iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VIII **Month of publication:** Aug 2023

DOI: https://doi.org/10.22214/ijraset.2023.55194

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Diseases Prediction Using Classification Algorithm

Rajneesh Thakur[1], Mansha[2], Pranjal Sharma[3], Dhruv[4]
Computer Science Engineering, Chandigarh University Mohali, India

Abstract: ***In recent years, machine learning has grown in popularity, with wide range of applications in medicine industries. Disease prediction is critical in healthcare since it aids in early detection and rapid response. In this paper, we give a detailed analysis comparing the performance of four common classification algorithms for disease prediction: Decision Tree, Random Forest, Naïve Bayes, and K-Nearest Neighbors (KNN). The study analyzes their effectiveness using a diverse dataset including medical records, symptom profiles, and patient demographics. The results indicate that Random Forest performs the best in terms of accuracy, followed closely by Naive Bayes. Decision Tree provides interpretability, while KNN demonstrates respectable prediction capabilities. The paper also explores the impact of feature selection and hyperparameter tuning on algorithm performance. The findings contribute to the field of disease prediction and can assist healthcare practitioners in selecting the most suitable algorithm for accurate predictions, leading to improved patient outcomes and resource allocation.***

Keywords: Classification Algorithms, Machine Learning, Diseases Prediction, Naïve Bayes

I. INTRODUCTION

In recent times, Over the last few years, there has been a significant increase in both the global patient population and the occurrence of various diseases, putting a burden on healthcare systems around the world. Unfortunately, this growth in demand has resulted in higher healthcare costs, driving increasing the cost of medical services in many countries. A doctor's visit is essential to begin therapy for the majority of diseases. However, technological developments and the availability of massive amounts of data give an opportunity to totally revolutionize the diagnostic practice. The examination of patient symptoms is an important part of disease diagnosis and prediction. By carefully analyzing symptoms and utilizing large datasets, algorithms may be able to provide accurate and cost-effective disease forecasts. It is impossible to overestimate the potential impact of such an approach on future medical care delivery. In our effort, we focused on precisely predicting diseases based on patient symptoms. To ensure resilience and dependability, we used four separates algorithms, each tailored to a different aspect of disease prediction. After thorough testing and analysis, we achieved a phenomenal accuracy rate of 92-95%. This level of precision demonstrates the applicability and promise of our technology for improving medical diagnostics. To improve usability and accessibility, we developed an interactive interface that allows for fluid interaction with the system. This user-friendly design makes it straightforward for patients and healthcare professionals to navigate and submit relevant symptom information, greatly accelerating the diagnosis process. Furthermore, we worked hard to clearly depict and communicate the results of our effort and study. By displaying forecasts and statistics We attempt to give clear and concise information to medical experts in order to assist them understand and make decisions based on our findings. Despite our great gains, it is critical to remember that a successful medical diagnostic system cannot be measured merely by accuracy. Sensitivity, specificity, and the ability to handle a wide range of illnesses and symptoms are all important factors to consider. As a result, it is critical that we evaluate our system on a regular basis using a variety of representative datasets to ensure its generalizability and dependability.

Working closely with medical specialists during the development and evaluation phases remains a primary goal. Their knowledge and comments are critical for enhancing our algorithms, ensuring system security, and increasing overall performance.

II. LITERATURE REVIEW

This review of the literature investigates the use of categorization algorithms in predictive modelling for a variety of healthcare challenges. Diabetes, heart disease, brain infection, Alzheimer's disease, and chronic kidney disease (CKD) have all been identified as major issues by experts. [1] To solve these issues, they used classification algorithms to create prediction systems and improve diagnostic accuracy. Diabetes, heart disease, and brain infection were identified as major health issues by Dhiraj Dahiwade. Researchers used a variety of approaches to address these issues, including Naive Bayes, K-Nearest Neighbor (KNN), Decision Trees (DT). They applied these algorithms to patient EHR data to estimate the likelihood of getting Alzheimer's disease. Qian, X built a successful prediction method based on patient EHR data. Furthermore, under the Wearable 2.0 system, IM. Chen advocated the design of smart, machine-washable clothes.

In this case, the performance measurements employedwere FP and REC.

