International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

HEART DISEASE PREDICTION USING BIO INSPIRED ALGORITHMS

Himabindu Manchiraju (MCA Scholar), B V Raju College, Vishnupur, Bhimavaram, West Godavari District, Andhra Pradesh, India, 534202.

Y. Srinivasa Raju, B V Raju College, Vishnupur, Bhimavaram, West Godavari District, Andhra Pradesh, India, 534202.

Abstract:

Heart related diseases or cardiovascular diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques andanalyze their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers.

1. INTRODUCTION

According to a report by McKinsey [1], 50% of Americans have one or more chronic diseases, and 80% of American medical care fee is spent on chronic disease treatment. With the improvement of living standards, the incidence of chronic disease is increasing. The United States has spent an average of 2.7 trillion USD annually on chronic disease treatment. This amount comprises 18% of the entire annual GDP of the United States. The healthcare problem of chronic diseases is also very important in many other countries. In China, chronic diseases are the main cause of death, according to a Chinese report on nutrition and chronic diseases in 2015, 86.6% of deaths are caused by chronic diseases. Therefore, it is essential to perform risk assessments for chronic dis- eases. With the growth in medical data [2], collecting electronic health records (EHR) is increasingly convenient [3]. Besides, [4] rst presented a bio-inspired high-performance heterogeneous vehicular telematics paradigm, such that the collection of mobile users' health-related real-time big data can be achieved with the deployment of advanced hetero- geneous vehicular networks. Chen et al.proposed a healthcare system using smart clothing for sustainable health monitoring. Qiu et al. [8] had thoroughly studied the het- erogeneous systems and achieved the best results for cost minimization on tree and simple path cases for heteroge- neous systems. Patients' statistical information, test results and disease history are recorded in the EHR, enabling us to identify potential data-centric

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

solutions to reduce the costs of medical case studies.Wang et al. [9] proposed an efcient ow estimating algorithm for the telehealth cloud system and designed a data coherence protocol for the PHR(Personal Health Record)-based distributed system. Bates et al. [10] proposed six applications of big data in the eld of health- care. Qiu et al. [11] proposed an optimal big data sharing algorithm to handle the complicate data set in telehealth with cloud techniques. One of the applications is to identify high-risk patients which can be utilized to reduce medical cost since high-risk patients often require expensive healthcare. Moreover, in therst paper proposing health- care cyber-physical system [12], it innovatively brought for- ward the concept of prediction-based healthcare applications, including health risk assessment. Prediction using traditional disease risk models usually involves a machine learning algorithm (e.g., logistic regression and regression analysis, etc.), and especially a supervised learning algorithm by the use of training data with labels to train the model [13], [14]. In the test set, patients can be classified into groups of either high-risk or low-risk. These models are valuable in clinical situations and are widely studied [15], [16]. However, these schemes have the following characteristics and defects. The data set is typically small, for patients and diseases with specific conditions [17], the characteristics are selected through experience. However, these pre-selected characteristics maybe not satisfy the changes in the disease and its influencing factors.

With the development of big data analytics technology, more attention has been paid to disease prediction from the perspective of big data analysis, various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk clas- sication [18], [19], rather than the previously selected characteristics. However, those existing work mostly considered structured data. For unstructured data, for example, using convolutional neural network (CNN) to extract text characteristics automatically has already attracted wide attention and also achieved very good results [20], [21] . However, to the best of our knowledge, none of previous work handle Chinese medical text data by CNN. Furthermore, there is a large difference between diseases in different regions, primarily because of the diverse climate and living habits in the region. Thus, risk classification based on big data analysis, the fol- lowing challenges remain: How should the missing data be addressed? How should the main chronic diseases in a certain region and the main characteristics of the disease in the region be determined? How can big data analysis technology be used to analyze the disease and create a better model?

