

## Comparative Analysis of Various Algorithms on Multi -Diseases Using Machine Learning

**S M Hassan Rizvi, Anjana Mishra, Shubham Kumar Singh, Chirag Kumar Jain**  
Department of Computer Science & Engineering, C.V. Raman Global University, India

Correspondence Author: **Anjana Mishra**

**Abstract :** There are numerous machine learning approaches that can perform predictive analytics on vast volumes of data in a range of businesses. Although using predictive analytics in healthcare is challenging, it will eventually help practitioners make quick choices about the health and treatment of patients based on vast amounts of data. Globally, diseases including liver disease, diabetes, kidney diseases, cancer and heart-related diseases are responsible for a large number of fatalities, however the majority of these deaths are the result of improperly timed disease check-ups. Due to a lack of medical infrastructure and a low doctor-to-population ratio, the aforementioned issue exists. According to data, India has a doctor-to-population ratio of 1:1456 compared to the WHO's suggested ratio of 1 doctor to 1000 patients, demonstrating a physician shortage. If not identified early, diseases including diabetes, liver, kidney, cancer and heart disease pose a risk to humanity. As a result, many lives can be saved by early detection and diagnosis of these disorders. The main goal of this research is to use machine learning classification algorithms to anticipate dangerous diseases. Diabetes, heart disease, liver, cancer and heart diseases are all covered in this study. Our team developed a medical test online application that uses the idea of machine learning to make predictions about various diseases in order to make this run smoothly and be accessible to the general public. Our goal in this effort is to create a web application that uses machine learning to forecast numerous ailments, such as liver, diabetes, kidney, cancer and heart disorders.

**Keywords :** Logistic Regression, Support Vector Machine (SVM), K-Nearest - Neighbours (KNN), Chronic Kidney Disease (CKD), Random Forest

## 1. Introduction

The leading causes of death in today's society are heart disease, cancer, kidney diseases, liver diseases and diabetes. A set of disorders that damage your heart are collectively referred to as heart disease. Heart disease includes conditions including arrhythmias (problems with heart rhythm), coronary artery disease, and congenital heart defects (the defects of the heart you are born with). Instead of "cardiovascular disease," the term "heart disease" is frequently used. The term "cardiovascular disease" typically refers to illnesses that damage your heart's rhythm valves or muscles, as well as conditions that cause heart attacks, angina (heart discomfort), or stroke. According to a report in "Science Direct," of India's almost 11 million annual deaths, 28 percent are attributable to cardiovascular disease. The "Centres for Disease Control and Prevention" estimate that in the US, a cardiovascular disease death occurs every 36 seconds.

Diabetes is a condition that develops when your blood glucose level, commonly known as blood sugar, is extremely high. Your main source of energy comes from your blood sugar, which is obtained from the food you eat. Insulin is a hormone secreted by the pancreas that helps fuel metabolic processes by removing glucose from diet. According to the "International Diabetes Federation," diabetes kills 42 lakh people worldwide every year, and it costs over 760 billion USD (as a part of health expenditure). Epidemiology of Diabetes reports that over 10 lakh people in India lose their lives to diabetes each year, and the "Indian Heart Association" predicts that by 2035, approximately 11 crore people would have the disease.

Chronic kidney disease (CKD) is the inability of the kidneys to carry out their normal blood-filtering role, among other things. Chronic refers to the slow deterioration of kidney cells over a protracted length of time. This illness causes severe renal failure, which results in the body accumulating a lot of fluid and the kidneys losing their ability to filter blood. This causes the body's potassium and calcium sodium levels to rise alarmingly. High concentrations of these chemicals cause a number of additional health problems in the body. CKD in many cases leads to permanent dialysis or kidney transplants. A history of kidney disease in the family also leads to high probability of CKD. Literature shows that almost one out of three people diagnosed with diabetes have CKD. Literature also presents evidences of early identification and care of CKD [15] can improve the quality of the patient's life.

The liver is the largest organ of the body and it is important for digesting food and releasing the toxic element of the body. Alcohol consumption and viruses cause liver injury, which puts a person in danger of dying. There are many different kinds of liver diseases, including cirrhosis, liver cancer, tumours, and hepatitis. Among them, liver diseases and cirrhosis are the leading killers. Consequently, one of the major health issues in the globe is liver disease. Around 2 million individuals worldwide pass away from liver disease [16] each year. According to the Global Burden of Disease (GBD) project, which was released in the journal BMC Medicine, one million people died

from cirrhosis and one million had liver cancer in 2010. The biomedical field's use of machine learning for the diagnosis and prognosis of liver disease has had a significant effect.

