IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

DL (Diagnostic Logistic)

Ankita Singh, Khushi Tyagi, Lavish Sadh, Manan Singh Ms. Vernika Singh
Raj Kumar Goel Institute of Technology, Ghaziabad

Abstract: *In today's world, people [7] face various health problems due to their lifestyle choices and environment. It is challenging to make accurate predictions because of the complexity of the situation. Coping with illnesses is one of the most difficult tasks. Medical sciences gather a substantial amount of data each year. This data can be analyzed to benefit patients' early care, and the abundance of medical information has contributed to more accurate analyses. Data mining helps to uncover hidden patterns in vast amounts of medical data, enabling the identification of diseases based on patient symptoms. Algorithms, such as K Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest, are utilized to accurately forecast diseases.*

Keywords: KNN.

I. INTRODUCTION

To make an accurate diagnosis in general disease prediction, various factors such as a person's lifestyle choices and medical test results are taken into consideration. This approach helps to predict the likelihood of a general disease and whether the risk is low or high [6].

Computers are now more intelligent thanks to machine learning, which also gives them the ability to think.

Many analysts believe that learning is essential to generate insights. These techniques are available in various forms, such as unsupervised, semi-supervised, supervised, reinforcement learning, evolutionary learning, and deep learning. These techniques are utilized to classify vast amounts of data with great accuracy.

This paper's goal is to examine or predict diseases on the basis of different symptoms.

Therefore, we employ SVM (support vector machine) ,K-Nearest Neighbor (KNN) and Random forest algorithm.

The daily increase in data makes it crucial to use it for accurate disease prediction. However, processing large amounts of data is also crucial in general, which is why data mining plays a significant role.

Machine learning makes it simple to classify huge datasets. Understanding how to diagnose patients correctly through clinical examination and assessment is essential. Inadequate information management has had an impact on the quality of the data association.

A legal method must be found to focus and process information in a viable and effective manner as data volumes increase. In order to build a classifier that can separate the data depending on different criteria, various machine learning software are used.

Classifiers are used to divide the dataset into two or more classes and analyze medical data for disease prediction. Machine learning has become ubiquitous, and people may unknowingly use it multiple times throughout the day.

II. METHODOLOGY

1) K-Nearest Neighbour Algorithm

K-Nearest Neighbour is a supervised learning algorithm that is simple to implement. It assumes that new and existing cases are comparable, and new instances are classified based on their similarity to existing categories. By using the K-NN method, fresh data can be quickly and accurately sorted into an appropriate category. Although it is more frequently used for classification, this approach can be utilized for both classification and regression issues. Since it saves the training dataset rather than instantly learning from it, the technique is also referred to as a lazy learner algorithm.During the training phase, the KNN algorithm simply saves the dataset and classifies fresh data based on its similarity to the training data.

For example[8], if we have a picture of a creature that resembles both a cat and a dog, we can use the KNN method to identify it. The KNN model will search for features in the new data set that are similar to those in the photographs of cats and dogs, and based on those features, it will classify the data as belonging to either the cat or dog group.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

