International Journal of Engineering & Technology, 7 (3) (2018) 1019-1023

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

doi: 10.14419/ijet.v7i3.12654

Research paper

Intelligent health risk prediction systems using machine learning: a review

Mr. Santosh A. Shinde [1] *, Dr. P. Raja Rajeswari [2]

[1] *Research Scholar, Department of CSE, KLEF, Guntur, Andhra Pradesh, India*
[2] *Professor, Department of CSE, KLEF, Guntur, Andhra Pradesh, India*
**Corresponding author E-mail:shinde.meetsan@gmail.com*

Abstract

Humans are considered to be the most intelligent species on the mother earth and are inherently more health conscious. Since Centuries mankind has discovered various proven healthcare systems. To automate the process and predict diseases more accurately machine learning methods are gaining popularity in research community. Machine Learning methods facilitate development of the intelligence into a machine, so that it can perform better in the future using the learned experience. Machine learning methods application on electronic health record dataset could provide valuable information and predication of health risks.

The aim of this research review paper are four-fold: i) serve as a guideline for researchers who are new to machine learning area and want to contribute to it, ii) provide state-of-the-art survey of machine learning, iii) application of machine learning techniques in the health prediction, and iv) provides further research directions required into health prediction system using machine learning.

Keywords:*Electronic Health Records; Health Risk Prediction; Machine Learning; Risk Prediction Model.*

1. Introduction

Health informatics is one of the greatest challenges and machine learning is one of the growing fields in the computer science. Electronic health records (EHR) information contains data of large number of patients. The challenge with the huge data is how to analyze the data and infer the intelligence. This is also known as Rapid Learning Health Care (RLHC) where a large amount of data that is continuously growing needs to be analyzed to predict accurate model for prediction of health risks and diseases. Challenges in representing patient data and analyzing it to produce meaningful information are hurdle for predictive modelling using EHRs. Recent progress of machine learning algorithms in designing and applying it on EHR dataset has shown promising results. However, majority of techniques rely on labelled data for accurate prediction of results. The state-of-the-art machine learning models helps to solve clinical task which were difficult to solve using the traditional methods [20], [21]. The goal of machine learning methods to learn and improve over the time can be used for predictions.

a) Observations: Mining of electronic health records (EHRs) could yield insightful information that can help to improve health risk predictions, thus improving the clinical decisions. A main objective of precision medicine is to develop quantitative models for patients that can be used to predict health status so also help prevent disease or disability [15].

Recent research studies demonstrates thatthe use of EHRs has enabled data-driven prediction of drug effects and interactions [18], detection of type 2 diabetes subgroups [12], [16]. However, predictive models based on the machine learning methods have not been widely used in clinical decision support systems [2], [5], [10].

Machine learning deals with the problems of extracting features from the input dataset to solve predictive tasks such as decision support, forecasting, diagnosis of a disease, detecting virus anomalies, etc. The main challenge is to infer knowledge that is relevant structural patterns on input dataset. In addition, biomedical data sets are full of incompleteness, dirty data, noisy data, unwanted data, etc [9] that makes application of fully automated methods difficult.

b) Contributions: To summarize, this paper makes followingcontributions:

- Classifies categories of machine learning methods.
- Classifies state-of-the art research in health risk prediction systems or disease prediction systems.
- Provides future research directions.

The rest of this paper is organized as follows: Section 2 provides introduction to machine learning. Section 3 presents detailed literature review of machine learning methods in health risk predictions. Section 4 provides future research directions, and Section 5 concludes the paper.

2. Background

This section describes the machine learning landscape

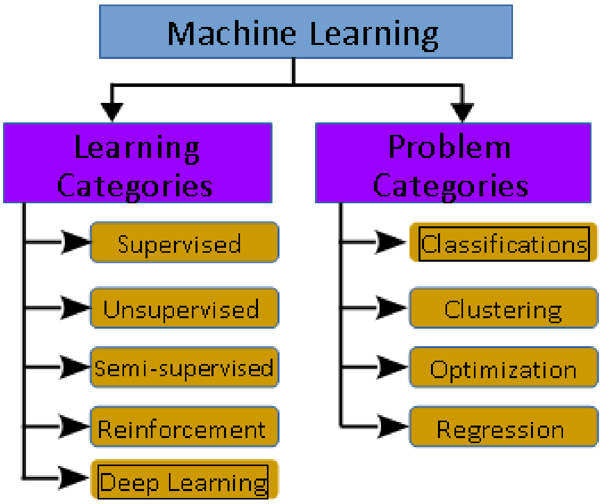

Fig. 1: Concept Map of Machine Learning Categorization.

and basic concepts. Machine Learning methods facilitate development of the intelligence into a machine, so that it can perform better in the future using the learned experience.