Cancer had one of the greatest mortality rates ever noted worldwide. The only reason of 87 lakh deaths in 2015 was cancer. Breast cancer has one of the highest documented mortality rates in terms of deaths, just behind lung cancer. According to a Times of India story from the 19th of August 2020, approximately 2.4 lakh breast cancer cases are to be expected to be the most common site in India by 2025. Diabetes is a condition that develops when your blood glucose level, also known as blood sugar, is extremely elevated. Your main source of energy comes from your blood sugar, which is obtained from the food you eat. Insulin is a hormone secreted by the pancreas that helps fuel metabolic processes by removing glucose from meals. According to the "International Diabetes Federation," diabetes is responsible for 42 lakh fatalities worldwide and costs the US economy 760 billion dollars annually to treat. In India, over 10 lakh people die annually due to diabetes (Epidemiology of Diabetes), and according to the "Indian heart Association" nearly 11 crore individuals will end up suffering from diabetes by 2035. The previous versions of the AI models for medical examinations are focused on a single illness for each inquiry. Like one check if you're looking into diabetes, one if you're looking into cancer, one if you're looking into skin infections, etc. A single inquiry cannot typically carry out more than one infection expectation. With our suggested system, you can make predictions about heart disease[4,16,17,19], cancer, and diabetes using a single user interface that unifies numerous diseases. In this study, we conduct the prediction of multiple diseases using machine learning classification algorithms like Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest-Neighbours (KNN)[2,3] and Random Forest Classifier, and apart from predicting these diseases we will be adding two more diseases in the API namely, kidney diseases and liver diseases.

## 2. Relevant Work

The analysis of earlier models that were put forth to forecast the illnesses connected to our proposed work is covered in this part. The detection of different diseases has been the subject of numerous studies. They have used different data mining methods to effectively forecast a number of diseases.

[1] Indukuri Mohit, K. Santhosh Kumar, Uday Avula Kumar Reddy and Badhagouni Suresh Kumar (2021) used 769 instances of the Pima dataset with features like pregnancies, blood pressure, body mass index, etc. are used to test a variety of classification methods, including SVM, Logistic Regression, Decision Tree, KNN, and Random Forest. The classification algorithm Random Forest has been reported to have the highest accuracy at 74.4%, and the KNN has been reported to have the lowest accuracy at 71.3% in this study. Data mining techniques were discussed in the paper.

- [2] Olga Tsave, Ioanna Chouvrada, Nicos Maglaveras, Ioannis Vlahavas, and others (2017), covered diabetes sub-classification on the dataset which was gathered by a survey using Google Forms, and they used Naïve Bayes and SVM classification algorithms on it. The capacity of similar strategies to extract patterns and develop models from data is what gives these approaches their power and success. They recorded accuracy of 64.92 for SVM and 60.44 for Naive Bayes.
- [3] Fikirte Girma Woldemichael and Sumitra Menaria (2018) provided a comprehensive description of the Diabetes and used the Framingham dataset from Kaggle to apply classification algorithms like SVM, Decision Tree, Naive Bayes, and K-Nearest Neighbors. The prediction of the chance of heart disease was compared by the authors among different machine learning algorithms. The KNN classification algorithm's reported accuracy in this study is 83.60%.
- [4] Purushottam, Richa Sharma and Dr. Kanak Saxena (2016) used the application of Knowledge Extraction based on Evolutionary Learning address cardiovascular disease in their work on the customised dataset collected from different sources. The top-down decision tree was constructed. In each phase, a test for the actual node (beginning with the root node) is chosen that best divides the given cases by classes. The work's highest stated accuracy is 86.3%.
- [5] M. Chinna Rao, K. Ramesh, and G. Subbalakshmi (2011) addressed the extraction of hidden information from a heart disease dataset that can respond to complex queries in their presentation of a decision support system for heart disease prediction using the Naïve Bayes classification method in their work by means of various parameters including age, sex, blood pressure and blood sugar and achieved 64.3% accuracy.
- [6] Noreen Fatima, Li Liu, Sha Hong, Haroon Ahmed (2020) outlined the study's previous models and suggested work on the cancer forecast using data mining and machine learning techniques like SVM, KNN, Logistic Regression, Naïve Bayes and ANN that can accurately predict cancer in large health records. Following a comparison of various methods on the Wisconsin Diagnostic Breast Cancer, we determined that the machine learning algorithm SVM is the best fit for predicting breast cancer.
- [7] Ch. Shravya, K. Pravallika, and Shaik Subhani (2019) did relative study on the implementations of models using Logistic Regression, SVM and KNN on the UCI dataset. They also used (PCA) Principal Component analysis to study the results and used dimensionality reduction to clearly describe their findings. After overall analysis, they considered SVM model as well suited for prediction with an accuracy of 92.7%.
- [8] Nikita Rane, Rucha Kanade, Sulochana Devi and Jean Sunny (2020) proposed a comparison of algorithms like Naïve Bayes, Random Forest, ANN, KNN, SVM and Decision Trees on Wisconsin Diagnostic Breast Cancer.
- [9] Dilip Singh Sisodia and Deepti Sisodia (2018) uses three classification algorithms, Decision Tree, Naïve Bayes and SVM on PIDD dataset. The performances were compared on parameters like Precision, Accuracy, Measure and Recall. Naïve Bayes shows best performance with 76% accuracy.

