IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

# Comparative Study on Multi Disease Predictors

**Dr. R.Venkata Ramana chary, Vijaykumar Mantri, Dr. Divanu Sameera, Kuncharam Manideep, Sara Srikanth, Julakanti Yashwanth**

Professor, Department of Information Technology, BVRIT, Narsapur, India

Associate Professor, Department of Information Technology, BVRIT, Narsapur, India

Associate Professor, Department of Information Technology, BVRIT, Narsapur, India

B. Tech, Department of Information Technology, BVRIT, Narsapur, India

B.Tech, Department of Information Technology, BVRIT, Narsapur, India

B.Tech, Department of Information Technology, BVRIT, Narsapur, India

**ABSTRACT:** In the current years, the development of machine intelligence (AI) and the continuous origin of AI's research in the healing field have allowed the public to visualize the wonderful prospects of the integration of AI and healthcare. The vehement deep education field among the ruling class has proved greater potential in uses to a degree affliction guess and drug response guess. From the primary logistic reversion model to the machine learning model, and before to the deep knowledge model contemporary, the accuracy of healing ailment guess has been steadily upgraded, and the accomplishment entirely aspects has likewise existed considerably improved. This item presents a few elementary deep learning foundations and a few prevailing afflictions and summarizes the open ocean education indicator methods equivalent to various ailments.

Point out a succession of problems in the current affliction prognosis and create a prospect for the future development. It aims to explain the influence of deep knowledge in ailment prediction and shows the extreme equivalence between deep education and the healing field from now on the incident. The unique feature ancestry orders of deep education means can still play an important duty from now on in healing research.

Advancements in analytical models, the chance of GPU fittings, and cloud foundation started to play a pivotal part in healthcare practices and research. It has many forms and techniques to collect, survive, resolve, and envision large books of constructions, and unorganized and semi-organized dossiers. Data Science plays an essential act in medical fields accompanying better support for disease and curing the afflictions.

In this project, we are going to forecast diversified afflictions like a tumor, Diabetes, heart, liver, sort, sickness, and pneumonia by utilizing various machine learning models.

**KEY WORDS:** Artificial Intelligence, Data Science, GPU fittings, Machine Learning.

## I. INTRODUCTION

It is a web application that helps to predict the disease based on the parameters. Here it takes input from the user such as text/image data. And that input will be tested with the trained machine learning algorithm and the output will be displayed as per the results.

The benefit of using the application is that patients can take get the testing records without visiting the hospital or consulting the doctor. It is also beneficial to the doctors to decrease their pressure by just saying the test results. The main purpose of the application is to provide their test results based on the values in the test report. It is a very smart, economic, and efficient platform for getting the e-opinion without visiting the hospital

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

## II. RELATED WORK

We went through a lot of different research papers to understand all the previous work done on the project that we have undertaken. We have understood the following inferences.

[1] Shafiq's Muhammad Abdulhamid, Muhammad Shafiq Abd Latiff, Haruna Chiroma, Oluwafemi Osho, Gaddafi Abdulsalam, Adamu I. Abubakar, and Tutu Herawan. By using different machine learning algorithms like SVM and Bayesian classifiers.

[2] Lutfun Nahar Lota B M Mainul Hossain. Having the SVM algorithm gives better accuracy but suffers from implementation complexity.

[3] Abonyi, J., Seifert, F.: Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recogn. Lett. 14(24), 2195–2207 (2003)

[4] Ster, B., Dobnikar, A.: Neural networks in medical diagnosis: comparison with other methods. In: Proceedings of the International Conference on Engineering Applications of Neural Networks, pp. 427–430 (1996)

[5]Salama, G., Abdelhalim, M.B., Zeid, M.: Experimental comparison of classifiers for breast cancer diagnosis, 180–185 (2012). https://doi.org/10.1109/ICCES.2012.6408508

Apart from the above Papers, we have also got to know how to design a machine learning algorithm using best practices in terms of naming conventions, project modularity, and even database design during our literature survey. All these things are quite essential for creating an efficient, readable, and scalable model that any organization can rely on it.

