International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT24103217

Machine Learning for the Multiple Disease Prediction System

Kumar Bibhuti B. Singh[1], Ashutosh Sharma[2], Ashish Verma[2], Ranjeet Maurya[2], Dr. Yusuf Perwej[3]

[1]Assistant Professor, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, Uttar Pradesh, India

[2] B.Tech Scholar, Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, Uttar Pradesh, India

[3]Professor, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, Uttar Pradesh, India

ARTICLEINFO

Article History:

Accepted: 15 June 2024

Published: 30 June 2024

Publication Issue

Volume 10, Issue 3

May-June-2024

Page Number

673-684

ABSTRACT

Disease prediction, which aims to identify individuals who are at risk of contracting specific diseases, is a crucial component of healthcare. Recently, machine learning algorithms have shown to be effective tools in the fight against illness prediction due to their superior ability to sort through large datasets in search of complex patterns. The development of Machine Learning (ML) in the contemporary healthcare period has created new opportunities for the diagnosis and treatment of chronic illnesses. In this paper we are proposes a complete Multiple Disease Prediction System that makes accurate predictions of diabetes, cancer, and heart disease using machine learning algorithms. The system's purpose is to analyse intricate medical datasets and find trends and risk factors related to these illnesses. The system uses cardiovascular data analysis and logistic regression to detect heart disease and provide a probabilistic evaluation of heart health. Convolutional Neural Networks, which evaluate medical imaging to find malignancies with high precision, are used to simplify cancer detection. Finally, Support Vector Machines are used to predict diabetes by taking into account a variety of metabolic and genetic indicators to evaluate. Making it simpler for people to detect their own health issues with just their symptoms and exact vital signs is the aim of this project. The proposed approach improves both the predictive power and precision of sickness.

Keywords: Symptoms, Data-Driven, K-Nearest Neighbour (KNN), Healthcare, Random Forest, Disease Prediction, Support Vector Machine (SVM), Machine Learning.

I. INTRODUCTION

One of the most important aspects of human life is the ability to predict diseases. One of the most important aspects of treating a disorder is predicting an individual's health early on. It has always been predominantly the fault of a medical expert. Thus, innovation that improves logistics is essential to the healthcare industry [1]. The medical industry is cantered on innovation. It drives the development of novel medications, remedies, and therapies [2]. Innovation is another factor that keeps the medical industry on the leading edge. The healthcare industry has a tonne of potential for expansion [3]. Innovation is necessary to progress in a number of fields. Among these are increasing the effectiveness of medical operations, finding novel treatments for diseases, and enhancing patient care.

One way to innovate in the medical field in this digital age is to digitalize medical procedures [4]. The cost of consultations is prohibitive and the demand on doctors' time is excessive, which are two of the most important issues facing the healthcare business [5]. This problem is mostly highlighted by using patient symptoms as input for disease prediction. Nowadays, it is usual practice in healthcare [6] to send patients to a family doctor first, who may diagnose common health conditions based on the patient's history and symptoms, and then refer them to a specialist [7].

Machine learning-based multiple illness prediction holds the potential to transform healthcare by facilitating more precise and [8] individualised diagnosis, timely interventions, and more efficacious therapies. This strategy does, however, have several drawbacks and restrictions, such as the requirement for representative and varied data, the possibility of algorithmic bias, and the requirement for an ethical and transparent implementation process [9]. Despite these difficulties, the subject of multiple illness prediction using machine learning [10] is developing quickly and has a lot of potential for the future of healthcare. It is anticipated that machine learning algorithms will advance in sophistication and accuracy as more data becomes accessible and technology develops further, improving patient outcomes and general health. Machine learning (ML) is experiencing rapid growth and finds applications across various domains within computer science. It involves extracting meaningful insights from large datasets. ML techniques are utilized in industries, marketing, medical diagnosis [11], and various scientific fields. Medical data analysis, in particular, benefits significantly from ML techniques, which are widely employed in analysing medical datasets. Classification methods are particularly relevant in medical data analysis, as they categorize datasets into predefined categories and predict future outcomes based on the data. Due to their high accuracy and performance, classification methods are preferred in healthcare applications [12]. Machine learning encompasses various forms, including regression, clustering, and classification.

II. RELATED WORK

The study titled "Diagnosis of Parkinson's Disease using Artificial Neural Network" by Anila M et al. [13] aimed to showcase how speech analysis could aid in detecting Parkinson's disease. Various machine learning algorithms, including ANN, Random Forest, KNN, SVM, and XG Boost, were employed to identify the optimal model. Error rates were computed, and performance metrics were evaluated for each model. However, a limitation of this study is its focus solely on artificial neural networks (ANNs) with two hidden layers. While sufficient for basic datasets, this approach may overlook the potential benefits of utilizing deeper neural networks or alternative feature selection methods. A number of models for illness prediction [14] using patient symptoms are provided by some research articles, as indicated in the introductions. The most well-liked and precise models are listed here. For the classification of illnesses based on symptoms, Jianfang et al. [15] used SVM, or Support Vector Machine [16]. Although it takes more time, the SVM model is efficient for illness prediction [17]. Cancer Assessment and Forecasting Tumour categorization, early diagnosis, and prognostic evaluation are all aspects of cancer prediction that make use of machine learning. The examination of genetic markers, clinical data, and histological pictures accomplishes this [18]. Predicting outbreaks and tracking the spread of diseases are two ways in which machine learning models contribute to infectious disease surveillance. Epidemiological and social media data are frequently used in these models.