Fig. 1 shows two ways to categorize machine learning methods namely: Learning Categorization and Learning Problems Categorization. Machine learning algorithms or techniques categorization using learning techniques contains five sub-fields namely: Supervised learning, Unsupervised learning, Semi-supervised learning, Reinforcement learning, and Deep learning.

a) Supervised Learning: Supervised learning methods work on known expectations on the labelled input dataset. Methods in this category aim at establishing a relationship between the input and output attributes of the labelled dataset. Labelled data helps to build more reliable and accurate model, however, it is computationally expensive process.
b) Unsupervised Learning: These methods target to analyze the structure of data in the provided input unlabelled dataset and build mapping between the input and output attributes, whereas output attributes are unknown before the analysis.
c) Semi-supervised Learning: These methods used both labelled and unlabelled datasets to develop models for intelligence inference.
d) Reinforcement Learning: These methods goal is to maximize the rewards from the result. That is reinforcement learning method produces a sequence of decisions that help to acquire highest rewards.
e) Deep Learning: These methods focus on unifying artificial intelligence with machine learning. It works on common data to provide meaningful insights. It works on input dataset that has less labelled data and solves

problems classified under semi-supervised learning to build complex neural network models.

Machine learning algorithms or techniques are also classified using learning problems as: Classification, Clustering, Optimization, and Regression.

a) Classification: It is a grouping technique that depends on the given value of target and dataset. According to the provided target value it quantifies and classifies the dataset.
b) Clustering: It is a technique that takes only dataset as an input and identifies interesting patterns to derive intelligence. As compared to classification, in the clustering target value is not provided as an input or it is an unknown parameter.
c) Regression: In this technique intelligence or information is derived from the past learning experience. An equation is derived that matches with most of the data points and the cases where data doesn't fit with the curve, those data points we discard. This technique is known as regression.
d) Optimization: It is a method to improve the performance of the system in terms of various attributes.

3. Literature review

In this section, we discuss existing research efforts that predict health risks or health diseases using various techniques.

3.1. Binary classification techniques

A Plethora of Health Risk Systems is available in the literature. However, Most of the research in the initials days focused on developing Disease Risk Prediction Models using Machine Learning for a single candidate Disease. These were mostly the Binary Classification problems which given a medical records dictates whether a person is suffering from specified disease or not. Problems of this sort are called as Single Label, Single Class Classification Problems.

Gaudinat et al. [7] propose a method to categorize health documents using HON code principles. Wu et al. [27] propose a model to quantify the value of EHRs in breast cancer prediction. Watcharapasorn et al. [26] propose a model to predict the mortality rate of undergone surgery patient by using Chiang Rai Nutrition Assessment information (CNA) using various methods such as J48, ADTree and KNN.

LaFreniere et al. [13] propose model to predict hypertension based on neural network. The authors used CPCSSN data set, which is a rich dataset containing huge samples of 1, 85,371 patients and 1, 93,656 controls. In all 11 factors are considered for training the network which includes Birth Year, Gender, BMI, Diastolic and Systolic BP, Low and High Density Lipoprotein, Triglycerides, cholesterol, Micro-albumin and Urine Albumin Creatinine Ratio. Interestingly the authors discussed that the number of samples is also an important parameter in building an accurate machine learning based prediction model.

Influenza is informally known as the Flu. It is caused by RNA Viruses and is a contagious disease. Attaluri et al. [1] developed a model for the prediction of influenza virus antigenic types. The neural network based Model classifies antigenic types such as H1, H3, and H5 and hosts such as Human, Avian, and Swine. Murota et al. [17] presents the case of the 2009 pandemic-flu response in Japan. Shankar et al. [23] propose a model for recognition of diseases and prediction of Swine Flu. It is difficult for humans to differentiate the patients with common cold and swine flu disease infection. Thakkar et al. [24] propose a model based on the collected data for Swine Flu Prediction using Nave Bayes. We live in the era of Social Networking. Social Networking sites such as Facebook and twitter and plays a very significant role in our day today lives in communication, information exchange, group communications and expression of our sentiments. Twits on Twitters have potential in discovering and understanding the spread of epidemics. Grover et al. [8] propose a method to detection and prediction of swine flu disease using Twitter micro-blogging website dataset.

Tuberculosis (TB) is a dreadful and infectious disease caused by the inhalation of bacteria Mycobacteria tuberculosis. Early prediction of TB is therefore crucial. Jamie et al. [14] propose a multiple-instance learning based method for prediction of TB disease using chest X-rays dataset. Cao et al. [4] used a deep convolutional neural network (CNN)-based model for classification of TB X-ray images. Cui et al. [3] reported a comparison of PLS, SVM (with linear and quadratic kernels), linear discriminant analysis (LDA), and random forest (RF) on clinical metabonomics data.

3.2. Multiclass classification techniques

The multilabel classification problem is converted into a multiclass classification problem by following two ways:

1) Algorithm Adaptation: They can be adapted, extended and customized to solve the multilabel learning tasks based on basic Machine Learning algorithms.

2) Problem Transformation: They convert the multilabel learning problem into a single-label classification problem. There are 3 methods of Problem Transformation namely:
a) Binary-Relevance(BR) Method
b) Label Power-set(LP) Method
c) Pair Wise Method

These methods are described as:-
a) Binary-Relevance (BR) Method: It basically utilizes the one-against-all idea. It converts the multilabel problem into N classification problems, where N represents the cardinality of the label set. However it fails to utilize correlations among the labels.
b) Label Power-set (LP) Method: It converts a multilabel problem into a single-label problem by transforming the label set of each instance into an atomic and unique label li, e.g. the multilane set x, y, x would become the single label xyz. It overcomes the limitations of the BR method but whenever the size of the single instance is large the time complexity is worst in performance.
c) Pair-wise Method: It is based on the round-robin classification. All pairs of labels are encompassed by the classifier and eventually a majority voting algorithm is used in order to mix all the classifiers.

Multilabel Classification Methods are as follows:-
i) RAkEL Method: Random k-Labelsets (RAkEL) method for multilabel classification proposed by G. Tsoumakas et al. [25]. It uses Label Power sets in order to train to train on groups of smaller, randomly selected sets of labels, which are of size k, using different classifiers on groups of LPs. Eventually for deciding the target values it applies the criteria of majority voting rule. If the average of the predictions is greater than specific threshold value then the label is chosen as true for that instance.
ii) The ELPPJD method:proposed by Runzhi et al. [22] is an Ensemble Label Power-set Pruned datasets Joint Decomposition (ELPPJD) method to solve the multilabel classification problem for the disease prediction.

Recently, researchers have undertaken work involving the comorbid Disease Prediction, which deals with investigating given a set of medical records; whether a person is suffering with multiple health risks and vulnerable to multiple diseases at the same time.
Deep Learning Architectures are also gaining more and more popularity in recent days because of their potential to deliver state-of-art accuracy in predictive analytics.
Andrew Maxwell et al. [29] focused light on application of DNN for multi-label classification of intelligent health risk prediction. The paper quotes that Deep Learning architectures are even outperforming humans for their applicability for multi-label classification. The proposed DNN uses grid search technique in order to optimize the hyper parameters to get the best solution and lets the network converge on a model that suits the data. The performance of the network is compared with other traditional classifiers in terms of Accuracy, Precision, Recall and F-score and it was found that the F-score of DNNs is little less than Random Forest but otherwise higher than other classifiers. Interestingly it was found that DNN outperforms RAkEL method. As far as the Activation function is concerned the sigmoid function gives better results as compared with ReLU function. The accuracy of the network is also tested with varied number of units per Layer and it was found that 35 Units per Layer gives promising results. Essentially speaking the DNNs is the future of multilabel classification being capable of adapting to the original data and learns in the most optimized way. The authors conclude that accuracy in terms of chronic disease forecasting still remains a major research opportunity in the future.
While discovering the probability of having suffering from multiple health disorders, it is necessary to take into account the temporal relationship between medical events, such as medical checkups followed by diagnosis, medication and further complications and timestamp the medical events as medical history of the patient, and usage of this temporal data for training the machine learning models and deriving useful insights out of it and evolving the learning process should also be considered for more accurate and reliable health risk prediction models in future.
Zhang et al. [28] propose a deep learning method for predicting comorbid risk by analyzing longitudinal EHR
data.

3.3. Mental health analysis

According to WHO 1 among 4 people in the world will be affected by mental or neurological disorders at some point in their lives [31]. Inappropriate diagnosis of mental health leads to wrong treatment and sometimes premature death. Considering US alone, around 12 million people are diagnosed incorrectly per year [30]. The diagnosis of Mental Health in most of the cases depends on Psychiatrists interrogating a patient by asking a set of questions and interpreting the answers. This task in certain cases depends on the experience, intuition and quite a trail and basis method. The severity of Mental Disorders is noteworthy which in few cases leads to acute depression, hypertension and leading even towards suicides. About 8, 00,000 people commit suicide worldwide every year, of these 1, 35,000(17 %) are from India alone, a nation with 17.5% of world’s population [32].
Existing machine learning approaches has made attempts to predict the quality of sleep, sleep disorders, bipolar disorder, hypertension, stress levels and other mental diseases. However, each with its own limitations and state-of-the art comprehensive generalized machine learning models are still in evolution and presents a wider scope for researchers to look into this interesting area of research.
GhassanAzar et al. [30] presented a Novel GA based approach for diagnosing the Mental Diseases. Their approach is kind of Dictionary based and combines the best of Manual and automated reasoning. Patient will be interrogated by the Psychiatrist and the keywords are extracted from the answers given by the patients which are then intelligently mapped using Genetic Algorithm based approach with the DSM-IV-TR Manual Dictionary. The system automatically suggests the diagnosis and treatment. This is a semi-automated method which can be used in a controlled environment in the presence of a Psychiatrist. A. R. Subhani et al. [33] proposed a machine learning model in order to detect the Mental Stress at different Levels. In all 4 Levels of Stress are considered. EEG Signals are recorded for the subjects in a controlled and stressed environment in order to train the model. EEG based Multi Level Stress detection is perhaps a very rare approach adopted by the authors and delivered the performance accuracy of 83.46%. ArkaprabhaSau and Ishita Bhakta [34] used Machine Learning for predicting the Anxiety and Depression in elderly patients. They collected data from 520 subjects from a hospital in Kolkata. The data contained variety of parameters like age, social status, family history of depression, pains in body parts, diabetes etc. Ten different classifiers were employed and achieved an accuracy of 90 per cent (SMO Algorithm) and 91 per cent accuracy using RF algorithm.
Another interesting work was carried out by NJaques et al. [35] for predicting user’s next day Stress Level, Health conditions and Mood using Machine Learning. They used Multitask Learning (MTL) and Domain Adaptation (DA) approaches. The datasets used were - (i) Physiology data recorded using Affectiva Q Sensors including skin temperature, skin conductance etc. (ii) Location using GPS of mobile (iii) Cell Phone data like SMS, calls and screen usage (iv) surveys of social interactions and other activities (v) Whether and (vi) Mood and Wellbeing Labels (self-reported). The results are promisingly good and model can be extended to a more generalized settings.
Smartphones can be potentially useful for its use in knowing the location, physical activities performed by the user, daily distance travelled and calories burnt. The Challenges and Opportunities discussed by G. Mikelsons et al. [36] for the use of Smartphone

data in Psychological State Prediction is another significant milestone.

The study carried out at Darmouth College by Rui Wang et al. [37] in developing the StudentLife dataset for evaluating and predicting the mental wellness, Performance in Studies and Behavioural trends is a breakthrough in using Smartphones in building the unique standard dataset of its own type. It has opened up new doors of research for the research community in this sub-domain. They collected student data for 10 weeks duration in a semester. Different traits recorded were activity, conversation, sleep, location, stress, mood, social interactions, exercise, behaviour and others. Statistical Analysis on the collected data was carried out leading to interesting observations. The dataset is publically available for the research community. Apart from Electronic Health Records (EHR), Bioinformatics also presents abundant opportunities of research in terms of Gene Sequences studies and their classification. S. Zaman and RizoanToufiq [38] implemented Condon Based BPNN for the classification of Hypertension Gene Sequences. Using NCBI they collected the gene sequences. It is noteworthy that their results varied with the variation in the number of samples.

Sleep Science has grabbed the attention of researchers since long and Sleep is an essential part of Human life. Many Psychological conditions and parameters are related to the quality of sleep and lack of sleep can in turn leads to other physiological problems in human body. Nowadays Wearable devices and sensors are been researched out extensively and it is the future trend in many scientific endeavours. AartiSathyanarayana et al. [39] had developed a framework for predicting the Quality of Sleep using the RAHAR Algorithm. They used the Wearable Sensors Apple Watch and it is possible to compare it with Actigraph used in clinical settings. They employed Adaptive Boosting, Logistic Regression, Random Forest and SVM. The Mental Disorder which has alterations in terms of depression and manic/hypomanic stages popularly known as Bipolar Disorder is a severe mental disorder encountered in some patients. The patient's mood is swinging sometimes and the mood change can vary depending on the conditions and the severity of the disorder. ECG based prediction of Bipolar Disorder mood changes proposed by G. Valenza et al. [40] is another potential area for further studies and extensions. The authors proposed a PHYCHE system which uses some wearable devices and the data collected needs further statistical analysis for Mental Health. From a wearable T-shirt the ECG signals were recorded from subjects wearing them whole night and the HRV features of the signals are used for further analysis. Principal Component Analysis was used and finally SVM Classifier decides the mood state. On an average the prediction accuracy is 69%. John Devapriam et al. [41] presented an article concerning In-patient services for people with intellectual disability and mental health or behavioural difficulties. The article discusses Comorbid mental health problems. Mental Health disorders also demonstrate co morbidities with other health ailments. N. Satyanarayana et al. [42] ruled out that Age, Anger and Anxiety has temporary effects on Blood Pressure Raise in patients. Other factors like obesity also might have effects on BP increase effect, the study of future scope as per the authors quoting.

4. Future research directions

This section provides a few research directions.

- Electronic Health Records (EHR) entries include various features that can improve breast cancer risk prediction. In addition, developing statistical models to estimate the probability of breast cancer will help clinicians identify individuals at higher risk.
- Various encoding techniques are used in electronic health records, such as the International Classification of Disease (ICD) to encode diagnoses. These encoding techniques contain a hierarchical structure that represents relationships among the codes, which may improve the effectiveness of predictive models [6], [11].
- Another interesting direction to investigate is to propose risk prediction models using random forest method of machine learning, and compare the performance between random forest method and LASSO.
- Wearable sensors are becoming popular in health care systems [19]. Therefore, collecting and analyzing a stream of data from wearable computing devices at real time to perform predictions on chronic diseases.
- Performance evaluation of various models such as rainforest model, SVM, etc used for health record predictions.
- Study of various factor and complex disease patterns to build models for predictions using deep learning and ANN techniques.

5. Conclusion

In this paper, we focus on exploiting machine learning methods and its applications in prediction of health risks and healthcare systems.

The objectives of this research review paper are fourfold: i) serve as a guideline for researchers who are new to machine learning area and want to contribute to it, ii) provide state-of-the-art survey of machine learning, iii) application of machine learning techniques in the health prediction, and iv) provides further research directions required into health prediction system using machine learning.

References

[1] P. K. Attaluri, Z. Chen and G. Lu, "Applying neural networks to classify influenza virus antigenic types and hosts," *2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, Montreal, QC, 2010, pp. 1-6.https://doi.org/10.1109/CIBCB.2010.5510726.

[2] R. Bellazzi and B. Zupan," Predictive data mining in clinical medicine: Current issues and guidelines," International Journal of Medical Informatics, Volume 77, Issue 2, 2008, Pages 81-97.https://doi.org/10.1016/j.ijmedinf.2006.11.006.

[3] Y. Cao, X. Cui, T. Chen, M. Su, A. Zhao, X. Wang, Y. Ni, and W. Jia, "Comprehensive comparison of classifiers for metabolic profiling analysis," *2010 3rd International Conference on Biomedical Engineering and Informatics*, Yantai, 2010, pp. 2311-2315.https://doi.org/10.1109/BMEI.2010.5639754.

[4] Y. Cao, C. Liu, B. Liu, M. J. Brunette, N. Zhang, T. Sun, P. Zhang, J. Peinado, E. S. Garavito, L. L. Garcia, and W. H. Curioso , "Improving Tuberculosis Diagnostics Using Deep Learning and Mobile Health Technologies among Resource-Poor and Marginalized Communities," *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, Washington, DC, 2016, pp. 274-281.https://doi.org/10.1109/CHASE.2016.18.

[5] D. Dahlem, D. Maniloff, and C. Ratti, "Predictability bounds of electronic health records," In *Scientific Reports volume 5*, 2015.

[6] J. Fan, Y. Wu, M. Yuan, D. Page, J. Liu, I. M. Ong, P. Peissig, and E. Burnside, "Structure-leveraged methods in breast cancer risk prediction," *Journal ofMachine LearningResearch*, Volume 17, 2016, Pages1-15.

[7] A. Gaudinat, N. Grabar, and C. Boyer, "Machine learning approach for automatic quality criteria detection of health web pages," In *12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*, 2007.

[8] S. Grover and G. S. Aujla, "Twitter data based prediction model for influenza epidemic," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2015, pp. 873-879.

[9] S. S. Holzinge Andreas., "Biomedical informatics: discovering knowledge in big data," In *Springer, New York*, 2014.https://doi.org/10.1007/978-3-319-04528-3.

[10] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," In *Nature Reviews Genetics volume 13*, 2012.

[11] I. Kamkar, S. K. Gupta, D. Phung, and S. Venkatesh, “Stable feature selection for clinical prediction,” *J. of Biomedical Informatics*, Volume 53, pp.277 – 290, Feb. 2015.https://doi.org/10.1016/j.jbi.2014.11.013.

[12] By Li Li, Wei-Yi Cheng, Benjamen S. Glickberg, Omri Gottesman, Ronald Tamler, Rong Chen, Erwin P. Bottinger, Joel T. Dudley, “Identification of type 2 diabetes subgroups through topological analysis of patient similarity,”In *Sci. Transl. Med. 7*, 2015.https://doi.org/10.1126/scitranslmed.aaa9364.

[13] D. LaFreniere, F. Zulkernine, D. Barber and K. Martin, "Using machine learning to predict hypertension from a clinical dataset," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, 2016, pp. 1-7.https://doi.org/10.1109/SSCI.2016.7849886.

[14] J. Melendez, B. van Ginneken, P. Maduskar, R. H. H. M. Philipsen, K. Reither, M. Breuninger, I. M. O. Adetifa, R. Maane, H. Ayles, and C. I. Snchez, "A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays," in *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 179-192, Jan. 2015.https://doi.org/10.1109/TMI.2014.2350539.

[15] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “Deep patient: An unsupervised representation to predict the future of patients from the electronic health records,” In *Scientific Reports volume 6*, 2016.

[16] R. Miotto and C. Weng, ”Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials,”*Journal of the American Medical Informatics Association*, Volume 22, pp.141–150.https://doi.org/10.1093/jamia/ocu050.

[17] T. Murota, A. Kato and T. Okumura, "Emergency management for information systems in public health a case study of the 2009 pandemic-flu response in Japan," *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, Mannheim, 2010, pp. 394-399.https://doi.org/10.1109/PERCOMW.2010.5470637.

[18] T. N. P, Y. P. P., D. R., and A. R. B., “Data-driven prediction of drug effects and interactions,” In *Sci. Transl Medicine, Volume4*, 2012.

[19] K. Patil, Q. Jawadwala and F. C. Shu, "Design and Construction of Electronic Aid for Visually Impaired People," in *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 172-182, April 2018.https://doi.org/10.1109/THMS.2018.2799588.

[20] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potentials,” In *Health Information Science and Systems, volume 2, Issue1*, 2014.

[21] W. J. Roy and W. F. Stewart, “Prediction modeling using ehr data: challenges, strategies, and a comparison of machine learning approaches,” In *Medical care, volume 48, Issue 6, 2010, pp. 106-113*.

[22] L. Runzhi, L. Wei, and L. Yusong, “An ensemble multilabel classification for disease risk prediction, “In *Journal of Healthcare Engineering*, 2017.