## III. EXISTING SYSTEM

In the existing system, the model the accuracy is not up to the mark. Only one disease is present in the existing system, and where the user needs to check each disease each time. Prediction of the disease results in not being accurate and efficient to use. In the present system, people cannot use the application for multiple diseases is not there, where every timea user needs to visit manually either the other or physical needs to visit a doctor.

## IV. PROPOSED SYSTEM

Advancements in analytical models, availability of GPU hardware, and cloud infrastructure began to play a pivotal role in healthcare practices and research. It has numerous tools and techniques to archive, manage, analyze, and predict large volumes of structures, and unstructured and semi-structured data. Data Science plays a vital role in medical fields with better support for diagnosis and cure for the diseases. In this research, a comparative study on conventional models and modern methods is carried out using 11 different machine learning models for classification. The response or target variable is the diagnostic observation having two classes viz benign (not cancerous) or malignant (cancerous). The dataset used is The Wisconsin Diagnostic Breast Cancer (WDBC) which is taken from the UCI machine learning repository. The performances of the models are analyzed using the validation matrices, precision, recall, accuracy, ROC curve, and F1 Score. In this research study, we have built the ensemble model using the base estimator as a Random Forest classification model, Principal Component Analysis (PCA), Extra Gradient Boosting (XGBoost), and bagging techniques. Among 10 conventional models, the highest accuracy is obtained for the combined model which is 98.8%. A modern method tSNE is used for dimensionality reduction along with the feature engineering method called the k-best method and the base model for the classification is the Random Forest model is used. This combined approach resulted in the best accuracy of 99%. We have obtained a little improved score using PCA and tSNE methods than the reference articles considered in this research.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

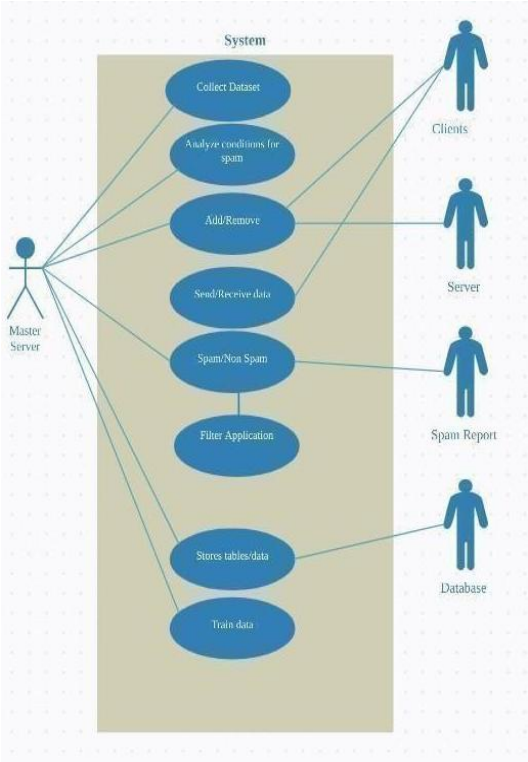



Fig: A basic flow of an application.

As shown in the figure, patients or users can add the details to the application, and then they can predict their output results.

## V. SIMULATION RESULTS

The model helps to predict the different diseases based on the report. For this, we have collected the dataset from Kaggle and trained the model,and tested the model using different reports of the patients.

Datasets:

1. **Cancer Data Set:** For the cancer dataset we have collected data from the Kaggle which contains 25000 rows[as a data] and it contains 30 columns [as a parameter] which help to detect whether a person suffering from cancer and the level of cancer of the patient.

2. **Heart Data Set:** For the Heart dataset we have collected data from the Kaggle which contains 18800 rows [as a data] and it contains 12 columns [as a parameter] which help to detect whether a person suffering from any heart disease or not.

3. **Kidney Data Set:** For the Kidney dataset we have collected data from the Kaggle which contains 15000 rows [as a data] and it contains 18 columns [as a parameter] which help to detect whether a person suffering from any kidneydisease or not.

4. **Diabetics Data Set:** For the diabetic dataset we have collected data from the Kaggle which contains 35000 rows [as a data] and it contains 9 columns [as a parameter] which helps to detect whether a person suffering from any diabetic disease or not.

