Grenze International Journal of Engineering and Technology, June Issue

E-Health System for Disease Prediction based on Machine Learning

Honey Jain[1], Rohan Benhal[2] and Tanmayee Parbat[3]
[1, 3]B.E IT, Maeer's MIT, Pune
Email: tanmayee.parbat01@gmail.com
[2]BBA IT, Maeer's MIT, Pune

***Abstract*—The health reports of the persons including diagnostics information and medical prescriptions are provided in the form of test based case notes due to this the previous health conditions and the medicines used by the person are not known when he visits the hospital later. But storing all the health information of a person in the cloud as the soft copy reduces this problem. To achieve this each and every hospital, dispensary, labs must have an internet connection for registration of patient's data, each patient will be identified by the unique Health Id and all the data of the patient will be stored in the cloud and the data can be accessed by only the particular patient. In order to prevent and treat illness, it is critical to perform an accurate and timely analysis on any health-related problem. The ability to diagnose disease by obtaining all information from a linked Health ID combined with Machine Learning techniques will improve the system's ability to detect diseases. We believe that our diagnostic model can operate like a doctor in the earlier diagnosis of this disease, allowing for timely treatment and the preservation oflife.**

***Index Terms*— cloud computing, disease prediction, health id generation, machine-learning algorithm.**

I. INTRODUCTION

In order to prevent and treat illness, it is critical to perform an accurate and timely analysis on any health- related problem. It is possible that the standard method of diagnosis will not be adequate in the event of a severe illness. The development of a diagnosable disorder method machine learning-based (ML) algorithms for the predictions of just about any disease can aid in the development of a more definitive reading than that of the traditional processes now in use.

The goal of this project should be to provide and develop an appropriate foundation for an Internet of Things-based smart health system. In order to address the fundamental concerns of IoT-based intelligent health systems, the framework takes a layered approach. It also provides a comprehensive method of data gathering from either the physician to cloud computing, which can be accessible from anywhere in the world. Section A describes the numerous diseases as well as the parameters that can be used to forecast the development of diseases. Section 2 describes the system's goal, whereas section 1 describes the system's purpose.

A. Hypothesis behind selection of different parameters

The following clinical parameters were discovered for each of the diseases while observing the various aspects

Grenze ID: 01.GIJET.8.2.318

for the detection of heart disease, cancer, and diabetes throughout the process of monitoring the various features for the detection of these diseases:

Heart disease Parameters
Age, education, gender, does the person smoke, cigarettes per day, medication of blood pressure, any prevalent stroke, prevalent hyper-tension, diabetes, total cholesterol level, systolic BP, diastolic BP, body mass index, glucose levels, heart rate, etc.

Cancer Parameters
Age, Gender, Air Pollution, Alcohol use, Dust Allergy, Occupational Hazards, Genetic Risk, chronic Lung Disease, Balanced Diet, Obesity, Smoking, Passive Smoker, Chest Pain and Coughing of Blood, etc.

Diabetes Parameters
Glucose Before fasting, Glucose Anytime, Age, Gender, Blood Pressure, Family member with Diabetes past or present, BMI, Pregnancies, percentage of occurring diabetes, Polyuria, Polydipsia, sudden weight loss, weakness, Polyphagia, Genital thrush, visual blurring, Itching, Irritability, delayed healing, Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Function, etc.
Based on these characteristics, it is determined that some characteristics, such as BMI, glucose levels, genetic history, weight loss, smoking behaviours, and so on, are shared by a variety of diseases. While performing providing a good between the characteristics of various diseases including their respective dependencies, the following table was discovered.

B. Objectives

- Detail survey on disease prediction system by various researchers.
- To diagnose the disease by accessing all information from linked Aadhar Card ID.
- Use of Machine Learning algorithm for accurate diagnosis of disease.
- To provide privacy and security for all sensitive data.
- Use of cloud database for 100 percentages of availability of data at any time.