Ankita Tyagi's goals were to improve CKD identification, reduce the influence of risk factors, better understanding of causes and consequences, and evaluate new therapy options.[15] To accomplish the objectives, Tyagi utilized the Chi-square approach, data mining tools, and Classification algorithms, such as Support Vector Machines (SVM), K- Nearest Neighbors (KNN), Decision Trees (DT). were utilized in the disease prediction system. The evaluation of these models was based on the performance metric of Precision (PRE).

In a similar vein, P. Hamsa and Gayathri also identified an issue with statistical prediction models in the evaluation domain, as they often fall short in delivering high-quality outcomes. In response to this problem, Gayathri devised a hybrid solution that combines a fuzzy expert system with various algorithms including Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), and Support Vector Machines (SVM). It is worth noting that Naive Bayes demonstrated an impressive success rate of 95% in their study. Classification accuracy in identifying the diabetes state.[11] Due to the increasing fatality rate, Archana Singh concentrated on the precise diagnosis and forecast of heart related disorders. To create a solution, machine learning concepts and methodologies were used, including algorithms such as SVM and KNN. True Positive (TP) and False Positive (FP) were the performance indicators employed in this study. Pankaj Chittora AL. [9] addressed the risingprevalence of chronic kidney diseases and emphasized the need for early diagnosis to prevent premature deaths. The study utilized the SMOTE technique to balance the dataset, and performance metrics such as TP, FP, Precision (PRC), and Recall (REC) were employed to evaluate the machine learning solution. In summary, these studies have utilized classification algorithms to address healthcare issues such as diabetes, heart disease, brain infection, Alzheimer's disease, and chronic kidney disease. The proposed solutions have demonstrated improved detection, prediction, and diagnosticaccuracy, thereby contributing to the advancement of healthcare practices.[6]

In this paper, the use of machine learning and data mining techniques in research on diabetes is reviewed. Decision trees, support vector machines, and neural networks are only a few of the several diabetes management and prediction techniques covered by the authors. Additionally, they emphasize how crucial feature selection and data pretreatment are to making correct predictions. The obstacles and potential prospects for machine learning in diabetes research are covered in the paper's conclusion.

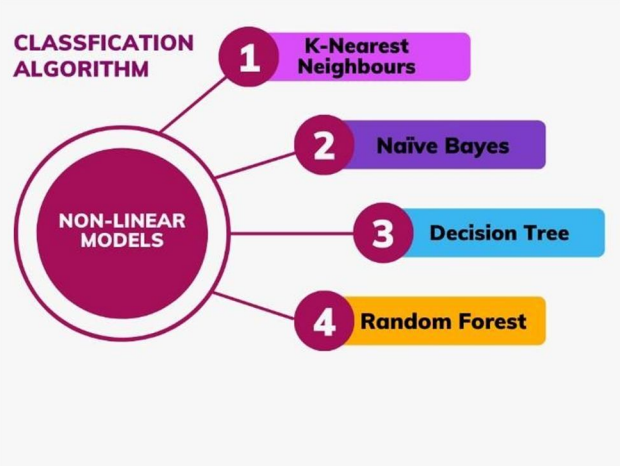


Fig.1 Classification Algorithms

III. PROPOSED SOLUTION

The study article on disease prediction classification algorithm's suggested solution is the creation of a novel strategy to precisely identify diseases based on the symptoms provided by patients. By utilizing machine learning strategies and a classification algorithm, the main goal is to improve the diagnosis procedure.[5] In order to identify the most likely sickness or condition, the system will concentrate on gathering crucial patient data, such as their name and symptoms. In order to accomplish this, a strong machine learning model will be created utilizing a vast dataset that includes a variety of diseases and related symptoms. To verify the model's accuracy and dependability, it will go through a thorough training and validation process. The system will be able to locate patterns and connections within the collection of symptoms by utilizing the power of machine learning, enabling accurate disease prediction. Furthermore, the system will incorporate a user-friendly interface designed to facilitate the input of patient data and symptoms by medical practitioners.[2] The classification algorithm will analyze the data after the symptoms are submitted into the system and offer a list of likely diseases, ranked by likelihood, after it has processed the data.

