PREDICT THE DIFFERENT DISEASE USING KNN AND CNN

P. Nagaraj[1],Dasugari Tharun Goud[2],Gundala Anirudh Goud[2], Mandaloju Shravya[2], Pisati Hima Reddy[2]

[1]Associate Professor, Department of CSE, Sreyas Institute of Engineering and Technology,

[2]B.Tech Students, Department of CSE, Sreyas Institute of Engineering and Technology,

ABSTRACT - The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Medical science has large amount of data growth per year. Due to increase amount of data growth in medical andhealthcare field the accurate analysis on medical data which has been benefits from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge amount of medical data. We proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) machine learning algorithm for accurate prediction of disease. For disease prediction required disease symptoms dataset. In this general disease prediction, the living habits of person and checkup information consider for the accurate prediction. The accuracy of general disease prediction by using CNN is 84.5% which is more than KNN algorithm. And the time and the memory requirement is also more in KNN than CNN. After general disease prediction, this system able to gives the risk associated with general disease which is lower risk of general disease or higher.

Keywords *– Machine Learning, Data Analysis, CNN, KNN, Disease Prediction.*

1. INTRODUCTION

Nowadays, a promising trend in healthcare is to move routine medical checks and other health care services from hospital to the home environment. With those patients gets health care more easily especially in case of emergencies. Moreover, hospitals can reduce their burden by shifting the possible and easy tasks to the home environment. One major advantage is in reduction of expenditure. Patients could avoid the fees charged by hospital each time they went to visit doctor. Therefore, it is urgent that in the near future a trending technology need to be implemented in the health industry to develop advanced health care techniques and technologies and use them for the easy monitoring of patients from anywhere else. Patient monitoring include checking the physical conditions of the patient and their medication details. If the right medicines are taken at right time there are less chances that the condition of a patient getting worse. So poor medication adherence is a major problem for the population and medicine providers.

2. LITERATURE SURVEY

We streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real- life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction. We propose a new convolutional neural network (CNN)-based multimodal disease risk prediction 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 with 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 algorithm [1]. The interactive patient risk prediction method, which actively queries medical experts with the *relative similarity* of patients. We explore our method on both benchmark and real clinic datasets, and make several interesting discoveries including that querying relative similarities is effective in-patient risk prediction, and sometimes can even yield better prediction accuracy than asking for absolute questions [2]. This article we propose a Wearable 2.0 healthcare system to improve QoE and QoS of the next generation healthcare system. In the proposed system, washable smart clothing, which consists of sensors, electrodes, and wires, is the critical component to collect users' physiological data and receive the analysis results of users' health and emotional status provided by cloud-based machine intelligence [3]. A cyber-physical system for patient-centric healthcare applications and services, called Health-CPS, built on cloud and big data analytics technologies. This system consists of a data collection layer with a unified standard, a data management layer for distributed storage and parallel computing, and a data-oriented service layer. The results of this study show that the technologies of cloud and big data can be used to enhance the performance of the healthcare system so that humans can then enjoy various smart healthcare applications and services [4]. We focus on the problem of data sharing obstacles in cloud computing and propose an approach that uses dynamic programming to produce optimal solutions to data sharing mechanisms. The proposed approach is called Optimal Tele health Data Sharing Model (OTDSM), which considers transmission probabilities, maximizing network capacities, and timing constraints. Our experimental results have proved the flexibility and adoptability of the proposed method [5].

3. EXISTING SYSTEM

A CNN-MDRP algorithm for a disease prediction from a large volume of hospital's structured and unstructured data. Using a machine learning algorithm (Neavi- Bayes) Existing algorithm CNNUDRP only uses a structured data but in CNN-MDRP focus on both structured and unstructured data the accuracyof disease prediction is more and fast as compared to the CNNUDRP. Here they consider big data. Medical science has large amount of data growth per year. Due to increase amount of data growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge amount of medical data.