Predicting Mental Health problems Machine learning has been the subject of research into the prediction of mental health problems including anxiety and depression. Surveys of patients, analyses of text, and data collected by wearable sensors all contribute to the data set [19]. Issues of Ethics and Personal Data Protection A new field of study is developing to address ethical issues with transparency, algorithmic bias, and data privacy [20]. A rising concern is checking that illness prediction models don't break any ethical rules [21]. The resource consumption, including medical and long-term care expenditures, is the objective variable of the study in [22]. A medical care forecasting model utilizing a random forest machine learning method [23] is also included. More than a hundred types of data, including health promotion initiatives, clinical testing, and medical procedures, are incorporated into this approach. For classification, the model uses the mean decrease Gini, and for regression, it makes use of the mean square error (MSE) [24]. Furthermore, a study introduced two machine learning techniques for predicting diabetes patients: the Random Forest algorithm for classification and the XGBoost algorithm for a hybrid approach. Interestingly, the XGBoost [25] algorithm demonstrated better performance, achieving an accuracy rate of 74.10%. Lastly, researchers explored multiple machines learning techniques, including support vector machine, logistic regression, Random Forest [26], Decision Tree, gradient boost, K Nearest Neighbor, and Naïve Bayes algorithm, to predict diabetes. Among these methods, Naïve Bayes exhibited promising results.

III. Problem Identification

A large number of machine learning models now in use for health care analysis focus on a single illness at a time. As an illustration, the first is for liver analysis, the others are for cancer analysis and lung disorders of such kind. A person must visit many websites in order to anticipate more than one sickness. A single study cannot forecast more than one illness under any common framework. The accuracy of some of the models is worse than others, which can impact on patient care. An organisation must install many models in order to analyse the health reports of its patients, which adds to the expense and time involved. A number of the current systems take into account extremely few parameters, which might lead to inaccurate findings. The analysis revealed several risk factors associated with cardiovascular disease, such as hypertension, elevated cholesterol levels, tobacco use, and diabetes. A risk score may be computed using these risk variables to estimate the probability that a person will acquire cardiovascular disease. To identify risk variables and generate a risk score that may be utilized for managing and preventing illness, conventional statistical approaches are employed. 3.1.1

IV.Disadvantages of Existing System

This is one of the main issues with machine learning systems. Inaccurate predictions and incorrect diagnoses may result from biased or insufficient training data utilized to build the system. Because their data could not be well-represented in the training set, this is particularly troublesome for underrepresented communities. When a machine learning model is trained too closely, it develops an excessive level of specialization in its prediction of a particular dataset. This might lead to decreased accuracy and poor generalization to new data. A lot of machine learning algorithms are "black boxes," making it challenging to figure out how they make their predictions. This could be difficulty in the medical field, where it's critical to be able to describe the process used to arrive at a diagnosis. Some diseases are rare, which means that there may not be 17 enough data available to train a machine learning model accurately. This can limit the effectiveness of the system for predicting such diseases. Implementing machine learning systems for healthcare can be expensive and time-consuming. Hospitals and clinics may need to invest in new hardware, software, and staff training to implement these systems effectively

V. Proposed System

In this proposed system involved analysing a multiple disease patient dataset with proper data processing. Different algorithms were used to train and predict, including Decision Trees, Random Forest, SVM, and Logistic Regression [27], ad-aboost. Why Predicting more than one illness at once in a multi-disease model eliminates the requirement to switch between models in order to make a disease prediction [28]. To address data bias, a proposed system would use a diverse range of training data, including data from underrepresented populations, to ensure that the system can accurately predict diseases across all groups. The system would use algorithms that are robust to overfitting and have high accuracy on unseen data. This could be achieved by using techniques such as regularization and cross-validation. To address the lack of interpretability of machine learning models, the proposed system would use explainable AI techniques to provide clear and understandable reasons for its predictions. This would increase the trust and acceptance of the system among healthcare providers and patients.

VI. Suggested System Architecture

System architecture refers to the conceptual model outlining a system's behaviour, structure, and other pertinent aspects. It involves a formal description and representation of the system, arranged to facilitate understanding of its behaviours and structures, known as an architectural description shown in figure 1. A system architecture typically comprises subsystems and system components that collaborate to form the entire system. Architectural description languages are formalized languages used to describe system architecture. In the context of machine learning, where computer systems can learn without explicit programming, the author of this work has utilized three machine learning algorithms: KNN, Naïve Bayes [29], and Logistic Regression. The architectural diagram highlights the key functional connections and offers a high-level overview of the system's principal components.

