ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

BIO-INSPIRED ALGORITHMS FOR HEART DISEASE PREDICTION

K.UmaRani[1], Om Pattewar[2], Lakshman Kumar[3], L.Anjaneyulu Naik[4], M.Kaveri[5]

Assistant professor, Department of Computer Science and Engineering[1]

Student, Department of Computer Science and Engineering[2,3,4,5]

Sree Dattha Group of Institutions, Sheriguda, Telangana. [1,2,3,4,5]

ABSTRACT

Cardiovascular Diseases (CVDs) have emerged as a leading global cause of mortality, presenting a significant public health challenge worldwide, including in India. Timely and accurate diagnosis is crucial for effective treatment. Machine Learning (ML) algorithms have proven highly effective in automating the analysis of complex medical datasets, aiding healthcare professionals in diagnosing heart-related conditions. This paper provides a comprehensive survey of ML models used for this purpose, evaluating their performance critically. Specifically, supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), and ensemble models are reviewed in depth.

Index Terms: Cardiovascular Diseases, Machine Learning, Supervised Learning Algorithms, Diagnosis, Healthcare Technology.

I.INTRODUCTION

Chronic diseases constitute a significant healthcare challenge globally, with implications for both health outcomes and economic expenditures. In the United States, chronic diseases affect 50% of the population and consume 80% of healthcare spending, totaling approximately 2.7 trillion USD annually, or 18% of the GDP. Similarly, in China, chronic diseases are the primary cause of death, accounting for 86.6% of all deaths in 2015 according to a national report on nutrition and chronic diseases. The collection of electronic health records (EHRs) has become increasingly feasible with the growth of medical data, facilitating the identification of data-driven solutions to reduce healthcare costs.

Recent advancements include bio-inspired heterogeneous vehicular telematics paradigms for real-time collection of mobile users' health data using advanced vehicular networks [4]. Additionally, innovative healthcare systems like smart clothing for continuous health monitoring have been proposed by Chen et al. [5]. These systems record patients' statistical information, test results, and medical histories in EHRs, offering potential cost-saving solutions for medical case studies. Wang et al. [6] introduced an efficient flow estimation algorithm for telehealth cloud systems and designed data coherence protocols for Personal Health Record (PHR)-based distributed systems. Bates et al. [7] identified six key applications of big data in healthcare, emphasizing the role of predictive analytics in managing high-risk patients and reducing medical costs.

Predictive healthcare applications traditionally employ machine learning algorithms, such as logistic regression and supervised learning techniques, to classify patients into high-risk and low-risk groups based on labeled training data [8]. However, these methods often rely on manually selected features that may not adapt well to evolving disease characteristics.

With the advent of big data analytics, there is a growing interest in automatically selecting features from large datasets to improve risk classification accuracy [9], [10]. While structured data have been extensively studied, there is increasing attention towards handling unstructured data, such as using Convolutional Neural Networks (CNNs) for automatic extraction of text features [11], [12]. Notably, there is a lack of research applying CNNs to Chinese medical text data, highlighting a gap in the current literature.

Challenges remain in addressing missing data and determining the primary chronic diseases and their regional characteristics. To address these challenges, this paper proposes a novel approach combining structured and unstructured data in healthcare. Firstly, a latent factor model is used to reconstruct missing data from medical records obtained from a hospital in central China. Secondly, statistical methods are employed to identify major chronic diseases prevalent in the region. Thirdly, hospital experts are consulted to extract relevant features from structured data, while unstructured text data are automatically processed using CNN algorithms. Finally, we introduce the CNN-based multimodal disease risk prediction (CNN-MDRP) algorithm, integrating structured and unstructured features to develop an enhanced disease risk model. Experimental results demonstrate that CNN-MDRP outperforms existing methods in disease risk prediction.

This introduction sets the stage for understanding the significance of chronic diseases, the role of big data analytics in healthcare, and outlines the approach proposed in this paper to address current challenges in disease risk assessment and prediction.