This will help additional diagnostic procedures or therapies. The gathered patient data and disease predictions will also be kept in a SQL database to guarantee easy data administration and accessibility. Future study and analysis will greatly benefit from this information, which will enable the exploration oftrends, patterns, and potential improvements in disease prediction.[3] The research intends to considerably increase the precision and effectiveness of disease diagnostics by putting this suggested fix into practice. Earlier therapies and better patient outcomes can result from early and precise disease detection. By offering a dependable and adaptable tool for disease prediction, this system has the potential to completely transform the healthcare sector, ultimately benefiting patients as well ashealthcare professionals.

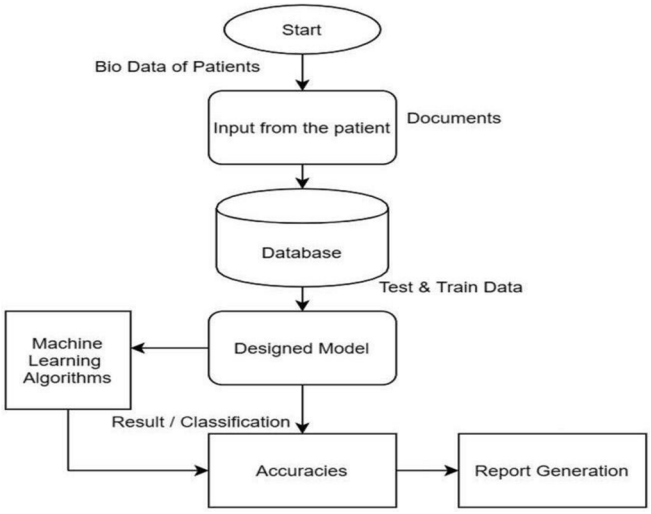


Fig.2 Block Diagram of Working of Model [15]

A. Machine Learning

What exactly is "machine learning"? Every day, people ask this question. Simply said, machine learning is a field of artificial intelligence that is quickly developing. It enables computers to automatically learn from experience, analyses enormous amounts of data, and produce insightful results without the necessity of explicit programming. It stands out for its ability to spot intricate patterns, adjust to shifting circumstances, and make wise predictions or conclusions.[7]

"Machine learning is the extraction of knowledge from data based on algorithms created from training examples." -Emanuel Diamant.

B. Types Of Classification Algorithms

Classification algorithms are categorized as linear and nonlinear algorithms:

1) Linear algorithms

Linear algorithms are ones that can be described mathematically by a linear equation. This indicates a linear relationship between the attributes and the target parameter. The link between height and weight, for example, is linear.[11] Algorithms that are linear Types:

a) *Logistic Regression:* A statistical technique called logistic regression is used to address binaryclassification issues. It forecasts the likelihood of an instance belonging to a specific class.[10]

b) *Support Vector Machines (SVM):* SVM is an adaptable technique that may be used for classification as well as regression. It finds the best hyperplane for differentiating several classes or predicts the value of a continuous targeted parameter.[13]

2) Non-Linear Algorithms

Non-linear algorithms, on the other hand, cannot be represented by a linear equation. This means that A non-linear relationship can be seen inthe association between the characteristics and the goal variable. For example, the relationship between blood pressure and age is non-linear.[14] Non-linear algorithmtypes:

a) *Decision Tree Classification:* It is regarded as a highly successful and adaptable classification tool. It is utilized in picture categorization and pattern recognition. Due to its exceptional adaptability, this approach is utilized for classification in highly intricate issues. Furthermore, it demonstrates proficiency in handling challenges involving multiple dimensions. The structure consists of three components, namely theroot, nodes, leaves.[8]

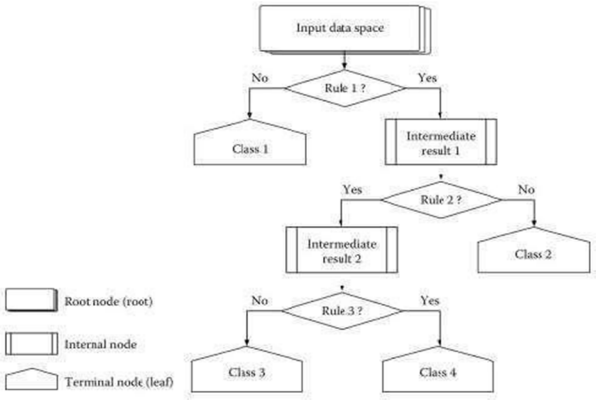


Fig.3 Decision Tree Flow Chart

b) *Nave Bayes:* A nonlinear framework based on the Bayes theorem. The Bayes theorem is an algebraic equation that calculates the likelihood of an event occurring given the likelihood of other events occurring. The likelihood of each characteristic existing in a given class is assumed to be independent of the probability ofthe existence of the other qualities in the class by NaiveBayes. This assumption is known as the naive Bayes assumption.