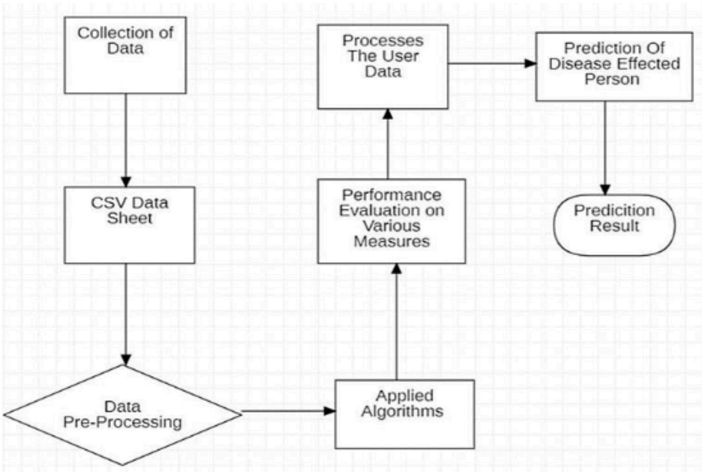


Figure 1: The System Architecture

A sequence diagram, also known as an interaction diagram, illustrates the sequence and relationship among various elements as they interact with each other. It serves as a valuable tool for software engineers and business experts alike, aiding in the documentation of existing processes or the understanding of requirements for new systems.

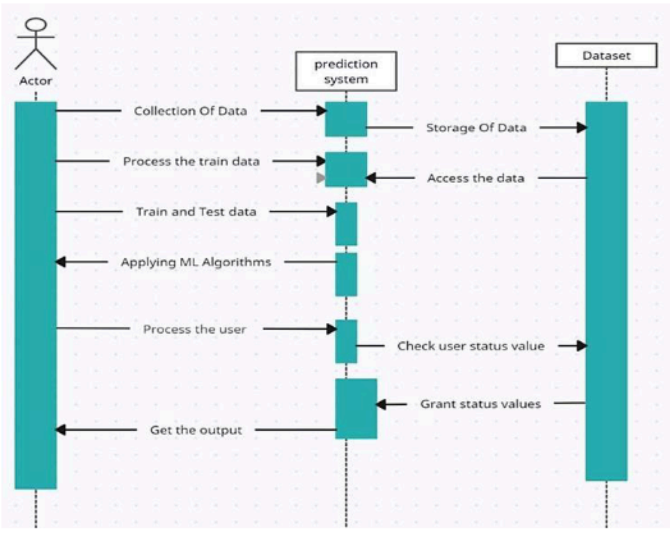


Figure 2: The Sequence Diagram

Additionally, sequence diagrams may be referred to as event diagrams or event scenarios. Sequence diagrams are mostly used when moving from needs described as use cases to the next, more formalized level of elaboration. It is common practice to refine use cases into one or more sequence diagrams. From the figure 2 a sequence diagram the prediction system can collect the data from actor and store the data in dataset. Prediction system processes the train data and access the data from dataset then prediction system uses the train and test data and apply ML [30] algorithms and check user status value and grand status values then get the output.

In the deployment diagram, the hardware intended for running the program is depicted, providing a static view of the system deployment. It outlines the nodes and their connections, establishing the software deployment strategy for the hardware. This diagram bridges the gap between the software architecture designed and the physical system architecture where the program will function as a node [31]. Communication channels are represented to illustrate the connections between multiple nodes involved in the system. A deployment diagram for multiple disease prediction comprises components such as the disease dataset, data preprocessing module, ML [32] algorithms module, and predictive model module. The user interface collects input data from the disease dataset, processes it using ML algorithms, and predicts the disease using the predictive model.

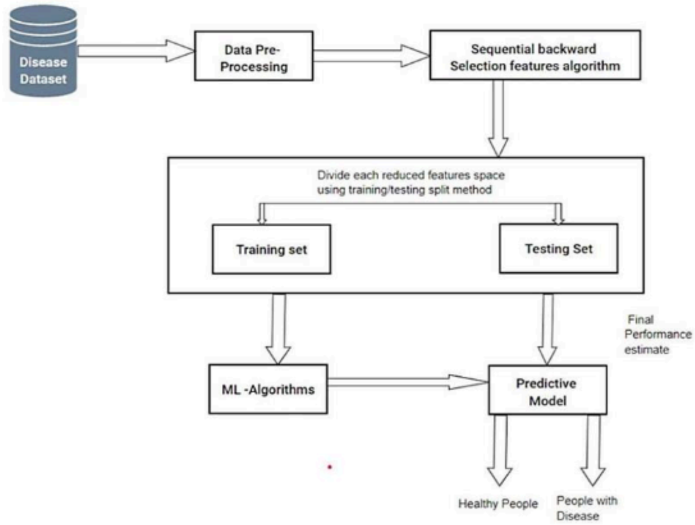


Figure 3: The Deployment Diagram

VII. Implementation

A technical specification or method is realized as a program, software components, or other computer system through computer programming and deployment. This process is known as implementation. Standards or specifications may have a large number of implementations. In object-oriented programming, a concrete class that implements an interface is a specific situation. The initial step in developing the prediction system is selecting the training and testing datasets. In this research, both training and testing datasets were utilized. Attributes in a dataset are characteristics utilized by systems for prediction. Examples include age, gender, sex, and heart rate. Preprocessing is essential to obtain meaningful results from machine learning algorithms [33]. For instance, the Random Forest technique cannot handle datasets with null values. Therefore, null values from the original raw data must be addressed. To handle certain classified values, we can use the following code to transform them into binary values represented as "0" and "1" for our project. There are two methods for balancing imbalanced datasets. Both oversampling and under sampling are present [34]. Reducing the size of the data collection achieves equilibrium under sampling dataset. When there is sufficient data, this method is taken into consideration. Increasing the dataset's size is how Over Sampling achieves dataset balance. This

Procedure.