To solve these problems, we combine the structured and unstructured data in healthcare eld to assess the risk of disease. First, we used latent factor model to reconstruct the missing data from the medical records collected from a hospital in central China. Second, by using statistical

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362
IJARST

knowledge, we could determine the major chronic diseases in the region. Third, to handle structured data, we consult with hospital experts to extract useful features. For unstruc- tured text data, we select the features automatically using CNN algorithm. Finally, we propose a novel CNN-based multimodal disease risk prediction (CNN-MDRP) algorithm for structured and unstructured data. The disease risk model is obtained by the combination of structured and unstructured features. Through the experiment, we draw a conclusion that the performance of CNN-MDPR is better than other existing methods.

2. SYSTEM DESIGN

UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

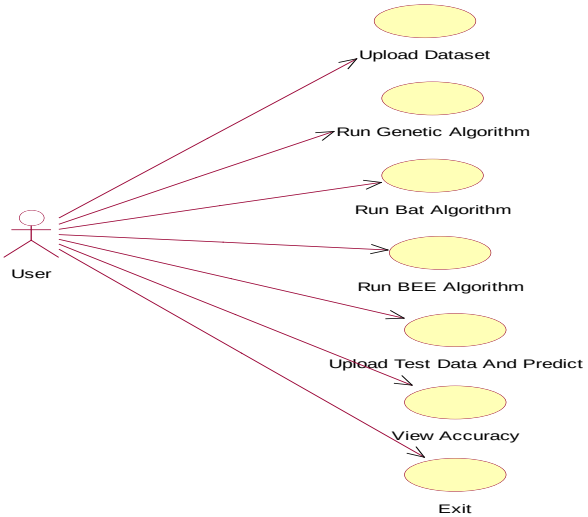
The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

USE CASE DIAGRAM:

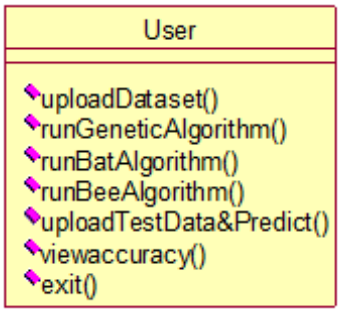
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



CLASS DIAGRAM:

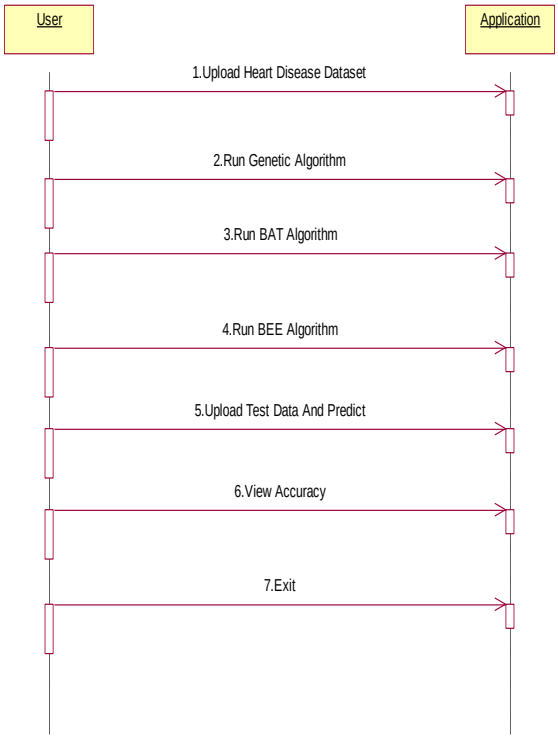
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information

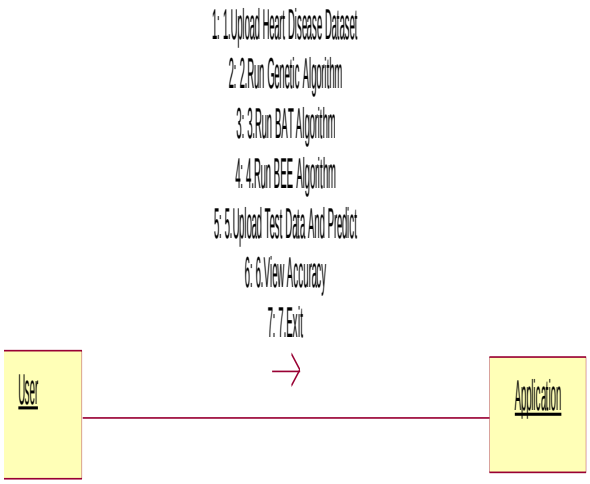


SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



Collaboration Diagram:



3. TEST RESULT

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

interact with this. The future holds a lot to offer to the development and refinement of this project.