II. LITERATURE SURVEY

- P. B. Jensen, L. J. Jensen, and S. Brunak proposed that mining electronic health records (EHRs) has become increasingly important for improving research applications and clinical care. This review discusses methods and challenges in leveraging EHR data for research and clinical purposes, highlighting opportunities for advancing healthcare through data mining and analysis. It examines trends in EHR utilization, including data privacy concerns and integration with emerging technologies such as machine learning and artificial intelligence (AI), to enhance patient care outcomes and healthcare system efficiencies.

- D. Tian, J. Zhou, and Y. Wang proposed a dynamic and self-adaptive network selection method tailored for multimode communications in heterogeneous vehicular telematics environments. The method aims to enhance communication reliability and efficiency by dynamically selecting optimal network modes based on real-time vehicle and network conditions. It

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

introduces a novel algorithm that integrates vehicle mobility patterns and network performance metrics, enabling seamless transitions and improving overall system robustness in diverse telematics scenarios.

- M. Chen, Y. Ma, and J. Song proposed that smart clothing integrates human physiology monitoring with cloud computing and big data analytics, enabling continuous and sustainable health monitoring. This paper explores the design, implementation, and benefits of smart clothing systems in enhancing healthcare delivery and personal health management. It discusses sensor integration technologies, data transmission protocols, and cloud-based analytics platforms that support real-time health monitoring and personalized healthcare interventions, emphasizing the potential for smart clothing to revolutionize preventive healthcare practices.

- J. Wang, M. Qiu, and B. Guo present a framework for enabling real-time information services in telehealth systems leveraging cloud-based big data platforms. The framework aims to improve the efficiency and effectiveness of telehealth services by integrating real-time data analytics and information-sharing capabilities over scalable cloud infrastructures. It discusses the architecture's scalability, security measures, and data processing techniques to support timely medical decision-making and patient management, illustrating how cloud-based big data platforms can enhance telehealth system performance and accessibility.
- D. W. Bates, S. Saria, and L. Ohno-Machado explore the application of big data analytics in healthcare to identify and manage high-risk and high-cost patients. They discuss methodologies, challenges, and potential benefits of using advanced analytics to improve patient outcomes and optimize healthcare resource allocation. The paper highlights case studies and outcomes from healthcare organizations implementing big data analytics, emphasizing the role of predictive modeling, data-driven interventions, and population health management strategies in enhancing care coordination and reducing healthcare costs.

III.PREVIOUS WORK

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality [1]. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15,2017. Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India have lost up to $237 billion, from 2005-2015, due to heart related or Cardiovascular diseases . Thus, feasible and accurate prediction of heart related diseases is very important.

Medical organisations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately.

IV.PROPOSED MODEL

In this project student want to detect heart disease from dataset using Bio Inspired 4 features optimizing algorithms such as Genetic Algorithm, Bat, Bee and ACO. Here ACO algorithm is design in python to solve Travelling Salesman Problem to find shortest path and it cannot be implemented with heart disease dataset, so I am implementing 3 algorithms called Genetic, Bat and Bee.

Bio inspired algorithms design to optimized features used in dataset for training classification algorithms to increase prediction accuracy, sometime some datasets may have irrelevant values inside dataset and those irrelevant attributes or values may degrade classification accuracy so using optimize algorithms we can reduce features (attribute values) from dataset. This optimize algorithms will be applied on dataset to check whether all values are related to dataset or not, if any attribute found unrelated then it will removed from dataset.