7.1 Parkinsons Disease Prediction

The Parkinson Disease prediction module is one of the core of a multiple Disease prediction system. It uses data about the Effected and normal people data preferences to generate the result of the patient [35]. It performs the Different machine algorithms likeKNN, XGBoost, SVM, RANDOM FOREST, etc

7.1.1 Attribute Information

1. name: Recording number and ASCII topic name.
2. MDVP: Fo(Hz): Vocal Fundamental Average (in Hz).
3. MDVP: Fhi(Hz): The highest fundamental frequency of the voice.
4. MDVP: Flo(Hz): Minimum Vocal Fundamental Frequency (in Hz).
5. Jitter Metrics:
MDVP: Jitter(%): Percentage of fundamental frequency variation.
MDVP: Jitter(Abs): Absolute jitter value.
MDVP: RAP: Relative amplitude perturbation.
MDVP: PPQ: Five-point period perturbation quotient.
Jitter: DDP: Derivative of amplitude perturbation.
6. Shimmer Metrics:
MDVP: Shimmer: Amplitude variation.
MDVP: Shimmer(dB): Shimmer in dB.
Shimmer: APQ3: Three-point amplitude perturbation quotient.
Shimmer: APQ5: Five-point amplitude perturbation quotient.
MDVP: APQ: Eleven-point amplitude perturbation quotient.
Shimmer: DDA: Difference in amplitude perturbation.
7. NHR and HNR: Measurements of the voice's noise-to-tone component ratio.
8. status: Subject's health status, categorized as (0) for healthy or (1) for Parkinson's.
9. Nonlinear Dynamical Complexity Metrics:
RPD: Nonlinear measurement of underlying frequency change.
D2: Nonlinear dynamical complexity metric.
10. DFA: Exponent of signal fractal scaling.
11. Nonlinear Frequency Change Metrics:
spread1: Spread of nonlinear measures of fundamental frequency change.
spread2: Spread of nonlinear measures of fundamental frequency change.
PPE: Pitch period entropy. Comparison of Models

7.1.2 Comparison of Models

	Model Name	Train Accuracy(%)	Test Accuracy(%)	AUC Score
0	Logistic Regression	82.524272	74.576271	0.887879
1	Decision Tree Classifier	83.980583	88.135593	0.910606
2	AdaBoost	83.980583	88.135593	0.854545
3	Random Forest Classifier	99.029126	84.745763	0.841667
4	kNN	100.000000	98.305085	0.966667
5	SVM	100.000000	94.915254	0.992424
6	XGBoost	100.000000	91.525424	0.956061

We can say that KNN Model is good for our dataset but SVM giving more AUC.

7.1.3 Classification Report

```
knn = KNeighborsClassifier(n_neighbors=1)
knn.fit(X_train, y_train)

y_pred_knn = knn.predict(X_test)

print(knn.score(X_train, y_train))
print(knn.score(X_test, y_test))

print(metrics.classification_report(y_test, y_pred_knn))
```

```
1.0
0.9830508474576272
              precision    recall  f1-score   support

           0       1.00      0.93      0.97        15
           1       0.98      1.00      0.99        44

    accuracy                           0.98        59
   macro avg       0.99      0.97      0.98        59
weighted avg       0.98      0.98      0.98        59
```

7.2 Diabetes Disease Prediction

Using several supervised machine learning techniques, the prediction seeks to identify a patient's likelihood of developing diabetes early on. It uses data about [36] the Effected and normal people data preferences to generate Whether person is affected or not from a particular Disease.

7.2.1 Attribute Information

1. Pregnancies
2. Glucose
3. Blood pressure
4. Skin Thickness
5. Insulin
6. BMI
7. Diabetes Pedigree Function
8. Age

7.2.2 Comparison of Models

```
# Accuracy on test set
print("Logistic Regression: " + str(accuracy_logreg * 100))
print("K Nearest neighbors: " + str(accuracy_knn * 100))
print("Support Vector Classifier: " + str(accuracy_svc * 100))
print("Naive Bayes: " + str(accuracy_nb * 100))
print("Decision tree: " + str(accuracy_dectree * 100))
print("Random Forest: " + str(accuracy_ranfor * 100))

Logistic Regression: 71.42857142857143
K Nearest neighbors: 78.57142857142857
Support Vector Classifier: 73.37662337662337
Naive Bayes: 71.42857142857143
Decision tree: 68.18181818181817
Random Forest: 75.97402597402598
```

7.2.3 Classification Report

```
# Classification report
from sklearn.metrics import classification_report
print(classification_report(Y_test, Y_pred_knn))

              precision    recall  f1-score   support

         0.0       0.81      0.87      0.84       100
         1.0       0.72      0.63      0.67        54

   micro avg       0.79      0.79      0.79       154
   macro avg       0.77      0.75      0.76       154
weighted avg       0.78      0.79      0.78       154
```