- [10] G.M. Irfan, M.H. Rashid, T. Tazin, S. Bourouis and M.M. Khan (2021) focused on the dataset from CKD Prediction Dataset. Three distinct models for CKD prediction were trained using the LR, Decision Tree (DT), and k-NN algorithms. And in comparison, to the DT (96.25%) and KNN (71.25%), the LR obtained highest accuracy (97%).
- [11] M.A. Islam, S. Akter, M.S. Hossen, S.A. Tisha and S.A. Hossein (2020), investigated the performance of the Nave Bayes (NB), RF, LR and Decision Stump models on a customised dataset from CKD Dataset. Z-score is used for the normalization purpose for predicting the risk of CKD, with obtained accuracy of 93.9 % for Naïve Bayes, 98.88% for Random Forest, 94.76% for Logistic Regression.
- [12] S.Y. Yashfi, M.A. Islam, N. Sakib, T. Islam, M. Shahbaaz, S.S. Pantho and Pritilata (2020) proposed a method for predicting CKD risk based on 455 patients' data from the UCI Machine Learning Repository and a real-time dataset from Khulna City Medical College. RF and ANN were trained and tested using 10-fold cross-validation on the data. The RF and ANN attain accuracy of 97.12% and 94.5%, respectively.
- [13] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive (2020) used logistic regression and decision tree algorithm on Wisconsin Diagnostic Dataset which contains 569 instances and the highest accuracy obtained is 95.1% using Decision Tree Classifier.
- [14] N. H. Priya, N. Gopikarani, and S. S. Gowri (2021) applied various algorithms namely, K-Nearest Neighbours, Decision Tree, Support Vector Machine, Random Forest, and multilayer perceptron on Framingham dataset from Kaggle. Each algorithm works on the model to fit and train it using a training dataset, and the model is tested using a new dataset for each algorithm, and the highest accuracy obtained is 89% by Random Forest on UCI Dataset.
- [15] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, Z. Leonowicz, R. Gono and others (2021) proposed a model in which several algorithms like ANN, C5.0, Chi-square automation interaction detector, LR, Linear SVM with penalty  $L_1$  &  $L_2$ . Features selected for classification are Correlation based feature selection, wrapper method feature selection and least absolute shrinkage. LSVM with penalty  $L_2$  gave highest accuracy with 98.46% when tested on CKD Disease Dataset from UCI repository.
- [16] Ketan Gupta, Neda Afreen, Divyarani D and NasminJewani (2022) studied and analysed the prediction of liver sickness in patients. To increase efficiency, the data was cleaned using several techniques, including imputation of missing values with the median, dummy encoding, and outlier elimination. Several classification algorithms, including logistic regression, decision trees, random forests, and KNN, were used in this study. The highest accuracy obtained is 63% using Random Forest Classifier model.
- [17] Senthilkumar Mohan, ChandrasegarThirumalai and Gautam Srivastava (2019) for heart disease with the hybrid random forest with a linear model (HRFLM) produces an improved performance level in the prediction of cardiovascular diseases with an accuracy of 88.7%.

[18] Adil Hussain Seh and Dr. Pawan Kumar Chaurasia (2019) uses different deep learning and classification techniques like ANN, Weighted Fuzzy Rules, ANN Fuzzy and Naïve Bayes to predict the diseases on The Cleveland Heart Disease Dataset containing 303 records and Statlog Heart Disease Dataset consist of 270 records. The highest accuracy obtained is 91.10% using ANN Fuzzy algorithm.

[19] G. Subbalakshmi, K. Ramesh, M.C. Rao (2011) developed a Decision Support in Heart Disease Prediction System (DSHDPS) applying the Naïve Bayes data mining modelling technique. It predicts the likelihood of patients developing heart disease based on medical characteristics such as age, gender, blood pressure, and blood sugar with accuracy of 73%.