4. LIMITATIONS

- Low accuracy
- Insufficient Data Set
- Over fitted or under fitted model

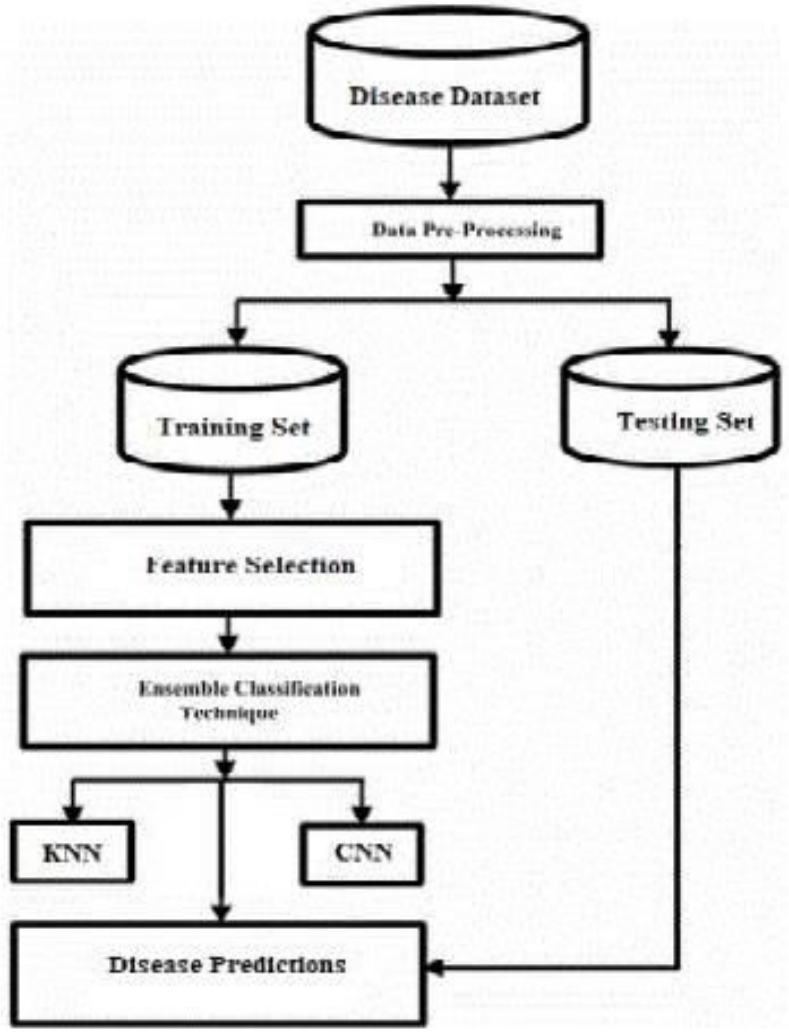
5. PROBLEM STATEMENT

Diseases are leading cause of morbidity and mortality worldwide, accounting for approximately one third of all deaths. Prevention of The Diseases requires timely identification of people at increased risk to target effective dietary, lifestyle or drug invention. Over the past two decades, numerous prediction models have been developed, which mathematically combine multiple predictors to estimate the risk of developing. To reduce this Kind of Diseases and for the in-time treatment, we have created a Disease Prediction model.

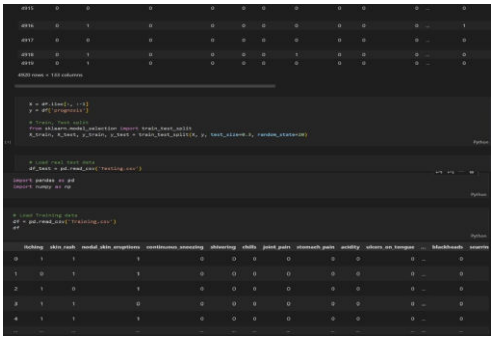
6. PROPOSED SYSTEM

We proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithm for accurate prediction of disease. For disease prediction required disease symptoms dataset. In this general disease prediction, the living habits of person and checkup information consider for the accurate prediction. Initially we take disease dataset from UCI machine learning website and that is in the form of disease list with its symptoms. After that preprocessing is performed on that dataset for cleaning that is removing comma, punctuations, and white places. And that is used as training dataset. After that feature extracted and selected. Then we classify that data using classification techniques such as KNN and CNN.