c) *Random Forests:* It is a method of supervised learning that can be used to solve regression and classification problems. These algorithm's four crucial steps are as follows:

- It chooses random samples of data from the dataset.
- For each sample dataset chosen, it generates decision trees.
- All potential results will now be tallied and decided upon.
- The final prediction will be determined and providedas the classification outcome.

d) *K-Nearest Neighbor (KNN):* This simple yet successful classification and regression approach uses K-Nearest Neighbor (KNN). It forecasts an instance's class or value using the majority vote or average of its neighboring examples in the feature space.[12]

3) *Linear Vs Non-Linear Algorithm For Diseases Prediction*

Why non-linear algorithm prefers over linear ones for disease prediction? Non-linear algorithms are preferred overlinear algorithms for disease prediction via because non- linear algorithms can capture complex relationships and interactions among numerous components that contribute todisease development, allowing for more accurate and nuanced predictions. Linear algorithms, on the other hand, presume a linear relationship between predictors andoutcomes, which may be insufficient to represent the complexity of diseases and their risk factors.

IV. IMPLEMENTATION

In this project, we utilized several commonly used libraries and environments for database analysis and model building.The project leveraged the following libraries:

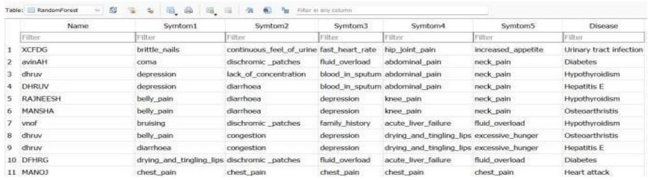
1) *Tkinter:* Tkinter is a standard Python GUI library that allows for the rapid development of graphical user interfaces. It provides various widgets such as buttons, labels, entry fields, check boxes, and list boxes. In this project, Tkinter was employed to create an interactive GUI for our model. The GUI included widgets such as message boxes, buttons, labels, option menus, text fields, and titles.

2) *NumPy:* NumPy is a well-known scientific computing library written in Python. It provides you with sophisticated tools for working with multidimensional arrays. NumPy's primary goal is to efficiently handle multidimensional homogenous arrays. It can generate, manipulate, and process arrays with total, mean, standard deviation, max, min, and other functions. The array processing features of NumPy make it ideal for data handling in our project.

3) *Pandas:* Pandas is a popular Python data analysis toolkit that provides enhanced performance through its backend code, which is implemented in C or Python. Pandas introduces data structures called Series and Data Frames, which are used to store and manipulate data. Series represents a one-dimensional array, while Data Frames represent a two-dimensional data structure. These data structures provide efficient storage and enable various operations on the data. Data Frames make it easier to work with attributes and results. In our project, Pandas played a crucial role in data analysis and management. In our project, we used Pandas Data Frames extensively for handling datasets, data manipulation, and preprocessing tasks.
4) *Scikit-learn (sklearn):* scikit-learn is an open-source Python toolkit that provides a wide range of machine learning algorithms, data preprocessing techniques, cross- validation methods, and visualization tools. It includes support for classification, regression, and clustering tasks, offering algorithms such as support vector machines, random forest classifiers, decision trees, Gaussian naive Bayes, and K-nearest neighbors. In our study, we utilized the built-in classification approaches of scikit-learn, such as decision trees, random forest classifiers, K-nearest neighbors, and as well as naïve Bayes. To analyses the efficacy of our models, we employed scikit-learn's verification cross-validation capabilities and visualization features such as classification reports, confusion matrices, and accuracy scores.
5) *Jupyter Notebook:* Alongside the mentioned libraries, we employed We'll be using Jupyter Notebook as our core platform for development. Jupyter Notebook is open-source software that is free that allows users to create and share notebooks with real-time code, equations, graphics, and text. It supports a wide range of computer languages, including Python, and offers a collaborative environment for activities like data analysis and model construction. We were able to boost our efficiency and collaboration by using Jupyter Notebook.