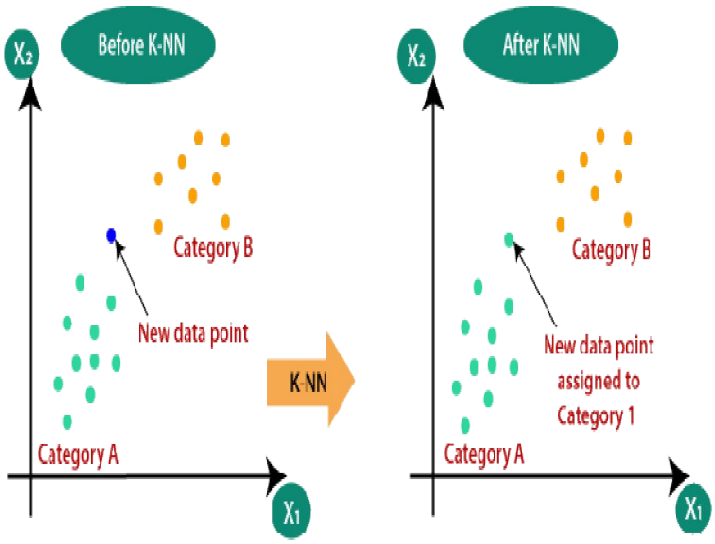


Figure 1.1 KNN algorithm

2) Support Vector Machine

It is primarily used for Classification issues. The SVM algorithm aims to establish a decision boundary that can divide n-dimensional space into classes to classify new data points in the future. The optimal decision boundary is a hyperplane. SVM selects the extreme points that help to create the hyperplane. These extreme instances are represented by support vectors, which give the Support Vector Machine method its name. For example, a decision boundary or hyperplane is used to classify two distinct hyperplanes, as shown in the diagram below.

We can use the example we used for the KNN classifier to understand SVM. For instance, let's say we want to tell a cat from a dog, and we spot a strange cat that also looks like a dog. The SVM technique allows us to create such a model. We will first train our model with multiple images of cats and dogs so that it is accustomed to the different characteristics of cats and dogs before testing it with this strange species.

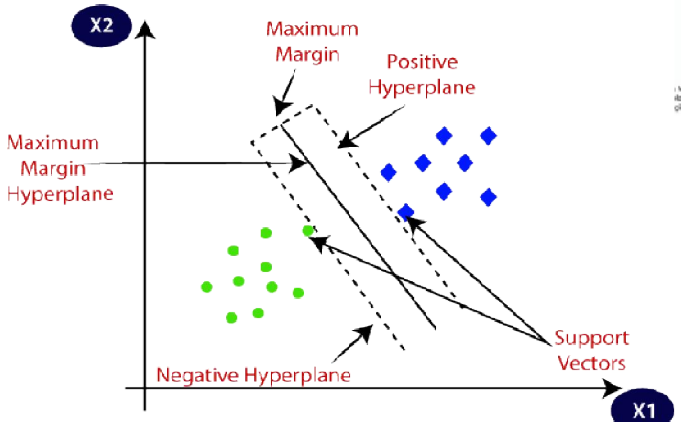


Figure 2.1 SVM algorithm

3) **Random Forest Algorithm**

Data scientists frequently employ the random forest method since it is well-liked in the field. This algorithm is used to solve classification and regression issues. Building decision trees from various samples and utilizing their average for classification or majority vote for regression is how the method operates.

The Process of the Random Forest Algorithm

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

Step 1: - The Random Forest algorithm constructs each decision tree in the model using a subset of data points and a subset of features.This process is repeated multiple times, resulting in a collection of decision trees that are used for classification or regression. By using different subsets of data and features, the Random Forest model helps to reduce overfitting and improve the accuracy of predictions.
Step 2: For each and every sample, a unique decision tree is created.
Step 3: Output will be produced by each decision tree.
Step 4: For classification and regression, the final result is evaluated using the majority vote or averaging.