V.SYSTEM ARCHITECTURE DIAGRAM

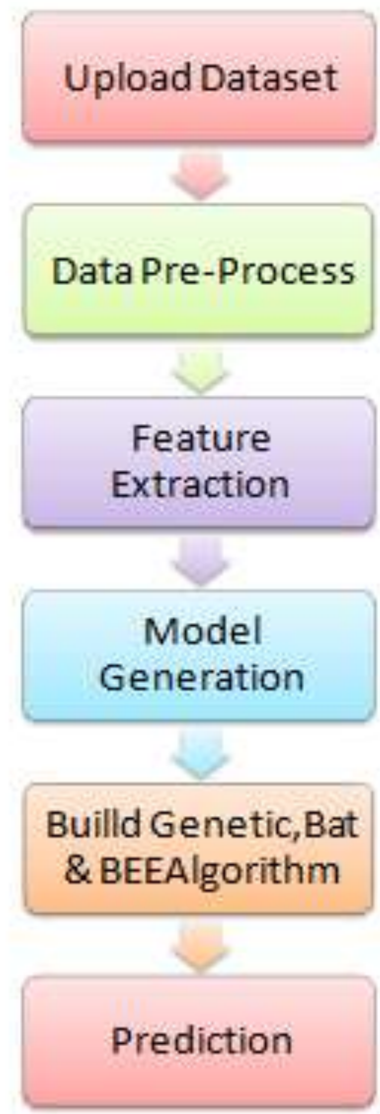


Figure.1 System Architecture

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

VI.IMPLEMANTATION:

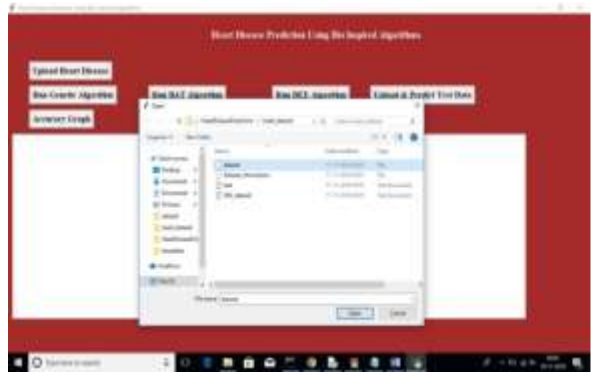
- **Upload Heart Disease Data:** Upload the dataset containing information about heart disease.
- **Run Genetic Algorithm:** Apply a Genetic Algorithm (GA) to the data to perform feature selection or optimization.
- **Run BAT Algorithm:** Use the BAT Algorithm to further analyze the data or improve prediction models.
- **Run BEE Algorithm:** Implement the Bee Algorithm for another layer of data processing or model enhancement.
- **Upload & Predict Test Data:** Upload a test dataset and make predictions using the trained model.
- **Accuracy Graph:** Generate and display a graph showing the accuracy of the different algorithms.

Exit: Close Application

VII.RESULTS

To run this project double click on 'run.bat' file to get below screen

In above screen click on 'Upload Heart Disease' button and upload heart disease dataset. See below screen



In above screen uploading dataset file, after uploading will get below screen

Now click on 'Run Genetic Algorithm' button to run genetic algorithm on dataset and to get its accuracy details. While running this algorithm u can see black console to see feature selection process, while running it will open empty windows, u just close all those empty windows except current window

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

In above screen for GA accuracy, precision and recall we got 100% result. Now click on 'Run Bat' algorithm button to get its accuracy

In above screen for BAT we got 45% accuracy, now click on 'Run BEE Algorithm' button to get BEE accuracy

In above screen for BEE we got 38% accuracy, now click on 'Upload & Predict Test Data' button to upload test data and to predict it class

In above screen I am uploading test file which contains test data without class label, after uploading test data will get below screen

In above screen application has predicted disease stages. Now click on 'Accuracy Graph' button to view accuracy of all algorithms in graph format

In above graph x-axis represents Algorithm Name and y-axis represents accuracy of those algorithms

VIII.CONCLUSION

In this paper, we introduce a novel convolutional neural network-based multimodal disease risk prediction (CNN-MDRP) algorithm that leverages both structured and unstructured data from hospital sources. To the best of our knowledge, no existing work has addressed the integration of these two data types in the realm of medical big data analytics. Our proposed algorithm achieves a prediction accuracy of 94.8%, outperforming several

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

standard prediction algorithms. Additionally, it demonstrates a faster convergence speed compared to the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

IX.FUTURE ENHANCEMENT

Moving forward, further research can explore enhancing the CNN-MDRP algorithm by integrating more advanced deep learning architectures to improve prediction accuracy and robustness across diverse medical datasets. Additionally, incorporating real-time data streaming capabilities could enable continuous monitoring and immediate intervention for high-risk patients. Exploring the application of reinforcement learning techniques could optimize decision-making processes within the healthcare system, enhancing personalized treatment strategies. Finally, expanding the CNN-MDRP framework to encompass a broader range of medical conditions beyond cardiovascular diseases could extend its utility in predictive healthcare analytics.

