International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

HUMAN MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING

P. Anirudh[1], M. Rajesh[2], M. Shalem[3], P. Srinivas[4], Dr. R.R.S.Ravi Kumar[5], Dr K.S.R.K Sarma[5]

[1,2,3,4]Dept. of Computer Science & Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad –500075, India

[5]Asst.Professor, Dept. of Computer Science & Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad –500075, India

[5]Professor, Dept. of Computer Science & Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad –500075, India

Abstract:
Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. Like one analysis for diabetes analysis, one for Heart analysis, one for lung diseases, one for kidney analysis like that. There is no common system where one analysis can perform more than one disease prediction. In this proposing system which is used to predict multiple diseases by using Flask. In this system used to analyze Diabetes analysis, heart disease analysis, Liver disease and Kidney analysis, the accuracy of medical disease prediction has been continuously improved, and the performance in all aspects has also been significantly improved. It aims to clarify the effectiveness of machine learning in disease prediction and demonstrates the high correlation between machine learning and the medical field in future development. The performance of the proposed system is evaluated on the scales of accuracy. The results reveal the effectiveness of our proposed methodology in predicting multiple diseases in comparison to other benchmark methods. The disease prediction is accomplished based on the features extracted from the raw dataset. To implement multiple disease analysis used machine learning algorithms and Flask Python pickling is used to save the model behavior. The importance of this analysis is to analyze the maximum diseases, so that to monitor the patient's condition and warns the patients in advance to decrease mortality ratio. Our research proposed to study machine learning algorithms like, Decision Tree (DT), Support Vector Machine (SVM), KNN, Random Forest (RF), Naive Bayes, logistic regression.

Keywords: Machine learning, Flask, Decision Tree, Support Vector Machine (SVM), Random Forest LR, KNN.

1. Introduction

In terms of data collecting and processing, healthcare is one of the most worrisome industries. With the advent of the digital errand technological advancements, a vast quantity of multidimensional data on patients is created, including clinical factors, hospital resources, illness diagnostic information, patients' records, and medical equipment. The enormous, dense, and complex data must be processed and evaluated in order to extract knowledge for effective decision making. Medical data mining offers a lot of potential for uncovering hidden patterns in medical data sets. By identifying significant patterns and detecting correlations and relationships among many variables in huge databases, the use of various data mining tools and machine learning approaches has changed healthcare organizations. It serves as an important instrument in the medical sector, providing and comparing existing data for the future course of action. This technology com-bines multiple analytic methodologies with modern and complex algorithms, allowing for the exploration of massive amounts of data. It is used in healthcare to gather, organize, and analyze patient data in a systematic manner. It may be used to identify inherent inefficiencies and best practices for providing better ser-vices, which may lead to improved diagnosis, better medicine, and more

DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

successful treatment, as well as a platform for a deeper knowledge of the mechanisms in practically all elements of the medical domain. Overall, it assists in the early detection and prevention of disease epidemics by searching medical databases for pertinent information. The process of determining a condition based on a person's symptoms and indicators is known as medical diagnosis. In the diagnostic process, one or more diagnostic procedures, such as diagnostic tests, are performed. Diagnosis of chronic illnesses is avital issue in the medical industry since it is based on many symptoms. It is a complex procedure that frequently leads to incorrect assumptions. When diagnosing illnesses, the clinical judgment is based mostly on the patient's symptoms as well as the physicians 'knowledge and experience. Furthermore, when medical systems tems evolve and new treatments become available, it becomes more difficult for physicians and doctors to stay up with the cur-rent innovations in clinical practice. For effective therapy, medical practitioners and doctors must be well-versed in all pertinent diagnostic criteria, patient history, and a mix of medication therapy. However, mistakes are possible since they make judgments instinctively based on information and experience gained from past experience with patients. Because of factors such as multi-tasking, restricted analysis, and memory capacity, their cognitive capacities are restricted. As a result, it is difficult for a physician to make the right judgment on a consistent basis if he is not sup-ported by clinical tests and patient history information. Even experienced physicians can benefit from a computer-aided diagnostic system in making sound medical judgments. Thus, medical professionals are very interested in automating the diagnosis process by integrating machine learning techniques with physician expertise. Data mining and machine learning approaches are making significant efforts to intelligently translate accessible data into valuable information in order to improve the diagnostic process's efficiency. Several studies have been conducted to explore the use of machine learning in terms of diagnostic abilities. It was dis-covered that, when compared to the most experienced physician, who can diagnose with 79.97% accuracy, machine learning algorithms could identify with 91.1% correctness . Machine learning techniques are explicitly used to illness datasets to extract features for optimal illness diagnosis, prediction, prevention, and therapy.