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.

To implement this algorithms I am using Heart disease dataset which contains 14 attributes and 4 class labels where 0 refers to No heart Disease and 1 refers to stage1 disease and 2 and 3 refers stage 3 and 4 disease.

Below are some values from dataset to train algorithms

age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,class
63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.0,0.0,2.3,3.0,0.0,6.0,0
67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.0,1.0,1.5,2.0,3.0,3.0,2
67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.0,1.0,2.6,2.0,2.0,7.0,1
37.0,1.0,3.0,130.0,250.0,0.0,0.0,187.0,0.0,3.5,3.0,0.0,3.0,0

First records contains dataset column names and remaining records are the values of dataset. In last column we have class values as 0, 2, 1 and 3 as disease stage.

Test dataset also contains record values but it will not have class labels and application will apply that test values on train dataset to predict it class labels. Some values from test dataset.

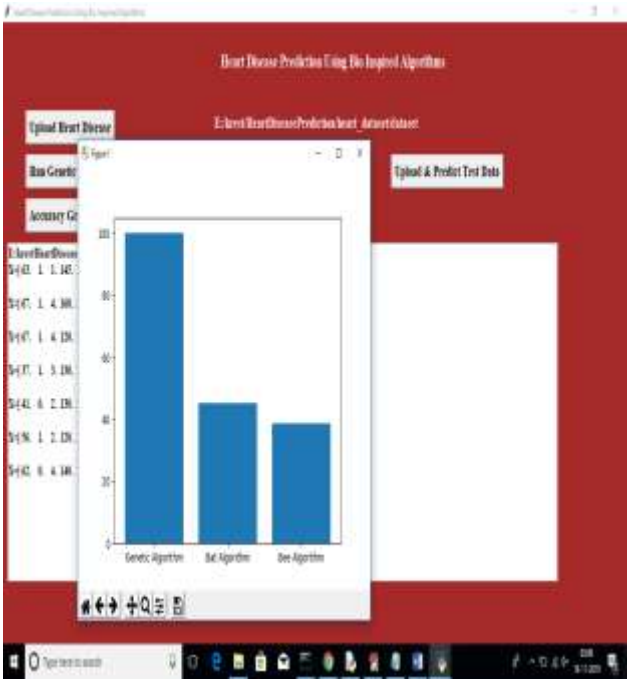
age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal
63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.0,0.0,2.3,3.0,0.0,6.0
67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.0,1.0,1.5,2.0,3.0,3.0
67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.0,1.0,2.6,2.0,2.0,7.0

In above test dataset we can see there is no class name and application will predict it.

All this files are available inside 'heart_dataset' folder.

4. OUTPUT RESULT

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362



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

5. CONCLUSION

In this paper, we propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

6. 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 1691–1700.

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.

[21] W. Yin and H. Schutze, ''Convolutional neural network for paraphrase identification,'' in Proc. HLT-NAACL, 2015, pp. 901–911.

[22] N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, ''Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2015, pp. 855–864.

[23] S. Zhai, K.-H. Chang, R. Zhang, and Z. M. Zhang, ''Deepintent: Learning attentions for online advertising with recurrent neural networks,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1295–1304.

[24] K. Hwang and M. Chen, Big Data Analytics for Cloud/IoT and Cognitive Computing. Hoboken, NJ, USA: Wiley, 2017.

[25] H. Chen, R. H. Chiang, and V. C. Storey, ''Business intelligence and analytics: From big data to big impact,'' MIS Quart., vol. 36, no. 4, pp. 1165–1188, 2012.

[26] S. Basu Roy et al., ''Dynamic hierarchical classification for patient riskof-readmission,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2015, pp.