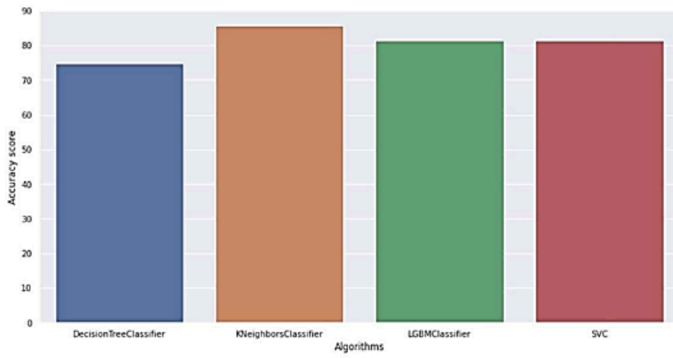
7.3 Heart Disease Prediction

It uses data about the Effected and normal people data preferences to generate the result of the patient [37]. It performs the Different machine algorithms like KNN, XGBoost, SVM, RANDOM FOREST, etc. This seeks to forecast using various supervised machine learning techniques.

7.3.1 Attribute Information

1. Age
2. Sexual
3. Types of Chest Pain
4. Blood pressure at rest
5. Cholesterol serum
6. Blood sugar fasting
7. Resting Heart Diagram Outcome
8. Peak heart rate attained
9. Workout Lessened Angina
10. Fluorescence-colored vessels

7.3.2 Accuracy Results



VIII. Outcome

The findings of the suggested system, which can forecast disease more quickly, accurately, and reliably than the current approach, are shown in this section. A variety of machine learning algorithms are used to get the desired outcomes. It will indicate whether or not the patient [38] has a disease based on the disease they have selected when they add the appropriate parameter. The parameters will display the range of required values; if a value is outside of this range, invalid, or empty, a warning message will appear, requesting that the proper value be added. It will indicate whether or not the patient [39] has a disease based on the disease they have picked when they add the disease-related parameter. The parameters will display the range of required values, if the value is outside of this range, invalid, or empty, a warning message will appear, requesting that the right value be added [40].