II. LITERATURE REVIEW

A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun in paper [1] proposed system that uses a database that contained information about patients having heart disease. Authors extracted features using three selection algorithms that are relief, minimum redundancy, and maximum relevance (mRMR), and least absolute shrinkage and selection operator, which was cross- verified by the K-fold method. The extracted features were sent to 6 different machine learning algorithms and then it was classified based on the presence or absence of heart disease. On the Cleveland heart disease dataset with all of its attributes, the authors evaluated the performance of various machine learning techniques, including logistic regression, k-nearest neighbour, artificial neural network, support vector machine, Naive Bayes, and decision tree. Relief, mRMR, and LASSO were utilised to choose relevant features in the second study, which was conducted using feature selection algorithms. Performances of third-generation classifiers were evaluated based on a set of selected features. In addition, the k-fold cross-validation approach was employed to test the results. Performance evaluation measures were used to monitor the effectiveness of classifiers in order to determine their effectiveness. Before applying classifiers, all features were normalised and standardised to ensure that they were comparable.
In their publication [2], the authors Tülay Karaylan and zkan Kiliç proposed a computer - aided diagnostic system that makes the use of artificial neural network multilayer perceptron algorithm to forecast heart illness. To construct the neural network, 13 clinical features were used as input, and the neural network was then trained using the back-propagation algorithm approach to determine the nature and extent of coronary heart disease with or without an accuracy of 95%. Heart disease is predicted using input variables that appear to be disease risk factors derived from a dataset, as well as output variables that are categorised into categories such as "disease absence" and "disease presence," among other things. Heart disease predictors use output data that fall into a classification, such as "disease absence" and "disease presence," to make predictions about the condition. Unsupervised learning problems, such as the difficulties in forecasting heart disease, are referred to as "supervised learning problems." Because the output variables are derived from the program's course, the forecasting of heart disease is characterised as "supervised learning of the identify and respond." Heart disease prediction system was developed using the Backpropagation Algorithm, which is a common Artificial Neural Network learning technique. The algorithm was used to predict heart disease. Given that it is the sole technique

that can be used to forecast nonlinear relationships, the Backpropagation Algorithm is the most effective classification algorithm for the prediction of cardiovascular disease.

Michalis V. Karamouzis Dimitrios I. Fotiadis in paper [3] proposed there have been numerous applications of machine learning (ML) techniques in cancer research, which would include artificial neural networks (ANNs), bayesian networks, support vector machines (SVMs), and decision trees (DTs). These techniques have resulted in more effective and accurate decision-making across the board.

Authors R. Delshi howsalya devi1, p. sreevalli, keerthana, prathyusha, m.asia in paper [4] proposed a general disease prediction based on symptoms of the patient. They used the random forest neural network approach for disease prediction in order to make an accurate prediction of the disease. Disease symptom datasets are essential for disease forecasting purposes. Throughout this comprehensive prediction and diagnosis, the person's lifestyle habits as well as checkup information are taken into account for an accurate forecast. Initially, we obtain a disease dataset from the UCI machine learning website, which is in the form of a disease list with symptoms associated with each ailment. A preprocessing step is then conducted on the dataset in order to clean it up, which includes eliminating commas, punctuation, and white spaces. And this is utilised as a training dataset for the algorithm. Following that, the feature was extracted and selected. Then, using classification techniques such as the random forest algorithm, we sort and organise the data.

Authors Naipeng Dong, Hugo Jonker, and Jun Pang performed review on Challenges in eHealth in paper [5]. They claim that a number of factors contribute to their success. Nowadays, healthcare is predicated on the ability to access medical records at any time and from any location. The adoption of the cloud infrastructure in healthcare makes it easier to share and integrate medical records.. However, while the cloud services paradigm has some advantages, it also poses a number of privacy and security risks to health-related data. Essentially, cloud computing providers should address security risks in the cloud in order to increase the level of confidence between patients and healthcare professionals. Hybrid Cloud IaaS “Host” PaaS “Build” SaaS “Consume” On premises Intranet High security Behind the firewall Off premises Internet Low Security Outside the firewall Private Cloud Public Cloud Relationship between delivery and service models.

Authors Sahandi, R. & Alkhalil, Adel & Opara-Martins, Justice in paper [6] explains about cloud. Authors propose review on use of Cloud in e-health in. Because the cloud may be used to store, manage, secure, distribute, and preserve electronic health records (EHRs), laboratories information systems, pharmacy information systems, and medical photographs, the cloud is becoming increasingly popular among healthcare professionals. Patients will receive better care overall as a result of the availability of up-to-date health records and the ongoing contact between multiple health care providers. Security, privacy, and trust concerns are the most significant barriers to widespread use of cloud computing by healthcare providers, aside from a lack of standards and legislation, as well as interoperability issues.

Authors Sathya, M. & Madhan, S. & Jayanthi, K in paper [7] present an idea of a service model in technological and economic views for the comfort of patients and also the open challenges in implementing IoT in real world medical field.

Authors Chaitrali S. Dangare and Sulabha S. Apte in paper [8] proposed Study of Heart Disease Prediction System using Data Mining Classification Techniques. Currently, many hospitals handle healthcare data through the use of healthcare information systems, which, according to them, include vast amounts of data that may be utilised to extract hidden information for the purpose of producing intelligent medical diagnoses. The primary goal of this project is to develop an Intelligent Heart Illness Prediction System that really can diagnose heart disease based on a historical heart database, which will be used in future research. Medical terms including such sex, hypertension, and cholesterol, as well as 13 input attributes, are used in the development of this system. Two additional attributes, namely obesity and smoking, are used in order to obtain more accurate results, as these character traits are believed to be key risk factors for heart disease. The classification techniques used in data mining, such as Neural Networks, Decision Trees, and Naive Bayes, are employed.

In their study [9], the authors Mostafa Haghi, Saeed Danyali, Sina Ayasseh, Ju Wang, Rahmat Aazami, and Thomas M. Deserno used simultaneous monitoring to provide a full evaluation of the subject's health in a randomised controlled trial. Air pollutants, such as poisonous and hazardous gases, as well as specific matter, have an impact on diseases such as cardiovascular disease, lung cancer, and respiratory disease. The dangers of being exposed to such contaminants have already been extensively researched and documented. Scientific innovation has been carried out to determine the impact of air pollution on the brain and the development of mental disease (e.g., depression). The findings of the study demonstrate a relationship between respondents' psychological well-being and air pollution. In other words, exposure to increased levels of air pollution reduces the amount of time that the person spends outside. As a result of reducing physical activity outside,

psychological anguish is exacerbated, with restricted exposure to the sun and feelings of loneliness among the consequences.

In their study [10], the authors Mohammadzadeh, N., Gholamzadeh, M., and Saeedi conduct a systematic here on implementation of wearable technology sensors are used to monitor the vital signs of patients in epidemics, as well as a description of the technology. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) technique, this systematic review was carried out in this study. This method was developed by Moher et al. and is considered to be one of the most effective methods for authors to use when doing systematic reviews.

III. Proposed System

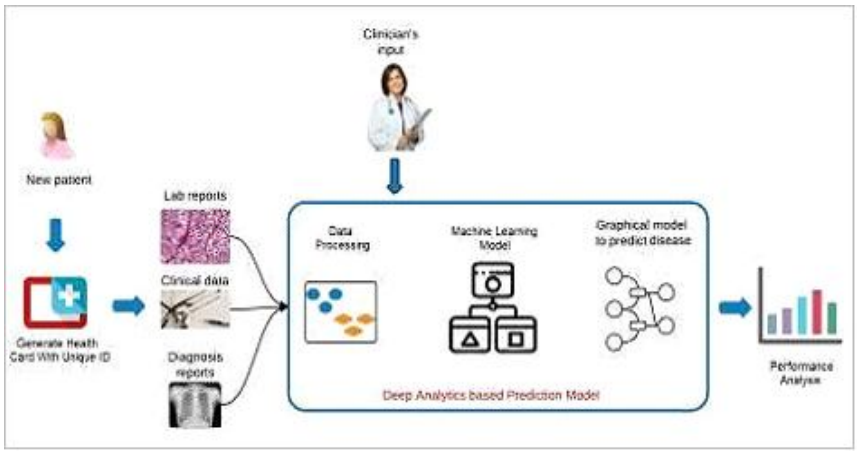

Figure 1: Disease prediction system model using machine-learning model

Fig. 1 shows the proposed system for disease prediction. Initially Unique Health Id is generated for each patient and all data and reports can be search using that Health ID. The doctor may not be available always when needed. But, in the modern time scenario, according to necessity one can always use this prediction system anytime. The symptoms of an individual along with the age and gender can be given to the ML model to further process. After preliminary processing of the data, the ML model uses the current input, trains and tests the algorithm resulting in the predicted disease. The proposed system uses dataset consisting of gender, symptoms, and age of an individual was pre-processed and fed as an input to different ML algorithms for the prediction of the disease. The different ML models used for prediction of disease. The outcome of the models is the disease as per the symptoms, age, and gender is given to the processing model. Following are the important modules of proposed system.

IV. Result and Discussion

From the data in Table 1, it can be shown that the suggested algorithm is extremely effective in terms of identifying individuals with heart disease. Due to the fact that if the initial weights are removed, the algorithm behaves similarly to a conventional neural network, this is the case. This neural network classifies input depending on the parameters provided and generates outputs that are not dependent on the settings of other networks.

