



Design and Implementation in Web Application of Effecting Heart Disease and Diabetes for Predicting Covid-19

Kanchana. A¹, Mangammagari Panithavya², Meenakshi. G³, Sandhiya. K⁴

Assistant Professor¹, UG Scholar^{2,3,4}

Department of Computer Science and Engineering
Panimalar Engineering College, Chennai, India

Abstract:

Heart disease and Diabetes is one among the foremost vital causes of mortality within the world nowadays. Prediction of COVID -19 could be a critical challenge within a space of clinical knowledge analysis. Machine learning (ML) has been shown to be effective in aiding in creating selections and predictions from the big amount of knowledge made by the health care trade. We've conjointly seen ML techniques being employed in recent developments in numerous areas of Internet of Things (IoT). Numerous studies offer solely a glimpse into predicting COVID-19 with ML techniques. Performance and accuracy of the applied algorithms is mentioned and compared. Comparison of the various machine learning techniques employed in this study reveals which algorithm is best fitted for prediction of COVID-19.

Keywords: Heart Disease, Diabetes, ML(Machine Learning), Covid-19.

I. INTRODUCTION

Heart is one of the most vital organs for the proper functioning of our body. According to a survey by WHO, 31% of the worldwide deaths every year occurs due to Cardiovascular Diseases (CVDs). Also, more than 75% of these deaths occur in low and middle income countries including India. The main challenge is to accurately predict the existence of CVDs inside human body. The older techniques have not been very successful in efficiently predicting the heart diseases. Many medical instruments are available in the market for the prediction of heart diseases but there are some drawbacks of these instruments like they are very costly, they are not efficient enough for predicting heart diseases. Age, Sex, Blood Pressure, Cholesterol, Blood Sugar, Diabetes, etc. and some lifestyle factors like obesity, eating unhealthy food, less physical activity, smoking, consumption of alcohol, etc. are some of the major risk factors that leads to heart diseases. Most of the lifestyle risk factors are controllable. In the last few decades, medical science has used the technological advancements very well to improve the quality of healthcare. These advancements in technology have paved ways for accurate diagnosis and prediction of Diabetes is a condition where person's body is unable to balance glucose-insulin level after various prandial modes. The patients of diabetes have increased exponentially from past few years due to their unbalanced diet and unhealthy lifestyle. There are around 463 million diabetic people around the world who can be benefited by smart healthcare technologies to improve their quality of life. Diabetes may lead to heart disease, kidney infection, blindness and nerve damages. It has been observed that severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) has infected the diabetic patients largely among the others. People with diabetes were also more susceptible to previous pandemic known as Middle East respiratory syndrome (MERS) and severe Acute Respiratory Syndrome (SARS) type

of corona virus along with H1N1 type of severe influenza during 2009 SARS-CoV-2 has affected mainly to old-age people and persons having underneath health conditions. The various case studies reported diabetes as the major pre-existing comorbidity among COVID-19 patients.

II. LITERATURE REVIEW

Himanshu Sharma, M A Rizvi.[1] has proposed a Prediction heart disease using Machine Learning. In this paper, typically explains the Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart disease is the Leading cause of death worldwide. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using two main Machine Learning Algorithm namely Decision Tree Algorithm and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

Amit M. Joshi, Urvashi P. Shukla, Saraju P. Mohanty,[2] has done a paper on Smart Healthcare for Diabetes during COVID-19. In this paper explains the The diabetic patients are at higher risk from novel coronavirus disease 2019 (COVID-19) that spreads through Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). There are around 20% to 50% of COVID-19 cases had diabetes across the different re-gions from the world. This article discusses recommendations and associated risk for diabetic patients to balance glyncemic profile during COVID-19 outbreak. It also discusses the case study of various countries with impact of COVID-19 for diabetic

patients. It presents emerging smart healthcare that can potentially safeguard against COVID-19.