2. PROBLEM STATEMENT

In the current healthcare systems, most machine learning models are designed to predict only a single disease at a time, such as diabetes, heart disease, or kidney disease. This creates a limitation because patients often suffer from multiple diseases simultaneously, and separate systems are required for each prediction.

Additionally, traditional diagnosis mainly depends on doctors' experience, patient symptoms, and clinical tests, which can sometimes lead to delays, errors, or inaccurate predictions, especially when handling large and complex medical data.

There is no unified system that can:

Analyze multiple diseases together

Provide fast and accurate predictions

Assist doctors in decision-making using data-driven insights

Therefore, there is a need to develop a single integrated system that can predict multiple diseases efficiently using machine learning techniques. This system should be capable of analyzing patient data, identifying patterns, and providing early diagnosis to improve treatment and reduce risks.

3.IMPLEMENTATION

The machine learning models used in this project are:

- Logistic Regression
- K-Nearest Neighbor (KNN)
- Support Vector Machine (SVM)
- Decision Tree

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

- Random Forest

3.1 Data Set Description

There are 13 variables in this data set:

- 8 categorical variables,
- 4 continuous variables, and
- 1 variable to accommodate the loan ID

id	diagnosis	radius_me	texture_m	perimeter_	area_mea	smoothne	compactn	concavity_	concave p	symmetry_	fractal_dir	radius_se	texture_se	perimeter_	area_se	smoothne	compactn	concavity_	concave p	symmetry_	fractal_dir	radius_wo	te
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193	25.38	
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532	24.99	
84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.004571	23.57	
84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.009208	14.91	
84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.005115	22.54	
843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.005082	15.47	
844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91	0.004314	0.01382	0.02254	0.01039	0.01369	0.002179	22.88	
84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856	50.96	0.008805	0.03029	0.02488	0.01448	0.01486	0.005412	17.06	
844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.005731	0.03502	0.03553	0.01226	0.02143	0.003749	15.49	

Pregnancie	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesPe	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0

id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
0	48	80	1.02	1	0		normal	notpreser	notpreser	121	36	1.2			15.4	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	7	50	1.02	4	0		normal	notpreser	notpresent		18	0.8			11.3	38	6000		no	no	no	good	no	no	ckd
2	62	80	1.01	2	3	normal	normal	notpreser	notpreser	423	53	1.8			9.6	31	7500		no	yes	no	poor	no	yes	ckd
3	48	70	1.005	4	0	normal	abnormal	present	notpreser	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
4	51	80	1.01	2	0	normal	normal	notpreser	notpreser	106	26	1.4			11.6	35	7300	4.6	no	no	no	good	no	no	ckd
5	60	90	1.015	3	0			notpreser	notpreser	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	yes	no	ckd
6	68	70	1.01	0	0		normal	notpreser	notpreser	100	54	24	104	4	12.4	36			no	no	no	good	no	no	ckd
7	24		1.015	2	4	normal	abnormal	notpreser	notpreser	410	31	1.1			12.4	44	6900	5	no	yes	no	good	yes	no	ckd
8	52	100	1.015	3	0	normal	abnormal	present	notpreser	138	60	1.9			10.8	33	9600	4	yes	yes	no	good	no	yes	ckd

Table 1 Structure of the data sets.