We used Jupyter Notebook's capabilities to write and execute code, analyze and visualize data, and document our analysis process. Its notebook style allowed us to integrate code cells, markdown cells for written explanations, and visualizations in a single document, making it easier to show and share our work.

Furthermore, we used DB Browser for SQL (formerly known as SQLite Database Browser) to save the results ofseveral algorithms. An excellent visual, open-source tool for generating, developing, and modifying SQLite compatible database files is called DB Browser for SQL. For managing databases, making tables, running SQL queries, and showing data, it features an intuitive user interface. We were able to efficiently store and organize the data generated by our numerous algorithms by using DB Browser for SQL.



	Name	Symtom1	Symtom2	Symtom3	Symtom4	Symtom5	Disease
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	XCFDG	brittle_nails	continuous_feel_of_urine	fast_heart_rate	hip_joint_pain	increased_appetite	Urinary tract infection
2	avinAH	coma	dischromic _patches	fluid_overload	abdominal_pain	neck_pain	Diabetes
3	dhruv	depression	lack_of_concentration	blood_in_sputum	abdominal_pain	neck_pain	Hypothyroidism
4	DHRUV	depression	diarrhoea	blood_in_sputum	abdominal_pain	neck_pain	Hepatitis E
5	RAJNEESH	belly_pain	diarrhoea	depression	knee_pain	neck_pain	Hypothyroidism
6	MANSHA	belly_pain	diarrhoea	depression	knee_pain	neck_pain	Osteoarthristis
7	veoF	bruising	dischromic _patches	family_history	acute_liver_failure	fluid_overload	Hypothyroidism
8	dhruv	belly_pain	congestion	depression	drying_and_tingling_lips	excessive_hunger	Osteoarthristis
9	dhruv	diarrhoea	congestion	depression	drying_and_tingling_lips	excessive_hunger	Hepatitis E
10	DFHRG	drying_and_tingling_lips	dischromic _patches	fluid_overload	acute_liver_failure	fluid_overload	Diabetes
11	MANOJ	chest_pain	chest_pain	chest_pain	chest_pain	chest_pain	Heart attack

Fig.4 Saving the data throughout the project

Overall, the combination of Jupyter Notebook and DB Browser for SQL improved our productivity, allowing us to successfully produce, analyze, and save data throughoutthe project.

A. Dataset

The University of Columbia study at New York PresbyterianHospital in 2004 provided the dataset used in this project fordisease prediction. http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html

It covers 150 disorders, with an average of 810 symptoms for each. 70% of the data utilizedfor training takes into consideration all input components.The symptoms associated with the condition are indicated as1 and remain as 0. It contains information about numeroussymptoms and diseases. The classification system predictsand displays the disease associated with the selected symptom in a text box after selecting a symptom and clickinga button. The dataset contains useful information for developing and accessing disease prediction models.

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0232680_shortness of breath
		UMLS:C0013833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0030990_sweat^UMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0438716_pressure chest
UMLS:C0011847_diabetes	1421	UMLS:C0013810_polyuria
		UMLS:C0085602_polydipsia
		UMLS:C0232680_shortness of breath
		UMLS:C0008031_pain chest
		UMLS:C0004093_asthenia
		UMLS:C0027497_nausea
		UMLS:C0085619_orthopnea
		UMLS:C0034642_rale
		UMLS:C0030990_sweat^UMLS:C0700590_sweating increased

Fig.5 Dataset containing information for developing and accessing disease predictionmodels.

Using the training set, and its performance is evaluated usingthe testing set. To finds the best accurate disease forecast system, many classifying methods, including Decision Trees, Random Forest, and Naive Bayes, are used to the dataset [16]. Following preprocessing of data, feature selection or extraction may be undertaken to limit the number of variables and select the most significant for disease prediction. This can help enhance the performance of the classification algorithm and reduce the risk of overfitting. Additionally, hyperparameter adjustment can be used to improve the performance of the chosen classification method. This entails tweaking the algorithm's parameters to discover the ideal combination of parameters that maximizesits accuracy on the testing set.

