International Journal of Engineering Sciences & Emerging Technologies, July. 2021.
ISSN: 22316604 Volume 10, Issue 6, pp: 126-136 ©IJESET

A SURVEY ON EFFICIENT USE OF DATA MINING AND MACHINE LEARNING IN HEALTHCARE APPLICATIONS

Aishwarya Hastak[1], Prangana Kashyap[1], Nupur Surana[1], Bhushan Inje[2]
[1]Department of Computer Engineering, SVKM's NMIMS, MPSTME, Shirpur
aishwarya.hastak19@nmims.edu.in, kashyapprangana24@gmail.com,
nupur.surana56@nmims.edu.in
[2]Asst. Prof., Department of Computer Engineering, SVKM's NMIMS, MPSTME, Shirpur
bhushan.inje@gmail.com

ABSTRACT

With the continuous advancement in technology, data mining and machine learning have found applications in various industries, including but not limited to agriculture, automobile, e-commerce, marketing, security, banking. Massive amounts of data have been collected in the healthcare industry over the past years using the EHR systems; data analysis with the help of machine learning models can help in making predictions and gaining helpful insights into many unsolved and untapped areas of healthcare. This paper aims to summarize a bunch of ideas and innovative applications of these technologies in healthcare, including early disease prediction, health trackers, symptom diagnosis. The use of various machine learning algorithms and data mining tools and their efficiencies also have been studied. The accuracy of various algorithms used in early disease prediction is compared.

KEYWORD: *Data mining, Machine Learning, Healthcare, Early disease prediction, Smart Healthcare Systems.*

1. Introduction

Data mining is a technique widely used in extracting knowledge from vast databases. It mainly deals with processing unstructured data and finding out patterns to gain valuable insights. Raw data can be converted into useful information with the help of Data Mining [1]. Several data mining mechanisms have been discussed in this survey.

Machine learning uses various calculation techniques to make effective predictions. It requires a target variable that needs to be predicted (dependent variable), based on other values (independent variable). Oftentimes data mining and machine learning are overlapped because they use the same algorithms. These technologies have prominently helped businesses in finding issues, analyzing their performance and achieving their objectives, through the process of knowledge discovery in databases. [2]

The healthcare industry has collected vast amounts of data for a very long time, through the use of EHR systems. This data can be put to beneficial use to gain knowledge and insights that would be useful in developing effective treatment plans, diagnosis tools, tracker systems, insurance advisors and many such applications. [3]

1.1. Data mining in healthcare

Data mining is one of the most essential and popular techniques that has been incorporated by many healthcare organizations. It is not easy to analyze big and complex healthcare data manually. Technology provides very efficient methods through data mining to handle these data. It uses various other various methodologies like machine learning etc. to overcome such problems and helps to transform big data into useful information, which then helps in decision-making. The handling of big data is done by utilizing various data mining tools like Weka, Knime, R programming, Keel, Orange. [4]

Doi: 10.5281/zenodo.5139537

The challenge faced in performing data mining is the handling of this overwhelming amount of data. The task of preprocessing, cleaning and storing needs to be done correctly, to avoid any inconsistencies, incorrect predictions and complications in accessing the data.

1.2. Data Analysis in healthcare

We need a systematic and strategic approach for everything, which is provided by data analysis in healthcare. For the improvement of health system management, diagnosis, engagement of patient data analysis is a powerful tool. This can help patients of healthcare organizations to avoid risks related to health by taking useful decisions. It uses techniques that include qualitative, statistical, quantitative, predictive, descriptive etc. Unlike Data Mining, it focused more on the process of transforming, cleaning and evaluating data with the help of logical and analytical reasoning.

1.3. Machine learning in healthcare

Machine learning performs an essential role in human health prognosis. It consists of numerous algorithms, which helps it to operate through useful mathematical computations and helps in imaging and predict human body abnormalities. In this field, machines are used since with the help of machines, scanning in a human body can be easily done, images will help in early disease prediction. Algorithms in Machine Learning have proven to be effective to date [5]. A few of the widely used machine learning algorithms in healthcare that have so much are Support Vector Machine (most standard algorithm in healthcare), Random Forest Algorithm (a non-parametric algorithm), Naive Bayes (based on Bayes Theorem), Logistic Regression (used for prediction). With the help of such algorithms, it has become easy to work on massive datasets in healthcare and go beyond the ordinary capabilities.