8.1 Diabetes Prediction

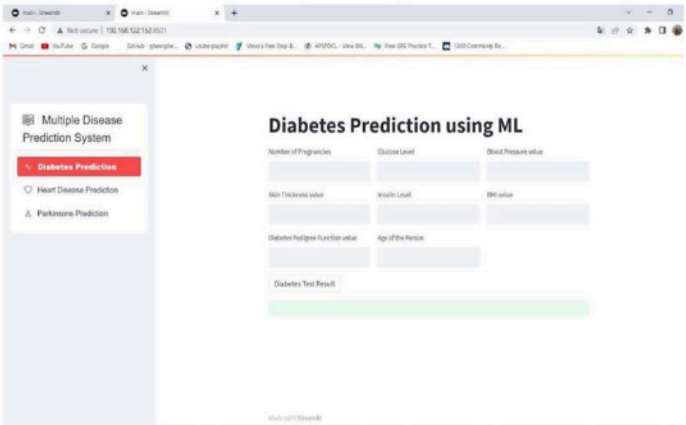


Figure 4: The Diabetes Prediction Home Page

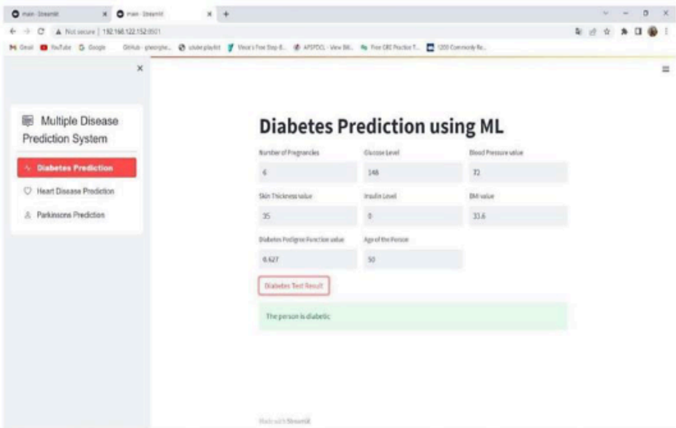


Figure 5: The Diabetes Prediction Result Page

8.2 Heart Disease Prediction

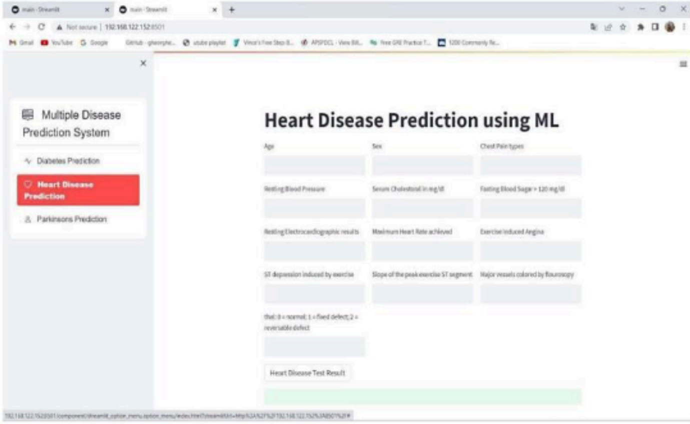


Figure 6: The Heart Disease Prediction Home Page

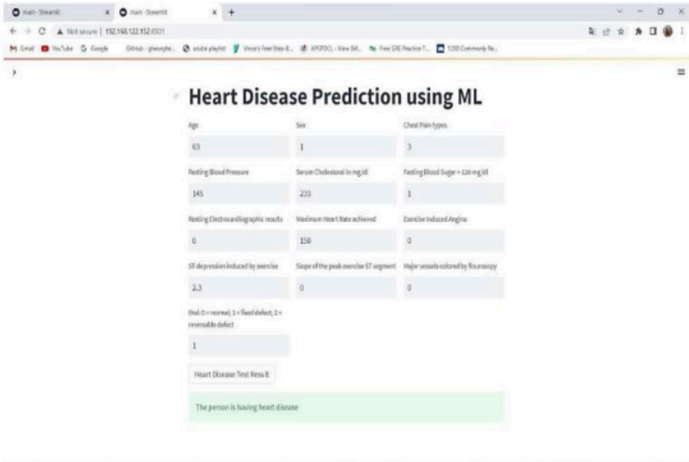


Figure 7: The Heart Disease Prediction Result Page

8.3 Parkinson's Disease Prediction

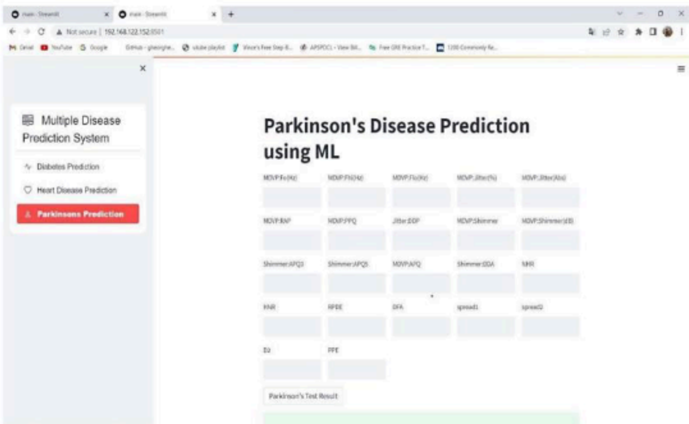


Figure 8: The Parkinson's Disease Prediction Home Page

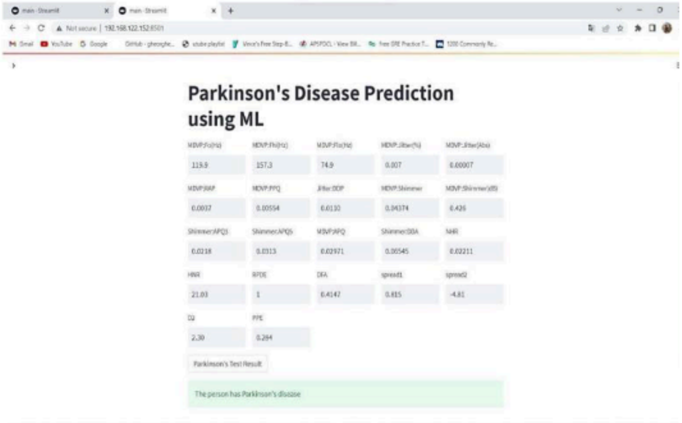


Figure 9: The Parkinson's Disease Prediction Result Page

IX. Conclusion

Because of their lifestyle choices and the condition of the earth, people in the modern day are vulnerable to a wide range of ailments. Early detection and prediction are essential to preventing many disorders from progressing to their most severe manifestations. For doctors, detecting ailments by hand might be difficult at times. Machine learning-based multiple illness prediction is a promising medical approach that has the potential to revolutionize disease identification and treatment. By analyzing large volumes of patient data, machine learning algorithms can uncover patterns and connections that may not be immediately apparent to human physicians. This approach can lead to earlier diagnoses, better treatments, and improved patient outcomes. However, the use of machine learning in healthcare faces challenges and limitations, such as the risk of bias and the need for diverse and representative data. Ongoing research and development in this field are addressing these issues, gradually unlocking the full potential of machine learning for predicting multiple diseases. Machine learning algorithms are expected to grow more complex and accurate as technology advances and more data becomes accessible, which will improve patient outcomes and enable more individualized treatment. Healthcare could undergo a change in the future thanks to the exciting field of research on machine learning for multiple ailment prediction.