### 3. Methodologies

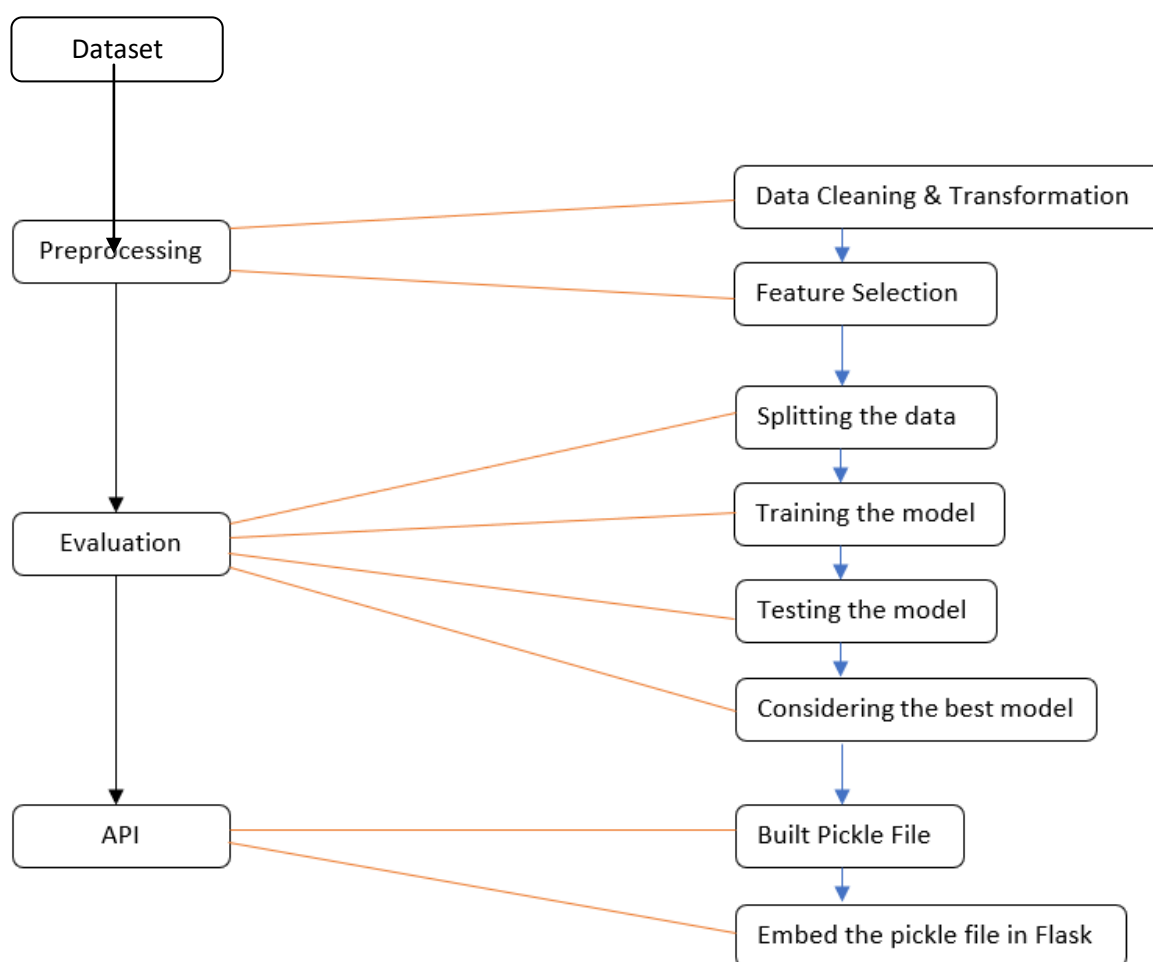


Fig 1: Functional Diagram

The methodology used in the proposed study is described in this fig 1. As previously mentioned, the goal of our work is to create a web tool that uses machine learning models to identify diseases like CKD, liver diseases, cancer, diabetes, and heart diseases. The following are the machine learning techniques applied in our suggested research:

### 3.1 Logistic Regression

Because the logistic function is used in this method, it is known as logistic regression. The logistic function, which has a "S" shape and is plotted between 0 and 1, was created for statistical functionalities. Logistic regression employs equations similar to linear regression for the purpose of representation.

Logistic Regression Equation

$$Y = 1/(1+EXPO(-value)) \quad \text{---(1)}$$

The value of the output, usually referred to as  $y$ , is predicted by linearly combining the input values (typically referred to as  $x$ ) and co-efficient (Beta).

Logistic Regression Equation

$$y = EXPO(u_0 + u_1 * x) / (1 + EXPO(u_0 + u_1 * x)) \quad \text{--(2)}$$

$y$  is predicted outcome,  $u_0$  is intercept or bias, and  $u_1$  is single input coefficient value.

The possibility of first-class (or can be termed as default class) is predicted by logistic regression models. For instance, if we are developing a model to predict a person's gender based on height, the default class could be male, which can be written as –

$$P(\text{gender}=\text{male}|\text{height}) \quad \text{--(3)}$$

Probabilities must be converted into binary numbers, either 0 or 1. The logistic function converts probabilities into predictions. The model could be built as:

$$y = EXPO(u_0 + u_1 * x) / (1 + EXPO(u_0 + u_1 * x)) \quad \text{--(4)}$$

On solving further, we get the equation as:

$$\ln(p(x)/1-p(x)) = u_0 + u_1 * x \quad \text{--(5)}$$

Odds of first-class or default class refers to the left-hand size calculation (ratio). The chance of an event is divided by the probability of its complement event to determine the odds.

$$\ln(\text{odds}) = u_0 + u_1 * x \quad \text{--(6)}$$

Logistic regression makes it straightforward to make predictions and put them into practise. Assume, for instance, that we are trying to determine a person's gender based on their height. Let's say that individual is 150 inches tall, and the assumed coefficients are  $u_0 = -100$  and  $u_1 = 0.6$ .

$$y = EXPO(u_0 + u_1 * X) / (1 + EXPO(u_0 + u_1 * X)) \quad \text{--(7)}$$

$$y = EXPO(-100 + 0.6 * 150) / (1 + EXPO(-100 + 0.6 * 150))$$

$$y = 0.0000453$$

The chance of a male is almost zero.