3.2 Packages

3.2.1 NumPy

Numpy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ND array very easy. Arrays are very frequently used in data science, where speed and resources are very important.

3.2.2 Pandas

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008. Pandas allows us to analyze big data and make conclusions based on statistical theories. Pandas can clean messy data sets and make them readable and relevant. Relevant data is very important in data science.

3.2.3 Matplotlib

Matplotlib is a low level graph plotting library in python that serves as a visualization utility. Matplotlib was created by John D. Hunter. Matplotlib is open source and we can use it freely. Matplotlib is mostly written in python.

3.2.4 Missingno

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

Missingno is an excellent and simple to use Python library that provides a series of visualization's to understand the presence and distribution of missing data within a pandas data frame.

3.2.5 Seaborn

Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

3.2.6 SciPy

SciPy is a scientific computation library that uses NumPy underneath. SciPy stands for Scientific Python. It provides more utility functions for optimization, stats and signal processing. Like NumPy, SciPy is open source so we can use it freely. If SciPy uses NumPy underneath, why can we not just use NumPy SciPy has optimized and added functions that are frequently used in NumPy and Data Science.

3.2.7 Scikit-learn

Sklearn is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

3.2.8 Joblib

Joblib is a set of tools to provide lightweight pipelining in Python. In particular: transparent disk-caching of functions and lazy re-evaluation (memoize pattern) easy simple parallel computing Joblib is optimized to be fast and robust on large data in particular and has specific optimizations for numpy arrays.

3.2.10 Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

3.2.11 Streamlit

Streamlit is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.

3.2.12 Web Application

This webapp was developed using Flask Web Framework and was deployed on Heroku server. The models used to predict the diseases were trained on large Datasets. All the links for datasets and the python notebooks used for model creation are mentioned below in this readme. The webapp can predict following Diseases:

Diabetes
Liver Disease
Heart Disease
Kidney Disease
Liver Disease
Pneumonia Disease
Stroke Disease
Malaria Disease

4.METHODOLOGY

4.0. Machine learning:

Machine Learning is a kind of algorithm that permits software applications to become more accurate in predictability without being explicitly programmed. a subset of AI supports the thought that a system can learn from data, identify the pattern and make decisions to urge optimal solutions

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

with minimum human intervention. There are two sorts of ML algorithms, supervised machine learning algorithms and unsupervised machine learning algorithms.

4.1 Logistic Regression:

Logistic Regression model is a Machine Learning classification method (algorithm) that is used to forecast or predict the probability of a categorical dependent factor. In a logistic regression model, the dependent variable is a binary that contains data coded as 1 (yes, etc.) or 0 (no, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

Logistic Regression is one among most popular useful models for categorical data, especially for binary response data in data modelling. Unlike rectilinear regression models, logistic regression models can directly predict probabilities (values that are restricted to the (0,1) interval); furthermore, these probabilities are well-calibrated in comparison to the possibilities predicted by other classifier models, like Naive Bayes.

Logistic regression preserves the marginal probabilities of the training data. The multiplier of the model also gives some hints about the relative importance of every input variable. Let us consider again the bank of Kigali Loan dataset where the payment status falls in two categories (completely paid or other) as the figure below shows that the probability of defaulting falls between 0 and 1. □(□) = □ (□= 1/□) . . . . . □ 4. □Also we can use Linear Regression as equation □(□) = □ + □1 □ . . . . . □ 4. □

By using prediction approach (Completely paid=0, Default=1) While working with Logistic Regression we use the logistic function in order to avoid that the probability of □(□)would go beyond 0 and 1.

□(□) = □ 0+□1 □1 + □ 0+□1 □……Equation 4.