X. REFERENCES

[1]. Md R. Hoque and M. Sajedur Rahman, "Predictive modelling for chronic disease: machine learning approach," in Pro ceedings of the 2020 the 4th International Conference on Compute and Data Analysis, pp. 97–101, CA, USA, 2020

[2]. Y. Perwej, Dr. Faiyaz Ahamad, Dr. Mohammad Zunnun Khan, N. Akhtar, "An Empirical Study on the Current State of Internet of Multimedia Things (IoMT)", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), ISSN (Online) 2394-2320, Volume 8, Issue 3, Pages 25 - 42, 2021, doi: 10.1617/vol8/iss3/pid85026

[3]. Littell, C.L., "Innovation in medical technology: Reading the indicators", Health Affairs, Vol. 13, No. 3, (1994), 226-235. https://doi.org/10.1377/hlthaff.13.3.226

[4]. Y. Perwej, Mohammed Y. Alzahrani, F. A. Mazarbhuiya, Md. Husamuddin, "The State-of-the-Art Cardiac Illness Prediction Using Novel Data Mining Technique", International Journal of Engineering Sciences & Research Technology (IJESRT), ISSN: 2277-9655, Volume 7, Issue 2, Pages 725-739, 2018, DOI: 10.5281/zenodo.1184068

[5]. Mobeen, A., Shafiq, M., Aziz, M.H. and Mohsin, M.J., "Impact of workflow interruptions on baseline activities of the doctors working in the emergency department", BMJ Open Quality, Vol. 11, No. 3, 2022

[6]. Ahmed, S., Szabo, S. and Nilsen, K., "Catastrophic healthcare expenditure and impoverishment in tropical deltas: Evidence from the mekong delta region", International Journal for Equity in Health, Vol. 17, No. 1, 1-13, 2018

[7]. Roberts, M.A. and Abery, B.H., "A person-centered approach to home and community-based services outcome measurement", Frontiers in rehabilitation Sci., Vol. 4, 2023

[8]. Y. Perwej, “An Evaluation of Deep Learning Miniature Concerning in Soft Computing”, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, Volume 4, Issue 2, Pages 10 - 16, 2015, DOI: 10.17148/IJARCCE.2015.4203

[9]. Dragana Miljkovic et al, “Machine Learning and Data Mining Methods for Managing Parkinson’s Disease” LNAI 9605, pp 209-220, 2016

[10]. M. A. E. Van Stiphout, J. Marinus, J. J. Van Hilten, F. Lobbezoo, and C. De Baat, “Oral health of Parkinson’s disease patients: a case-control study,” Parkinson’s Disease, vol. 2018, Article ID 9315285, 8 pages, 2018

[11]. Y. Perwej, “The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents”, Transactions on Machine Learning and Artificial Intelligence (TMLAI), which is published by Society for Science and Education, United Kingdom (UK), ISSN 2054-7390, Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.863

[12]. Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, “A Machine Learning Approach to Predicting Diabetes Complications,” Healthcare, vol. 9, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/healthcare9121712.

[13]. Anila M, Dr. G. Pradeepini,” A Review on Parkinson’s Disease Diagnosis using Machine Learning Techniques “, International Journal of Engineering Research & Technology, Vol. 9 Issue 06, June-2020

[14]. Y. Perwej, “Unsupervised Feature Learning for Text Pattern Analysis with Emotional Data Collection: A Novel System for Big Data Analytics”, IEEE International Conference on Advanced computing Technologies & Applications (ICACTA'22), SCOPUS, IEEE No: #54488 ISBN No Xplore: 978-1-6654-9515-8, Coimbatore, India, 2022, DOI: 10.1109/ICACTA54488.2022.9753501

[15]. Cao, J., Wang, M., Li, Y. and Zhang, Q., "Improved support vector machine classification algorithm based on adaptive feature weight updating in the hadoop cluster environment", PloS One, Vol. 14, No. 4, 2019

[16]. Y. Perwej, S. A. Hann, N.t Akhtar, “The State-of-the-Art Handwritten Recognition of Arabic Script Using Simplified Fuzzy ARTMAP and Hidden Markov Models”, International Journal of Computer Science and Telecommunications (IJCST), Sysbase Solution (Ltd), UK, London, ISSN 2047-3338, Volume, Issue 8, Pages, 26 - 32, 2014

[17]. Hamidi, H. and Daraee, A., "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases", International Journal of Engineering, Transactions B: Applications, Vol. 29, No. 7, 921-930, 2016

[18]. Maurano, M., Humbert, R., Rynes, E., Thurman, R., Haugen, E., Wang, H., Reynolds, A., Sandstrom, R., Qu, H., Brody, J., 2012. Systematic Lo calization of Common Disease-Associated Variation in Regulatory DNA. Science, 337, pp. 1190- 1195. https://doi.org/10.1126/science.1222794.

[19]. Kumar, A., 2021. Disease Prediction and Doctor Recommenda tion System using Machine Learning Approaches. International Jour nal for Research in Applied Science and Eng. Tech.. DOI: 10.22214/IJRASET.2021.36234.

[20]. Y. Perwej, Firoj Parwej, Nikhat Akhtar, “An Intelligent Cardiac Ailment Prediction Using Efficient ROCK Algorithm and K- Means & C4.5 Algorithm”, European Journal of Engineering Research and Science (EJERS), Bruxelles, Belgium, ISSN: 2506-8016 (Online), Vol. 3, No.

12, Pages 126 – 134, 2018, DOI: 10.24018/ejers.2018.3.12.989

[21]. Patil, K., Pawar, S., Sandhyan, P., Kundale, J., 2022. Multiple Disease Prognostication Based On Symptoms Using Machine Learning Techniques. ITM Web of Conferences. https://doi.org/10.1051/itmconf/20224403008.

