



DISIGNING DISEASE PREDICTION MODEL USING MACHINE LEARNING APPROACH

Sinchana C S, Sushmitha H C, Shivaleela S

Student, Student, Assistant Professor

Master Of Computer Application

Dr. Ambedkar Institute of Technology Bangalore, India

Abstract: In the field of medical diagnostics, machine learning is gradually becoming more common. This is primarily due to improvements in disease categorization and recognition systems, which can provide information to enable medical specialists in the early discovery of fatal diseases, leading to a noticeably higher patient survival rate. In this study, we use three alternative classification methods, each with its unique set of benefits, to forecast disease in three separate disease databases that are accessible in the UCI repository (heart, lung cancer, COPD, and diabetes). Backward modeling utilizing each dataset's features were chosen using a p-value test. The study's conclusions are in favor of employing machine learning to identify diseases early. The study 'Preventive Healthcare and Corporate Female Workforce' also said that long hours and working under strict deadlines cause up to 75% of working women to suffer from depression or general anxiety disorder, compared to women with lesser levels of psychological demand at work. Machine learning can be used in the medical industry to diagnose, detect, and forecast a variety of diseases. This paper's major objective is to give clinicians a tool for early heart disease detection. In turn, this will assist in giving patients appropriate care and preventing negative outcomes.

I INTRODUCTION

According to the Centers for Medicare and Medicaid Services, 3.3 trillion dollars were spent on healthcare nationwide in 2016. Half of Americans suffer a number of chronic diseases \$10,348 per person. Early detection of prevalent diseases such as breast cancer, diabetes, coronary artery disease, and others tumors can aid patients to manage and lower their risk of passing away from these diseases. Thanks to developments in machine learning and artificial intelligence, a variety of classifiers and clustering techniques are being used to do this. The approaches taken in this paper's methodology shows how to use machine learning algorithms to forecast diseases such as Diabetes, which is characterized by high blood glucose or blood sugar levels, breast cancer, which is a relatively frequent condition among women, and heart disease, which is the leading cause of death in the United States. The UCI machine learning library [1] provides access to the datasets used to create the predictive models in this study, and they can be downloaded. In India, 10% of adults suffer from hypertension while the country is home to 25-30 million diabetics. Three out of every 1,000 people suffer a stroke. The number of deaths due to heart attack is projected to increase from 1.2 million to 2 million in 2010. The data is cleaned before being imported in CSV format disease dataset. After data managing and attribute selection, machine learning algorithms are used to predict illness. These algorithms include Logistic Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), and Adaptive Boosting. Their accuracy is compared to find the best model for that disease dataset. Python 3.7 was used for all of the analysis and visualization.

II LITERATURE SURVEY

1. Prediction of cardiopathy Using Machine Learning Algorithms

Health care has enormous data of information, we need to process the data using certain techniques, data processing is one in all the techniques often used in machine learning algorithms, cardiopathy is that one of the most leading explanation for death worldwide. Using this machine learning algorithm technique predicts the emerging possibilities of cardiopathy. The results of this methods provide the chances of occurring cardiopathy in terms of percentage. The datasets used are classified in terms of medical field of attributes like input and output data to split using model selection method, this technique valuates those parameters using processing classification technique. The datasets are predict the output using python programming.

2. Prediction of diabetes supported personal life indicators.

Diabetes Mellitus or Diabetes has been described as equal to Cancer and HIV (Human Immunodeficiency Virus). It increases when the persons there are high glucose levels over a protracted period. Recently, it's been noted as a risk for develop Alzheimer, and variety one cause for blindness & nephritis. Prevention of the disease is additionally a important topic for research within the healthcare community people.

3. Research on machine learning for specific disease prediction utilizing the principle of component analysis.

World Health Organization give the report the majority of people died in heart diseases. This will continue in future, we will face approximately 23.6 million people will die from cardiopathy upcoming 2030. The healthcare industry collects a number of data of cardiopathy in some cases we can't solve the exact issue due to confidentiality. During the project, study of Principal Component Analysis has been carried out that identifies lowest numerous characteristics and we can predict using Algorithms for supervised machine learning. The aim of this project is to predict techniques for supervised machine learning to forecast cardiopathy, data processing has been done and split the data with trained data and fit the values and finally predict the output result. Nowadays diabetes is a life threatening disease because small children also affected by this disease if we know machine learning algorithm we can pass the particular symptoms and we can perform prediction.