### 3.2 K-Nearest Neighbors (KNN)

KNN is a machine learning method used for both categorization and regression. Because it requires numerous iterations to achieve the highest level of accuracy, the algorithm is regarded as computationally costly. This method uses supervised machine learning, in which the data is labelled and the algorithm learns to forecast the outcome based on the input data. Even with big and noisy training data, the algorithm still performs well. The sample is split into training and test datasets by the algorithm. The model building and training process uses the training information. The model

developed predicts the test results. Now, we calculate the distance between the test location and the prepared k nearest element values. There are various ways to determine the distance between the elements using different formulae like Minkowski Distance, Euclidean Distance and Manhattan Distance.

- Minkowski Distance:

$$S(i, j) = \sqrt[q]{|u_{i1} - u_{j1}|^q + |u_{i2} - u_{j2}|^q + \dots + |u_{ik} - u_{jk}|^q} \quad \text{--- (8)}$$

- Euclidean Distance:

$$q = 2$$

$$S(i, j) = \sqrt{|u_{i1} - u_{j1}|^2 + |u_{i2} - u_{j2}|^2 + \dots + |u_{ik} - u_{jk}|^2} \quad \text{---(9)}$$

- Manhattan Distance:

$$S(i, j) = |u_{i1} - u_{j1}| + |u_{i2} - u_{j2}| + \dots + |u_{ik} - u_{jk}| \quad \text{---(10)}$$

The most popular method for calculating the values of the distance test sample and trained data is the Euclidean method.

The parameter K, which represents the number of closest peers, is indicated. Finding the finest k worth to achieve the model's highest level of precision is a challenging task. There is no pre-defined measurable method to identify the k worth and achieve astounding accuracy. The only way to determine k values with impressive accuracy is by using the brute force technique, which necessitates determining accuracy for various k values. For the prediction, the K values of the neighbours from 1 to 20 are taken into account, along with the neighbour who provides the greatest accuracy.

### 3.3 Support Vector Machine (SVM)

One of the supervised learning techniques is SVM. Both classification and regression research use this algorithm. Data is plotted in n-dimensional space using this method (uses coordinates). SVMs come in both linear and nonlinear varieties. We use the linear SVM classifier algorithm in our model because we are working with linearly separable data. Finding the best hyperplane, which serves as the boundary that separates classes into groups, is how classes are classified. In two dimensions, the line is a hyperplane. In two-dimensional space, the line is adequate to divide the classes.

For example, consider the equation  $S_0 + (S_1 * U_1) + (S_2 * U_2) = 0$

$B_0$  and  $B_1$  are the factors, and  $B_2$  is the line's intercept. The input points are  $K_1$  and  $K_2$ . Classification is done using this line. The equation returns a number that is greater than zero above the line, and the resulting data value falls under category "0". The data point falls into category "1" if it is below the line and the number the equation returns is less than zero. It is challenging to classify a point that yields a value that is near to zero. The gap between the line and the nearest data point is referred to as the margin. On the off chance that it has the greatest advantage, the ideal line can separate the classes. This line is known as maximal margin hyperplane. The vertical distance between the line and the closest highlight is used to measure this margin. The data values at these locations are known as support vectors, and they are



important for both describing the line and building the classifier. Support vectors are used to identify and support hyperplanes.

### 3.4 Decision Tree

Algorithms for supervised learning include the computation of Decision Trees. A decision tree method can be used to handle regression and classification problems, unlike other supervised learning algorithms. Making a training model that can be used to forecast a class or estimate of objective factors by incorporating choice standards drawn from earlier data (training data) is the general idea behind using decision trees.

### 3.5 Random Forest

A group of unpruned classification-based trees makes up Random Forest. Given that it is insensitive to dataset noise and has a very low risk of overfitting, it exhibits amazing performance in terms of a variety of real-world issues. It operates more quickly than many other tree-based algorithms and usually increases accuracy for testing and validation data. Individual decision tree algorithm forecasts are combined to form random forests. When building a random tree, there are several options for adjusting the efficiency of the random forest.

### 3.6 Dataset Used

#### *Heart Dataset*

We used the "Heart Disease Dataset" by UCI to forecast the occurrence of cardiac diseases. 13 medical predictor characteristics and 1 target feature make up this dataset. These characteristics are chol, cp, trestbps, age, fbs, sex, restecg, exang, slope, thal, ca, oldpeak, and thalach. 75 attributes and 303 occurrences make up the dataset.