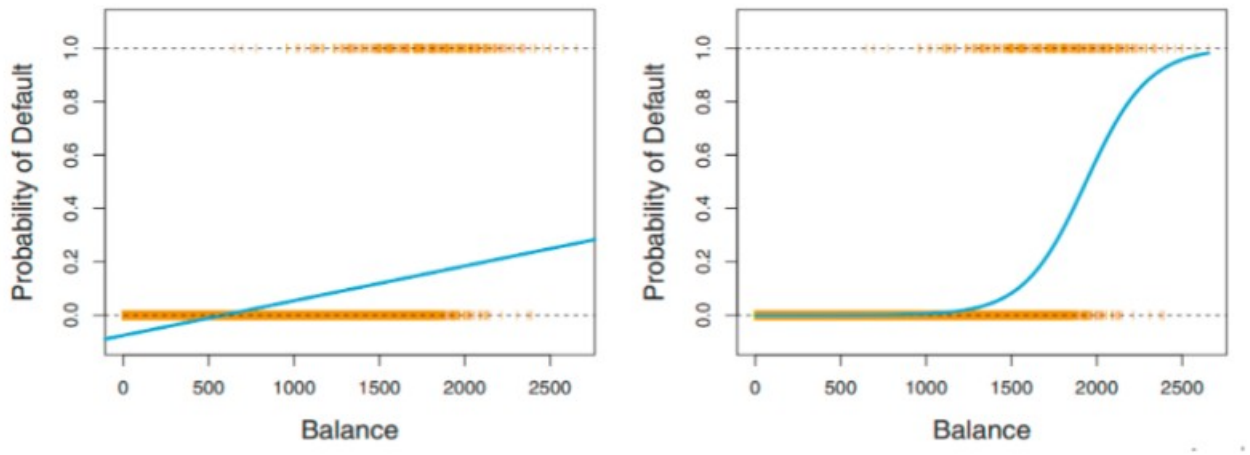


Figure 1 Probablity distribution

4.2 Decision Tree classifier :

Decision trees work as classification models, and they have two steps; the first step is called the learning step while the second step is called prediction steps. In the learning step is where the machine learns from the given dataset and develops based on it. The prediction step is where the machine uses the patterns from the learning step to predict the response of the given dataset. Decision trees are one the most useful types of supervised learning algorithms that are mostly used for classification problems. Furthermore, it works for both continuous and categorical dependent variables. In this algorithm, we split the dataset into two or more homogeneous sets. This is done based on most significant independent/attribute features to make as recognizable groups as possible. Decision Tree is a tree where each node represents a feature, each branch left or right represents decision (rule), and each leaf represents an outcome (categorical or continuous value). There are

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

quite a handful of articles about decision trees. Some articles give you a detailed explanation about decision trees, including information on what's a decision tree, how to generate trees, how to do pruning, and why we should use decision trees. Decision trees model in the form of tree structure and work as classification or regression models. Decision trees break the given dataset in smaller and smaller subsets.

4.3 Support Vector:

Machine It is a classification method. In this model, we plot each data item as a point in n-dimensional space where n is the number of independent factors you have with the value of each factor being the value of a particular coordinate. Support 11 Vector Machine (SVM) is a supervised learning method with the goal of constructing a hyper-plane in a high-dimensional space, which could be used to segregate different populations. A very good support vector machine needs to create multiple hyperplanes for multi-classification or hyperplane which maximize the distance to the nearest training data points of any margin support vectors or class as this would reduce (lower) the generalization error of the classifier. This could be expressed with the following optimization problem: In this research I will use the supervised machines, where we have an input variable and an output variable; Algorithms are used to learn the mapping function, from input to output. Supervised machine learning techniques mainly classified in two subgroups, classification and regression. Classification deals with discrete outputs and regression model runs with continuous outputs. Support vector machines (SVM), Random Forest (RF), Logistic Regression, Decision Tree are the most popular and widely used supervised algorithms. For example, if we only had two features like Hair length and weight of an individual, we'd first plot these two variables in 2-dimensional space where each point has 2 coordinates (these coordinates are known as Support Vectors)

