IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116134

A Study of Machine-Based Smart Disease Prediction Systems in the Health Care Domain

Kajal Nande[1], Dr. Manoj L. Bangare[2], Ravindra Honaji Borhade[3]

PG Scholar, Department of Information Technology, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune[1]

Associate Professor, Department of Information Technology, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune[2]

Ph.D. Research Scholar, Sinhgad College of Engineering, Department of E&TC, Savitribai Phule Pune University, Pune[3]

Abstract: People presently suffer from several of diseases as a result of their lifestyle and the surroundings. As a result, being able to predict illness at an early stage is critical. Doctors, on the other side, find it difficult to make specific estimates based on symptoms. For accurate disease prediction, existing systems use the KNN and CNN machine learning algorithms. Illness symptom collection is essential for disease prediction. Existing system gives lots of time to execute as well as it does not gives accurate results. Because of this we propose our system by using CNN algorithm for better accuracy and to reduce time execution.

Keywords: CNN, KNN, disease prediction, data processing, machine learning.

I. INTRODUCTION

Artificial intelligence has enhanced computers' intellect and given them the ability to think. Various AI research consider machine learning to be a subfield. It is critical to understand the right diagnosis of patients through clinical examination and evaluation. Computer-assisted decision-making systems may become essential for making persuasive judgments. The health-care sector creates a large amount of data, such as clinical assessments, patient reports, cures, follow-up appointments, medication, and so on. It is tough to effectively arrange.

As a result of inadequate information management, the quality of the data association has suffered. As the volume of data expands, an appropriate approach for concentrating and processing information in a viable and efficient manner is essential. One of the numerous machine learning software is used to develop a classifier that can categorise data based on its attributes. The data collection has been divided into two or more classes. These classifiers are used in medical data analysis and disease prediction.

Disease similarity computation significantly relies on disease-related biological data in addition to illness traits, which has facilitated the development of new bioinformatics technologies. In the majority of previous investigations, disease terminology from the Illness Ontology (DO), the HPO, and the GO was utilised to calculate disease similarity. Healthcare is extremely important in our daily lives. With good therapy, health disorders can be diagonalized and prevented at an early stage. Various curing equipment such as CT, MRI, PET, and others can quickly identify problems present inside our bodies or beneath the skin. S. L. Bangare et al. [11-14] have worked in the health care related projects using machine learning. N. Shelke et al. [15], S. Gupta et al. [16] and G. Awate et al. [17] also showcased their machine learning work. Furthermore, certain uncommon disorders such as heart attack and heart stroke can be easily prevented in their early stages.

II. LITERATURE SURVEY

Sr. no.	Author Name, Year	Outline	Advantages
1	Wenxing et al. [1], IEEE Access/2019,	This study provided a novel deep-learning-based wholly hybrid guidance algorithm that predicts the patient's likely	1) It considers order own circle of relative's participants similarly to low order combination of illness among

DOI: 10.17148/IJARCCE.2022.116134

		condition based purely on the patient's medical records and serves as a connection between patients and professionals.	illness features, 2) Improved comprehensiveness.
2	Dahiwade et. al. [2], IEEE Xplore/2019	Proposed well-known sickness prediction, in which individual living behaviour and checkup information are not overlooked for an acceptable forecast.	1) Minimal time commitment 2) The lowest possible price 3) Ailment prediction accuracy is 84.5 percent.
3	Xu, Z. et. al. [3], IEEE/2019	The detection performances are confirmed through the use of an actual case take a look at primarily based totally on 3 yrs of scientific the Hong Kong Hospital Govt's history	1) Comfortably includes the co morbidity incorporating the community into a Bayesian approach 2) Displays advanced predicting abilities.
4	Repaka, A. N. et. al. [4], IEEE/2019,	This paper centered on heart sickness prognosis with the aid of using thinking about preceding statistics and data. To obtain this SHDP become constructed thru NB on the way to are expecting hazard elements regarding coronary heart sickness.	1) Accuracy is 89.77% despite decreasing the characteristics. 2) In compare to earlier encrypting algorithms, AES' general quality is generally safe.
5	Gao, J. et. al. [5], IEEE/2020,	A technique for predicting disease similarity using node instance learning was proposed.	1) The appropriate chemical data source's problem forecasting.
6	Mathew, et. al. [6], IEEE/2019,	These studies defined a healthcare chatbot which can take the position of doctor’s conventional sickness diagnostic and remedy inspiration methods. A chatbot can carry out the position of a doctor.	1) This approach aids in the reduction of daily examinations. 2) It recognizes the signs and symptoms and provides an accurate diagnosis. 3) The use of a chatbot does not necessitate the assistance of a physician. 4) Less expensive
7	Maurya, A. et.al. [7], IEEE/2019,	The method is proposed for CKD patients with ML system to automate the categorization of chronic kidney disease into severity-based stages.	1) Detects and recommends diets that will benefit both doctors and patients.
8	Yi Zhang et. al. [8], IEEE/2019,	The researchers investigated a new two-stage prediction model (DRW-BNSP).	1) Increase disease associations. 2) Better understand the pathogenesis of complex diseases