Sanchayita Dhar, Krishna Roy, Tanuree Dey, Pritha, Datta, Ankur Biswas.[3] has proposed on A Hybrid Machine Learning Approach for Prediction of Heart Diseases. In this paper explains the Heart diseases are the chief cause of death all over the world over the last few decades. To avoid heart disease or coronary illness and discover indications early, individuals over 55 years must have a total cardiovascular checkup. Researchers and specialists developed various intelligent techniques to improve capacity of the health care professionals in recognition of cardiovascular disease. In cardiovascular disease finding and treatment, single data mining strategies are giving the reasonable precision and accuracy. Nevertheless the usage data mining procedure be capable of reducing the number of test that is required to be carried out. In order to decrease the Figure of deaths from heart diseases there has to be a quick and efficient detection technique providing better accuracy and precision. The aim of this paper is to present an efficient technique of predicting heart diseases using machine learning approaches. Hence we proposed a hybrid approach for heart prediction using Random forest classifier and simple k-means algorithm machine learning techniques. The dataset is also evaluated using two other different machine learning algorithms, namely, J48 tree classifier and Naive Bayes classifier and results are compared. Results attained through Random forest classifier and the corresponding confusion matrix shows robustness of the methodology.

Md. Razu Ahmed, S M Hasan Mahmud, Md. Altab Hossin, Hosney Jahan, Shek Rashed Haider Noori.[4] has done a paper on A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms. In this paper, Heart disease prediction and detection has long been considered as a critical issue. Early detection of heart disease is an important issue in health care services (HCS). In growing amount of health care systems, patients are offered expensive therapies and operation that is quiet expensive for developing countries. Recently, heart disease is a prominent public chronic disease, ex. it's a growing concern in the US. The main reason of these diseases are tobacco consumption, bad life style, lack of physical activity and the intake of alcohol. Therefore, there is a need for the cloud based architecture that can efficiently predict and track health information.

GG Ladha.[5] has done a paper on A computation analysis to predict diabetes based on data mining. In this paper, explains the report suggested from different health organization shows the alarming condition due to diabetes worldwide. Different researchers around the globe have investigated it on different parameters and the investigation is going on for the early stage detection. The main objective of this paper is explore and elaborate the methodological prospective to predict the diabetes based on the dataset presented. This exploration is helpful in providing us a direction to find out the research gaps, so that future efficient diabetes framework can be developed. This computation analysis also provides us the parametric exploration along with the knowledge of the attributes and the way of implementing the classification framework.

D.Shetty.[6] has done a paper on Diabetes disease prediction using data mining. In this paper explains the Data mining is a subfield in the subject of software engineering. It is the methodical procedure of finding examples in huge data sets including techniques at the crossing point of manufactured intelligence, machine learning, insights, and database systems. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

III. EXISTING SYSTEM

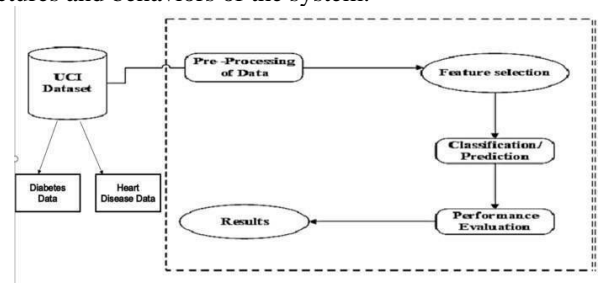
Before, the Doctors only view the report to convey the result to the patients. There are some problems appeared while seeing the laboratory details, they can't predict it properly. There is some difficulty in existing project, they created for some other purpose to test for different disease prediction, but According to covid 19, Diabetes and heart disease result is very important to predict, whether the person will be affected by covid-19 or not.

IV. PROPOSED SYSTEM

The prime objective of this project is to find the factors impacting the health of the society with the greater accuracy level and giving the good result. We have used python and pandas operations to perform heart disease and Diabetes classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation, and the results with improved accuracy. This software is used to predict the covid by checking the heart disease and diabetes report.

V. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



At first, we getting the data of Diabetes and Heart Disease from UCI Dataset, then we are Pre-Processing of Data. After that we

are doing the Feature Selection for classification / Prediction. Then the Performance Evaluation Occurs for the result.

VI. METHODOLOGY

1.Dataset Description: The objective of the dataset is to predict whether or not a patient has diabetes, based on certain diagnostic

measurements included in the dataset. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, chest Pain, Cholesterol, FBP, ECG and soon.

