ISSN: 0976-0172
Journal of Bioscience
AndTechnology

Creating a Parkinson's Disease Prediction System: A Look at Machine Learning Techniques

Sumit Das[1]*, Tanusree Saha[1], Ira Nath[1] and Dipansu Mondal[2]

[1]JIS College of Engineering, Kalyani, India.
[2]University of Kalyani, Kalyani, India.

Article Info

Received: 17-01-2020 *Revised: 06-02-2020* *Accepted: 04-03-2020*

ABSTRACT

There is great potential for the use of Machine Learning (ML) methods to solve problems in many different industries, especially in the medical and healthcare professions. We aim to use ML approaches to solve the long-standing problems of Parkinson's disease (PD) prediction and therapy in this project. Inconsistent therapy and inaccurate PD prediction have plagued the field for a long time. By creating a prediction approach that is unique to PD datasets, our study hopes to lessen the impact of these difficulties. We do this by thoroughly investigating and contrasting the performance of several ML systems designed for binary classification problems with respect to PD prediction. We want to reduce inconsistencies and related dangers by establishing a reliable approach for correctly investigating and diagnosing PD by comparing and contrasting the performance of various algorithms. Our research shows that ML methods are crucial for creating reliable decision-support systems for PD prediction. Our work offers valuable insights for the creation of improved medical applications by combining the findings of numerous algorithms. It also helps to fill up research gaps. In sum, the study provides an in-depth analysis of ML methods for PD prediction, demonstrating how they might enhance diagnostic procedures and patient results. In addition to expanding our knowledge of PD, our study highlights the revolutionary power of ML in solving difficult medical problems.

KEYWORDS :
Medical Care, Machine Learning, Predictive Modelling, and Parkinson's Disease are some of the topics covered bythisarticle.

INTRODUCTION:

Machine learning (ML) has become a hot topic in tech because of its cutting-edge features and complex technical applications, which are influencing many different industries1. Machine learning (ML) is a subfield of AI that excels at extracting useful information from a wide variety of data sources and formats. Problems still exist, despite its competence in handling large datasets.Especially in the healthcare industry, here precise data categorization continues to be a significant challenge3. The use of ML algorithms in healthcare has great potential, especially for illness prediction. These algorithms show incredible predictive power by finding complex patterns in medical data, which

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

highlights their potential to change healthcare.

Recent literature reviews have focused on algorithms in healthcare, highlighting.

Our primary objective in doing this research is to learn more about machine learning (ML) algorithms as they pertain to healthcare, with an emphasis on illness prediction and, more specifically, Parkinson's disease. Accurate prediction and early identification pose particular hurdles in the case of Parkinson's disease (PD), a neurodegenerative disorder characterised by motor and non-motor symptoms. The diagnosis of PD is still challenging, with many cases including miss-predictions and pre-diagnostic mistakes, even if medical technology has advanced. In light of these difficulties, we suggest a thorough examination of ML approaches to binary classification, with a focus on accuracy. We want to find the best way to build a predictive system for PD diagnosis by thoroughly analysing and contrasting several ML algorithms. We are primarily focused on developing a systematic approach to reliably predicting who will develop PD and reducing diagnostic errors and miss-predictions. Our findings might help improve PD prediction approaches, which would improve early identification and intervention tactics. That's why it's important. We want to improve patient outcomes and quality of life by using ML algorithms to provide a path towards more accurate and dependable PD diagnosis.
With a focus on PD prediction, this work aims to bridge the gap between ML techniques and healthcare applications. With careful research and testing, we want to contribute to the area by shedding light on how to make the most of ML algorithms in health tracking and diagnosis. Review of the Literature

Machine Learning (ML) is all the rage now because of how cutting-edge it is and all the complex ways it can be used technologically.5. When it comes to neurological diseases like Parkinson's disease (PD), ML has a lot of potential in healthcare. Tremors, stiffness, and sluggish movement are motor signs of Parkinson's disease, a degenerative neurological disorder.

movementPD's meteoric rise in incidence throughout the years has had far-reaching consequences for public health across the world. One of the most financially devastating long-term health problems is neurodegenerative dementia, of which Alzheimer's disease (AD) is the most common form. The potential societal and patient welfare effects of automated Alzheimer's disease diagnosis and care are substantial. As a direct result of cognitive decline8, language difficulty is one of the most common signs of Alzheimer's disease. Gait kinematics and other spatiotemporal features are crucial for PD diagnosis9. It is common for PD to go undiagnosed, which delays treatment, since symptoms and disease progression may vary widely. Possible non-invasive, cost-effective, and more accessible methods of detecting Parkinson's disease might be realised via the use of machine learning and speech analysis. The accuracy and reliability of these techniques in clinical settings10, however, must be ensured by further study and validation. Researchers have been looking at using ML algorithms to enhance PD detection's accuracy and efficiency in order to tackle this difficulty. Researchers may build