2. Predictive Modelling and Quantitative Analysis on Healthcare Data

Data mining applications [7] find their use in the field of healthcare though areas that include effectiveness evaluation of treatment, healthcare and management of customer relationships and abuse and fraud detection, identifying risk factors connected with the start of diabetes.

Predictive modelling can be used by hospitals to determine whether the patient can pay their bill, and then guide them to other organizations or facilities that can aid them. It can be used by an individual to estimate their healthcare expense for the next year beforehand. This information can especially help in choosing the right health plan and insurance scheme.

2.1. Predicting Healthcare cost

Predictive modelling is a very effective weapon in this field. Health is not only the problem but sometimes while healthcare providers do not gather financial info from the patients, there is an information shortage which makes it problematic to estimate in the event that a specific patient-indebted person is probably going to pay his/her bill [8]. Neural network of the Logistic regression and the collective model made the best arrangement exactness and the choice tree was the best in characterizing cases that are "good" as per computer simulation for the previous problem. We also need to provide practical approaches for leading and computing analyses of qualitative statistics applicable for researchers in health services. Data Sources and Design is a technique where we depict existing qualitative literature to define practical methods for qualitative data analysis.

This technique can further be extended to predicting one’s healthcare expenses over a period of time. One such example is HealthSCOPE (Healthcare Scalable Cost Prediction Engine), which is a framework for investigating archived and present-day costs of medical services and to make predictions of future expenses. HealthSCOPE can be utilized by people to approximate the costs of their medical services in the coming year. Furthermore, HealthSCOPE upholds a populace based view for statisticians and insurance providers who need to approximate the long run expenses of a populace supported historically stored data on claims, a common situation for Accountable Care Organizations (ACOs).

Doi: 10.5281/zenodo.5139537

With the help of interactive data processing framework, clients can see claims, use HealthSCOPE to anticipate costs for the forthcoming year, intelligently select from many conceivable ailments, comprehend the variables that add to the cost, and analyze costs against recorded midpoints. System in the back-end contains cloud put together expectation administrations facilitated with respect to the infrastructure of Microsoft Azure that permit the simple sending of models encoded permit the simple arrangement of models encoded in Predictive Model Markup Language (PMML) and prepared utilizing either non-distributed environments or Spark MLLib or various.[14]

2.2. Intelligent Agents

Qualitative data analysis methods[9] have been discussed that apply inductive reasoning principle along with engaging pre-set code types to escort data analysis and clarification. These codes define structures according to their properties. For a massive volume of healthcare data, techniques of data mining such as Artificial Neural Network, Decision tree,Rule-based,Naïve Bayes and are used. A compelling technique for the withdrawal of significant examples from the coronary illness information distribution centers for the creative figuring of respiratory failure dependent on the assessed huge weightage, the regular examples having a worth more noteworthy than a predefined limit were chosen for the worth estimation of heart attack[10] has been provided.

The objectives are to be evaluated beside the proficient models. Also the complex tasks which are to be done by a healthcare manager will be done by Intelligent agents[11] following the strategic approach. Agent data mining info structure is used to generate data mining diagnostic support with the following specifications: Remote data access network, End-user interface, Diagnostic support, Data mining engine and strategic assistance A methodology development has been discussed to extract patterns of transitions between adverse effects after implanting Left Ventricular Assist Device in patients with advanced heart disease. These transitions will be extracted through temporal mining.[12]

3. Using Data mining for early disease prediction

One heavily researched application of data mining and machine learning in healthcare is for early disease predictions and prognosis. Numerous factors like the patient's history, background, environment, lifestyle are taken into account and predictions are made. This is especially helpful because the patient can then take precautionary measures and prevent the occurrence of the disease by making changes in his lifestyle and habits. Data Mining uses different types of analysis that include regression, association rule learning, clustering, and classification.