**Algorithms:**

1. **Logistic Regression:** Logistic regression is a process of representing the probability of an individual outcome given an input variable. The most universal logistic regression models a twofold consequence; something that can take two principles in the way that true/false, yes/no, on this wise.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

2. **Decision Tree:**A decision seedling is a sequential diagram-like structure at which point each in-house node symbolizes a "test" on an attribute (like whether a chance of something happening happens suddenly luck), each arm portrays the outcome of the test, and each leaf bud shows a class label (resolution captured subsequently computing all attributes).

3. **Gaussian Naïve Bayes:** Gaussian Naive Bayes supports constant expensive features and models each as abide by a Gaussian (common) dispersion. An approach to constitute a simple model search out acquire that the dossier is described by a Gaussian dispersion accompanying no co-difference (liberated dimensions) 'tween ranges.

4. **Gradient Boosting Classifier:** Gradient boosting classifiers are a group of machine intelligence algorithms that integrate many weak knowledge models to generate a forceful predicting model. Decision trees are usually used when action gradient boosting.

5. **Linear Discriminant Analysis:** Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function Analysis is a dimensionality contraction method generally pre-owned for supervised categorization questions. It is not new for forming a distinct camp that is segregating two or more classes. It is used to project the looks in larger dimension scope into a lower measure scope.

6. **Support Vector Machine:** Support Vector Machine(SVM) is a supervised machine intelligence algorithm used for both categorization and reversion. Though we mention regression complications also appeals best suitable for categorization. The objective of the SVM algorithm can find a hyperplane in an N-spatial room that categorizes the dossier points. The measure of the hyperplane depends upon the number of features. If the number of recommendation countenance is two before the hyperplane is just flattery. If the number of recommendation appearances is three, therefore the hyperplane enhances a 2-D plane. It enhances troublesome to assume when the number of visages surpasses three.

7. **Random Forest:** Random Forest is a Supervised Machine Learning Algorithm namely used widely in Classification and Regression complications. It builds decision trees on various samples and takes their most chosen categorization and the average for fear of regression.

8. **K-Nearest Neighbour:** K-Nearest Neighbours are individuals of ultimate basic still essential categorization algorithms in Machine Learning. It belongs to the directed learning rule and finds severe application in pattern acknowledgment, dossier excavating, and intrusion discovery. It is widely not necessary in corporal scenarios because it is non-parametric, intention, it does not make some latent presumption about the distribution of dossier (as opposite to additional algorithms such as GMM, that acquire a Gaussian distribution of the likely dossier).

9. **XG Boost:** XGBoost, which means Extreme Gradient Boosting, is an adaptable, delivered gradient-boosted decision tree (GBDT) machine intelligence atheneum. It supports parallel tree boosting and is the superior machine intelligence study for regression, categorization, and rating problems. It's essential to understand of XGBoost to first grasp the machine learning ideas and algorithms that XGBoost builds upon directed machine intelligence, conclusion timbers, ensemble education, and slope pushing. Supervised machine intelligence uses algorithms to train a model to find patterns in a dataset accompanying labels and visage and therefore uses the prepared model to envision the labels on a new dataset's features.

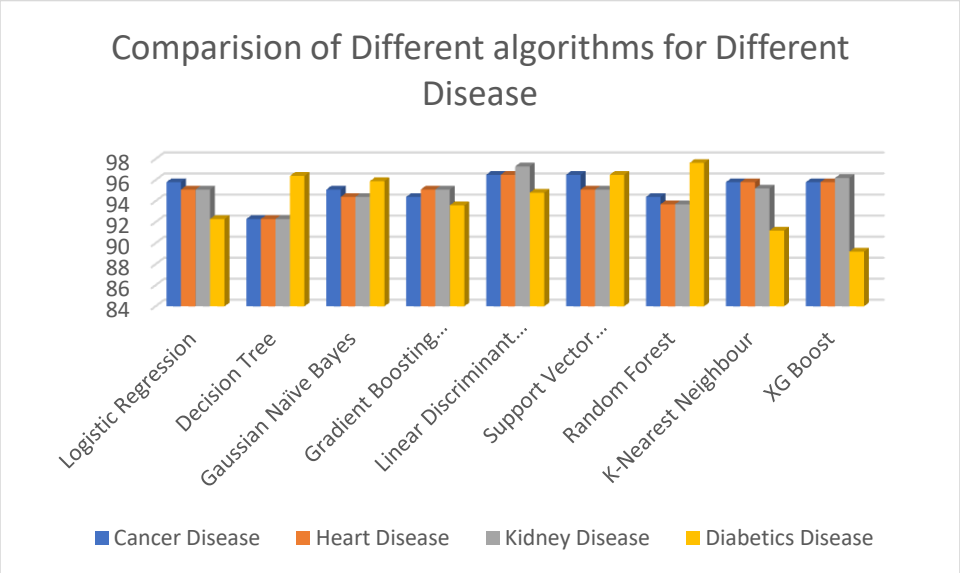
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