ISSN: 0976-0172
Journal of Bioscience AndTechnology

accurate and understandable models for Parkinson's disease prediction by combining supervised learning approaches with appropriate feature selection strategies applied to speech datasets. With any luck, this undertaking will pave the way for earlier diagnosis and intervention11. In the context of Parkinson's disease, several academic investigations have explored the domain of speech dysfunction. A wide range of variables impacting speech problems have been examined in these research, as have the effectiveness of different treatment strategies and the effects of speech impairment on other aspects of patients' well-being.12. Ongoing research aims to enhance our understanding of the underlying processes of speech and swallowing difficulties in Parkinson's disease, with a particular emphasis on developing more effective treatments. This includes investigations into novel therapeutic approaches like transcranial magnetic stimulation (TMS) and deep brain stimulation (DBS) as well as efforts to decipher the complex role of brain circuits in motor control. therapeutic magnetic stimulation (TMS) with the goal of improving swallowing and speaking abilities13. Predicting and diagnosing diseases using ML and AI methods has recently attracted a lot of attention14. There has been a lot of excitement about AI applications recently, and the rise of deep learning (DL) has changed the AI tool landscape15. The purpose of this research is to examine the efficacy of PD detection algorithms16 by means of binary classification methods such as logistic regression, support vector machines, k-nearest neighbours, decision trees, and Random Forest. The main goal of XAI approaches is to make AI models more clear and understandable so that doctors can understand how the models arrived at their predictions. Because of the gravity of the consequences of medical choices, this is of the utmost importance. You may accurately determine the algorithms' accuracy and efficiency in Parkinson's disease prediction18 by following these steps and doing comprehensive tests on relevant datasets. In addition, medical diagnosis has been radically altered by the development of ML, especially deep learning (DL), which has allowed for the processing of massive amounts of data with hitherto unseen levels of accuracy19. Medical imaging, test results, and patient complaints are just a few of the sources of information that doctors use into the diagnosis process. Unfortunately, the need to include multimodal information has not been well addressed by current deep-learning models that aim to aid in diagnosis. There is potential to improve patient triage and speed up clinical decision-making via the development of unified multimodal transformer-based models. Healthcare providers are now able to make better, more timely diagnoses, prognoses, and treatment choices because to this technological innovation. In general, the use of machine learning and artificial intelligence in medical diagnosis represents a sea change in healthcare, since these tools will augment human doctors' knowledge and experience to better serve patients.

A technique

Here, the approach used is

The work highlights the need of using ML approaches in a systematic way to predict and identify PD, as seen in Figure 1. A predictive model is the end result of the methodology's various critical steps, which are all designed to help with data

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

inquiry and elucidation.
(i) Bringing in Required Libraries and Dependencies: That is the first thing to do when starting the ML pipeline21. We carefully analyse the selection and inclusion of necessary libraries to achieve optimum performance, since the behaviour and quality of an ML system are dependent upon the input characteristics.
(ii) Data Collected: Kaggle, a well-known site for data science contests and datasets, is the source of the dataset used in this work. Loading the dataset into the pandas data frame after capture allows for thorough data analysis. In order to get a feel for the make-up and organisation of the data, it is helpful to know how many rows and columns the dataset has. The completeness and integrity of the dataset is further evaluated using procedures like statistical analysis and the identification ofmissingvalues22.
(iii) Data preprocessing: This crucial step involves cleaning and organising raw data into a coherent structure that is ready for further analysis. In order to fix errors and improve the dataset's quality and consistency, operations including data cleansing, normalisation, and transformation are performed. (iv) EDA, or exploratory data analysis, is a critical technique for discovering trends, patterns, and outliers in the dataset. In order to guide future modelling and analysis efforts, researchers thoroughly examine the data to uncover important insights into underlying linkages and events. part (v)The process of feature engineering comprises taking raw data and transforming it into useful features that may be used to train and understand models. This procedure involves identifying and engineering significant variables to capture key aspects crucial to PD prediction. Characteristics act as crucial inputs into the ML pipeline, which allow the algorithm to discover significant correlations and patterns in thedata. (ii) Using a train-test split methodology26, the dataset is divided into two parts: the training set and the validation set. The training set is for training the model and estimating its parameters, whereas the validation set is for evaluating the model and its performance. Researchers may improve and optimise the model's performance by iteratively gauging the model's effectiveness and generalizability.
(vii) Standardising Data: When data is standardised, it is rescaled such that the characteristics follow a standardised scale, usually with a mean of 0 and a variation of 1. This process is also called normalisation. By making ensuring that feature distributions are consistent, this normalisation method reduces the impact of biases and value range differences. More stable and dependable predictions are made possible by standardisation, which improves the ML model's interpretability and stability 27. The following is a quick illustration of thealgorithms: One of the most important and powerful tools in machine learning is the K-Nearest Neighbour (KNN) algorithm. It is highly regarded for its non-parametric nature, simplicity, and adaptability. Apt for use in regression and classification, Classification prediction is where it really shines. Using the proximity principle, KNN sorts newly-imported data into coherent clusters or segments by lining them up with examples that have already been trained. Assigning the new observation to the classification that is most comparable to the current dataset is based on the premise that the two are similar. By giving precedence to closeness, the algorithm sorts incoming data according to how similar they are to nearby cases, which allows it to classify them. The fact that KNN can manage massive datasets with little impact on

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

classification accuracy and performance28 is noteworthy. Backend vector algorithm With an emphasis on the VC dimension and the empirical risk reduction notion, the Support Vector Machine (SVM) model embodies a potent machine learning strategy based on computational and statistical concepts. When dealing with problems involving pattern recognition, this approach has clear benefits, especially when dealing with situations involving few samples, data heterogeneity, and computational complexity. Standout features of SVM include its ability to avoid problems like the "curse of dimensionality" and reduce the likelihood of "over-learning," demonstrating resilience in the face of different types of complexity.