A survey analyzed disabled children and their families based on measures like age, gender etc. [6] It mentioned the term "tertiary prevention" that concentrates on already affected people. It focuses on disability improvement, preventing the delaying difficulties hence improving the quality of life. But prevention alone won't work, we need to have few curing measures.

3.1. Prediction of Diseases in Smart Health Care System using Machine Learning

A procedure of finding and breaking down various information designs from enormous crude datasets is data mining. The principal point of information mining is to extract the significant data from a far reaching dataset. The information mining accompanies a heap of bundles, for example, AI, insights and database framework. These components decide the effectiveness in Knowledge Discovery in the database process. KDD comprises different procedures, for example, data cleaning, information choice, data transformation, data change, information design looking lastly information portrayal. The information mining strategies that are mostly utilized are Association rule, Clustering, Classification, relapse and so on. The affiliation rule can be utilized to build up connection among two factors. Bunching is a strategy of gathering the structures subject to the closeness among them. Order is allocating things in assortment to aim sets of data. The relapse attempts for evaluation of the different model for discovering the connection among information with minimal mistake.

The outcomes are acquired by the investigation of various calculations in the medical services expectation. Major algorithms comprises Neural networks, Random forest, Support vector machines,

Doi: 10.5281/zenodo.5139537

logistic regression etc. The exactness is maximum in the neural networks among this, if properly trained by the sets of data.[13]

3.2. Analysis of Algorithms used for Early Disease Prediction

In this section, we analysed the accuracies of various algorithms that have been used in a few disease prediction applications.

The developers of a self check-up mobile app [17], gathered data from private clinics, and used the generated rules from C4.5 algorithm (decision tree algorithm), which is used as a classification algorithm, on Weka to analyse the collected data. They for the development process, used the Process Model for Healthcare (PMH) methodology. They performed predictive modelling using Logical Model Tree (LMT) and J48, and found that LMT gave predictions that are more accurate.

To reduce the errors in diagnosis, the Ensemble Gain ratio Feature Selection (EGFS) model can be used, which identifies relevant and most important features. After using Area Under Curve (AUC) to measure the accuracy, the accuracy obtained for EGFS is as high as 96.49%. [18]

Using Naïve Bayes Classifier model and independent variables (like BMI, pregnancies, Glucose level, Blood Pressure, Insulin level, Thickness of skin, gender, age); prediction of diabetes can be made with an accuracy of 96%. [16]

HDPM, An Effective Heart Disease Prediction Model for a Clinical Decision Support System proposed the use of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate the outliers, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbour (SMOTE-ENN) to balance the training data distribution and XGBoost to predict heart disease. This combination of algorithms was found to outperform all other independent models, with an accuracy of (95-98 %). [19]

Table 1. Study of accuracy of various models and algorithms for early disease prediction

Reference	**Algorithm/ Model**	**Accuracy**
P Anandajayam et al. [23]	Recurrent neural network (RNN)	97.6%
SJ Pasha et al. [18]	Ensemble Gain ratio Feature Selection (EGFS)	96.49%
KL Priya et al. [16]	Naïve Bayes Classifier	96%
Koh et al.[7]	Decision Tree *Sensitivity* *Specificity* *True positives* *True negatives*	 69.47% 96.64% 75.21% 95.58%
K Deepika et al. [24]	Support Vector Machine (SVM)	95.2%
Chen et al. [25]	CNN-MDPR	94.8%
L Ali et al. [26]	X2 statistical model and deep neural network(DNN) classification	93.33%
A Gavhane et al. [27]	Multi Layered Perceptron (MLP)	91%
MA Alim et al. [28]	Random forest with Stratified KFold cross-validation	86.94%
Dahiwade et al. [29]	Convolutional neural network (CNN)	84.5%