| Algorithm/Disease | **Cancer Disease** | **Heart Disease** | **Kidney Disease** | **Diabetics Disease** |
|---|---|---|---|---|
| Logistic Regression | 95.8 | 95.1 | 95.10 | 92.3 |
| Decision Tree | 92.3 | 92.3 | 92.3 | 96.4 |
| Gaussian Naïve Bayes | 95.1 | 94.4 | 94.4 | 95.9 |
| Gradient Boosting Classifier | 94.4 | 95.1 | 95.1 | 93.6 |
| Linear Discriminant Analysis | 96.5 | 96.5 | 97.3 | 94.8 |
| Support Vector Machine | 96.5 | 95.1 | 95.1 | 96.50 |
| Random Forest | 94.4 | 93.7 | 93.7 | 97.63 |
| K-Nearest Neighbour | 95.8 | 95.8 | 95.2 | 91.2 |
| XG Boost | 95.8 | 95.8 | 96.2 | 89.2 |

**Table -1 Comparison of different algorithms for different diseases**




Graph- That shows the comparative study between different algorithms on different disease

We have designed a website that helps to implement the models which we have designed the website takes the input from the user such as parameters and predicts the output based on thebest fit of the algorithm, The algorithm is chosen based on the diseases and chooses the best algorithm. As shown below-

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

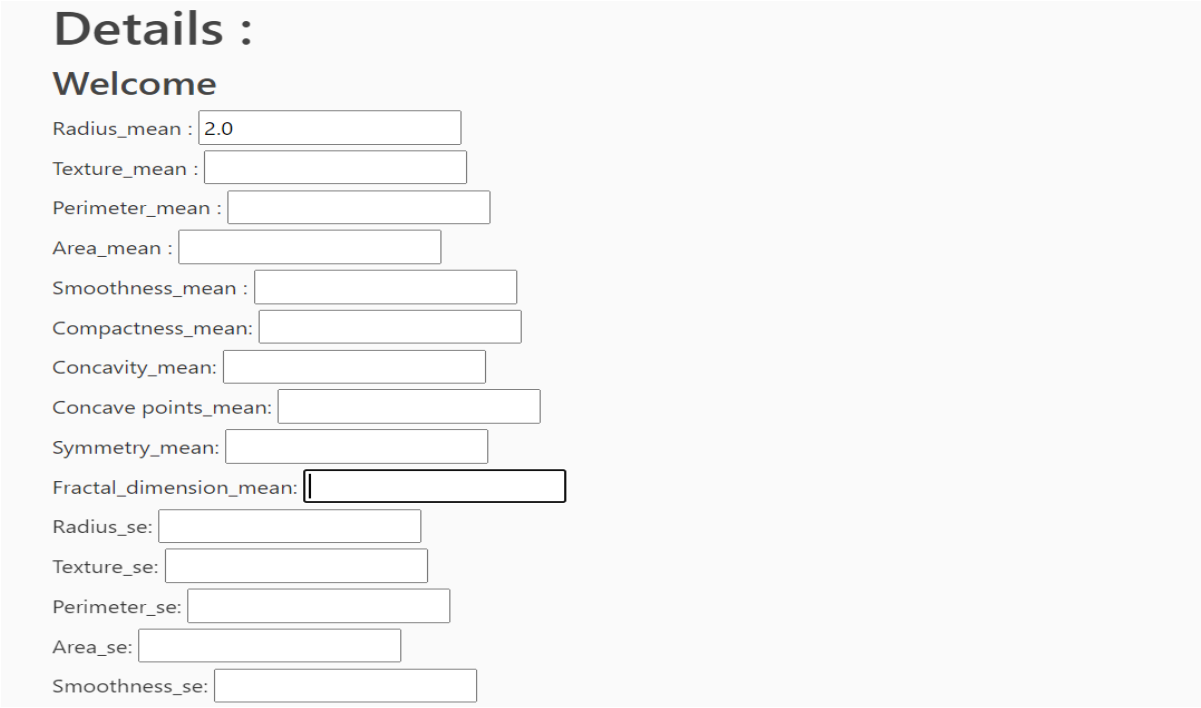



Fig-1.0

Fig-1.0 which takes all the inputs of the cancer data shown in the report and chooses the best algorithm from the above algorithms and predicts the output

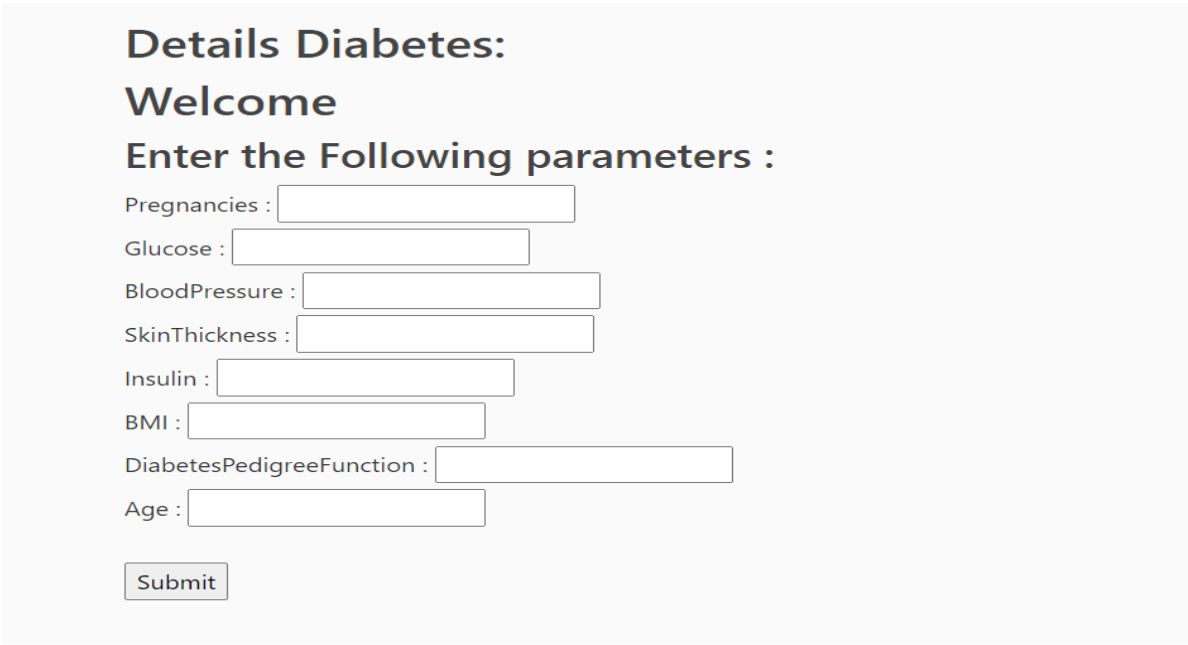



Fig-1.1

Fig-1.1 which takes all the inputs of the Diabetes data shown in the report and chooses the best algorithm from the above algorithms and predicts the output

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

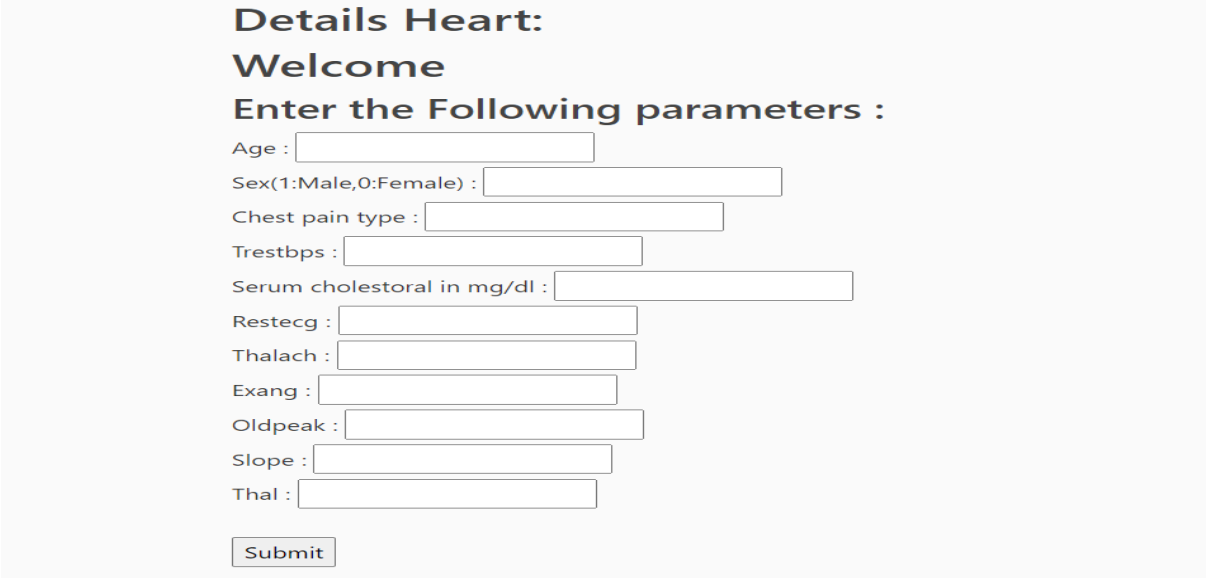



Fig-1.2

Fig-1.2 which takes all the inputs of the Heart data shown in the report and chooses the best algorithm from the above algorithms and predicts the output

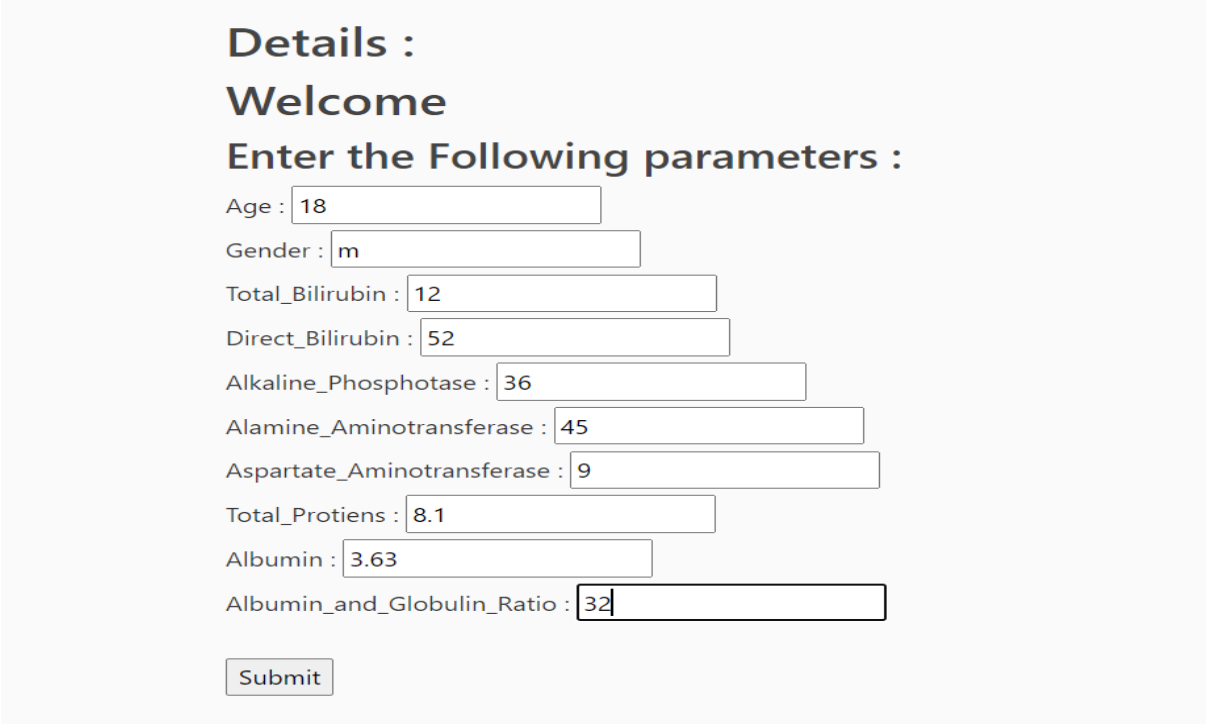



Fig-1.3

Fig-1.3 which takes all the inputs of the Liver data shown in the report and chooses the best algorithm from the above algorithms and predicts the output

For implementing the above models, we have designed a web application where patients can upload their details andpredict their output based on the chosen best algorithm by the machine. Which helps to remove the dependency on the doctors.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004163|

## VI. CONCLUSION AND FURTURE WORK

In today's world accuracy in a medical domain is very important. So, in order,we have implemented algorithms to study different diseases using different algorithms, and the best accuracy algorithm is picked and implemented [Table-1 shows the best algorithm to be picked and predicted]and we have only implemented only 4 diseases and now we can implement many other diseases and algorithms.

## REFERENCES

[1] Akash C. Jamgade, Prof. S. D. Zade, "Disease Prediction Using Machine Learning", Dept. of Computer Science and Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur, Maharashtra, India, e-ISSN: 2395-0056, pISSN: 2395-0072
[2] Harini D K, Natesh M, "PREDICTION OF PROBABILITY OF DISEASE BASED ON SYMPTOMS USING MACHINE LEARNING ALGORITHM", Dept. of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India, e-ISSN: 2395-0056, p-ISSN: 2395-0072
[3] Chen Wang, Thierry Pun, and Guillaume Chanel, "A Comparative Survey of Methods for Remote Heart Rate Detection From Frontal Face Videos", PMCID: PMC5938474, PMID: 29765940
[4] Chen Wang, Thierry Pun, and Guillaume Chanel, "A Comparative Survey of Methods for Remote Heart Rate Detection From FronRinkal Keniya, Aman Khakharia, Vruddhi Shah, Vrushabh Gada, Ruchi Manjalkar, Tirth Thaker, Mahesh Warang, Ninad Mehendale, "Disease prediction from various symptoms using machine learning", K. J. Somaiya College of Engineering, Mumbai, Initial Face Videos", PMCID: PMC5938474, PMID: 29765940
[5] Niklas Donges, "A Complete Guide to the Random Forest Algorithm", AI Expert, builtin.com, JUNE 16, 2019
[6] https://www.kaggle.com/uciml/indian-liver-patient-records
[7] https://www.kaggle.com/mansoordaku/ckdisease
[8] https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria
[9] Shafiq's Muhammad Abdulhamid, Muhammad Shafiq Abd Latiff, Haruna Chiroma, Oluwafemi Osho, Gaddafi Abdulsalam, Adamu I. Abubakar, and Tutu Herawan. By using different machine learning algorithms like SVM and Bayesian classifiers.
[10] Ster, B., Dobnikar, A.: Neural networks in medical diagnosis: comparison with other methods. In: Proceedings of the International Conference on Engineering Applications of Neural Networks, pp. 427–430 (1996)

## BIOGRAPHY

**Dr. R.Venkata Ramana chary** is a professor in the Information Technology Department, B V RAJU INSTITUTE OF TECHNOLOGY. He received Ph.D from GITAM university.
**Vijaykumar Mantri** is an Associate professor in the Information Technology Department, B V RAJU INSTITUTE OF TECHNOLOGY, He received M.Tech from VTU, Belgaum
**Dr Sameera Ch**is an Associate professor in the Information Technology Department, B V RAJU INSTITUTE OF TECHNOLOGY, He received M.Tech from JNTUH, Belgaum

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details