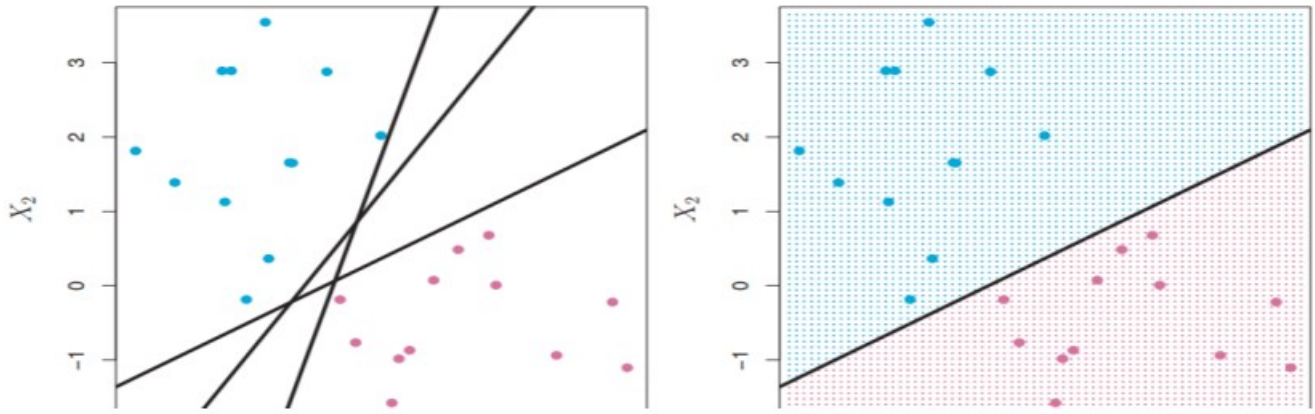


Figure 2 Variables in Support Vector Machine

4.4 Random Forest Random:

Forest model is a trademark term for an ensemble of decision trees. In Random Forest, we have a collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes. Each tree is planted & grown as follows If the number of cases in the training dataset is P, then a sample of P cases is taken at random but with 12 replacements. This sample will be the training dataset for growing the tree If there are N input features, a number n<<N input is specified such that at each node, n features are selected at
random out of the P and the best split on these m is used to split the node. The value of n is held constant during the forest growing

MSE = $\frac{1}{N}\sum$ □□=1 (□ − □) 2

Where N is the number of data points in from given dataset, f □is the value returned by the model

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

and y □is the actual value for data point □.

4.5 K Nearest Neighbors classifier

Is a simple machine learning algorithm that stores all available variables and classifies new variables based on a similarity measure (distance) KNN has been used in statistical estimation and pattern recognition already in the since1970's as a non-parametric technique. KNN classifiers use the distance to classify to class with its neighbors and this depends on the value of K. If K=1 means that the class is simply assigned to its neighbor by using distance function.

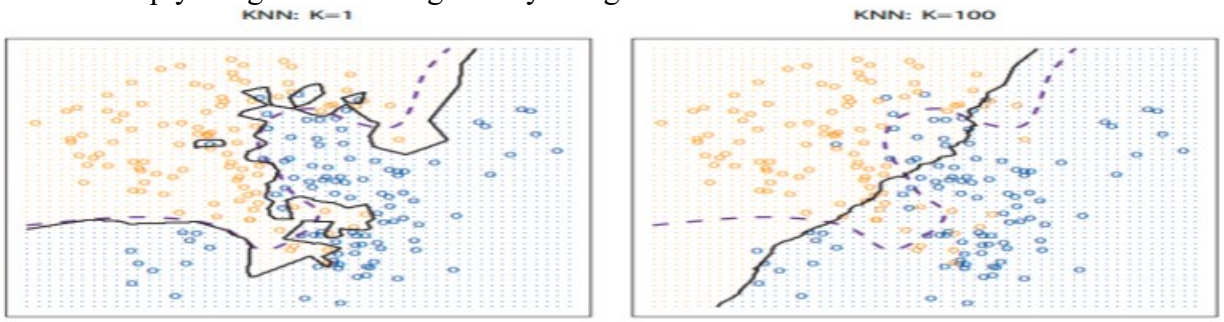


Figure 3 K Nearest Neighbors

Knowing the correct optimal value K is best by first controlling the data. In general, a high value of K is more precise because it reduces the overall noise in your data but there is no guarantee. Cross-validation is another way to consider a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3-10. Distance function