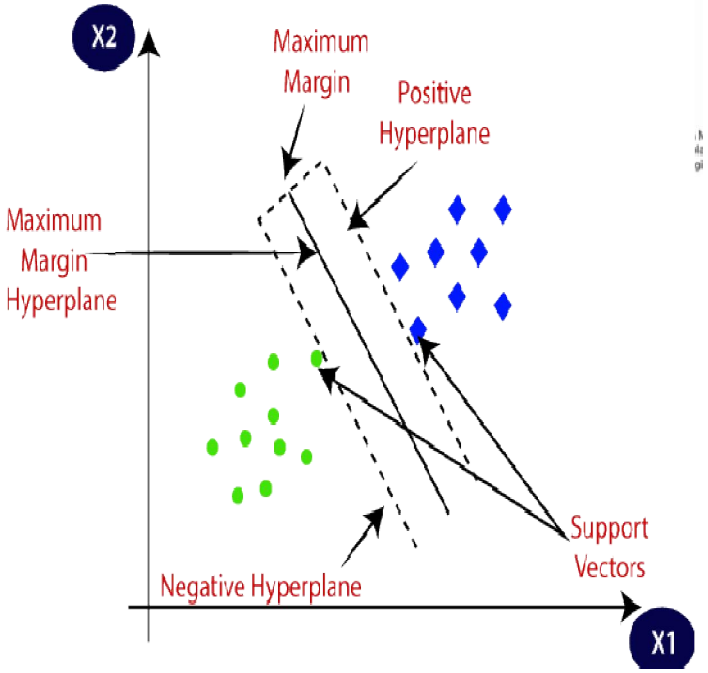


Figure 3.1 Random Forest algorithm

III. LITERATURE

Chronic diseases are a major problem in the healthcare industry everywhere in the world. The medical statement claims that chronic diseases are to blame for the rising death rate among people. Over 70% of the patient's income is spent on the disease's therapies. Therefore, reducing the patient's danger of passing away is really crucial. The development of medical research facilitates the acquisition of health-related data. The patient's medical history, demographic information, and results of medical analyses are all included in the healthcare data. Depending on the geographies and the types of environments in those places, the diseases that result could vary. Therefore, in addition to the disease information, the data set should include information about the patient's environment and residence.
The integration of information has accelerated the evolution of the healthcare sector in recent years.
Information Technology is a part of it. The goal of integrating IT into healthcare is to improve people's quality of life by making it more comfortable and inexpensive, similar to how cell phones did so. This might be making healthcare intelligent, such as the development of the smart ambulance, smart hospital infrastructure, and so on, which benefits patients and doctors in various ways.
IM.Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn gives a concept for a wearable 2.0 system where
The aim is to develop intelligent clothing that can contribute to the quality and quantity of healthcare services in the future. One example of this is an IoT-based data collection system developed by Chen, which uses sensor-based smart clothing that can be washed. This clothing is worn by the patient, and the data collected by the sensors is transmitted to a cloud-based system. The collected physiological data is analyzed to provide insights into the patient's health status, including their emotional state. The washable smart fabric is designed with sensors, electrodes, and wires to collect and transmit data accurately. This innovation has the potential to improve patient care and revolutionize the healthcare industry.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

It was able to capture the patient's physiological state with the aid of this cloth. And this data is utilized for the analysis. discussed the difficulties encountered when creating the wearable 2.0 architecture.

The current healthcare system is facing various challenges such as gathering physiological data, harmful psychological impacts, opposition to wireless body area networking, and sustainable collection of massive physiological data. In addition, several activities are performed on data, including analysis, monitoring, and prediction. To address these challenges, the author proposes Wearable 2.0's smart clothing, which consists of sensor integration, electrical-cable-based networking, and digital modules. These functional elements have numerous applications, such as chronic disease tracking, elderly care, and emotion management.

B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang [2] using the patient's EHR data, a risk prediction system for Alzheimer's disease was developed. In this scenario, an active learning approach was utilized to address a real problem faced by the patient. The active risk prediction system was designed to consider the likelihood of the patient developing Alzheimer's disease.

Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri [4] a health-based CPS (Cyber-Physical System) has been developed and deployed on the cloud to effectively manage the vast amount of biological data. Y. Zhang highlighted the rapid growth of data volume in the medical industry. The main challenge of managing large data sets is that they are generated quickly and stored in different formats. To address this challenge, the CPS system utilized two key technologies - big data and cloud computing. The system enabled various cloud-based operations such as data analysis, monitoring, and prediction. This technology can provide insights into how to handle and manage the vast volume of biological data on the cloud. The CPS system comprises of three layers - data collection, data administration, and data-oriented layers. The data collection layer collects data in a standardized format. The parallel computing and distributed storage data management layer provides efficient data management.

With the aid of the health-cps system, 8 different operations are carried out by this system. Additionally, the vast array of healthcare services that this system is aware of this.