DATASET	DESCRIPTION
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m) ²)
Diabetes Pedigree Function	Diabetes pedigree function
Age	Age (years)
Outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0.
Chest Pain	1-4 (1: Typical Angina, 2: Atypical Angina, 3: Non-anginal, 4: Asymptotic)
Cholesterol	Serum Cholesterol in mg/dl
Fasting Blood Sugar	Blood sugar content before food intake if >120 mg/dl.
ECG	Resting Electrocardiographic results , 0-1 (0: Normal, 1: Having ST-T wave)
Max Heart Rate	Maximum heart beat rate, Beats/min
Exercise Induced Angina	Has pain been induced by exercise, 0-1 (0: No, 1: Yes).
Old Peak	ST depression induced by exercise relative to rest,0-4.
Slope of Peak Exercise	Slope of the peak exercise ST segment, 1-3 (1: Up sloping, 2: Flat, 3: Down sloping)
Ca	Number of vessels colored by fluoroscopy, 0-3
Thala	Displays the thalassemia
Num	Diagnostics of Heart Disease

2. Data Pre-Processing

Diabetes disease and Heart Disease data is pre-processed after collection of various records. The dataset contains a total of 769

patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 763 patient records are used in pre-processing.

Pregnancy	Glucose	BloodPres	SkinThick	Insulin	BMI	DiabetesF	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
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0

	A	B	C	D	E	F	G	H	I	J
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	num
2	67	1	4	160	286	0	2	108	1	2
3	67	1	4	120	229	0	2	129	1	1
4	37	1	3	130	250	0	0	187	0	0
5	41	0	2	130	204	0	2	172	0	0
6	56	1	2	120	236	0	0	178	0	0
7	62	0	4	140	268	0	2	160	0	3
8	57	0	4	120	354	0	0	163	1	0
9	63	1	4	130	254	0	2	147	0	2
10	53	1	4	140	203	1	2	155	1	1
11	57	1	4	140	192	0	0	148	0	0
12	56	0	2	140	294	0	2	153	0	0
13	56	1	3	130	256	1	2	142	1	2
14	44	1	2	120	263	0	0	173	0	0
15	52	1	3	172	199	1	0	162	0	0
16	57	1	3	150	168	0	0	174	0	0
17	48	1	2	110	229	0	0	168	0	1
18	54	1	4	140	239	0	0	160	0	0
19	48	0	3	130	275	0	0	139	0	0
20	49	1	2	130	266	0	0	171	0	0
21	64	1	1	110	211	0	2	144	1	0
22	58	0	1	150	283	1	2	162	0	0
23	58	1	2	120	284	0	2	160	0	1

3. Feature Section

From among the 8 attributes of the data set, one attributes pertaining to age is used to identify the personal information of the patient. The remaining 7 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of diabetes disease.

4. Classification Modeling

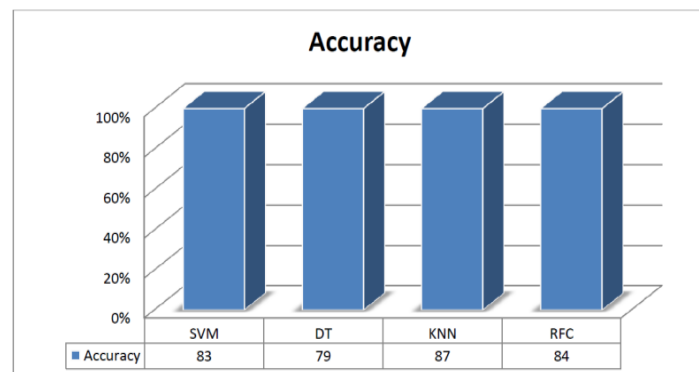
The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

- Support Vector Machine
- Decision Trees
- K Nearest neighbors
- Random Forest Classifier

5. Performance Analysis

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficiency of this model.

- ✓ K Nearest neighbors : 87.571
- ✓ Support Vector Classifier : 83.376
- ✓ Decision tree : 79.181
- ✓ Random Forest : 84.974



6. Algorithms

1. Support Vector Machine

- Support Vector Machine (SVM) is a supervised learning methodology which is used to analyse data used for classification and regression analysis. The training data is fed into an SVM training algorithm which then builds a model that classifies new examples to one category or the other, that makes SVM a non- probabilistic binary linear classifier.
- The accuracy score of SVM was found to be 83.0%.

1. Decision Tree

- Decision Tree algorithm belongs to the family of supervised learning algorithms. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data)
- The accuracy score of Decision Tree was found to be 79.0.

2. K Nearest Neighbours

- During testing, KNN classification algorithm has to find K-nearest neighbours of a new instance. This is time consuming if we do exhaustive comparison. K-nearest neighbours use the local neighborhood to obtain a prediction.
- The accuracy score of KNN was found to be 87.0%.

