INTERNATIONAL JOURNAL OF RESEARCH IN MEDICAL SCIENCES & TECHNOLOGY

e-ISSN:2455-5134; p-ISSN: 2455-9059

Employability of the Machine Learning Algorithms in the Early Diagnosis of Various Diseases

Teesha Ahuja
Bharati College, University of Delhi, New Delhi, India

Paper Received: 27th May 2022; **Paper Accepted:** 20th June 2022;
Paper Published: 29th June 2022

DOI: http://doi.org/10.37648/ijrmst.v13i01.015

How to cite the article:

Teesha Ahuja, Employability of the Machine Learning Algorithms in the Early Diagnosis of Various Diseases, IJRMST, January-June 2022, Vol 13, 158-163, DOI: http://doi.org/10.37648/ijrmst.v13i01.015

ABSTRACT

The most difficult task is accurately predicting disease. Environment and lifestyle factors contribute to a wide range of illnesses. As a result, it becomes a crucial task to predict disease earlier. On the other hand, the doctor finds it too difficult to predict symptoms accurately. Predicting the disease is important in using data mining to solve this issue. Medical science experiences significant annual data growth. Early patient care has benefited from accurate medical data analysis because of the growing amount of data in the medical field. Data mining uncovers hidden pattern information in a wide range of medical data by utilizing disease data. Based on the patient's symptoms, we proposed a general disease prediction. We use the machine learning algorithms K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) for accurate disease prediction. A dataset of disease symptoms was required for disease prediction. A person's lifestyle and checkup information are considered for an accurate prediction in this general disease prediction. CNN has a higher general disease prediction accuracy of 84.5% than the KNN algorithm. Additionally, KNN's memory and time requirements are higher than CNN's. This system can provide the risk associated with the prevalent disease, which can be either a lower or higher risk of the prevalent disease after general disease prediction.

INTRODUCTION

The application of data mining and machine learning techniques to health data, and patient treatment histories have made it possible to predict diseases for decades. Pathological data or medical profiles have been used in numerous works to predict specific diseases using data mining techniques. These methods attempted to anticipate the disease's recurrence. Additionally, some strategies attempt to predict the disease's control and progression. From the perspective of big data analysis, disease prediction has received more attention as big data technology has developed; To improve the accuracy of risk classification, a variety of studies have used an automatic selection of characteristics from a large amount of data rather than manually selecting the characteristics. To improve outcomes for patients, machine learning in healthcare is the primary focus. Correctly identifying and diagnosing a variety of diseases has become easier as a result of machine learning. Using effective multiple machine learning algorithms, predictive analysis aids in better disease prediction and patient treatment. The recent success of deep learning in various areas of machine learning has led to a shift toward machine

learning models that can learn rich, hierarchical representations of raw data with slight pre-processing and produce more accurate results.

The healthcare industry generates a significant amount of daily healthcare data by utilizing a patient's treatment history and health data, which can be utilized to predict potential future diseases. In the future, this buried information in healthcare data will be utilized for efficient healthcare decision-making. Additionally, these areas need to be improved by utilizing useful data in healthcare. Healthcare is one area where machine learning algorithms are used in this way. Medical facilities must be upgraded to make better decisions regarding the diagnosis and treatment options of patients. Humans are aided by machine learning in the healthcare industry in processing vast and intricate medical datasets and converting them into clinical insights. Physicians can then use this to provide additional medical care. When used in healthcare, machine learning, therefore, has the potential to improve patient satisfaction. Using treatment history of patient and health data, the k-the mean algorithm predicts diseases.

EXISTING SYSTEM

When training models to make predictions based on a standard disease risk model, a machine learning and supervised learning algorithm typically use labeled training data. Group test sets are used to classify patients into high-risk and low-risk categories.

Chen et al.'s system for using smart clothing to monitor health over time thoroughly researched heterogeneous systems and came up with the best cost-minimization strategies for tree and simple path cases involving heterogeneous systems. However, these models are extensively researched and only useful in clinical settings.