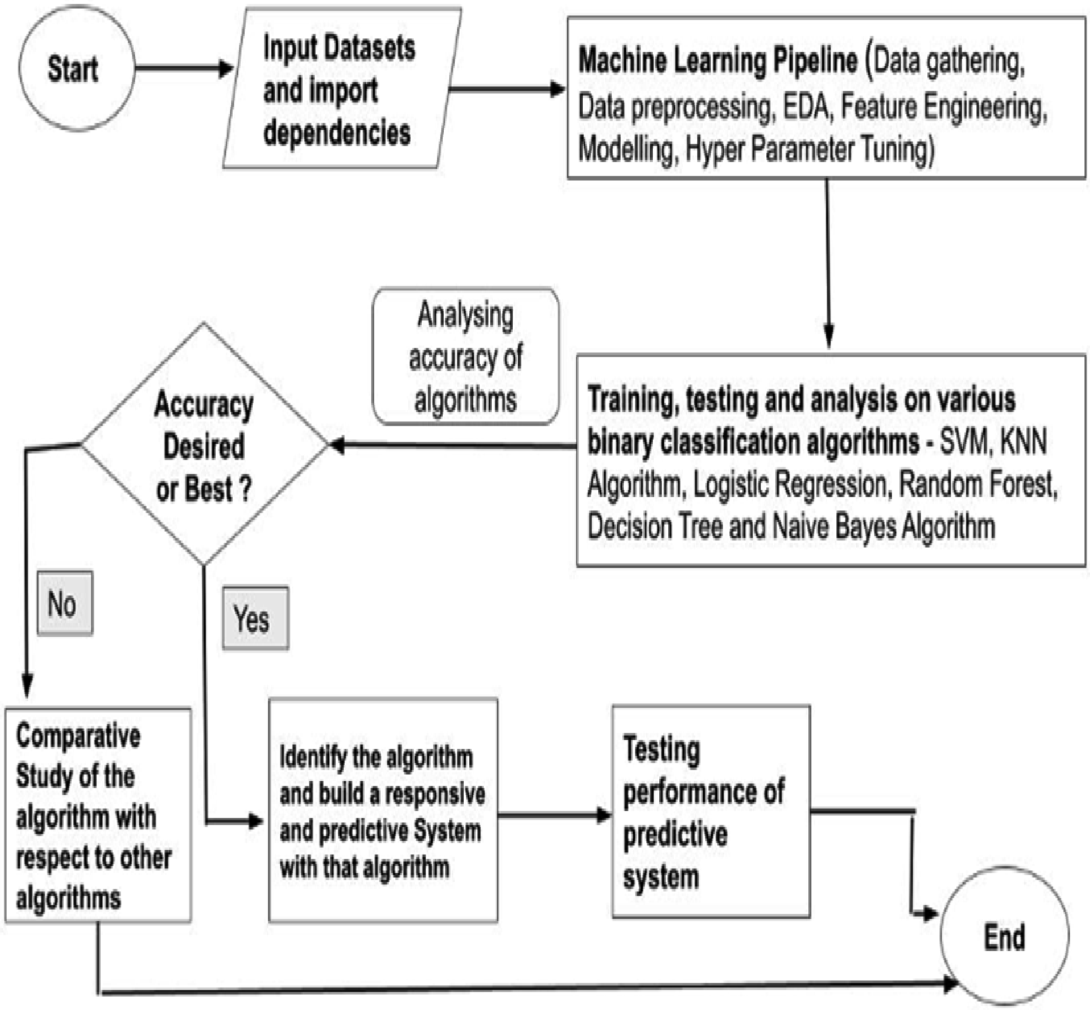


Fig.1. Proposed System

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

SVM has greatly advanced progress in many related disciplines, including pattern recognition, regression analysis, function estimation, time series forecasting, and more, thanks to its strong theoretical basis and clear mathematicalframework.

RationalRegression As a statistical method, logistic regression models the probability of discrete outcomes given a set of input variables. With its capacity to calculate the likelihood of a fresh sample falling into a certain category, it is an invaluable analytical tool, especially in classification jobs. Considering that different parts are classified in different ways...... Logistic regression, by making use of its features, helps a lot with the classification problems that are intrinsic to cyber securitydomains30. SimpleBayes This method is more common in text classification situations that need a large amount of training data, and it is a supervised learning approach that uses the Bayes theorem to handle classification. The Naive Bayes Classifier is an effective and fundamental method for classifying data. paves the way for the creation of efficient machine learning systems that can reliably forecast future outcomes. It makes predictions based on the probabilities linked to certain objects, as is typical with probabilistic classifiers. EvaluationStructure Additionally, it is a supervised learning technique that is useful for both classification and regression, albeit it is most often used for the former. The dataset's properties are represented by the internal nodes of this tree-like structure, the decision-making process is symbolised by the branches, and the ultimate wrapping up is provided by each leaf. Decision Nodes, which play a crucial role in decision-making processes with many branches, and Leaf Nodes, which show the results of individual choices without further branches, are the most important parts of a decision tree. Decision Trees facilitate testing, experimentation, and decision-making by capitalising on dataset attributes. Decision trees, which are visual representations of the problem-solving process, methodically investigate many avenues in order to arrive at possible answers.

ISSN: 0976-0172
Journal of Bioscience AndTechnology

Support Vector Machine Model

```
[ ] # training the SVM model with training data
    model.fit(X_train, Y_train)

    SVC(kernel='linear')
```

Evaluating the model

```
[ ] # accuracy score on training data
    X_train_prediction = model.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

Accuracy Prediction

```
[ ] print('Accuracy score of training data : ', training_data_accuracy)

    Accuracy score of training data :  0.8846153846153846
```

```
[ ] # accuracy score on training data
    X_test_prediction = model.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

```
[ ] print('Accuracy score of test data : ', test_data_accuracy)

    Accuracy score of test data :  0.8717948717948718
```

ISSN: 0976-0172
Journal of Bioscience AndTechnology

KNN Analysis

```
[ ] neigh = KNeighborsClassifier(n_neighbors=3)
```

```
[ ] neigh.fit(X_train, Y_train)

KNeighborsClassifier(n_neighbors=3)
```

```
[ ] X_train_prediction = neigh.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

```
[ ] print('Accuracy score of training data : ', training_data_accuracy)

Accuracy score of training data :  0.9743589743589743
```

```
[ ] # accuracy score on training data
    X_test_prediction = neigh.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

```
[ ] print('Accuracy score of test data : ', test_data_accuracy)

Accuracy score of test data :  0.8205128205128205
```

Fig. 3. Evaluation of KNN accuracy

Naive Bayes

```
[ ] gnb = GaussianNB()
```

```
[ ] gnb.fit(X_train, Y_train)

GaussianNB()
```

```
[ ] X_train_prediction = gnb.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