3. Random Forest Classifier

- Select random samples from a given dataset.
- Construct a decision tree for each sample and get a prediction result from each decision tree.
- Perform a vote for each predicted result.
- Select the prediction result with the most votes as the final prediction
- The accuracy of Random Forest Classifier was found to be 84.0%.

VI. CONCLUSION

At first, the four algorithms were implemented. Datasets were trained for all the algorithms individually. After this, all of them

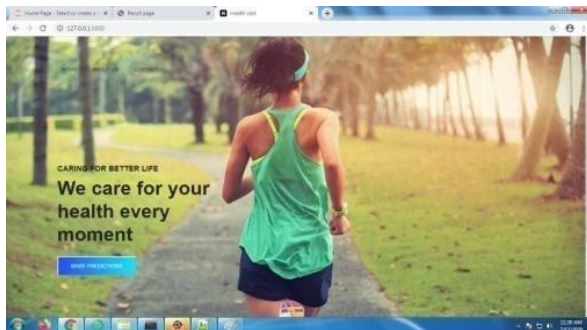
were tested. The most efficient algorithm was to be selected based on various criteria. We found out that KNN algorithm was the most efficient out of the four algorithms with an accuracy of 87.0%. Decision tree, Support Vector Machine and Random Forest Classifier had accuracy of 79.0%, 83.0% and 84.0% respectively. Thus, KNN algorithm was further implemented using a better user interface in form of a web application. For this HTML5, CSS, JS and Flask (Python's micro web-framework) were used. This would help the end users get a preliminary prediction about the condition of their heart and Diabetes. Since heart diseases and Diabetes are a major killer in India and throughout the world, application of a promising technology like machine learning to the initial prediction of Covid-19 will have a profound impact on the society. This will tell the user if they are at a risk and if they need to visit the doctor. This will help reduce the death rate due to covid. Hence by using the above approach successful analysis of heart diseases and Diabetes of the individual was performed and the result was obtained which predicted the risk of Covid-19 based on the parameters provided by the user.

VII. FUTURE ENHANCEMENT

By using the machine learning concept, newly trained dataset can be used for an even more accurate prediction system. Accounts can be created for each user and then by referring the past choice history of user's heart condition and Diabetes can be monitored to tell if there is any improvement or if the condition has deteriorated. Our future enhancement is, we going to develop it as Android app.

VIII. EXPERIMENTAL AND RESULTS

1. Home Screen



2. Input of Heart Disease

3. Input of Diabetes

4. Output of Heart Disease



5. Output of Diabetes

ID	NAME	AGE	SEX	DIABETES	DIABETES RISK
1	John	35	Male	Yes	High
2	Jane	45	Female	No	Low
3	Mike	55	Male	Yes	High
4	Sarah	65	Female	No	Low
5	David	75	Male	Yes	High
6	Emily	85	Female	No	Low
7	Robert	95	Male	Yes	High
8	Lisa	105	Female	No	Low
9	James	115	Male	Yes	High
10	Maria	125	Female	No	Low

IX. REFERENCES

- [1]. Himanshu Sharma, M A Rizvi. : "Prediction of Heart Disease using MachineLearning Algorithms: A Survey ", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 8 , 2017.
- [2].Amit M. Joshi, Urvashi P. Shukla, Saraju P. Mohanty, "Smart Healthcare for Diabetes during COVID-19", IEEE, 2020, <https://ieeexplore.ieee.org/document/9174644>.
- [3]. Sanchayita Dhar, Krishna Roy, Tanuree Dey, Pritha, Datta, Ankur Biswas.: "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", IEEE, 2018, <https://ieeexplore.ieee.org/document/8777531>.
- [4]. Md. Razu Ahmed, S M Hasan Mahmud, Md. Altab Hossin, Hosney Jahan, Shek Rashed Haider Noori.: "A Cloud Based

Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms", IEEE, 2018.

[5]. GG Ladha, "A computation analysis to predict diabetes based on data mining", IEEE, 2018.

[6]. Shetty, "Diabetes disease prediction using data mining", IEEE, 2017.

[7].G. Shanmugasundaram, V. Malar Selvam, R. Saravanan, S.Balaji.: "An Investigation of Heart Disease Prediction Techniques", IEEE, 2018.

[8].Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez.: "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", IEEE, 2017.

[9]. Zhilbert Tafa, "An intelligent system for diabetes prediction IEEE, 2020.

[10]. M Rout, "Prediction of Diabetes Risk based on Machine Learning Techniques", IEEE, 2020.