#### *Diabetes Dataset*

We used the "Diabetes.csv" by National Institute of Diabetes and Digestive and Kidney Diseases available in Kaggle to predict the occurrence of diabetes. This dataset contains 8 attributes such as Blood Pressure, Glucose, Skin Thickness, BMI, Insulin and others which all are numeric valued and have 768 instances.

#### *Cancer Dataset*

We used "Breast Cancer Wisconsin (Diagnostic) Dataset" by Kaggle for prediction the occurrence of cancer. One target feature and 31 medical predictor characteristics are included in this dataset. Several crucial qualities include diagnostic id, radius-mean, texture-mean, perimeter-mean, area-mean, smoothness-mean, compactness-mean, concavity-mean, and concavity points mean.

#### *Liver Dataset*

We used dataset from "UCI Machine Learning Repository". This dataset consists of 583 liver patient's data whereas 75.64% male patients and 24.36% are female patients. This dataset contains 11 parameters like age, gender, TB, DB, alkphos, sgpt, sgot, TP, alb and ag ratio.

#### *Kidney Dataset*

We used Chronic Disease Dataset downloaded from "UCI Machine Learning Repository". This dataset contains 25 attributes and 400 instances

#### 4. Results

The activities in this work are carried out in Google Colab and code is written in Python language. Logistic Regression, KNN, Random Forest Classifier, Decision Tree and SVM are the algorithms used in this study, and the accuracies are calculated using cross-validation with factor cv set to 10.

Our work has attained the highest levels of accuracy for Heart Diseases, Diabetes, Cancer, Kidney Diseases and Liver Diseases as 88.5%, 79%, 97%, 100% and 77.97% respectively using Logistic Regression technique for Heart Diseases and Random Forest for Diabetes, Cancer, Kidney Diseases and Liver Diseases. All the accuracies achieved by our model using different algorithms are being listed in the table 1,2,3 and 4 given below:

Table 1: Accuracies of Random Forest on different diseases

<b>Diseases</b>	<b>Random Forest</b>
<b>Heart Diseases</b>	86.9 %
<b>Diabetes</b>	79 %
<b>Cancer</b>	97 %
<b>Kidney Diseases</b>	100 %
<b>Liver Diseases</b>	77.97 %

Table 2: Accuracies of Logistic Regression on different diseases

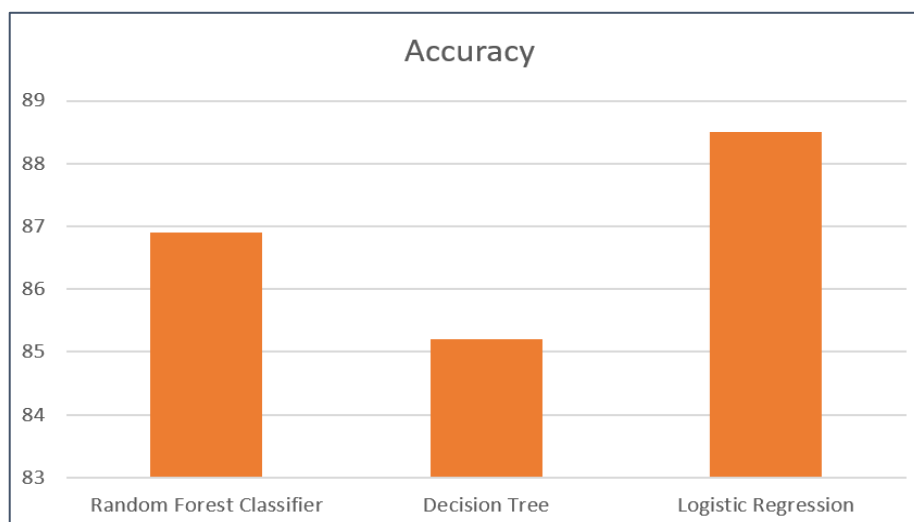
<b>Diseases</b>	<b>Logistic Regression</b>
<b>Heart Diseases</b>	88.5 %
<b>Diabetes</b>	77 %
<b>Cancer</b>	96 %
<b>Kidney Diseases</b>	96.8 %
<b>Liver Diseases</b>	69.4 %

Table 3: Accuracies of K Nearest Neighbour on different diseases

<b>Diseases</b>	<b>K Nearest Neighbour</b>
<b>Diabetes</b>	77 %
<b>Cancer</b>	95 %
<b>Kidney Diseases</b>	96.8 %
<b>Liver Diseases</b>	62.9 %

Table 4: Accuracies of Decision Tree on different diseases

<b>Diseases</b>	<b>Decision Tree</b>
<b>Heart Diseases</b>	85.2 %
<b>Cancer</b>	73 %
<b>Diabetes</b>	73 %