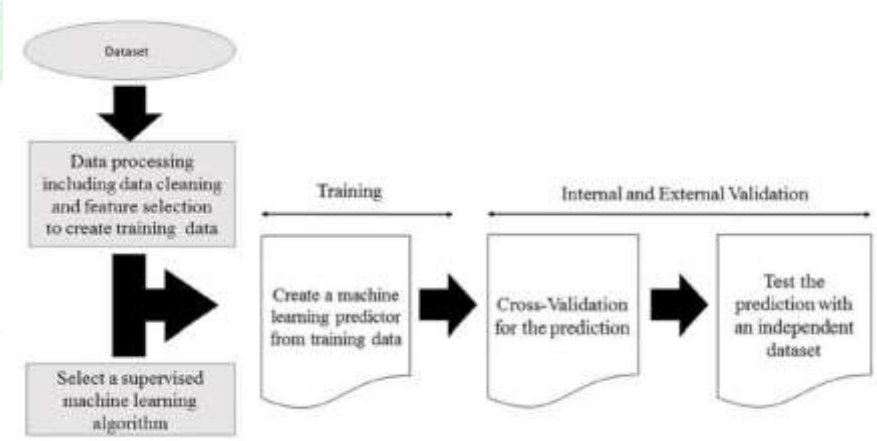
Bates and Co. propose six healthcare-related big data applications. EHR keeps track of patient statistics, test results, and disease history, allowing for the identification of potential data-centric solutions that reduce the cost of medical case studies. The subtypes of diseases are not predicted by the current systems. It is unable to predict people's conditions.

IV's predictions of diseases have needed to be more specific and specific. We have combined structured and unstructured data from healthcare fields to assess disease

risk in this paper. We consult hospital experts to identify useful features for handling structured data. Using the latent factor model to find the missing information in hospital-collected medical records. Additionally, we could identify the most prevalent chronic diseases in a specific community and region by utilizing statistical expertise.

Using the k-mean algorithm, we select the features for unstructured text data automatically. A k mean algorithm for structured and unstructured data is what we propose.

SYSTEM ARCHITECTURE



RESULTS

CONCLUSION

Higher accuracy is possible with the proposed system. We use structured and textual patient data based on the proposed k-mean algorithm. We combine the two data to determine that the accuracy rate can reach 95%. Medical big data analytics does not use either type of data in any of the systems or work currently in place. A K-Mean clustering method that works with both structured and unstructured data is what we suggest. Combining structured and unstructured features yields the disease risk model.

Financial support and sponsorship: Nil

Conflict of Interest: None

REFERENCES

[1] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, (2014), Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients, *Health Affairs*, vol. 33, no. 7, pp. 1123–1131.

[2] K.R.Lakshmi, Y.Nagesh, and M.VeeraKrishna, (2014), Performance comparison of three data mining techniques for predicting kidney disease survivability, *International Journal of Advances in Engineering & Technology*, Mar. 2014.

[3] Mr. Chala Beyene, Prof. Pooja Kamat, (2018), Survey on Prediction and Analysis of the Occurrence of Heart Disease Using Data Mining Techniques, *International Journal of Pure and Applied Mathematics*.

[4] Boshra Brahmi, Mirsaeid Hosseini Shirvani, (2015), Prediction and Diagnosis of Heart Disease by Data Mining Techniques, *Journals of Multidisciplinary Engineering Science and Technology*, vol.2, 2 February 2015, pp.164- 168.

[5] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, (2015), Incorporating temporal HE R data in predictive models for risk stratification of renal function deterioration, *Journal of biomedical informatics*, vol. 53, pp. 220–228.

[6] M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai, and R. Mishra, (2018), A proposed model for lifestyle disease prediction using support vector machine, *9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6.

[7] F. Q. Yuan, (2016), Critical issues of applying machine learning to condition monitoring for failure diagnosis, *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1903–1907.

[8] S. Ismaeel, A. Miri, and D. Chourishi, (2015), Using the extreme learning ma- chine (elm) technique for heart disease diagnosis, *2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015)*, pp. 1–3.

[9] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, (2018), Prediction of heart disease using machine learning, *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1275–1278.

[10] Y. Hasija, N. Garg, and S. Sourav, (2017), Automated detection of dermatological disorders through image-processing and machine learning, *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 1047–1051.

[11] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, (2019), Comparing different supervised machine learning algorithms for disease prediction, *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1– 16.

[12] R. Katarya and P. Srinivas, (2020), Predicting heart disease at early stages using machine learning: A survey, *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 302–305.

[13] P. S. Kohli and S. Arora, (2018), Application of machine learning in disease prediction, *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–4.