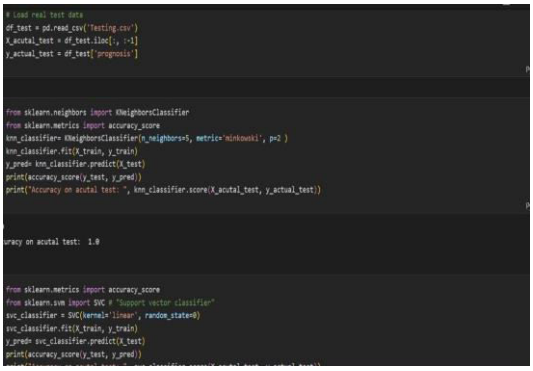
6.1 SYSTEM ARCHITECTURE:



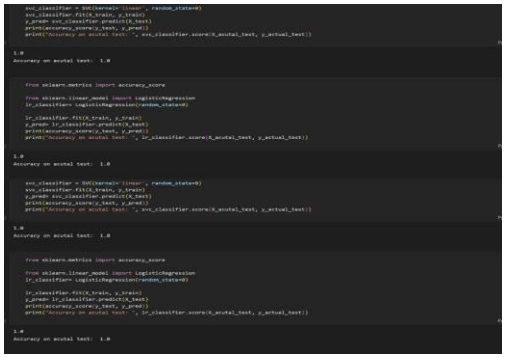
In this phase the designs are translated into code. Computer programs are written using a conventional programming language or an application generator, A dataset used for machine learning should be partitioned into three subsets — training, test, and validation sets. The proportion of training and a test set is usually 80 to 20 percent respectively. A training set is then split again, and its 20 percent will be used to form a validation set. In that 80 percent training data, we applied KNN, CNN algorithms and execute the code.



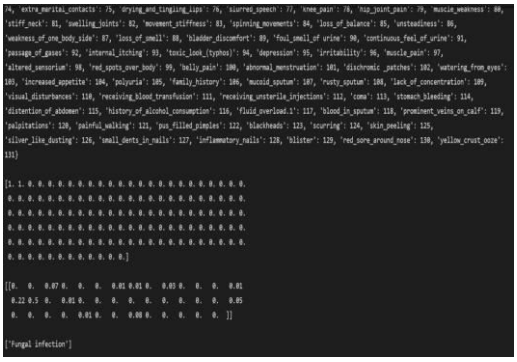
Execution code of KNN and CNN algorithm



The KNN algorithms accuracy



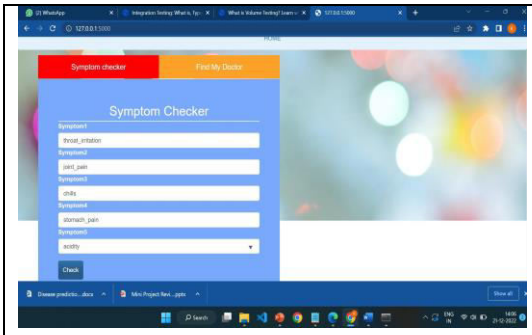
The CNN algorithms accuray



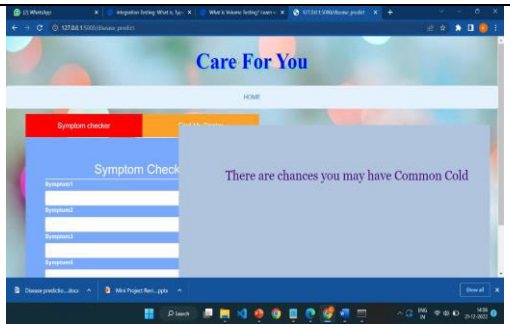
Prediction of Disease by KNN and CNN algorithm