[22]. T. Takura, K. H. Goto, and A. Honda, “Development of a predictive model for integrated medical and long-term care resource consumption based on health behaviour: application of healthcare big data of patients with circulatory diseases,” BMC Medicine, vol. 19, no. 1, pp. 1–16, 2021.

[23]. N. Akhtar, Saima Rahman, Halima Sadia, Y. Perwej, “A Holistic Analysis of Medical Internet of Things (MIoT)”, Journal of Information and Computational Science (JOICS), ISSN: 1548 - 7741, Volume 11, Issue 4, Pages 209 - 222, 2021, DOI: 10.12733/JICS.2021/V11I3.535569.31023

[24]. L. Beretta and A. Santaniello, “Nearest neighbor imputation algorithms: a critical evaluation,” BMC Medical Informatics and Dec.Making, vol. 16, no. 3, pp. 197–208, 2016.

[25]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 785-794, 2016

[26]. Shobhit Kumar Ravi, Shivam Chaturvedi, Dr. Neeta Rastogi, N. Akhtar, Y. Perwej, “A Framework for Voting Behavior Prediction Using Spatial Data”, International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume 10, Issue 2, Pages 19-28, 2022, DOI: 10.55524/ijircst.2022.10.2.4

[27]. Ren, Q., Cheng, H. and Han, H., "Research on machine learning framework based on random forest algorithm", in AIP con. proceedings, AIP Publishing Vol. 1820, 080020, 2017

[28]. Shobhit Kumar Ravi, Shivam Chaturvedi, Dr. Neeta Rastogi, N. Akhtar, Y. Perwej, “A Framework for Voting Behavior Prediction Using Spatial Data”, for published in the International Journal of Innovative Research in Computer Science & Technology (IJIRCST), Volume 10, Issue 2, Pages 19-28, 2022, DOI: 10.55524/ijircst.2022.10.2.4

[29]. N.Akhtar, Devendera Agarwal, “An Efficient Mining for Recommendation System for Academics”, International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878 (online), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020, DOI: 10.35940/ijrte.E5924.018520

[30]. Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, “A Machine Learning Approach to Predicting Diabetes Complications,” Healthcare, vol. 9, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/healthcare9121712

[31]. D. J. Park, M. W. Park, H. Lee, Y.-J. Kim, Y. Kim, and Y. H. Park, “Development of machine learning model for diagnostic disease prediction based on laboratory tests,” Scientific Reports, vol. 11, no. 1, pp. 1–11, 2021

[32]. N. Akhtar, H. Pant, Apoorva Dwivedi, Vivek Jain, Y. Perwej, “A Breast Cancer Diagnosis Framework Based on Machine Learning”, International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Volume 10, Issue 3, Pages 118-132, May-June-2023, DOI: 10.32628/IJSRSET2310375

[33]. Shweta Pandey, Rohit Agarwal, Sachin Bhardwaj, Sanjay Kumar Singh, Y. Perwej, Niraj Kumar Singh, “A Review of Current Perspective and Propensity in Reinforcement Learning (RL) in an Orderly Manner” , the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 9, Issue 1, Pages 206-227, 2023, DOI: 10.32628/CSEIT2390147

[34]. E Benmalek, J Elmhamdi and A. Jilbab, "UPDRS tracking using linear regression and neural

network for Parkinson's disease prediction", International Journal Of Emerging Trends and Technology In Com. Science, vol. 4, pp. 189-193, 2015

[35]. Y. Perwej, Mohammed Y. Alzahrani, F. A. Mazarbhuiya, Md. Husamuddin, "The State-of-the-Art Cardiac Illness Prediction Using Novel Data Mining Technique", International Journal of Engineering Sciences & Research Technology (IJESRT), ISSN: 2277-9655, Volume 7, Issue 2, Pages 725-739, 2018, DOI: 10.5281/zenodo.1184068

[36]. C.R. Pereira, D.R. Pereira, F.A. Da Silva, C. Hook, S.A. Weber, L.A. Pereira, et al., "A step towards the automated diagnosis of Parkinson's disease: Analyzing handwriting movements", Computer-Based Medical Systems (CBMS) 2015 IEEE 28th International Symposium on, pp. 171-176, 2015

[37]. W. Wang, J. Lee, F. Harrou and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning", IEEE Access, vol. 8, pp. 147635-147646, 2020

[38]. Syed Hassan Adil, Mansoor Ebrahim, Kamran Raza, Syed SaadAzhar Ali and Mansoor Ahmed Hashmani, Liver Patient Classification using Logistic Regression, IEEE

[39]. N. Akhtar, Nazia Tabassum, Dr. Asif Perwej, Y. Perwej," Data Analytics and Visualization Using Tableau Utilitarian for COVID-19 (Coronavirus)", Global Journal of Engineering and Technology Advances (GJETA), Volume 3, Issue 2, Pages 28-50, 2020, DOI: 10.30574/gjeta.2020.3.2.0029

[40]. A. Parcerisas et al., "A Machine Learning Approach to Minimize Nocturnal Hypoglycemic Events in Type 1 Diabetic Patients under Multiple Doses of Insulin", Sensors, vol. 22, no. 4, pp. 1665, Feb. 2022