□ = √∑□□=1 (□ − □) 2 …….…. Equation 4.5.a
□ *h*□ = ∑ □□=1 |□ −□ | ……...…. Equation 4.5.b
□ = (∑ □□=1 (|□ − □ |) □) 1 □……. Equation 4.5.c

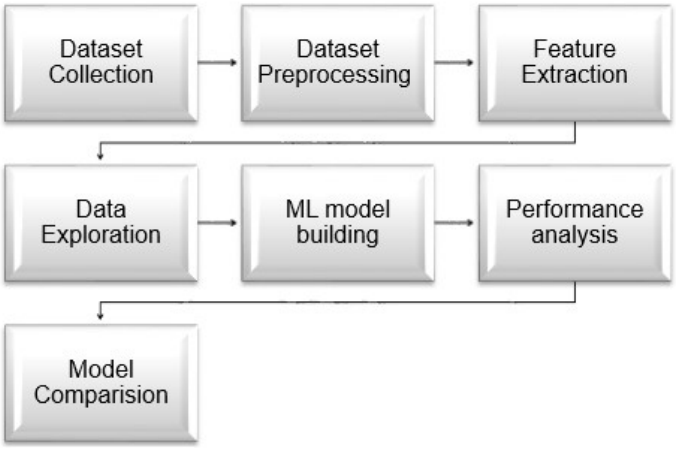


Figure 4 Architecture model

4.6 Data preprocessing

Data preprocessing is an important task to be done prior to analysis to get the data ready for analysis. As good data can only provide better results. In data preprocessing, the proposed system performs data cleaning, data imputation, data normalization, and transformation.

Data cleaning process removes null values and redundant attributes from the dataset. Implementation of the proposed model applies the sampling technique (PCA) on the preprocessed dataset to balance it. On this sampled data, the proposed system implements the Machine Learning Algorithms to check which algorithm suits better, which algorithm is suitable for prediction. This

DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

system also compares the accuracy of algorithms before and after feature selection to select the best algorithm that predicts the defaulters effectively. The architecture of the proposed system is given below. Python as a tool for data analysis will be used to clean data, split data into training set and test data set, training set is 80% while the test set is 20 %. we will use different models for the purpose of getting the best model to use, which will give high accuracy and less error.

4.7 Feature Extraction

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

Data Exploration

Data exploration definition: Data exploration refers to the initial step in data analysis in which data analysts use data visualization and statistical techniques to describe dataset characterizations, such as size, quantity, and accuracy, in order to better understand the nature of the data.

Choose a Models

Different algorithms are for various task, and you choose the proper one.

4.8 Train the Model

The goal of coaching is to answer an issue or make a prediction correctly.

4.9 Evaluate the Model

Uses some metric or combination of metrics to "measure" objective performance of the model. Test the model against previously unseen data. This unseen data is supposed to be somewhat representative of model performance within the world but still helps tune the model.

4.10 Performance analysis

In Machine Learning, the performance evaluation metrics are used to calculate the performance of your trained machine learning models. This helps in finding how better your machine learning model can perform on a dataset.

4.11 Model Comparison

Multiple model comparison is also called Cross Model Validation. Here the model refers to completely different algorithms. The idea is to use multiple models constructed from the same training dataset and validated using the same verification dataset to find out the performance of the different models.

4.12 Data Analysis

Different classification learning methods are heavily dependent on quantity and the quality of the data provided for training of the model. In this chapter, we will have an overview of the loan repayment data set from Bank of Kigali and perform exploratory data analysis in order to preprocess the data and improve the prediction results. The data will also be split into test sets (20%) and the training sets (80%). Training set is a data initial dataset that helps the program to understand how to learn and apply sophisticated technology. Also, training models determine the good values for all weights and bias from the labeled example. For the train dataset, the Machine learning algorithm builds models that examine many examples and attempts to find the model that minimizes the loss. The test data set is independent of the training data dataset, but it has the same probability distribution as the training dataset, and it is used to measure the performance. Training dataset will be used to fit the model, and the test dataset will be to evaluate the best model to get an estimation of generalization error.