```
[ ] print('Accuracy score of training data : ', training_data_accuracy)

Accuracy score of training data :  0.7243589743589743
```

```
[ ] # accuracy score on training data
    X_test_prediction = gnb.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

```
[ ] print('Accuracy score of test data : ', test_data_accuracy)

Accuracy score of test data :  0.6153846153846154
```

Fig. 4. Evaluation of accuracy and complexity for the Naive Bayes Algorithm

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

Decission Tree

```
[ ] clf = tree.DecisionTreeClassifier()

[ ] clf.fit(X_train, Y_train)

    DecisionTreeClassifier()

[ ] X_train_prediction = clf.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

[ ] print('Accuracy score of training data : ', training_data_accuracy)

    Accuracy score of training data :  1.0

[ ] # accuracy score on training data
    X_test_prediction = gnb.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

[ ] print('Accuracy score of test data : ', test_data_accuracy)

    Accuracy score of test data :  0.6153846153846154
```

Fig. 5. Evaluation of accuracy and complexity for the Decision Tree Algorithm

Logistic Regression

```
[ ] LG = LogisticRegression(random_state=0)

[ ] LG.fit(X_train,Y_train)

    LogisticRegression(random_state=0)

[ ] X_train_prediction = LG.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

[ ] print('Accuracy score of training data : ', training_data_accuracy)

    Accuracy score of training data :  0.8717948717948718

[ ] # accuracy score on training data
    X_test_prediction = LG.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

[ ] print('Accuracy score of test data : ', test_data_accuracy)

    Accuracy score of test data :  0.8205128205128205
```

Fig. 6. Evaluation of Logistic Regression Accuracy

ISSN: 0976-0172
Journal of Bioscience AndTechnology

Random Forest

Its adaptability in tackling both regression and classification challenges has made it a notable ML approach within the area of supervised learning. Random Forest uses a number of classifiers along with the idea of ensemble learning to improve model performance when faced with complex issues. When faced with data-driven difficulties, this method takes use of the combined knowledge of several classifiers. The algorithm's use of ensemble-based methodologies33 highlights its effectiveness in solving a range of machine learning issues.

In each model, a dataset comprising labeled samples is utilized to explore the statisticalrelationship between attributes and objectives. The model demonstrating superior performance is subsequently evaluated against a distinct dataset, not employed during the training phase. This assessment ensures the creation and upkeep of a predictive model capable of generalizing wellto new data. Drawing upon the methodologies discussed above, this summary encapsulates the scientific pursuit of model development and validation

.

Table 1. Accuracy of training and testing across various algorithms

Algorithm Name	Training Data Accuracy	Testing Data Accuracy
SVM (Support Vector Machine)	88.46%	87.17%
KNN Algorithm	97.43%	87.05%
Naive Bayes Algorithm	72.43%	61.53%
Decision Tree Algorithm	100%	61.53%
Logistic Regression Algorithm	87.17%	82.15%
Ramdom Forest Algorithm	88.46%	82.05%

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

Random Forest

```
[ ] RF = RandomForestClassifier(max_depth=2, random_state=0)

[ ] RF.fit(X_train, Y_train)

    RandomForestClassifier(max_depth=2, random_state=0)

[ ] X_train_prediction = RF.predict(X_train)
    training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

[ ] print('Accuracy score of training data : ', training_data_accuracy)

    Accuracy score of training data :  0.8846153846153846

[ ] # accuracy score on training data
    X_test_prediction = RF.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

[ ] print('Accuracy score of test data : ', test_data_accuracy)

    Accuracy score of test data :  0.8205128205128205
```