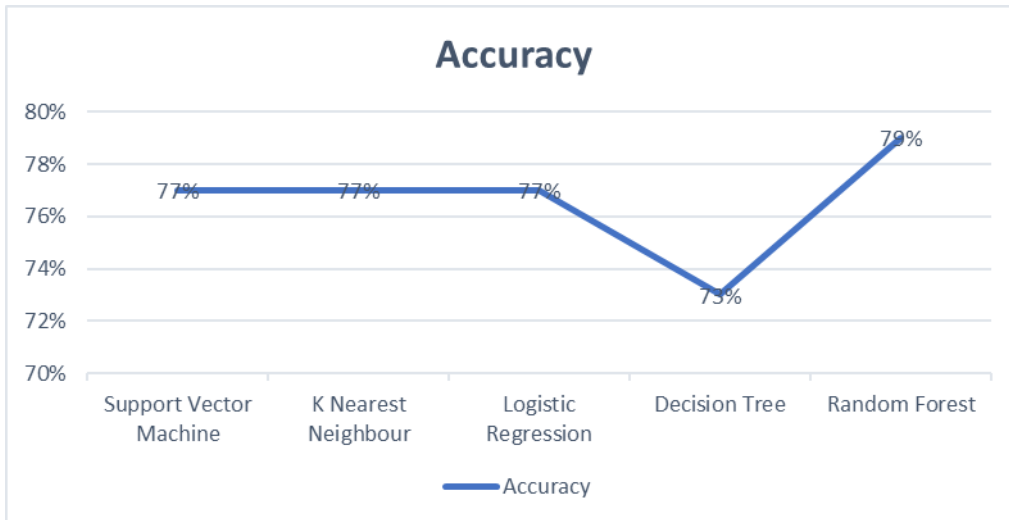


Fig 3: Accuracy values for Diabetes

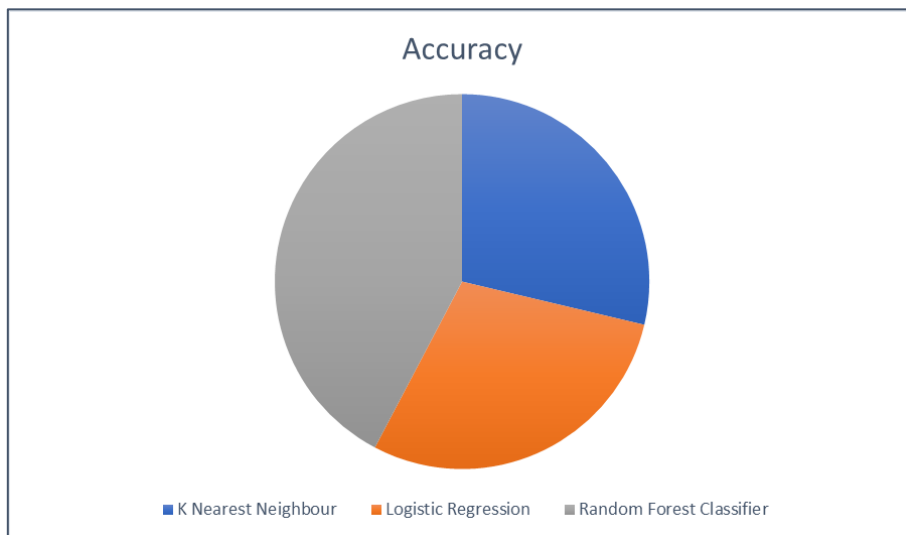


Fig 4: Accuracy values for Cancer

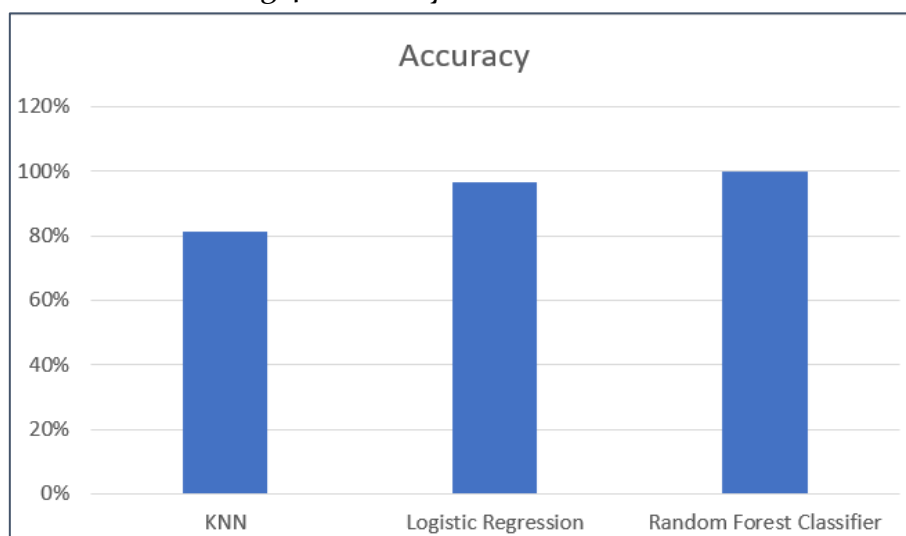


Fig 5: Accuracy values for CKD

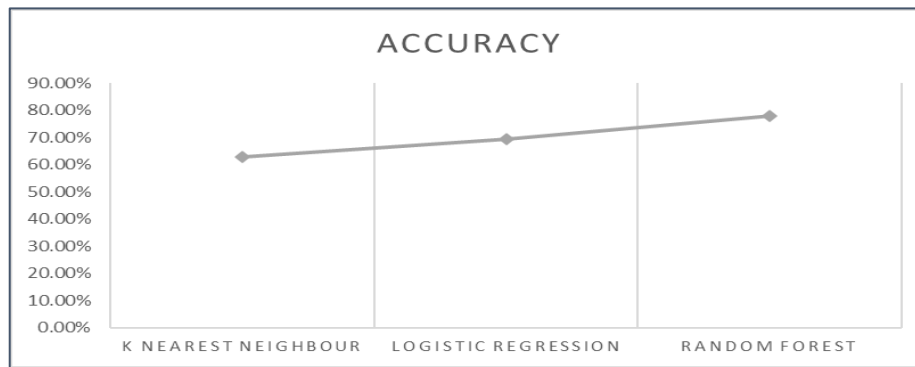


Fig 6: Accuracy values for Liver Diseases

## 5. Conclusion

The proposed effort integrates diabetes, heart diseases, kidney diseases, liver diseases and cancer into a single platform by deploying trained models utilising the lightweight Django API architecture. For training the models, five classification algorithms are used, with logistic regression providing good accuracy values for disease prediction of Heart and Random Forest providing best accuracy values for illness prediction of Diabetes, Cancer, Liver and Kidney Diseases.

In future, we can add more diseases in the existing API and we can make the interface user friendly, so that, user can use it more efficiently.

## References

1. Mohit, I., Santhosh Kumar, K., Kumar Reddy, U. A., & Suresh Kumar, B. (2021). An Approach to detect multiple diseases using machine learning algorithm. *Journal of Physics. Conference Series*, 2089(1), 012009.
2. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104–116
3. Woldemichael, F. G., & Menaria, S. (2018). Prediction of diabetes using data mining techniques. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 414–418.
4. Purushottam, K. Saxena, and R. Sharma, "Efficient heart disease prediction system," *Procedia Comput. Sci.*, vol. 85, pp. 962–969, 2016.
5. M. G. Subbalakshmi, M. Tech, M. K. Ramesh, A. Professor, M. M. Chinna, and R. M. Tech, "Decision support in heart disease prediction system using naive Bayes," *Psu.edu*.
6. N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020.
7. Shravya, C., Pravalika, K., & Subhani, S. (n.d.). Prediction of breast cancer using supervised machine learning techniques. *Ijitee.org*. Retrieved April 18, 2023.