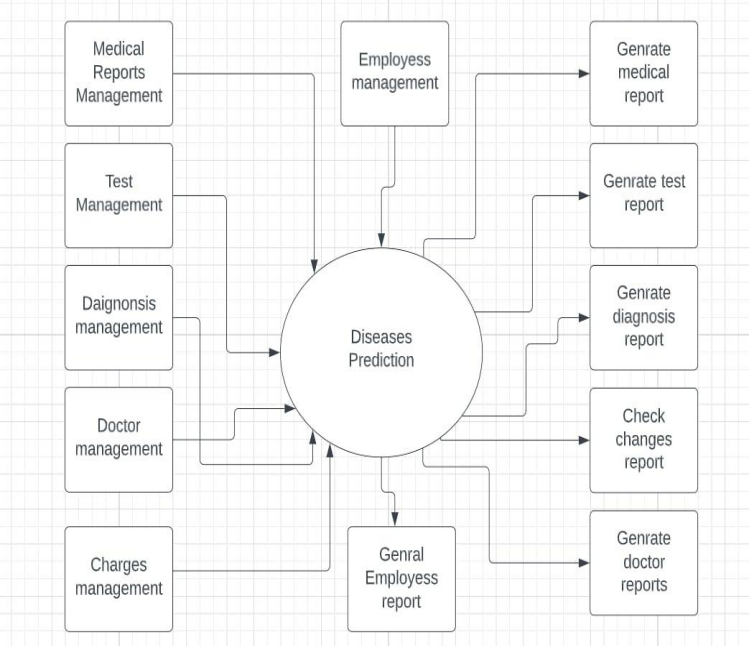
L. Qiu, K. Gai, and M.[5] Qiuthis study explored the effective management of a vast amount of hospital data using the cloud and proposed a telemedicine system. The author suggested enhancements to the telehealth system that primarily focused on cloud-based data exchange among all telehealth services. However, other issues such as network bandwidth and virtual machine switching can also hinder cloud data sharing. To improve data sharing using data sharing concepts, a cloud-based solution is recommended in this study. The proposed solution considers temporal restrictions, network capabilities, and transmission probability for better data sharing. A novel, optimal method for sharing large amounts of data was developed in this study, which provides users with the best approach for processing biomedical data.

IV. PROPOSED WORK

As we've shown in this paper, we can descry conditions using Machine literacy with the help of symptoms. There are different criteria we can descry the needed result similar to the system for prognosticating conditions, where we developed the top-position process of prognosticating conditions. It provides a simple summary of the entire system. It's intended to give a quick overview of ails, croakers, and reports while displaying the system as a single, high-position process, along with its connections to cases and croaker External realities.

* Managing all the Cases
* Managing all the Symptoms
* Managing all the conditions
* Managing all the Croakers
* Processing Case records and induce reports of all Cases.
* Processing Medicines records and induce reports.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023



V. RESULT DISCUSSION

A. Experimental Setup

1) The system was executed on a Windows 10 (64-bit) PC with an Intel Core i5-6200U processor running at a clock speed of 2.30 GHz and 8GB of RAM.

2) Dataset

Patient dataset retrieved from GitHub.

B. Comparison Results

In this section, the performance of three classification algorithms, KNN, SVM, and Random Forest Algorithm, is presented based on their accuracy and the time required for execution. The accuracy of these algorithms is compared for various thresholds, and the results are reported. It was observed that SVM outperformed KNN and Random Forest in terms of accuracy.