Fig. 7. Evaluation of Random Forest accuracy

ISSN: 0976-0172
Journal of Bioscience
AndTechnology

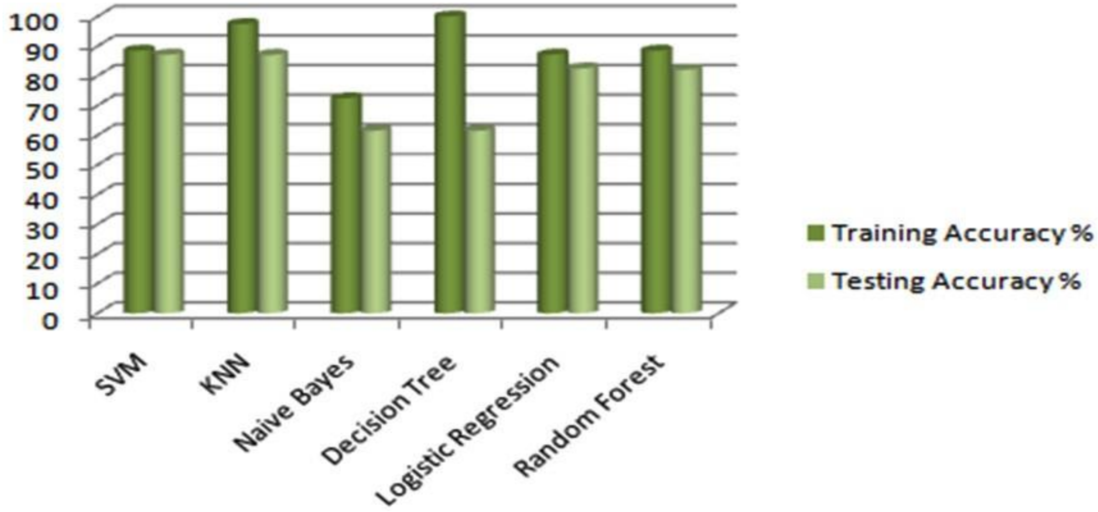


Fig. 8. Comparison of accuracy among various algorithms

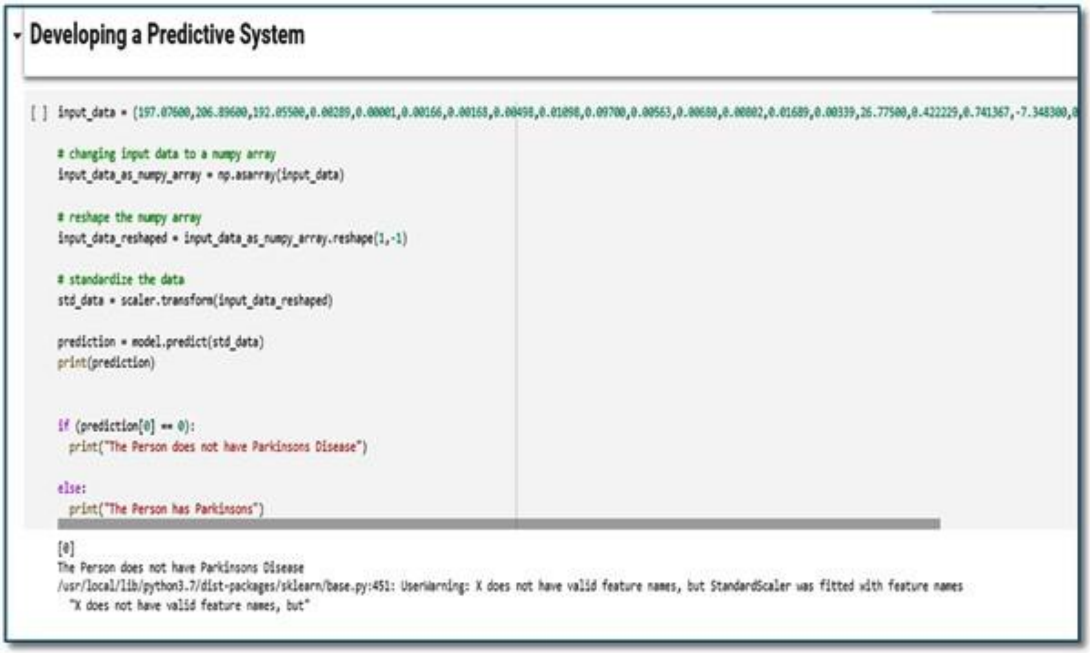
RESULTS AND DISCUSSION

By comparing and contrasting different binary classification techniques, this work adds to the body of knowledge. Extensive testing has shown that various algorithms achieve varying degrees of accuracy, highlighting the importance of accuracy as a measure of Machine Learning effectiveness. Importing libraries is the first step in the machine learning pipeline. Then, data homogeny and mock-up training are iterative phases of distinctive engineering methodologies. The prediction model for Parkinson's disease is built using sklearn techniques including SVM, KNN, Naive Bayes, and decision trees. Results from comparisons and analyses of complexity show that SVM approaches are the most effective in predictive modelling, as they get the best accuracy on testing datasets. As a result, a strong predictive model is created, which opens the door to better Parkinson's disease diagnosis and treatment. Systematic Variable Model (SVM) Achieving 88.46% accuracy on the training dataset and 87.17% accuracy on the testing dataset, the Support Vector Machine (SVM) model shows encouraging performance. These results demonstrate that SVM is a powerful tool for precise data classification. Figure 2 graphically displays an examination of SVM's accuracy and complexity, illuminating its performance characteristics. It seems that SVM has promise as a trustworthy prediction model, given its remarkable accuracy on both the training and testing datasets. There has to be more research and conversation around SVM to figure out what makes it work and how to makeevenbetter.MyNeighbourK

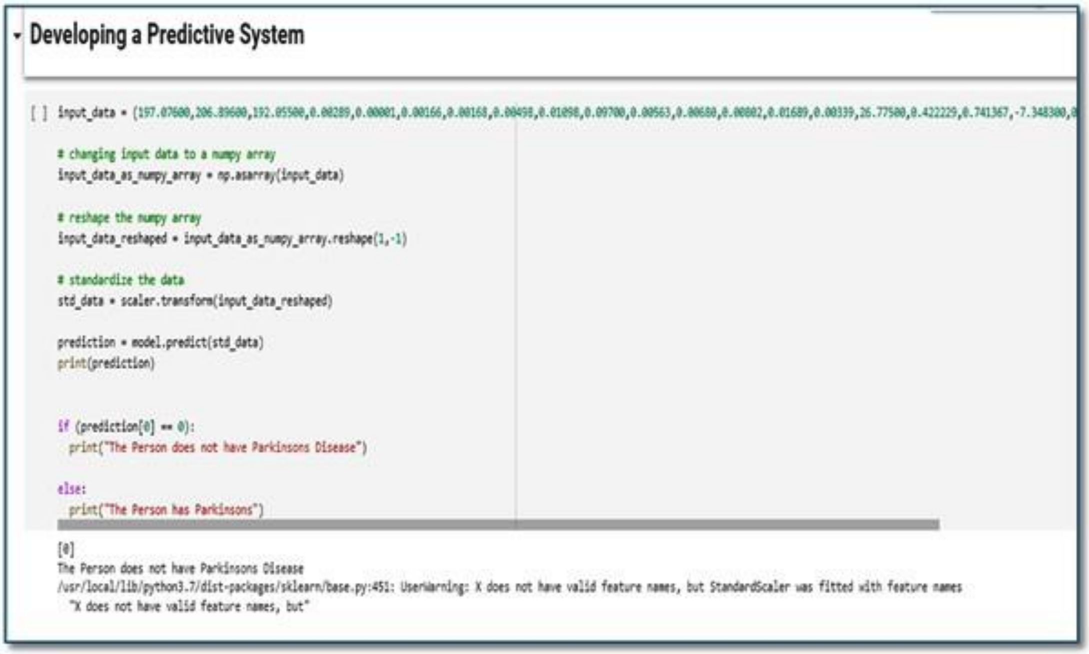
With training dataset accuracy of 97.43% and a testing dataset accuracy of 87.05%, the K-Nearest Neighbour (KNN) method shows impressive performance. These results show that KNN is quite reliable when it comes to categorising data points. To better understand KNN's performance characteristics, its accuracy and complexity are shown in Figure 3. It is worth mentioning that the model's excellent accuracy on the training dataset shows that it can efficiently identify patterns in the data. Nevertheless, more

