May 2022, doi:10.3390/diagnostics12051173.
- [13] W. Zhang; X. Han; S. Qiu; T. Li; C. Chu; L. Wang; J. Wang; Z. Zhang; R. Wang; M. Yang; X. Shen; Z. Li; L. Bai; Z. Li; R. Zhang; Y. Wang; C. Liu; X. Zhu (2022), "Analysis of Brain Functional Network Based on EEG Signals for Early-Stage Parkinson's Disease Detection," IEEE Access, vol. 10, pp. 21347–21358, 2022, doi:10.1109/ACCESS.2022.3150561.
- [14] O. M. Faruk; M. U. Mollah; M. A. Hossain (2018), "Diabetes Mellitus Prediction Using Machine Learning," Procedia Computer Science, vol. 139, pp. 225–232, 2018, doi:10.1016/j.procs.2018.10.035.
- [15] D. R. Robinson; R. A. Venkatraman; M. W. Hamilton (2017), "Logistic Regression Approach to Predicting Heart Disease Risk," International Journal of Medical Informatics, vol. 102, pp. 44–52, 2017, doi:10.1016/j.ijmedinf.2017.02.004.
- [16] R. Meena Prakash; P. Krishnaleela; R. Mahendran; T. Manickaraj; M. Mariyappan (2024), "Artificial intelligence based early diagnosis of heart disease using machine learning techniques," (Published in:) 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), IEEE.
- [17] L. D. Kessler; B. G. Aguiar (2019), "Heart Disease Prediction and Classification using Machine Learning Algorithms," Computers in Biology and Medicine, vol. 109, pp. 151–162, Dec. 2019, doi:10.1016/j.compbiomed.2019.05.012.
- [18] J. S. Yadav; A. Sharma; N. K. Singh (2020), "A Comprehensive Survey on Parkinson's Disease Diagnosis and Prediction Using Machine Learning," Computer Science Review, vol. 30, pp. 1–17, May 2020, doi:10.1016/j.cosrev.2018.12.003.
- [19] R. P. P. Bhat; R. G. K. Sharma; A. S. Mahajan (2018), "Parkinson's Disease Prediction using Voice Analysis: A Machine Learning Approach," Procedia Computer Science, vol. 132, pp. 135–142, 2018, doi:10.1016/j.procs.2018.05.069.
- [20] R. S. Rani; R. Agarwal; V. K. Sharma (2020), "Diabetes Prediction System using Ensemble Machine Learning Techniques," Journal of Computer Science and Technology, vol. 35, no. 3, pp. 483–495, May 2020, doi:10.1007/s11390-020-0288-3.
- [21] G. E. H. He; P. A. Menon; F. M. T. Kuss (2018), "Predictive Models for Early Diagnosis of Type 2 Diabetes Using Machine Learning," Health Informatics Journal, vol. 24, no. 3, pp. 340–348, Sep. 2018, doi:10.1177/1460458217717514.