X.REFERENCES

[1] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, The‘Big Data’Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System Reform Business Technology Office, 2016.

[2] M. Chen, S. Mao, and Y. Liu, ‘‘Big data: A survey,’’ Mobile Netw. Appl., vol. 19, no. 2, pp. 171–209, Apr. 2014.

[3] P. B. Jensen, L. J. Jensen, and S. Brunak, ‘‘Mining electronic health records: Towards better research applications and clinical care,’’ Nature Rev. Genet., vol. 13, no. 6, pp. 395–405, 2012.

[4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, ‘‘A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics,’’ IEEE Trans. Intell. Transp. Syst., vol. 16, no. 6, pp. 3033–3049, Dec. 2015.

[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ‘‘Wearable 2.0: Enable human-cloud integration in next generation healthcare system,’’ IEEE Commun., vol. 55, no. 1, pp. 54–61, Jan. 2017.

[6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, ‘‘Smart clothing: Connecting human with clouds and big data for sustainable health monitoring,’’ ACM/Springer Mobile Netw. Appl., vol. 21, no. 5, pp. 825–845, 2016.

[7] M. Chen, P. Zhou, and G. Fortino, ‘‘Emotion communication system,’’ IEEE Access, vol. 5, pp. 326–337, 2017, doi: 10.1109/ACCESS.2016.2641480.

[8] M. Qiu and E. H.-M. Sha, ‘‘Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems,’’ ACM Trans. Design Autom. Electron. Syst., vol. 14, no. 2, p. 25, 2009.

[9] J. Wang, M. Qiu, and B. Guo, ‘‘Enabling real-time information service on telehealth system over cloud-based big data platform,’’ J. Syst. Archit., vol. 72, pp. 69–79, Jan. 2017.

[10] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, ‘‘Big data in health care: Using analytics to

identify and manage high-risk and high-cost patients,'' Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014.

[11] L. Qiu, K. Gai, and M. Qiu, ''Optimal big data sharing approach for telehealth in cloud computing,'' in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2016, pp. 184–189.

[12] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, ''HealthCPS: Healthcare cyber-physical system assisted by cloud and big data,'' IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

[13] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, ''Localization based on social big data analysis in the vehicular networks,'' IEEE Trans. Ind. Informat., to be published, doi: 10.1109/TII.2016.2641467.

[14] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, ''Enhanced fingerprinting and trajectory prediction for iot localization in smart buildings,'' IEEE Trans. Autom. Sci. Eng., vol. 13, no. 3, pp. 1294–1307, Jul. 2016.

[15] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, ''Risk factors and risk assessment tools for falls in hospital in-patients: A systematic review,'' Age Ageing, vol. 33, no. 2, pp. 122–130, 2004.

[16] S. Marcoon, A. M. Chang, B. Lee, R. Salhi, and J. E. Hollander, ''Heart score to further risk stratify patients with low TIMI scores,'' Critical Pathways Cardiol., vol. 12, no. 1, pp. 1–5, 2013.

[17] S. Bandyopadhyay et al., ''Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data,'' Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1033–1069, 2015. [18] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, ''A relative similarity based method for interactive patient risk prediction,'' Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[19] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, ''Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration,'' J. Biomed. Inform., vol. 53, pp. 220–228, Feb. 2015.

[20] J. Wan et al., ''A manufacturing big data solution for active preventive maintenance,'' IEEE Trans. Ind. Informat., to be published, doi: 10.1109/TII.2017.2670505.