ISSN: 0976-0172
Journal of Bioscience AndTechnology

research is required to resolve any possible over-fitting or generalisation concerns, as shown by the somewhat reduced accuracy on the testing dataset. To improve the KNN algorithm's performance in practical settings, future data set for testing. Its capacity to properly categorise data items is shown by these outcomes. Naive Bayes' accuracy, as seen in Figure 4, sheds light on its performance characteristics. In spite of the algorithm's satisfactory performance on the training dataset, the studies may aim to fine-tune its parameters. Simple Bayes With a 72.43% accuracy rate on the training dataset and a 61.53% rate on the test dataset, the Naive Bayes Algorithm exhibits moderate performance. reduced precision on the test dataset indicates possible difficulties with extrapolation. To further understand this disparity and find ways to enhance the algorithm's performance, more research is required. Possible targets for future studies include enhancing the model's



ISSN: 0976-0172
Journal of Bioscience AndTechnology



Developing a Predictive System

```
[ ] input_data = (197.07600,206.89600,192.05500,0.00289,0.00001,0.00166,0.00168,0.00498,0.01098,0.09700,0.00563,0.00680,0.00802,0.01689,0.00339,26.77500,0.422229,0.741367,-7.348300,0

    # changing input data to a numpy array
    input_data_as_numpy_array = np.asarray(input_data)

    # reshape the numpy array
    input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

    # standardize the data
    std_data = scaler.transform(input_data_reshaped)

    prediction = model.predict(std_data)
    print(prediction)


    if (prediction[0] == 0):
      print("The Person does not have Parkinsons Disease")

    else:
      print("The Person has Parkinsons")

    [0]
    The Person does not have Parkinsons Disease
    /usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not have valid feature names, but StandardScaler was fitted with feature names
      "X does not have valid feature names, but"
```

ISSN: 0976-0172
Journal of Bioscience AndTechnology

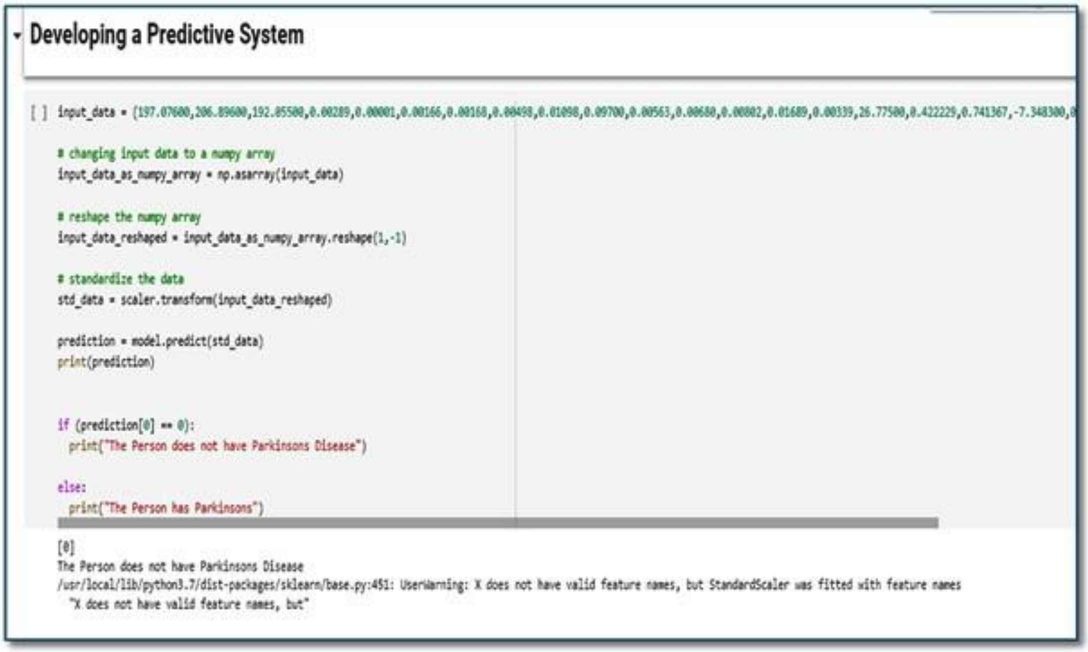


Fig. 9. Result of Predictive System

parameters or investigating other methods to improve its predictive power in various settings.

Approach using Decision Trees

On the training dataset, the Decision Tree Algorithm achieves an impressive 100% accuracy, showing that it is good at detecting complex patterns. A lower accuracy of 61.53% indicates that it performs rather poorly on the testing dataset. In Figure 5, we can see the Decision Tree's correctness, which helps us understand its performance indicators. The algorithm's flawless performance on the training dataset raises concerns about over-fitting or a lack of generalizability, since it achieves lesser accuracy on the testing dataset. To get to the bottom of things and figure out how to make the algorithm work better on new data, further research is required. Improving the Decision Tree Algorithm's prediction power across different datasets may be the subject of future studies that optimise its parameters or use ensemble approaches.

Predictive Modelling System

On the training dataset, the Logistic Regression Algorithm achieved an accuracy of 87.17%, demonstrating favourable performance.

as well as 82.15 percent on the test dataset. These results indicate that it is effective at correctly categorising data points. Figure 6 shows the accuracy of the Logistic Regression, which gives us a better idea of how it works. Further study is needed because to the somewhat reduced accuracy on the testing dataset, while the method performs well on both the training and testing datasets. Possible ways to make it better include playing around with the model's parameters or looking at feature engineering approaches to boost prediction accuracy. To fine-tune the Logistic Regression Algorithm for a variety of datasets and applications, more study is required.