To ensure reproducibility, the entire methodology should bedocumented, including the data collection process, preprocessing steps, feature selection or extraction techniques, classification algorithms used and theirparameters, and evaluation metrics used to assess the algorithm's performance.

Finally, the accuracy of the classification algorithm for disease prediction may vary based on the quality and quantity of patient data gathered, the preprocessing and feature selection techniques employed, and the classification algorithm chosen. As a result, additional research and development may be required to optimize the methodology for disease prediction utilizing classification algorithms. Once the algorithm has been chosen, it is utilized to predicta new patient's condition based on their symptoms. The system receives the patient's name and symptoms, and the algorithm guesses the disease based on the patterns identified in the training data. The prediction's accuracy is then assessed by comparing it to the actual disease diagnosed bya medical practitioner.

Overall, the disease prediction methodology employs collecting patient data, storing it in a SQL database, preprocessing the data, training and evaluating severalclassification algorithms, selecting the most accurate algorithm, and using it to predict the disease of a new patient based on their symptoms.

V. METHODOLOGY

Several phases are included in the process for the research study on disease prediction using classification algorithms. To begin, a patient information dataset is produced by gathering data from multiple sources such as hospitals, clinics, and medical research institutions. This information contains the patient's name, symptoms, and the ailment withwhich they have been diagnosed.

The dataset is then placed in a SQL database to facilitate dataprocessing and analysis. The data is then cleaned and alteredso that it is ready for analysis. This could include eliminating duplicates, addressing values that are missing, and encoding category variables, among other things.

The data is separated into training and testing sets after preprocessing. The training set is used for perfecting the classification algorithm, while the testing set is used to evaluate its performance. Many classification methods, such Decision Trees, Random Forest, and Nave Bayes, are used tothe dataset to discover the most accurate approach for diseaseprediction. The data is separated into training and testing setsafter preprocessing. The classification algorithm is trained

VI. RESULT

This program provides an automatic diagnostic method based on user input to save time and minimize expenses connected with the first diagnostic process.

The program accepts symptoms from the user and properly predicts diseases as output within the text field.

The technology forecasts diseases based on signs of infections or any discomfort experienced by the user. The Naive Bayesian algorithm is used for disease prediction. Extensive literature research has demonstrated that this approach produces great accuracy with massive datasets. The GUI offers labels for all probable diseases' symptoms. Symptoms are chosen carefully, and forecasts are formed. The dataset is split into 70% for training and 30% for datatesting. Training and testing are carried out within the GUI, and the he obtained results are available.

The same procedure is used for the Random Forest decision tree method and the K-Nearest neighbors (KNN) algorithm. The GUI provides disease labels as well as symptoms. The disease-specific symptoms are chosen, and predictions are created using the appropriate algorithm. However, because the algorithms are already known, thereis no need to explain them further.

A. Analysis of Algorithms on Training Data

In the training phase, the algorithms were exposed to the medical records of 41 patients who showed a combination of symptoms suggesting a vulnerability to disease. To mitigate the risk of overfitting, the training process considered 95 out of the total 132 symptoms. After the completion of training, each algorithm obtained a precision score.

Algorithm Used	Accuracy Score
Decision Tree	0.932927
Random Forest	0.936179
K Nearest Neighbour	0.932927
Naïve Bayes	0.942927

Fig6. Accuracy score of Algorithms

B. Graphical user interference Result

Fig.7 Disease Predictor Model using Classification Algorithm

The created GUI detects the user's symptoms. When the noneoption is selected, Users can choose a symptom from aprovided list, with the option to select up to five symptoms. Once the symptoms are selected, an algorithm is employed toevaluate the symptoms and determine the underlyinginfection based on predefined rules.

XYZ stated his symptoms as "chest pain", "backpain", "fast heart rate", "increased appetite" and "diarrhoea". The following are algorithm predictions:

1) Decision Tree: Diabetes
2) Random Forest: Diabetes
3) Naive byes: Hypertension
4) K Nearest Neighbor: Gastroenteritis

The overall purpose of this program is to develop an effective and affordable diagnostic system through automating disease prediction using user-supplied symptoms.