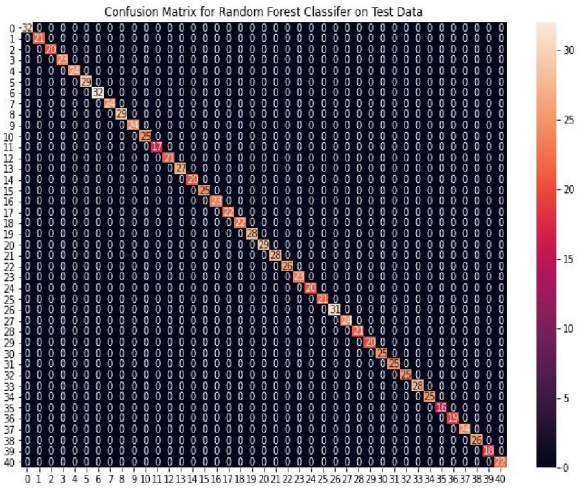


Figure 4.1 Confusion Matrix of Random Forest

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

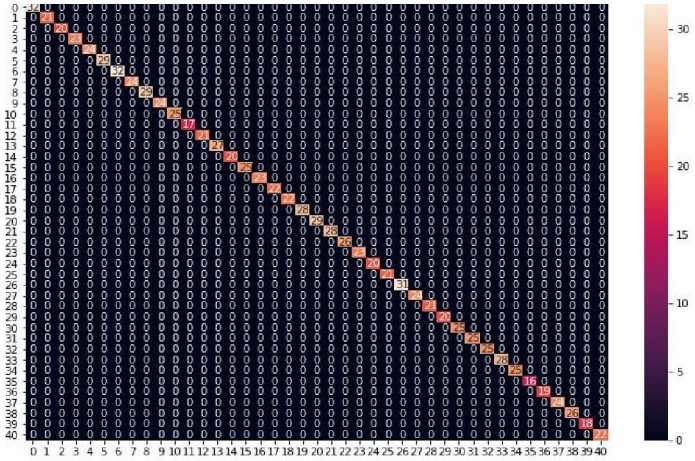


Figure 4.2 Confusion Matrix of SVM

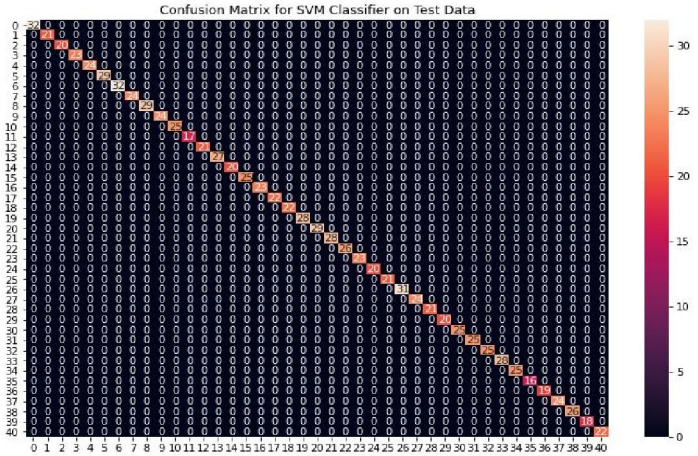


Figure 4.3 Confusion Matrix of KNN

VI. CONCLUSION

In this study, we proposed a comprehensive machine learning algorithm-based complaint prediction system. KNN, SVM, and Random Forest algorithms were utilized to classify case data since medical data is expanding rapidly in the modern medical world and needs to be reused to make accurate complaint predictions based on symptoms. By providing input of case records, we were able to generate an accurate general complaint prediction based on the position of the disease threat prediction. This technique allows for disease and risk prediction with minimal effort and expense. The performance of these algorithms was compared based on accuracy and processing time. The SVM algorithm outperformed the other two algorithms in terms of accuracy, and its processing time was lower than that of the other two algorithms. Therefore, in terms of accuracy and timing, SVM is considered the superior algorithm.

REFERENCES

1] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities"," IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity-based method for interactive patient risk prediction," Springer Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[3] IM. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," IEEE Common., vol. 55, no. 1, pp. 54–61, Jan. 2017.

[4] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health PS: Healthcare cyber physical system assisted by cloud and big data," IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

[5] S Mohan, C Thirumalai, G Srivastava Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, volume 7. Posted: 2019

[6] heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. Int. J. Intell. Eng. Syst, volume 12, issue 1. Posted: 2019

[7] A Mir, S N Dhage. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), p. 1 – 6. Posted: 2018

[8] M Mannerizing, M J Rahman, B Ahammed, M M Abedin. Classification and prediction of diabetes disease using machine learning paradigm. Health Information Science and Systems, volume 8, issue 1. Posted: 2020