DOI: 10.17148/IJARCCE.2022.116134

9	Pandey, H., &Prabha, S. [9], IEEE/2020,	This paper describes the IoT, which employs a pulse charge sensor and an Arduino to file real-time affected person data, that's eventually recorded the use of Thing Speak.	1) The proposed approach aids patients in early detection of cardiac disease. 2) It will be important as a way of mass screening in communities without hospital services.
10	Vijiya Kumar, et. al. [10], IEEE/2020,	Using the Random Forest approach, this study proposed a mechanism for more accurate early diabetes prediction for a patient.	1) When compared to other algorithms, the accuracy level is higher. 2) The system is capable of accurately, quickly, and accurately forecasting diabetic illness.

III. EXISTING SYSTEM

In this existing system author used KNN and CNN algorithms for the disease prediction system. However, this system takes a long time to execute and is inaccurate. Because of this we proposed our system by using CNN algorithm for gives the accurate results and system executes in time consuming process.

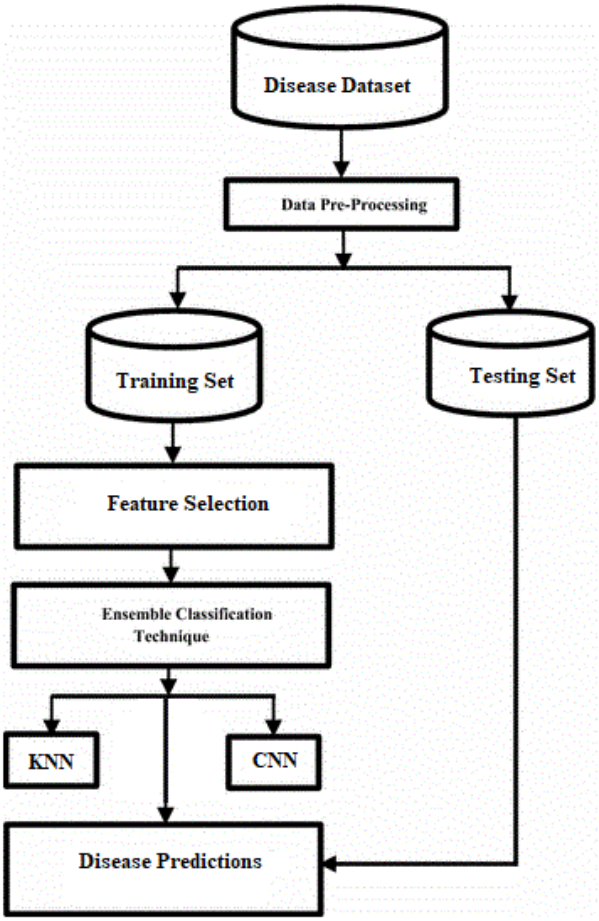

Fig 1Architecture of existing system

Initially, the system made use of data from the UCI machine learning website in the form of a disease list with symptoms. Following that, the dataset is cleaned, which includes removing commas, punctuation, and white spaces. This is the training dataset. After that, the feature was retrieved and chosen. They then categorise the data using techniques such as KNN and CNN. Using machine learning, we can reliably anticipate illness. To address all of these issues, we might recommend that our system use the CNN algorithm, which provides improved accuracy and saves time.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022
DOI: 10.17148/IJARCCE.2022.116134

IV. ALGORITHM

KNN
Common Distance Metrics:
Euclidean distance (continuous distribution):
$d(a,b) = \sqrt{\Sigma}(ai - bi)2$
Hamming distance (overlap metric):
bat (distance = 1)
toned (distance = 3)

Discrete Metric (boolean metric):
if x = y then d(x,y) = 0. Otherwise, d(x,y) = 1
Find the class from the list of KNN
Take the vote of the people of the KNN on different classifiers.
Weighted factor:
w =1/d (generalized linear interpolation) or 1/d2

V. CONCLUSION

We can see from this review that the present system identifies disease using the KNN algorithm. As the execution time increases, the accuracy of the present system decreases. To solve this issue, we developed a solution that employs the CNN algorithm to increase accuracy while simultaneously accelerating a time-consuming procedure.