VII. CONCLUSION AND FUTURE WORK

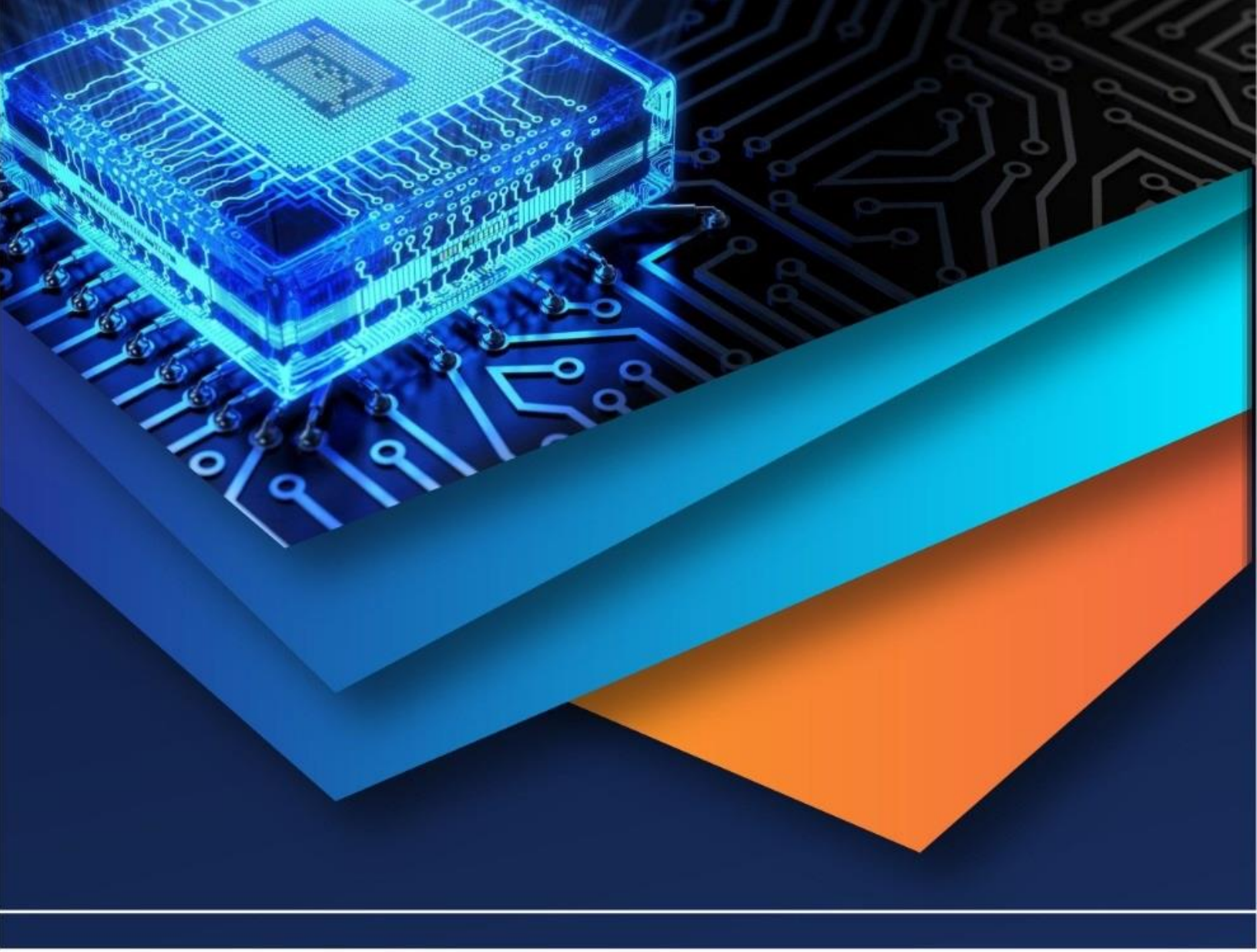
Finally, this report demonstrated the successful developmentof a disease prediction system using machine learning techniques. The system's goal is to forecast diseases based ongiven symptoms, with the goal of reducing the rush at hospital emergency rooms and alleviating the stress on medical staff. The system employs four distinct algorithms and obtained an average accuracy of roughly 94%, demonstrating its dependability in carrying out the desired duty.