4.13 Machine Learning Model

Ensemble learning is a machine learning paradigm in which a few learners are trained to solve the same problem with the goal of obtaining better predictive accuracy that could have been achieved from any of the constituent learning models alone. It is a well-established and widely employed

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

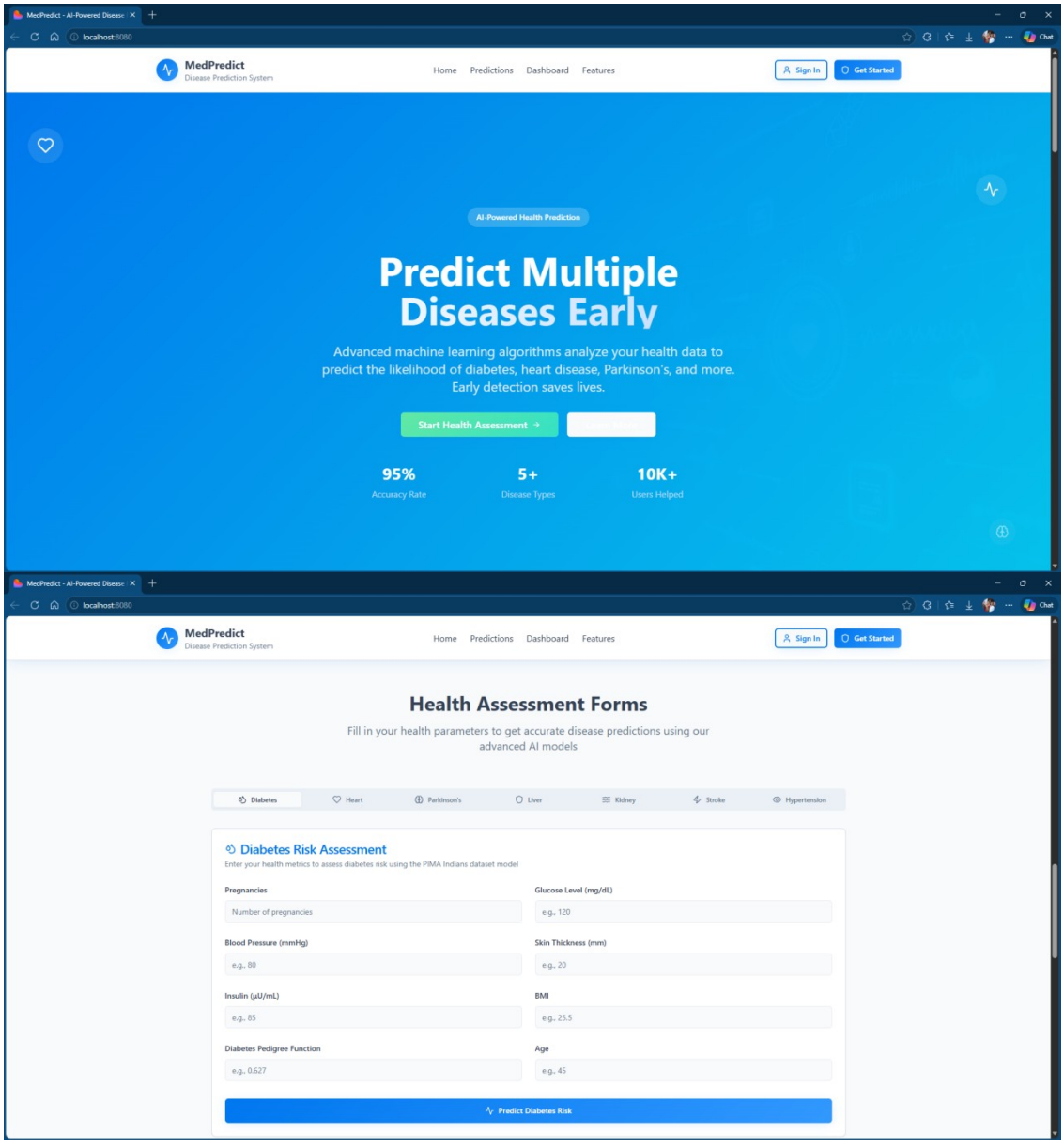
methodology designed to enhance the generalizable signal by averaging out noise from a diverse set of models.