An Algorithm for Random Forests

With a training-set accuracy of 88.46% and a testing-set accuracy of 82.05%, the Random Forest Algorithm shows high performance. These outcomes demonstrate how well it categorises data

ISSN: 0976-0172
Journal of Bioscience AndTechnology

points. Visualising the Random Forest's accuracy in Figure 7 gives useful insights into its performance characteristics. The algorithm's accuracy drops somewhat on testing datasets, but it does well on training datasets as well. as contrasted with the dataset used for training. More research into possible over-fitting or generalisation problems is warranted in light of this disparity. Improving the Random Forest Algorithm's prediction power across different datasets may be the subject of future studies that investigate ensemble approaches or aim to optimise the algorithm's parameters.

The results of these comparative analyses and investigations into the relative complexity of various categorization techniques show that the SVM achieves the highest level of accuracy on the test data, as shown in Table 1 and Figure 8. We have developed a prediction model that can identify people with Parkinson's disease as a consequence of our study. Our research showed that the SVM algorithm performed the best of all the algorithms we tested, with the KNN approach coming in a close second. This research highlights the need of ongoing efforts to better understand Parkinson's disease and smart systems. The results shown graphically in Figure 9, which compares the algorithms' performance, corroborate this finding.

Consequently, we have developed a smart prediction algorithm to identify people with Parkinson's disease. From all the algorithms that were tested in this research, the SVM algorithm performed the best, followed closely by the KNN Algorithm. The need of doing more study in the field of intelligent systems and advancements connected to Parkinson's disease is highlighted by these results, which also point to potential directions for future endeavours.

In summary The purpose of this study was to analyse algorithmic complexity and compare different algorithms. The goal was to create a predictive machine learning framework that could improve overall prediction accuracy and address discrepancies in Parkinson's disease prediction. Classification methods have been the subject of a great deal of research, but we have found that various approaches provide varied results when tested on different datasets. methods for improving algorithms and their performance.

This work's main strength is the systematic framework it provides for building the prediction system, which is based on extensive research and comparison of algorithms. The improvement of healthcare outcomes might be facilitated by this framework's ability to reduce obstacles in diagnosis and prediction. The system's dependence on datasets is a major flaw, however, since inaccurate or manipulated data could cause inaccurate predictions. Future study might focus on developing specialised systems that go beyond prediction tasks by investigating bidirectional methods of algorithm selection and development.

Still, developments in dataset quality and algorithmic improvements pose the greatest danger to our built system. These factors might make competing algorithms more accurate and predictive, reducing our system's effectiveness. To keep our prediction framework relevant and successful in the ever-changing field of machine learning, it will be vital to continuously monitor and update it.

ACKNOWLEDGMENTS

For their kind donation of necessary resources and facilities, the authors would like to thank the Management of JIS College of Engineering, JIS GROUP.

Possible bias

There is no conflict of interest, the

ISSN: 0976-0172
Journal of Bioscience AndTechnology

authors state.
Where to Get Money
We do not have any sources of financing.
Statement on Ethics Approval
No author has ever conducted research with humans or animals for this article.
The Role of the Authors
Every single person who took part in the research gave their informed permission, and all f the writers contributed equally.
Access the datasets used in this research at this URL: https://www.kaggle.com/. They are part of the Kaggle repository. Section 581 of Biosci., Biotech. Res. Asia, Volume 21, Issue 2, pages 569–582,(2024)the atasets/vikasukani/datasets/parkinsons-disease Referring to reference number 22, this dataset is mentioned.

REFERENCES

Abehh and Gohel's Machine Learning in Healthcare. The current genomics journal is published in 2021 and has the DOI: 10.2174/13892029226662107051243 pages 291–300. Second, Woodman RJ and Mangoni AA. An all-inclusive analysis of the current and future state of machine learning algorithms as they pertain to geriatric medicine. Publication: Ageing Clinical and Experimental Research, 2023, Volume 35, Issue 11, Pages 2363–2397, DOI: 10.1007/s40520-023-02552-2.
Three authors: Pudjihartono N, Fadason T, Kempa-Liehr AW, and O'Sullivan JM. The Use of Machine Learning for Disease Risk Prediction: A Survey of Feature Selection Techniques. Front Bioinforma, 2022, 2, doi:10.3389/fbinf.2022.927312.
4. Salhout SM. A thorough literature review on machine learning in healthcare strategic management. Arab Gulf Journal of Scientific Research 2023; pre-publication (pre-print). The DOI for this article is 10.1108/ AGJSR-06-2023-0252
5. Ijaz MF, Singla R, Kumar Y, Koul A. The use of AI in medical diagnosis: a comprehensive literature analysis, synthesis framework, and outline for future studies. Publication: Journal of Environmental Intelligence and Human-Computer Behaviour. January 13, 2022, online publication. 10.1127/s12652-021-03612-z(1972).
6. Machine Learning for the Diagnosis of Parkinson's Disease: A Literature Review. Front Ageing Neurosci. 2021;13. oi:10.3389/fnagi.2021.633752. (Mei J, Desrosiers C, Frasnelli J.) 7. M. Heinrich, W. Oertel, D. Surmeier, and F. Geibl. A possible treatment target in Parkinson's disease is mitochondrial dysfunction. The published version of this article is "Mol Neurodegener" from 2023 and the DOI is 10.1186/s13024-023-00676-7.
Eighth, Ammar RB and Ayed YB. Language-based indicators for the early diagnosis of Alzheimer's disease. Computer Science Procedia, 2020, 176: 763–770. Published online: 2020/09/071 by Proc. 9. Yousef R, Hashemi G, Hashemi E, and Mubarak I.Machine learning algorithms to detect Parkinson's illness using speech signal properties. The article is published in Front Artif Intell and has the DOI: 10.3389/frai.2023.1084001.
Y. Rahmatallah, A. Iyer, and A. Kemp wrote the paper. An AI approach to analysing audio recordings for the diagnosis of Parkinson's disease. Publication date: 2023 Jan 1; DOI: 10.1038/s41598-023-47568-w. 11. The Detection of Parkinson's Disease Through the Use of Filter Feature Selection and Al-AM, Salim, and Saeed's Work a Genetic Algorithm with Ensemble Learning. Article published in 2023 in Diagnostics with the DOI 10.3390/diagnostics13172816. 12.