4. Diseases prediction by using machine learning.

Nowadays day to day life the people facing various diseases that are related to person habitual activities and it is difficult to predict whether the person is affected by initial or final stages. Due to that accurate prediction is not possible for particular disease symptoms it is difficult task without using machine learning algorithm we can't get the right prediction of disease is that the upcoming challenging task face the affected people with particular diseases. Now-a-days science technology has developed infinite number of innovation ideas in that some cases also we can solve the issues in medical growth and healthcare field so we can use machine learning algorithm to predict various stages or final level we can save the person's life.

III EXISTING SYSTEM

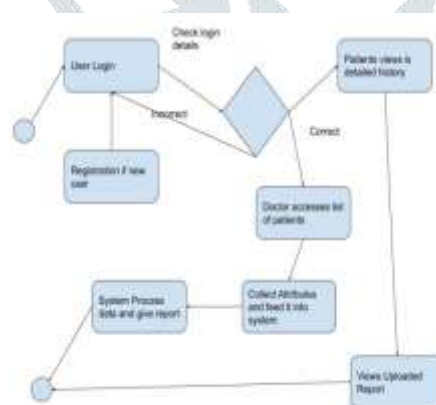
People today experience a wide range of ailments as a result of their lifestyle choices and the environment. As a result, it becomes essential to forecast illness in its early stages. However, based just on symptoms, clinicians find it difficult to make exact predictions. Accurately predicting illness is the hardest barrier to overcome. Data mining is essential in disease prediction to address this issue. Early patient treatment has benefited from accurate analysis of medical information because of the volume of data used in healthcare sector disciplines. Massive volumes of medical data are mined using illness data to find hidden patterns. We suggested a method for predicting sickness in general based on the patient's symptoms.

DISADVANTAGES OF EXISTING SYSTEM

1. It is crucial to know how to make a correct diagnosis of a patient through clinical examination and evaluation. Decision assistance systems that rely on computers may become vital for making compelling choices.
2. The health-care Industry creates a lot of data on clinical assessments, patient reports, diagnoses, follow-up visits, medication, and other topics. It takes a lot of planning to get it right.
3. The quality of the data association has been harmed as a result of poor information management. Increased data volume necessitates a lawful method of concentrating and processing data in a viable and efficient manner.
4. Medical data is growing at an exponential rate, making its use for predicting the correct disease a critical duty. However, processing huge data is extremely challenging, making it extremely difficult to anticipate the patient's sickness.

IV ACTIVITY DIAGRAM

Activity diagrams are related to the actions of the particular module. In this diagram we control the data for to access only for authenticated users.



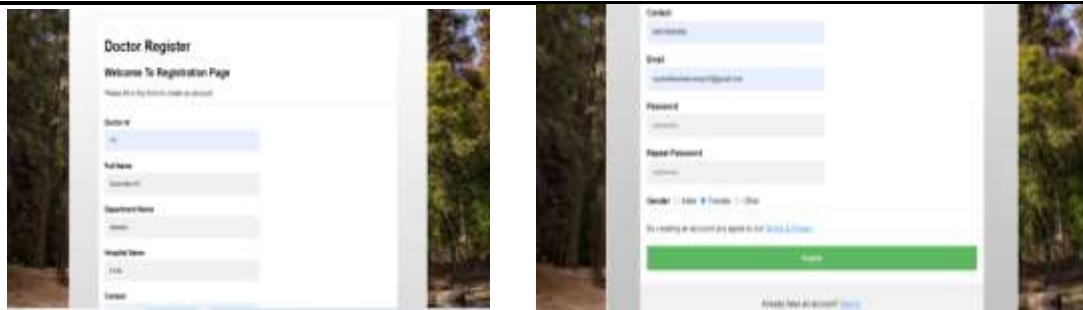
V ER DIAGRAM

An entity-relationship diagram (ERD) is relation between the attributes and behavior of particular module. In this diagram we are going to mention all the type of modules with attribute and behavior functionalities.