The benefits of utilizing machine learning for disease prediction are numerous, including earlier identification and diagnosis, more timely treatments, and better treatment outcomes. The capacity of the system to save user-entered data in a database enables for future modifications and the development of improved versions of such systems. Various algorithms, such as decision trees, random forests, naive bayes, [17]and deep learning models, have been addressed and effectively applied to a wide range of diseases throughout the research. The system also has an intuitive user interface and a variety of visual representations of collected data and findings. In the future, it is important to continue exploring and comparing alternative classification methods in order to enhance disease prediction models. Furthermore, the system could benefit from more comprehensive and diverse datasets to improve accuracy and generalizability. It would also be beneficial to perform comprehensive evaluations and validations of the system using real-world patient data. Furthermore, continued study and collaboration with medical professionals can help refine and expand the system's capabilities, making it an even more trustworthy disease prediction tool.

REFERENCES

[1] Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing disease prediction model using machine learning approach." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.
[2] Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." 2018 Fifth international conference on parallel, distributed and grid computing (PDGC). IEEE, 2018.

[3] Hamsagayathri, P., and S. Vigneshwaran. "Symptoms based disease prediction using machine learning techniques." 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV). IEEE, 2021.
[4] Singh, Archana, and Rakesh Kumar. "Heart disease prediction using machine learning algorithms." 2020 international conference on electrical and electronics engineering (ICE3). IEEE, 2020.
[5] Chittora, Pankaj, et al. "Prediction of chronic kidney disease-a machine learning perspective." IEEE Access 9 (2021): 17312-17334.
[6] Mir, Ayman, and Sudhir N. Dhage. "Diabetes disease prediction using machine learning on big data of healthcare." 2018 fourth international conference on computing communication control and automation (ICCUBEA). IEEE, 2018.
[7] Kanchan, B. Dhomse, and M. Mahale Kishor. "Study of machine learning algorithms for special disease prediction using principal of component analysis." 2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC). IEEE, 2016.
[8] Ayeldeen, Heba, et al. "Prediction of liver fibrosis stages by machine learning model: A decision tree approach." 2015 Third World Conference on Complex Systems (WCCS). IEEE, 2015.
[9] Barik, Shekharesh, et al. "heart disease prediction using machine learning techniques." Advances in Electrical Control and Signal Systems: Select Proceedings of AECSS 2019. Springer Singapore, 2020.
[10] Mir, Ayman, and Sudhir N. Dhage. "Diabetes disease prediction using machine learning on big data of healthcare." 2018 fourth international conference on computing communication control and automation (ICCUBEA). IEEE, 2018.
[11] Jaspreet Singh, Shruti Agarwal, Piyush Kumar, Kashish, Divyansh Rana, Rohit Bajaj. "Prominent Features based Chronic Kidney Disease Prediction Model using Machine Learning", 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), 2022
[12] Matura, R., Thakur, R. & Prathimesh (2023). Comparing the Performance of Different Supervised Learning Algorithms. Journal of Artificial Intelligence and Capsule Networks, 5(1), 52-68.
[13] doi:10.36548/jaicn.2023.1.00
[14] Gokul, S., M. Sivachitra, and S. Vijayachitra. "Parkinson's disease prediction using machine learning approaches." 2013 fifth international conference on advanced computing (ICoAC). IEEE, 2013.
[15] Mathew, Rohit Binu, et al. "Chatbot for disease prediction and treatment recommendation using machine learning." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2019.
[16] Shankar, VirenViraj & Kumar, Varun & Devagade, Umesh & Karanth, Vinay & Kumaraswamy, Rohitaksha. (2020). Heart Disease Prediction Using CNN Algorithm. SN Computer Science 1 10.1007/s42979-
[17] 020-0097-6.
[18] Gupta and M. K. Gupta, "Prediction of Diseases Using Different Machine Learning Approaches," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 712-717, doi: 10.1109/ICIEM54221.2022.9853132.
[19] R. Kumar, P. Thakur and S. Chauhan, "Special Disease Prediction System Using Machine Learning," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 42-45, doi: 10.1109/COM-IT-CON54601.2022.9850843.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY