International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.003

Patient Diagnostic System Using Binary Logistic Regression

DitipriyaGanguly
Department of Data Science and Business Systems,
SRM Institute of Scienceand Technology
dg7741@srmist.edu.in

Mriganka Jana
Department of Data Science and Business Systems
SRM Institute of Scienceand Technology
mj6564@srmist.edu.in

R Jayaraj
Department of Data Science and Business systems
SRM Institute of Scienceand Technology
jayarajr1@srmist.edu.in

***Abstract*—Medi-tracker is a web application using machine learning. Medi-Tracker serves doctors, medical representatives as well as patients to help understand their health condition after a blood test or after a diagnosis. Medi-Tracker was built with the intent to serve common people with an easy to access and understand user interface and reduce their cost to visit a doctor or stand in long queues to get the conclusion from their report. With an accuracy rate of 94% using Binary logistic regression Medi-Tracker is a python-based web application. According to a report even in 2022 many people can't afford a doctor's visit, so with the help of this just by undergoing a blood test they can have an AI based consultation at their fingertips. Our model is user-friendly and an end-to-end system for prediction. Our main purpose is to serve general people with easy medical assistance.**

Index Terms – Machine Learning, binary logistic regression, deep learning, .

I .INTRODUCTION

A. Artificial Intelligence

Artificial intelligence is the human-made intelligence for devices and machines. AI works as a helping hand for humans to perform tasks and help keep track of day-to-day activities in their lives. The term artificial intelligence was coined previously to defy machines that could mimic and perform human-like activities like "decision making" and "problem- solving". Artificial intelligence has come a long way in the present world of technology. Just giving a simple task to a machine and letting it finish the task with the desired output has saved a lot of time and manpower. Today a lot of sectors use artificial intelligence as their helping hand. From traffic control to weather prediction, vehicle manufacturing to cruise control, and autopilot in vehicles. From assembling small parts in a mobile device to making calls or setting reminders artificial intelligence has come a long way.

B. Machine learning

Machine learning, as illustrated in Fig.1, being a subset of AI (artificial intelligence) that deals with training a machine to perform a set of tasks based on the training data. Machine learning is a process to help machines to learn and provide solutions by predicting and performing simple statistical calculations. Machine learning is crucial in training machines and computers based on datasets and lets them perform actions and predictions based on the data. A machine is first trained with a dataset and after the training period is over the machine is given inputs and based on the training it can predict the output for the input received. In several fields machine learning algorithms are employed to execute tasks that would be difficult or impossible to perform with conventional techniques.

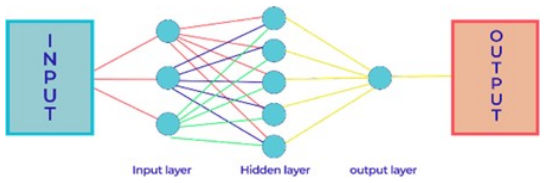

Fig. 1. Machine Learning

C. Deep Learning (DL)

DL or deep structured learning, as illustrated in Fig.2, is a broader part of machine learning. Where machine learning simply works with one hidden layer of the trained dataset, deep learning has more than one layer of the dataset to predict and perform tasks precisely. Deep learning is machine learning using artificial neural networks and representation learning.

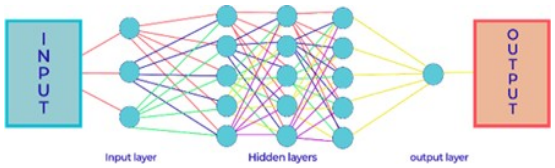

Fig. 2. Deep Learning

D. Artificial neural network (ANN)

ANN consists of a connection formed by interconnected nodes similar to neurons of animals(humans). As neurons transmit data from one part of the body to the brain and direct output to the preceptors an artificial neural network performs a similar task by sensing the environment and transmitting the data from the sensors to the processing unit and then from the processor back to the actuators to perform the output. An artificial neural network forms a network of interconnected nodes to share information and perform specific tasks layer by layer. Each layer is responsible for different tasks and duties. An artificial neural network is a neural network made artificially, these are a group of multiple preceptors or neurons at each layer that solves the tabular data, Image data, and testdata.

An Artificial neural network is capable of learning non-linear functions, it uses universal function approximations.

An ANN faces a few challenges such as -

- The number of trainable parameters increases, so does the size of the image.

DOI: rp-9788770040723.003

- It shows vanishing and exploding gradients.

AI is powering personal devices in our homes and offices similarly to electricity. AI will be the next big thing after the electricity that will hugely impact our daily life.

E. Medi - Tracker

The name of the project is **Medi-Tracker** as its main function is to track and predict whether a patient is having or has the chance to get a particular disease or not. This project uses Binary Logistic Regression to forecast the likelihood of disease in various organs of the body like the liver, kidney, heart, etc.

This project uses the datasets from Kaggle and other sources to train and link users via HTML forms and landing pages.

- Breast Cancer: Wisconsin Breast Cancer Dataset
- Diabetes: Pima Indian Diabetes Dataset
- Liver: Indian Patient Liver Records
- Heart: Heart Disease Dataset
- Kidney: Chronic Kidney Disease Dataset

II. Related Work

Patients have to consult a doctor each time they undergo a blood test or they have been through a diagnosis. Normal people are not able to understand what all these readings mean and how they are normal or not. So after each step or regular blood test patients or their family members have to consult a doctor just to understand the report. The existing process of undergoing a predicting diagnosis was a bit tiring and costly. Also, other patient diagnostic systems are not user-friendly, and accessing them for normal people is nearly impossible. The end-users so far were always doctors and medical representatives. Still, in 2022 many people will not be able to get proper healthcare services because of not the wide reach of doctors and hospitals. Using AI people can have a personal doctor at home. There is a scarcity of professionals like technicians, nurses, doctors and infrastructure. According to a World Health Organization report (WHO), there is only a 0.76:2.09 doctor to nurse ratio per 1000 people. Thus it can be said that there is an acute shortage of hospital beds is being faced by Indian Healthcare. Physical access to hospitals remains a significant barrier to both preventive and curative health care. High medical costs are incurred in India by approximately 63 million people and they face poverty due to the same. Using technology in a better way, healthcare can be made affordable for everyone.Previously, various studies were conducted for disease detection based on specific symptoms using various ML algorithms. Montoet. al. developed a SAS Statistical Analysis Software version 6.12 for influenza disease prediction [1]. A study of influenza patients with fever and at least two additional flu symptoms was conducted in the clinical trial, which involved a total of 3744 unvaccinated adults and adolescents. Eventually, the lab found that 2470 of the 3744 individuals were infected with influenza.. The model's accuracy rate was 79 percent. Karayilan et al. [2] devised a model for the heart disease pre- diction based on ANN and Backpropagation Neural Networks, two of the most prevalently used ANN learning techniques. 13 clinical features were fed as input and then trained with the help of backpropagation algorithm in order to decipher any heart disease with an accuracy of 95%. Chen et al. [3] streamlined a number of ML algorithms to accurately assess the chronic outbreak of disease. The amount of data gathered for training purposes was grossly inadequate. A latent factor model was being used in order to overcome this. A new multimodal disease risk prediction model (CNN-MDRP) based on convolutional neural networks was created. The algorithm's accuracy was roughly 94.8%. This study makes use of the actual hospital data taken from the hospital dataset. Chae et al. [4] analysed 80 infectiousdiseases using four different DL models: DNN, OLS, LSTM and ARIMA. All other models tested were outperformed by the DNN and LSTM models. Resul et al.[5] built an ensemble-based methodology to diagnose heart disease using SAS enterprise miner 5.2 in their study. Experiments on the heart disease dataset were carried out in order to fully automate the diagnosis of heart disease. Three independent neural network models were taken into consideration in order to create the ensemble model. There was an increment of neural network nodes of ensemble model, but further performance improvement wasn't noticed. The results of the experiment can be used to diagnose heart disease with an accuracy rate of 89.01%. Four ML algorithms are employed to process the input dataset: Naive Bayes, Random Forest, SVM and Simple CART. A classifier model is trained and tested for each algorithm, and the results obtained from them are compiled. SVM has the highest precision value of 0.784, while Random Forest has the lowest precision value of 0.756. With a higher accuracy of 79.13%, the SVM model outperformed the other methods in [6]. In this case, the "Pima Indians Diabetes Database", obtained from the "National Institute of Diabetes and Digestive and Kidney Diseases", is the dataset under consideration. Khourdifi et al. [7] employed the KNN model to predict and classify heart diseases. He used the UCI heart disease dataset and an accuracy of 99.7% was achieved. Sriram et al. [8] achieved a 90.26% accuracy with the in- corporation of the Random Forest model. The "Parkinson's disease voice dataset" was obtained from the "UCI Machine Learning repository at the Center for Machine Learning and Intelligent Systems". Parthiban et al. [9] applied automatic learning methods to diagnose cardiac related diseases present in patients diagnosed with diabetes. WEKA has been used to implement the Naive Bayes and SVM algorithms. The Chennai Research Institute's data set of 500 patients has been used. The Naive Bayes algorithm has an accuracy 74% whereas SVM has the highest accuracy of 94.60%. To detect heart disease, Chaurasia et al. [10] proposed using data mining approaches. The algorithms used in this case are bagging, Naive Bayes and J48. A data set on heart disease with 76 attributes is available from the UCI machine learning laboratory. Prediction is based on only 11 attributes. Naive Bayes achieved an accuracy of 82.31%. J48 provides an accuracy of 84.35%. Bagging yields an accuracy of 85.03%. Thus, on this data set, bagging has been able to provide a higher classification rate. Vembandasamy et al. [11] made use of the Naive Bayes algorithm for detection of cardiac diseases. The information used came from one of the Chennai's most renowned and

DOI: rp-9788770040723.003

leading diabetes research institutes. The dataset consisted of 500 patients. WEKA tool is used and 70% of the Percentage Split performs classification. Naive Bayes has an accuracy rate of 86.419%. A study waspresented by X. Liu et. al. [12] using a hybrid classification system devised upon the RFRS method in order to help in heart disease diagnosis. There are two subsystems for the proposed system: an RFRS featured selection system along with an classification system with an overall classifier. The maximum classification accuracy for the given model is of 92.59%. Maniruzzaman detected the diabetes disease by using certain ML algorithms [13]. The diabetes risk factors were identified using the Logistic Regression (LR). The model performed with an accuracy of 90.62%. Mohan et al.[14] put in place an effective heart disease prediction system. An accuracy of 88.4% was achieved using the HRFLM upon The Cleveland dataset. Daghrir et al. employs three ways to arrive at a conclusion on a given system: KNN, SVM, and CNN. Because of its efficiency and versatility, it outperforms KNN. Although an SVM classifier performed quietly, the CNN is regarded as a more strong and resilient classifier, giving an accuracy of 85.5%.

III. Proposed Method

Here we are proposing the solution to overcome long hospital cues and cut the costs of doctor visits just to know your health status after a blood test or diagnosis. Nowadays a blood test is easily available at the doorstep and the results are delivered via email or door to door. But reading and understanding the result data. We proposed a solution where anyone can easily access our web application for free and check for any potential harm they possess as per the blood report or not.

Our webpage is simple with an easy user interface. Any user with a little bit of knowledge can check the status of a patient or even the patient can check the status of his/her health via our web application.

Patients just have to put their data in their respective fields and click on predict to get their results in simple language. Our model is much more user-friendly than the previous models. Our model is an end-to-end system for prediction. Our main purpose is to serve general people with easy medical assistance. Our backend does not store the data of any patient so we don't ask users for their names or mobile number. The sole purpose is to help people and that is served through this model. By making use of web technology and AI people in remote areas can also get access to health care. Using technology in a better way, healthcare can be made affordable for everyone. We have used a light ML model that can load within no time without compromising the accuracy of the prediction. Light models quickly load which helps in better response to the page. This Web-App provides a wide range of predicting the contamination of multiple infectious diseases as in Diabetes, Kidney disease, Liver Infection, Heart Disease & Cancer. Though this app provides a wide rangeandhighspeedtheaccuracyhasnotbeencompromised.

Hyper-parameters have been selected very carefully so that the accuracy is not hampered.

As shown in Fig.3, Our system is always active and patients need not wait for doctor's visits or wait in queues. They can easily put in their report values and get the prediction for their health.

[start
{
Read details from html form
Process the details and check with the machine hearing model.pkl
Predict it is '0' or '1'
Check the value for '0' or '1' Send the output via html static
}
end]

Helping others is the only goal, and this methodology does that. People in faraway locations may now receive healthcare thanks to web technologies and AI. Healthcare may be made more accessible and inexpensive for everyone by utilising technology effectively. We have employed a quick-loading, lightweight ML model without sacrificing prediction accuracy. Light models load rapidly, improving the page's responsiveness.

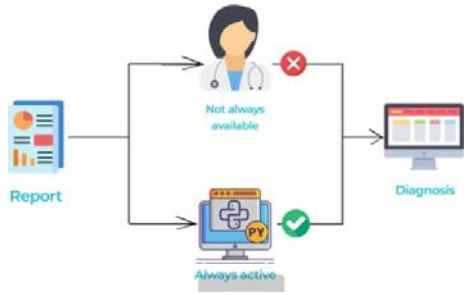

Fig. 3. Our model ideation

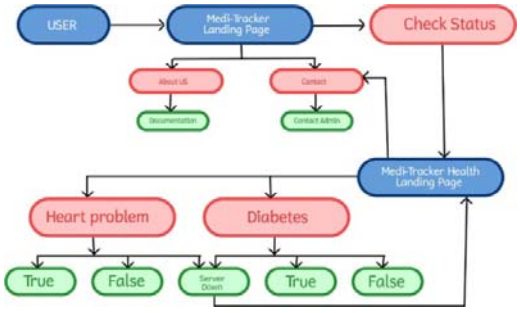

Fig. 4. Working of Model

We are using binary logistic regression, as shown in Fig. 4, to predict our patient's and user's health status.

In binary logistic regression, the predicted outcome has only two values (1 or 0) or (Yes or No)

It is the most utilized regression model where the value is either 1 or 0. Unlike fuzzy logic it has only 2 discrete outcomes.

If 32◦C is considered to be hot and any temperature below that is termed as cold then binary logistic regression would distinguish it as -

- 29◦C - Cold
- 31◦C - Cold

- 31.9◦C - Cold
- 32◦C - Hot

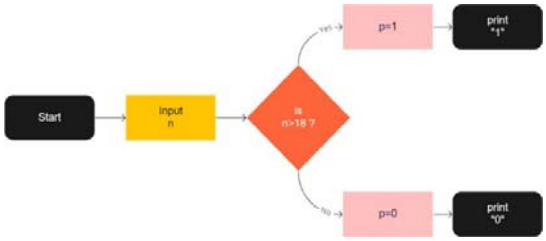

Fig. 5. Flow of Work of binary logistic regression

B. *Accuracy*

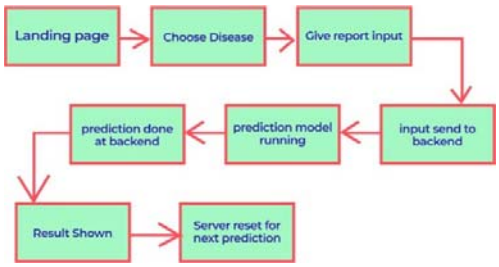

Fig. 6. Flow of Model

IV. EXPERIMENTS

A. *Training Details*

The Patient Diagnostic System, that is, Medi Tracker, was trained using logistic binary regression which gives discrete results as True or False, we gather data to and forth an HTML page in the form of a form, connected to the Python virtual environment in the background.

B. *Metrics*

1) *Accuracy:* The accuracy of the model is calculated by comparing the prediction of the model to the ground truth. The mathematical calculation of the overall accuracy is done using True Positives, True Negatives, False Positives, and False Negatives as the parameters.

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

C. *Experimental Results*

The dataset used was extracted from Kaggle to train the prototype, based on the trained process. When the input is given, the data has been successfully validated for the potential occurrence of the diseases with an accuracy of 96.97%.

V. CONCLUSIONS

All the previously developed machine learning algorithms and models could predict only one disease and had a low accuracy rate. In our study, we have attempted to retain a high accuracy while removing one significant flaw seen in earlier models used for the diagnosis of patient diseases- the low accuracy rate of predicting the patient's disease. We have incorporated one singular platform for all the diagnosis of a wide range of diseases, which increases the efficiency and convenience of use by regular people, with an accuracy of 95.62%.

REFERENCES

[1] A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, and J. Schweinle, "Clinical signs and symptoms predicting influenza infection," Archives of Internal Medicine vol. 160, no 21, p. 3243, 2000.

[2] O. Karayılan, and dKılı¸C, International Conference on Computer Science and Engineering (UBMK), IEEE, pp. 719–723, 2017.

[3] Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", IEEE Access 5, p. 8869, 2017.

[4] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data", International Journal of Environmental Research and Public Health, vol. 15, no. 8, p. 1596, 2018.

[5] Resul Das, Ibrahim Turkoglu, and AbdulkadirSengur, "Effective diagnosis of heart disease through neural networks ensembles", Expert Systems with Applications, vol. 36, issue 4, 2009.

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

[7] Y. Khourdifi, and M. Bahaj, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization", Int. J. Intell. Eng. Syst., vol. 12, no. 1, p. 242, 2019.

[8] Moshika, A., Thirumaran, M., Natarajan, B., Andal, K., Sambasivam, G., &Manoharan, R. (2021). Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access, 9, 74659-74673.

[9] G .Parthiban and S. K. Srivatsa, "Applying machine learning methods in diagnosing heart disease for diabetic patients," International Journal of Applied Information Systems, vol.3, no.7, pp.2249-0868, 2012.

[10] V. Chaurasia and S. Pal, "Data mining approach to detect heart diseases", International Journal of Advanced Computer Science and Information Technology, vol.2, no.4, pp.56-66, 2014.

[11] K. Vembandasamy, R. Sasipriya, and E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm", IJISET-International Journal of Innovative Science, Engineering & Technology, vol.2, pp.441-444, 2015.

[12] X. Liu, X. Wang, Q. Su, M. Zhang, Y.Zhu, Q. Wang, and Q. Wang, "A hybrid classification system for heart disease diagnosis based on the rfrs method," Computational and Mathematical Methods in Medicine, vol.2017, Article ID 8272091, p. 11, 2017.

[13] Rajesh, M., &Sitharthan, R. (2022). Image fusion and enhancement based on energy of the pixel using Deep Convolutional Neural Network. Multimedia Tools and Applications, 81(1), 873-885.

[14] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, p. 81542, 2019.

[15] J.Daghrir, L.Tlig, M.Bouchouicha,and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," In Proceedings of the 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sfax, Tunisia, 2–5 September 2020; IEEE: Manhattan, NY, USA, pp. 1–5, 2020.