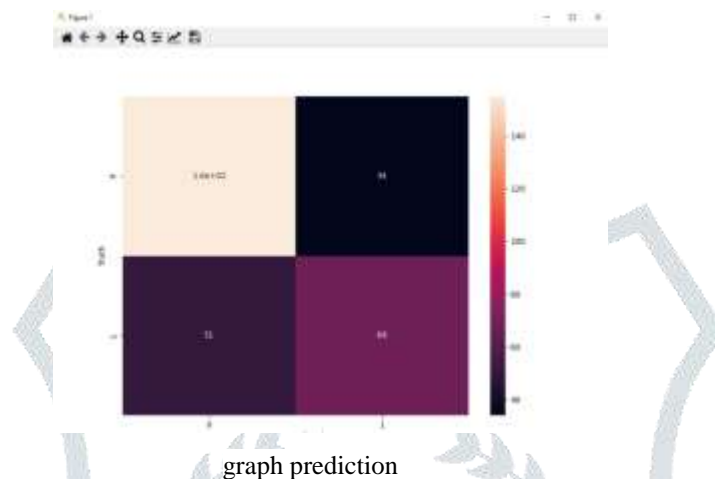


VI RESULT





registration successful



graph prediction

VII CONCLUSION

We accepted that the forecast made earlier did not employ a huge dataset [12] through our in-depth review of literature survey. Better prediction is ensured with a huge dataset. It also lacks a system for making recommendations. In the case of modest indicators of symptoms, we will make a recommendation to the patient on how to control or prevent the disease. The recommendations would be such that if followed, the patient would benefit. Thus, we will develop a system that will predict illness-affected patients using our knowledge base, which consists of a dataset of roughly 2000 patient records, and will also make recommendations based on the proximity of disease stages to the patients. We will use two algorithms to make predictions: Nave Bayes and K-Nearest Neighbor, and we will assess which approach delivers greater accuracy based on their performance criteria. This method, once created, can be utilised in the health-care industry to conduct medical checks on patients who have been harmed.

FUTURE ENHANCEMENTS

The proposed system can be developed in a variety of ways, and there is a lot of room for development. These include:

1. Improving the algorithms' accuracy.
2. Improving the algorithms to improve the system's efficiency and functionality.
3. Improving a few extra qualities to fight diabetes even more successfully.
4. To make it into a comprehensive medical diagnosis system for use in hospitals.

REFERENCES

- [1] Majumdar, P.K., and Sharma, M. 2009. occupational lifestyle diseases: A growing problem. *Indian Journal of Occupational and Environmental Medicine*, 13(3), pp. 109–112.
- [2] Obtainable [Online] at <https://www.dnaforensics.in/dna-test-cost-in-india> (Retrieved on June 27, 2018)
- [3] [3] In 2005, Suzuki, A., Lindor, K., St. Saver, J., Lymp, J., Mendes, F., Muto, A., Okada, T., and Angulo, P. Effects of alterations in lifestyle and body weight in non-alcoholic fatty liver disease. 1060–1066 in *Journal of Hepatology*, 43(6).
- [4] Naive Bayes prediction method for cardiac disease, Pattekari, S.A. and Parveen, A., 2012. 290–294 are included in *International Journal of Advanced Computer and Mathematical Sciences*, 3(3).
- [5] A. Anand and D. Shakti, 2015. based on a person's own lifestyle indicators, diabetes prediction. *First international conference on next-generation computing technologies (NGCT)*, 2015 (pp. 673-676). IEEE.
- [6] B.D. Kanchan and M.M. Kishor, 2016. investigation of machine learning techniques for component analysis-based disease prediction
- [7] *International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 2016 (pp. 5–10). IEEE.
- [8] Golbabaee, S., Soltanian-Zadeh, H., and Kazeminejad, A. 2017. Use rs-fMRI to diagnose Parkinson's disease using theoretical metrics and machine learning. (pp. 134–139) in *Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE.

[9] Milgram, M. Cheriet, and R. Sabourin, 2006. Which is more effective for handwriting recognition using SVMs: "one against one" or "one against all"? *Frontiers in Handwriting Recognition: Tenth International Workshop*, La Baule, France, Suvisoft, 2006.

[10] Hossain, R., S.H. Mahmud, M.A. Hossin, S.R.H. Noori, and H. Jahan, 2018. Using data mining techniques, PRMT predicts the risk factors for obesity in middle-aged people. 1068–1076 in *Procedia Computer Science*, vol. 132.

[11] Dr. Rashmi Phalnikar and Sayali Ambekar, 2018. Machine learning-based disease prediction, *International Journal of Computer Engineering and Applications*, volume 12, pages 1–6.