ISSN: 0976-0172
Journal of Bioscience AndTechnology

Diegroot C, Wiesman AI, Donhauser PW, et al. In Parkinson's disease, abnormal neurophysiological signalling is linked to difficulties with speech. In 2023, the journal "Npj Park Dis" published an article with the DOI: 10.1038/s41531-023-00495-z and the pages 1–13. Does N, Mitchell S, Sturgill L, Khemani P, and Sin MK make up Reference 13. Issues with Fluency and Gagging in Parkinson's Patients. Journal of Nurse Practise, 2022, 18, 848–851. doi:10.1016/j.nurpra.2022.05.019. 14. Parkinson's Disease with Dysphagia The Parkinson's Disease Foundation... The publication year is 2024. Retrieved on March 10, 2024. here: https://www.parkinson.org/library/fact-sheets/speech-swallowing

15. Kohler A, Memarzadeh K. The increasing use of AI in medical settings. Artif Intell Healthc released its findings online. 2020:25-60. doi:10.1016/B978-0-12-818438-7.00002-2. Publishing a research article

Alves AC, Bourbon M, Albuquerque J, Medeiros AM, Antunes M. 16. Investigation of the diagnostic accuracy of several classification algorithms in adult patients with familial hypercholesterolemia, including pre- and post-processing strategies for dealing with unbalanced data. The publication is titled "PLoS ONE" and was published in 2022 with the DOI:10.1371/journal.pone.0269713. The DOI is e0269713.

17. IBM: Definition of Explainable AI (XAI). The publication year is 2024. Retrieved on March 10, 2024. Visit https://www.ibm.com/topics/explainable-ai for more information;

18.A group of researchers including Dixit, Bohre, Singh, and others wrote the following article. A Thorough Examination of AI-Powered Models for the Diagnosis of Parkinson's Disease. Publication date: 2023-04-83, DOI: 10.3390/electronics12040783.

19.Sanyal MK, Das S. One example of a medical diagnosis of TB is machine intelligence diagnostic systems (MIDs). The article "Neural Comput Appl. 2020;32(19):15585-15595" and the associated DOI are the following: 10.1007/s00521- 020-04894-8.

20. A representation-learning model for clinical diagnostics that uses transformers and unifies the processing of multimodal input | Nature Biomedical Engineering. Retrieved on March 10, 2024. THIS PAGE: https://www.nature.com/articles/s41551-023-01045-x21.Machine Learning | Data Dependencies on Google for Developers. March 10, 2024, the publication year, was accessed. This video lecture is available at https://developers.google.com/machine-learning/crash-course/data-dependencies.

Methods for gathering and cleaning data - Qualcomm Developer Network, section 22. The publication year is 2024. Retrieved on March 10, 2024. To access the learning materials, go to https://developer.qualcomm.com/software/qualcomm-neural-processing-sdk.21.2 (DAS et al., 2024), 569–582 in Biosci., Biotech. Res. Asia, Article Number: 21(2), pre-processing for data collecting in artificial intelligence mobile learning

23. Machine Learning | Python Data Preprocessing - GeeksforGeeks. March 10, 2024, the publication year, was accessed. Data preparation, machine learning, and Python may be found at https://www.geeksforgeeks.org/.

24. Investigative Data Analysis: What Is It? | IBM. Originally published in 2024. Retrieved on March 10, 2024. Here is the link to the IBM website: https://www.ibm.com/topics/exploratory-data- analysis.

Domino Data Science Dictionary,

ISSN: 0976-0172
Journal of Bioscience AndTechnology

"Feature Engineering," accessed March 10, 2024, 25 years after publication in 2024. At https://domino.ai/data-science-dictionary/feature-engineering, you get all the information you need.
https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data (accessed March 10, 2024). 26. Training and Test Sets: March 10, 2024.
28. Jain V. K-Nearest Neighbour Algorithms. Vidhya, an analytics professional. Original publication date: January 31, 2022. Retrieved on March 10, 2024. Introducing KNN Algorithms: https://www. analyticsvidhya.com/blog/2022/01/introduction-to-knn-analytics/ ScienceDirect Topics published an overview on support vector machines in 2024. Retrieved on March 10, 2024. Viewed at: https://www.sciencedirect.com/topics/computer-science/support-vector-machine
30. Pant A. Logistic Regression: An Introduction. Medium. December 22, 2019 publication date. Retrieved on March 10, 2024. Welcome to the Introduction to Logistic Regression course at Forwards Data Science!
IBM published a page on naïve bayes classifiers in 2024, and on March 10, 2024, it was accessed. Visit https://www.ibm.com/topics/naive-bayes for more resources.
33. Learning using decision trees. The information was retrieved from Wikipedia on March 10, 2024. Accessed from: https://en.wikipedia.org/w/index.php?title=Decision_tree_learning&oldid=1211641048
Dr. Zhou's Random Forests and Regularisation dissertation was completed in 2022 at the University of Pittsburgh.
Splitting Data | Machine Learning. Google for Developers. Published 2024. Issue 4.4, Fall 2022, 27. Working Towards a "Standard Model" of Machine Learning. The link to the publicationis https://hdsr.mitpress.mit.edu/pub/zkib7xth/release/2, and the date of access is