Model Accuracies:

- Cancer Disease : 94%
- Heart Disease : 85%
- Kidney Disease :99%
- Liver Disease : 78%
- Malaria Disease : 94.4%
- Pneumonia Detection Disease : 94%
- diabetes Disease : 92%

From the finding we can see that Random Forest is the best model to be used there, but we can improve accuracy in selected models through algorithm tuning knowing that machine learning algorithms are driven by parameters. These parameters majorly influence the outcome of the learning process.

5.OUTPUT RESULTS



DOI:10.46647/ijetms.2026.v10i02.053 ISSN: 2581-4621

Fig 5.1 Home Page

REFERENCE

1. R. Manne, S.C. Kantheti, Application of artificial intelligence in **healthcare: chances** and challenges, Curr. J. Appl. Sci. Technol. 40 (6) (2021) 78–89, https://doi.org/10.9734/cjast/2021/v40i631320.
2. M. Sivakami, P. Prabhu. Classification of algorithms supported factual knowledge recovery from cardiac data set, Int. J. Curr. Res. Rev. 13(6) 161-166. ISSN: 2231-2196 (Print) ISSN: 0975-5241 (Online).
3. M. Sivakami, P. Prabhu. A Comparative Review of Recent Data **Mining Techniques** for Prediction of Cardiovascular Disease from Electronic **Health Records**. In: Hemanth D., Shakya S., Baig Z. (eds) Intelligent **Data Communication** Technologies and Internet of Things. ICICI 2019. LectureNotes on Data Engineering and Communications Technologies, vol 38.Springer, Cham 477-484. ISSN 2367-4512 ISSN 2367-4520 (electronic), ISBN978-3-030-34079-7 ISBN 978-3-030-34080-3 (eBook) 2020.
4. P. Prabhu, S. Selvabharathi. Deep Belief Neural Network Model for Prediction **of Diabetes** Mellitus. In 2019 3rd International Conference on Imaging, **Signal Processing** and Communication, ICISPC 2019 (pp. 138–142) Institute **of Electrical** and Electronics Engineers Inc. ISBN:9781728136639. 2019.
5. N. Jothi, N.A. Rashid, W. Husain, Data mining in healthcare – A review, ProcediaComput. Sci. 72 (2015) 306–313.
6. H. Polat, H. Danaei Mehr, A. Cetin. Diagnosis of chronic kidney disease based on support vector machine by feature selection methods, J. Med. Syst. 41(4) 201755.
7. K.B. Wagholikar, V. Sundararajan, A.W. Deshpande, Modeling paradigms **for medical** diagnostic decision support: a survey and future directions, J. Med.Syst. 36 (5) (2012) 3029–3049.
8. E. Gürbüz, E. Kılıç, A new adaptive support vector machine for diagnosis **of diseases**, Expert Syst. 31 (5) (2014) 389–397.
9. M. Seera, C.P. Lim, A hybrid intelligent system for medical data **classification, Expert** Syst. Appl. 41 (5) (2014) 2239–2249.
10. Y. Kazemi, S.A. Mirroshandel, A novel method for predicting kidney stone type using ensemble learning, Artif. Intell. Med. 84 (2018) 117–126.
H. Barakat, P. Andrew, Bradley, H. Mohammed Nabil Barakat, Intelligiblesupport vector machines for diagnosis of diabetes mellitus, IEEE Trans. Inf.Technol. Bio Med. J.