8. Sunny, J., Rane, N., Kanade, R., Devi, S., & Xavier Institute of Engineering. (2020). Breast cancer classification and prediction using machine learning. *International Journal of Engineering Research & Technology (Ahmedabad)*, 9(02).
9. D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms" *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, 2018.
10. Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S., & Khan, M. M. (2021). Comparative analysis for prediction of kidney disease using intelligent machine learning methods. *Computational and Mathematical Methods in Medicine*, 2021, 6141470.
11. Islam, M. A., Akter, S., Hossen, M. S., Keya, S. A., Tisha, S. A., & Hossain, S. (2020). Risk factor prediction of chronic kidney disease based on machine learning algorithms. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 952–957.
12. S. Y. Yashfi et al., "Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5.
13. P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for breast cancer prediction," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 796–801.
14. N. H. Priya, N. Gopikarani, and S. S. Gowri, "Analysis of heart disease prediction using machine learning techniques," in *Handbook of Artificial Intelligence in Biomedical Engineering*, 1st Edition., Series statement: Biomedical engineering: techniques and applications: Apple Academic Press, 2021, pp. 173–194.
15. P. Chittora et al., "Prediction of chronic kidney disease - A machine learning perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021.
16. K. Gupta, N. Jiwani, N. Afreen, and Divyarani, "Liver Disease Prediction using Machine learning Classification Techniques," in 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 221–226.
17. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
18. Hussain .Adil and Chaurasia P, (2019) A Review on Heart Disease Prediction Using Machine Learning Techniques, *International Journal of Management, IT and Engineering*, Volume : 9, Issue : 4, pp. 208–224.
19. Subbalakshmi, M. G., Tech, M., Ramesh, M. K., Professor, A., Chinna, M. M., & Tech, R. M. (n.d.). Decision support in heart disease prediction system using naive Bayes. *Psu.edu*.