7. RESULTS

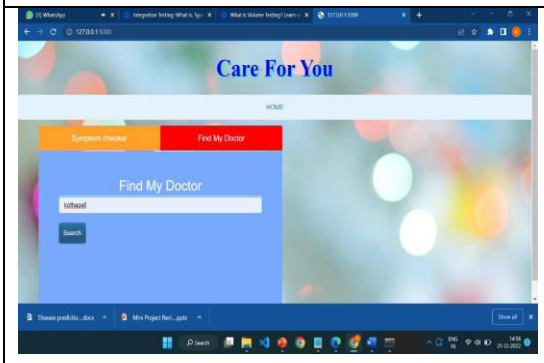
All the experimental cases are implemented in Java in congestion with Net beans tools and MySQL as backend, algorithms and strategies, and the competing classification approach along with various feature extraction technique, and run-in environment with System having configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM. Dataset Patient disease dataset downloaded from UCI machine learning website. Comparison Results and This section present the performance of the KNN and CNN classification algorithms in terms of time required and accuracy. It Shows accuracy Comparison of KNN and CNN algorithms for various Thresholds. X-axis shows Algorithm & Y-axis shows accuracy in %. CNN gives more accurate disease prediction than KNN.



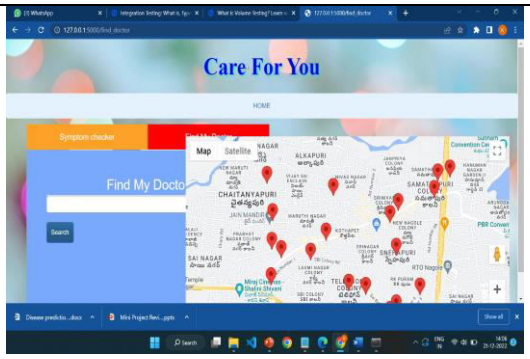
Input of symptoms to check for prediction



Output of the symptoms



Finding of doctor by giving location



Output of nearby available Doctors

8. CONCLUSION

In this research the target of the developed model is to perform well for the detection of bike riders and classification to detect the presence of the helmets in real time image and video. In the set target, the developed model reached up-to an accuracy of 74% and a speed of 1fps without GPU. In this, a minor drawback of the current developed model is that, it captures the image of all persons coming in the frame rather than bike riders, this is affecting the accuracy of the designed method because it should be applicable on the bike riders only. This implies that if a rider is not wearing a helmet but carrying it, a false positive is returned as the location of the helmet is not checked with respect to the person. A possible solution for this error is to use a detection algorithm in the second layer as well, which may compromise the speed and may require more sophisticated hardware.

9. FUTURE SCOPE

Automation of Disease Prediction saves the environment by using paper free work. To increase the accuracy and efficiency so that patients can get direct help. Management of disease related data. The applications in healthcare can have tremendous potential and usefulness. However, the success of healthcare hinges on the availability of clean healthcare data .In this respect, it is critical that the healthcare industry consider how data can be better captured, stored, prepared, and mined. Possible directions include the standardization of clinical vocabulary and the sharing of data across organizations to enhance the benefits of healthcare. Further, as healthcare data are not limited to just quantitative data, such as physicians' notes or clinical records, it is necessary to also explore the use of text mining to expand the scope and nature of what healthcare data mining can currently do.

10. REFERENCE

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik(2014), "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587.

[2] Girshick, R.B. (2015). "Fast R-CNN." 2015 IEEE International Conference on Computer Vision (ICCV), 1440-1448.

[3] Liu, Wei & Anguelov, Dragomir & Erhan, Dumitru & Szegedy, Christian & Reed, Scott & Fu, ChengYang & C. Berg, Alexander. (2016). SSD: Single Shot MultiBox Detector. 9905. 21-37. 10.1007/978-3-319-46448-0_2.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[5] Redmon, Joseph and Ali Farhadi. "YOLO9000: Better, Faster, Stronger." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 6517-6525.

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no. 6, pp. 1137-1149, 1 June 2017.

[7] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 6154-6162.

[8] Redmon, Joseph and Ali Farhadi. "YOLOv3: An Incremental Improvement." ArXivabs/1804.02767 (2018).

[9] G. Chandan, A. Jain, H. Jain and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1305-1308.

[10] K. Kim, P. Kim, Y. Chung and D. Choi, "Performance Enhancement of YOLOv3 by Adding Prediction Layers with Spatial Pyramid Pooling for Vehicle Detection," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6