Doi: 10.5281/zenodo.5139537

3.3. Data Mining Techniques and Tools

Information mining methods help in finding the covered up information in a gathering of illness information that can be utilized to investigate and anticipate the future conduct of ailments. Characterization is one the information mining methods, which appoints a class name to many unclassified cases. The fundamental goal of this paper is to think about the information mining devices based on their characterization precision. As indicated by the consequence of three information-mining devices utilized in this paper [15], it has been seen that various information mining apparatuses are outfitting various outcomes on the same informational index with distinctive characterization calculation. WEKA is indicating best arrangement exactness when contrasted with rapid miner and orange. In future, more malady dataset can be utilized for arrangement strategies and other information mining methods for example, bunching can be utilized to look at the exhibition of different information mining apparatuses.

The clinical consideration industry has seen an immense progression in making gigantic proportions of clinical data that delivered research in various fields. Attempts were done and are being done by the examiners in researching the clinical data so as to eliminate accommodating data by applying the specific developments. Electronic Health Record frameworks are the storage facilities of data which is the digitized plan of taking care of the clinical information. Electronic prosperity data is the snappiest creating data assortment. Mechanical movements in medical services data, digitizing prosperity records have achieved the quick move of the clinical consideration zone.

The area of Healthcare deals with a gigantic proportion of data and wishes to examine the information open and gives a higher plan which makes better decisions. One of the difficulties is determining the best approach to find helpful and significant data adequately among the huge measure of data accessible through the information mining strategies. Information preparing assumes a huge function in progressing and building up the new strategies which turn out viably for the large information accessible in medical care. [20]

The way towards acquiring patterns and connections among datasets is accomplished by using information investigation apparatuses, named as information mining produces substantial expectations. It helps in discovering examples and connections in the dataset. The client needs to contemplate the operations of the instruments and the calculations so they can dissect as an ideal strategy from the delivered outcomes. The determination of information mining apparatus and the advancement calculation will show an effect on the speed and precision.[21] [22]

4. Tracking Health using Self Checkup Apps

Data mining is essential in analysing patterns in data and performing clustering, classification and prediction. Accessible healthcare can be provided by analysing symptoms, providing a diagnosis, and suggesting treatment. By keeping track of the user's health activities such as heart rate, step count, a detailed health report can be generated. Data can be gathered from healthcare records, prescriptions, etc and Natural Language Processing can be used to perform analysis. Natural Language Understanding (NLU) algorithms can be used for a voice-assistant facility [30].

4.1. Need of Smart Self Checkup Applications

In the present world, where everything is available in the click of a few buttons, many people seek healthcare advice from the internet. When people are suffering through some minor illness, most people use google as a preliminary check for instant diagnosis.

In the year 2016, Forbes concluded through research that doctor's results are more accurate than any online tool. In mid-2018, NHS partnered with Amazon Alexa, to give users healthcare information through voice command. NHS commented in support of these tools and how they helped in providing people an instant diagnosis for common problems at any time of the day.

In a survey conducted by KolabTree in 2019, they found that 43% of the people look up their symptoms on the internet. A total of 59.7% of users turn to the internet to get health information. Additionally,

Doi: 10.5281/zenodo.5139537

they found that the majority of people that choose "Dr. Google" over an actual doctor belonged to a younger demographic. [30]

Historical healthcare data can be used to develop an application that would aid people in an initial check for a few common diseases. With the user's details (age, gender, current medications, and medical history) and symptoms as an input, it will then produce a list of plausible diseases the user might have (ranked based on their possibility), their cause and home treatments and when to contact a doctor.

In the current times when healthcare costs burn a hole in your pocket, we can provide a free medium to gain a fair understanding of your symptoms.

4.2 Smart Self Checkup Applications and Smart Assistants

Productive ways are expected to imagine the wellbeing status of an individual and the way of life, everyday decisions and medical care activities are influencing it. Current frameworks come up short on a complete interface for communication and investigation of gigantic and complex information and functions influencing the data. Upheld best in class information perception methods, we actualized and client tried a framework that envisions wellbeing information comprehensively after some time. The framework centers around the dynamic changes by utilizing a timetable of functions influencing the wellbeing status. We directed an inside and out client testing measure including studies, heuristics and perceptions to pass judgment on our framework. The outcomes show that our framework envelops an elevated level of User Satisfaction while giving a sufficient getting, connection and route of the data. [31]

Data mining is the way toward extricating data or finding usable information from complex data sets. Clinical field contains enormous heterogeneous information which is being changed over into valuable data by applying information handling strategies thus this advantageous information is utilized by the doctors to determine different infections to have palatable exactness during this work, creators have been centered around the determination of thyroid illnesses by utilizing neural net, vote group and stacking troupe techniques, where the classifiers are looked at regarding exactness, accuracy, review, and mistake rate. Exploratory outcome shows that Stacking group strategy has most elevated exactness than different strategies and furthermore has better forecast precision contrasted and the connected existing writing. [32]

Table 2. Classified work of reported literature related to healthcare smart assistant tasks and their unique features

Reference	Early disease prediction	Diagnosis / Treatment	Track / Monitor health	Doctor Consultation / Appointment	Features / Methods
M Gandhi et al. [33]	✓	✓	✓		SOS emergency trigger
R Amriitha et al. [34]	✓	✓			Indian food recommendation
A Aridarma et al. [35]			✓	✓	temperature, bmi, ecg sensors
D Dojchinovski et al. [36]			✓	✓	ECG sensor, voice assistant
C Gao et al. [37]				✓	book quick appointments

Doi: 10.5281/zenodo.5139537

Lixia Luan et al. [38]		✓	✓		Emotion recognition, exercise recommendation
IM Shofi et al. [39]	✓	✓			forward chaining method for diagnosis
V Kaviya et al. [40]	✓		✓	✓	Wearable technology
RB Mathew et al. [41]	✓	✓			Chat bot

5. Application of Deep Learning and Artificial Intelligence in Healthcare

Artificial Intelligence has been increasingly used in Smart Healthcare applications. It has the ability to recognize patterns in data automatically and produce effective laboratory reports and assessments. AI serves as an excellent decision support tool and can handle complex and vast data, thus saving time and efforts in modern healthcare.[42]

ScalpEye is a Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health. The system is accessible through a mobile application, which provides diagnose hair problems. The system takes 200x scalp images as input and the data is sent to their AI training server which is a cloud-based server. The server would provide probability results for one of four problems – dandruff, hair loss, oily hair and folliculitis. The cloud-based management server can also be utilized to store and keep track of the diagnosis and user records. This system is able to achieve an average precision of 97% - 99% [43].

FluSpider is a new vision of digital influenza surveillance system which makes use of Massive Data Mining techniques and Big Data technologies [44]. This is a digital surveillance system which tracks the visitors of web pages and estimates the spread of influenza across the world. It makes use of web crawlers on 15 different websites, collecting search data related to the virus. They then tracked the IP addresses of the website users and generated a real time map constructed from the total count of views per day in a country. It is capable of providing real time monitoring and information, well before the traditional information centers.

However, the accuracy of the results of data mining on healthcare data is highly dependent on the availability of clean and refined data. Various data mining algorithms used for prediction analysis are - Artificial Neural Networks, Naive Bayes, Logistic Regression, Decision Tree, Support Vector Machines. For disease prediction, Decision Trees are usually found to give more accurate and dependable results [45].

6. Risk Management and Use of Existing Strategies for Better Healthcare

Only maintaining health will not be enough, there is also a necessity of dealing with risks. A Risk prediction model can be used to predict such risks, where it takes a patient's symptoms as input to the model. Then a Risk assessment service [46] will provide a personalized risk report that contains the levels and factors and risks along with an intervention plan for promotion of good health.

To deal with such risks a tool named Cubists is used for building predictive models out of two million cases over a nine-year period. The objective is to gain insight into various aspects of California's [47] mortality and provide efficient services to customers.

A Privacy-Preserving Collaborative Model [48] learning plan has also been proposed named PCML comprising skyline computation. With PCML, medical care communities can safely become familiar with a worldwide conclusion model with their local discovering models in the assistance of the cloud; the touchy clinical information of each clinical help place is very much secured. In particular, with a Safe Multi-party Vector Correlation calculation (SMVC), every neighborhood determination model is encoded by its proprietors before being shipped off the cloud and can legitimately function without

Doi: 10.5281/zenodo.5139537

decoding. Nitty-gritty safety examination displays that PCML is able to oppose reliability dangers in the partially legit model. Also, PCML is executed together with clinical sets of data from the UCI AI store, broad recreation outcomes exhibit that PCML has proven to be proficient and is capable of actualizing viably.

However, along with all these, if we really want to improve, then we need to discuss the existing strategies too. As Multimodal information, driven methodology has risen as a fundamental activation for keen medical care frameworks with applications including sickness investigation to emergency, determination and therapy. Savvy medical care framework requires new requests for information from the executives and dynamic, which has propelled the quick advancement of clinical administrations utilizing figuring and new changes in the medical services industry. In this part, an extensive overview of existing strategies has been performed, which incorporates best in class techniques as well as the principal ongoing patterns inside the field. In particular, this survey centers around the sorts of dynamic cycles used in savvy medical services frameworks.

Primarily, changes that use multimodal affiliation mining with fine-grained information semantics in savvy medical services frameworks were presented. They reviewed the brilliant medical services arranged semantic discernment, semantic arrangement, substance affiliation mining, and talked about the upsides and downsides of those methodologies.

Furthermore, they examined approaches for multimodal information combination and cross-outskirt affiliation that are utilized in creating shrewd medical services frameworks. At last, special attention was given to panoramic interactive decision making, decision framework, and intelligent decision support systems. Smart Healthcare Systems are deployed in and prove to be salubrious to many fields, including privacy protection and knowledge discovery. [49]

7. Conclusion

In today's time, where healthcare costs burn a hole in our pockets, using data mining and machine learning methodologies eases the cost and effort required in managing healthcare facilities. It benefits the hospital staff as well as the patients. At customer level it saves time as well as cost and at staff level it enables them to make instantaneous real time decisions in less time and effort.

This paper consolidates several papers and applications of these technologies, including early disease prediction, self-diagnosis, health tracking, smart assistants and creating descriptive reports. Applications and efficiencies of various machine learning algorithms and data mining tools have been studied.

References

[1]. Sharma, Seema, Jitendra Agrawal, Shikha Agarwal, and Sanjeev Sharma. "Machine learning techniques for data mining: A survey." In *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1-6. IEEE, 2013.

[2]. Guruvayur, Sivaramakrishnan R., and R. Suchithra. "A detailed study on machine learning techniques for data mining." In 2017 International Conference on Trends in Electronics and Informatics (ICEI), pp. 1187-1192. IEEE, 2017.

[3]. Tekieh, Mohammad Hossein, and Bijan Raahemi. "Importance of data mining in healthcare: a survey." In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pp. 1057-1062. 2015.

[4]. Reddy, R. Pallavi, Ch Mandakini, and Ch Radhika. "A Review on Data Mining Techniques and Challenges in Medical Field."

[5]. Marr, Bernard. "How Big Data Is Changing Healthcare." Forbes. April 22, 2015. https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#4f365f828730.

[6]. Gordon Jr, R.S., 1983. An operational classification of disease prevention. Public health reports, 98(2), p.107.

[7]. Koh, H.C. and Tan, G., 2011. Data mining applications in healthcare. Journal of healthcare information management, 19(2), p.65.

Doi: 10.5281/zenodo.5139537

[8]. Zurada, J. and Lonial, S., 2005. Comparison of the performance of several data mining methods for bad debt recovery in the healthcare industry. Journal of Applied Business Research (JABR), 21(2).

[9]. Bradley, E.H., Curry, L.A. and Devers, K.J., 2007. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health services research, 42(4), pp.1758-1772.

[10]. Srinivas, K., Rani, B.K. and Govrdhan, A., 2010. Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2(02), pp.250-255.

[11]. Zaidi, S.Z.H., Abidi, S.S.R. and Manickam, S., 2002, June. Distributed data mining from heterogeneous healthcare data repositories: towards an intelligent agent-based framework. In Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002) (pp. 339-342). IEEE.

[12]. Movahedi, F., Zhang, Y., Padman, R. and Antaki, J.F., 2018, June. Mining temporal patterns from sequential healthcare data. In 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 461-462). IEEE

[13]. N. Shabaz Ali, G. Divya. "Prediction of Diseases in Smart Health Care System using Machine Learning". International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-5, January 2020

[14]. S. Woolf and L. Aron, Eds., U.S. health in international perspective: Shorter lives, poorer health. National Academies Press (US), 2013.

[15]. Ahmed, Kauser P. "Analysis of data mining tools for disease prediction." Journal of Pharmaceutical Sciences and Research 9, no. 10 (2017): 1886-1888

[16]. Priya, K. Lakshmi, Mourya Sai Charan Reddy Kypa, Muchumarri Madhu Sudhan Reddy, and G. Ram Mohan Reddy. "A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier." In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), pp. 603-607. IEEE, 2020.

[17]. Johari, Nur Aliyah Afiqah Mohd, Norizan Mohamad, and Norulhidayah Isa. "Smart Self-Checkup for Early Disease Prediction." 2020 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS). IEEE, 2020.

[18]. Pasha, Syed Javeed, and E. Syed Mohamed. "Ensemble Gain Ratio Feature Selection (EGFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction." In 2020 International Conference on Inventive Computation Technologies (ICICT), pp. 590-596. IEEE, 2020.

[19]. Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System." IEEE Access 8 (2020): 133034-133050.

[20]. Kogent Learning Solutions , Bill Franks , Wrok certified Big Data Analyst , Introducing Big Data Analytics and Predictive Modelling , Wiley publishers.

[21]. Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby. “Data Mining: A Prediction for students performance using classification method”, World Journal of Computer Application and Technology 2(2): 43-47, 2014.

[22]. Shweta Kharya, “Using Data Mining Techniques for Diagnosis and Prognosis of Cancer disease”, International Journal of Computer Science, Engineering and Technology(IJCSEIT), Vol.2, No. 2.

[23]. P Anandajayam, S Aravindkumar, P Arun and A. Ajith. “Prediction of chronic disease by machine learning”, pp. 1-6, IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019.

[24]. Deepika, Kumari, and S. Seema. "Predictive analytics to prevent and control chronic diseases." In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 381-386. IEEE, 2016.

[25]. Chen, Min, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. "Disease prediction by machine learning over big data from healthcare communities." Ieee Access 5 (2017): 8869-8879.

[26]. Ali, Liaqat, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed, and Javed Ali Khan. "An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}} $ Statistical Model and Optimally Configured Deep Neural Network." IEEE Access 7 (2019): 34938-34945.

[27]. Gavhane, Aditi, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. "Prediction of heart disease using machine learning." In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1275-1278. IEEE, 2018.

[28]. Alim, Muhammad Affan, Shamsheela Habib, Yumna Farooq, and Abdul Rafay. "Robust Heart Disease Prediction: A Novel Approach based on Significant Feature and Ensemble learning Model." In 2020 3rd

Doi: 10.5281/zenodo.5139537

International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-5. IEEE, 2020.

[29]. Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing Disease Prediction Model Using Machine Learning Approach." In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 1211-1215. IEEE, 2019.

[30]. Sriram, Ramya. "New Data Reveals How Many of Us Are Misdiagnosing Ourselves Using Google". Kolabtree Blog. August 15, 2019. https://www.kolabtree.com/blog/new-data-reveals-how-many-of-us-are-misdiagnosing-ourselves-using-google/.

[31]. World Health Organization. (2015) Noncommunicable diseases. [Online]. Available: http://www.euro.who.int/en/health-topics/ noncommunicable-diseases

[32]. S. Vijiyarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey", International Journal of Computer Applications & Information Technology, vol. II, no. I, pp. 17–21, 2013.

[33]. Gandhi, Meera, Vishal Kumar Singh, and Vivek Kumar. "IntelliDoctor-AI based Medical Assistant." In 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), vol. 1, pp. 162-168. IEEE, 2019.

[34]. Amriitha, R., M. Karpaga Meena, U. Fathima Shafa, and R. B. Vandhana. "CD-DIET: A Prediction and Food Recommendation System for Chronic Diseases."

[35]. Aridarma, Arga, T. L. Mengko, and Soegijardjo Soegijoko. "Personal medical assistant: Future exploration." In Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, pp. 1-6. IEEE, 2011.

[36]. Dojchinovski, Dimitri, Andrej Ilievski, and Marjan Gusev. "Interactive home healthcare system with integrated voice assistant." In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 284-288. IEEE, 2019.

[37]. Gao, Chunming, and Noriyuki Iwane. "Developing a Prototype for Evaluations of Healthcare Appointment and Registration Smart Assistant." In 2016 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 68-69. IEEE, 2016.

[38]. L. Luan, W. Xiao, K. Hwang, M. Shamim Hossain, G. Muhammad and A. Ghoneim, "MEMO Box: Health Assistant for Depression with Medicine Carrier and Exercise Adjustment Driven by Edge Computing," IEEE, 2020.

[39]. Shofi, Imam M., Luh Kesuma Wardhani, and Ghina Anisa. "Android application for diagnosing general symptoms of disease using forward chaining method." In 2016 4th International Conference on Cyber and IT Service Management, pp. 1-7. IEEE, 2016.

[40]. Kaviya, V., and G. R. Suresh. "INTELLIGENT WEARABLE DEVICE FOR EARLY DETECTION OF MYOCARDIAL INFARCTION USING IoT." In 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), pp. 1-4. IEEE, 2020.

[41]. Mathew, Rohit Binu, Sandra Varghese, Sera Elsa Joy, and Swanthana Susan Alex. "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning." In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 851-856. IEEE, 2019.

[42]. Kamruzzaman, M. M. "Architecture of Smart Health Care System Using Artificial Intelligence." In 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1-6. IEEE, 2020.

[43]. Chang, Wan-Jung, Liang-Bi Chen, Ming-Che Chen, Yi-Chan Chiu, and Jian-Yu Lin. "ScalpEye: A Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health." IEEE Access 8 (2020): 134826-134837.

[44]. Nouira, Kaouther, and Nesrine Ben Njima. "FluSpider as a new vision of digital influenza surveillance system: based on Big Data technologies and Massive Data Mining techniques." In 2020 International Multi-Conference on:"Organization of Knowledge and Advanced Technologies"(OCTA), pp. 1-10. IEEE, 2020.

[45]. Deepika, M., and K. Kalaiselvi. "A Empirical study on disease diagnosis using data mining techniques." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 615-620. IEEE, 2018.

[46]. Mei, Jing, Enliang Xu, Bibo Hao, Yuan Zhang, Yiqin Yu, and Shaochun Li. "Translational Health Informatics from Risk Prediction Modeling to Risk Assessment Service." In 2019 IEEE International Conference on Healthcare Informatics (ICHI), pp. 1-2. IEEE, 2019

[47]. Zhang, D., Ha, Q.L. and Lu, M., 2001, November. Mining California vital statistics data. In Proceedings 2001 IEEE International Conference on Data Mining (pp. 671-672). IEEE

Doi: 10.5281/zenodo.5139537

[48]. Wang, F., Zhu, H., Liu, X., Lu, R., Hua, J., Li, H. and Li, H., 2019. Privacy-Preserving Collaborative Model Learning Scheme for E-Healthcare. IEEE Access, 7, pp.166054-166065.

[49]. E. Barbi, F. Lagona, M. Marsili, J. W. Vaupel, and K. W. Wachter, ‘‘The plateau of human mortality: Demography of longevity pioneers,’’ Science, vol. 360, no. 6396, pp. 1459–1461, Jun. 2018.

Doi: 10.5281/zenodo.5139537