In the case of disease prediction in other cases, factors such as job status, obstetric history, nutritional status, and anaemia were all significant predictor factors. Disparities were found between psychosocial features such as job level, obstetric factors such as prenatal record, and nutritional intake, as well as between signs and symptoms of renal disease, according to the findings. As a result, accuracy is required for the classification of different parameters for the prediction of human disease, and feature extraction methods for disease prediction can be accurate when using the Deep Convolution Neural Network. As a result, an accurate classification of diseases was achieved using the Deep Convolution Neural Network.

V. Conclusion

The work presents the detail review of predicting the disease based on the symptoms, age, and gender of an individual patient. The various Machine Learning, deep learning models are applied for disease prediction by

TABLE I. ANALYSIS OF ACCURACY OF VARIOUS CLASSIFICATION ALGORITHMS

Algorithm	Accuracy (%)
Deep CNN	95%
Fuzzy CNN	96%
SVM	90%
CNN	91%
Proposed	97%

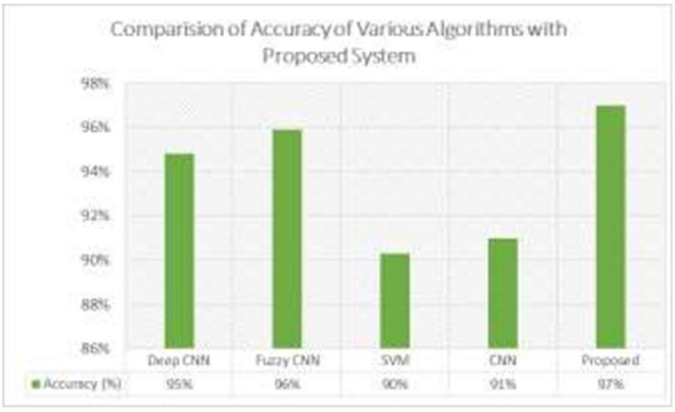

Figure 2. Analytical Comparison of various classification method with proposed system

researchers. As some models were dependent on the parameters, they could not predict the disease and the accuracy percentage was quite low. This work would help in lowering the cost required in dealing with the disease and would also improve the recovery process.

The development of increasingly advanced machine learning algorithms will be essential in the future to improve the accuracy of disease prediction. In addition, evidence - based manner should be adjusted more frequently after the training period in order to get potentially greater performance beyond the training phase. Furthermore, datasets should be increased to include more diverse demographics in order to minimise overfitting and improve the performance of the prediction that are implemented. Finally, in order to improve the performance of the learning methods, it is necessary to employ more relevant feature selection approaches.

REFERENCES

[1] A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, a hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mobile Information Systems 2018.

[2] T. Karayılan, O. Kılı¸c, in ¨ 2017 International Conference on Computer Science and Engineering (UBMK) (IEEE, 2017), pp. 719–723

[3] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis,
D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13, 8 (2015)

[4] R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm, 2012.

[5] N. Dong, H. Jonker, and J. Pang, "Challenges in eHealth: from enabling to enforcing privacy," in Foundations of Health Informatics Engineering and Systems. FHIES 2011. Lecture Notes in Computer Science, Z. Liu and A. Wassyng, Eds., pp. 195–206, Springer, 2011.

[6] European Network Information Security Agency, An SME Perspective on Cloud Computing, European Network Information Security Agency, Heraklion, Greece, 2009.

[7] Sathya, M. & Madhan, S. & Jayanthi, K.. (2018). Internet of things (IoT) based health monitoring system and challenges. International Journal of Engineering and Technology(UAE). 7. 175-178. 10.14419/ijet. v7i1.7.10645.

[8] Chaitrali S. Dangare and Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, 2012, 44-48.

[9] Seshadri Dhruv R., Davies Evan V., Harlow Ethan R., Hsu Jeffrey J., Knighton Shanina C., Walker Timothy A., Voos James E., Drummond Colin K. "Wearable Sensors for COVID- 19: A Call to Action to Harness Our Digital Infrastructure for Remote Patient Monitoring and Virtual Assessments", Frontiers in Digital Health,2020,10.3389/fdgth.2020.00008,2673-253X.

[10] Mohammadzadeh, N., Gholamzadeh, M., Saeedi, S. et al. The application of wearable smart sensors for monitoring the vital signs of patients in epidemics: a systematic literature review. J Ambient Intell Human Comput (2020).