REFERENCES

[1] Wenxing Hong, ZiangXiong, Nannan Zheng, Yang Weng, "A Medical-History-Based Potential Disease Prediction Algorithm", A Medical-History-Based Potential Disease Prediction Algorithm IEEE Access VOLUME 7, 2019, doi 10.1109/ACCESS.2019.2940644

[2] Dahiwade, D., Patle, G., &Meshram, E. (2019). Designing Disease Prediction Model Using Machine Learning Approach. 2019 Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xploredoi:10.1109/iccmc.2019.8819782

[3] Xu, Z., Zhang, J., Zhang, Q., & Yip, P. S. F. (2019). Explainable Learning for Disease Risk Prediction Based on Co morbidity Networks. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). doi:10.1109/smc.2019.8914644

[4] Repaka, A. N., Ravikanti, S. D., & Franklin, R. G. (2019). Design and Implementing Heart Disease Prediction Using Naives Bayesian. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862604

[5] Gao, J., Tian, L., Wang, J., Chen, Y., Song, B., & Hu, X. (2020). Similar Disease Prediction with Heterogeneous Disease Information Networks. IEEE Transactions on NanoBioscience, 1–1. doi:10.1109/tnb.2020.2994983

[6] Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019). Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862707

[7] Maurya, A., Wable, R., Shinde, R., John, S., Jadhav, R., &Dakshayani, R. (2019). Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning. 2019 International Conference on Nascent Technologies in Engineering (ICNTE). doi:10.1109/icnte44896.2019.8946029

[8] Y. Zhang, M. Chen, X. Xie, X. Shen and Y. Wang, "Two-Stage Inference for LncRNA-Disease Associations Based on Diverse Heterogeneous Information Sources," in IEEE Access, vol. 9, pp. 16103-16113, 2021, doi: 10.1109/ACCESS.2021.3053030.

[9] Pandey, H., &Prabha, S. (2020). Smart Health Monitoring System using IOT and Machine Learning Techniques. 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII).

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116134

doi:10.1109/icbsii49132.2020.9167660

[10] Vijiya Kumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S. (2019). Random Forest Algorithm for the Prediction of Diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). doi:10.1109/icscan.2019.8878802.

[11] S. L. Bangare, "Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images", Neuroscience Informatics, Volume 2, Issue 3, September 2022, 100019, ISSN 2772-5286, https://doi.org/10.1016/j.neuri.2021.100019 .

[12] S. L. Bangare, G. Pradeepini, S. T. Patil, "Implementation for brain tumor detection and three dimensional visualization model development for reconstruction", ARPN Journal of Engineering and Applied Sciences (ARPN JEAS), Vol.13, Issue.2, ISSN 1819-6608, pp.467-473. 20/1/2018 http://www.arpnjournals.org/jeas/research_papers/rp_2018/jeas_0118_6691.pdf

[13] S. L. Bangare, G. Pradeepini, S. T. Patil, "Regenerative pixel mode and tumor locus algorithm development for brain tumor analysis: a new computational technique for precise medical imaging", International Journal of Biomedical Engineering and Technology, Inderscience, 2018, Vol.27 No.1/2. https://www.inderscienceonline.com/doi/pdf/10.1504/IJBET.2018.093087

[14] S. L. Bangare, G. Pradeepini, S. T. Patil et al, "Neuroendoscopy Adapter Module Development for Better Brain Tumor Image Visualization", International Journal of Electrical and Computer Engineering (IJECE) Vol. 7, No. 6, December 2017, pp. 3643~3654. http://ijece.iaescore.com/index.php/IJECE/article/view/8733/7392

[15] N. Shelke, S. Chaudhury, S. Chakrabarti, et al. "An efficient way of text-based emotion analysis from social media using LRA-DNN", Neuroscience Informatics, Volume 2, Issue 3, September 2022, 100048, ISSN 2772-5286, https://doi.org/10.1016/j.neuri.2022.100048 .

[16] Suneet Gupta, Sumit Kumar, Shibili Nuhmani, Arnold C. Alguno, Issah Abubakari Samori et. al., "Homogeneous Decision Community Extraction Based on End-User Mental Behavior on Social Media", Computational Intelligence and Neuroscience, vol. 2022, Article ID 3490860, 9 pages, 2022. https://doi.org/10.1155/2022/3490860.

[17] Gururaj Awate, G. Pradeepini and S. T. Patil et al., "Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow",arXiv, https://doi.org/10.48550/arXiv.1806.10170