[23] M. Shankar, M. Pahadia, D. Srivastava, T. S. Ashwin and G. R. M. Reddy, "A Novel Method for Disease Recognition and Cure Time Prediction Based on Symptoms," *2015 Second International Conference on Advances in Computing and Communication Engineering*, Dehradun, 2015, pp. 679-682. https://doi.org/10.1109/ICACCE.2015.66.

[24] B. A. Thakkar, M. I. Hasan and M. A. Desai, "Health Care Decision Support System for Swine Flu Prediction Using Naïve Bayes Classifier," *2010 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, 2010, pp. 101-105.

[25] G. Tsoumakas, I. Katakis and I. Vlahavas, "Random k-Labelsets for Multilabel Classification," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079-1089, July 2011. https://doi.org/10.1109/TKDE.2010.164.

[26] P. Watcharapasorn and N. Kurubanjerdjit, "The surgical patient mortality rate prediction by machine learning algorithms," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Khon Kaen, 2016, pp. 1-5.https://doi.org/10.1109/JCSSE.2016.7748844.

[27] Y. Wu, E. S. Burnside, J. Cox, J. Fan, M. Yuan, J. Yin, P. Peissig, A. Cobian, D. Page, and M. Craven, "Breast Cancer Risk Prediction Using Electronic Health Records," *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, Park City, UT, 2017, pp. 224-228.https://doi.org/10.1109/ICHI.2017.62.

[28] J. Zhang, J. Gong and L. Barnes, "HCNN: Heterogeneous Convolutional Neural Networks for Comorbid Risk Prediction with Electronic Health Records," *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, Philadelphia, PA, 2017, pp. 214-221.https://doi.org/10.1109/CHASE.2017.80.

[29] A. Maxwell *et al.*, “Deep learning architectures for multi-label classification of intelligent health risk prediction,” *BMC Bioinformatics*, vol. 18, no. 14, p. 523, Dec. 2017.https://doi.org/10.1186/s12859-017-1898-z.

[30] G. Azar, C. Gloster, N. El-Bathy, S. Yu, R. H. Neela, and I. Alothman, “Intelligent data mining and machine learning for mental health diagnosis using genetic algorithm,” *IEEE Int. Conf. Electro Inf. Technol.*, vol. 2015–June, pp. 201–206, 2015.https://doi.org/10.1109/EIT.2015.7293425.

[31] Who-mental disorders affect one in four people. http://www.who.int/whr/2001/media_centre/press_release/en/. (Accessed on 04/26/2018).

[32] Suicide in india - wikipedia. https://en.m.wikipedia.org/wiki/Suicide_in_India. (Accessed on 04/26/2018).

[33] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, “Machine learning framework for the detection of mental stress at multiple levels,” *IEEE Access*, vol. 5, pp. 13545–13556, 2017.https://doi.org/10.1109/ACCESS.2017.2723622.

[34] A. Sau and I. Bhakta, “Predicting anxiety and depression in elderly patients using machine learning technology,” *Healthc. Technol. Lett.*, vol. 4, no. 6, pp. 238–243, 2017.https://doi.org/10.1049/htl.2016.0096.

[35] N. Jaques, S. Taylor, A. Sano, and R. Picard, “Predicting Tomorrow’s Mood, Health, and Stress Level using Personalized Multitask Learning and Domain Adaptation Ognjen (Oggi) Rudovic,” *J. Mach. Learn. Res.*, vol. 66, pp. 17–33, 2017.

[36] G. Mikelsons and M. Smith, “Towards Deep Learning Models for Psychological State Prediction using Smartphone Data : Challenges and Opportunities,” no. Nips, pp. 1–6, 2017.

[37] R. Wang, F. Chen, Z. Chen, T. Li, and G. Harari, “StudentLife : Assessing Mental Health , Academic Performance and Behavioral Trends of College Students using Smartphones,” 2014.

[38] S. Zaman and Rizoan Toufiq, “Codon Based Back Propagation Neural Network Approach to Classify Hypertension Gene Sequences,” pp. 443–446, 2017.

[39] A. Sathyanarayana, J. Srivastava, and L. Fernandez-Luque, “The Science of Sweet Dreams: Predicting Sleep Efficiency from Wearable Device Data,” *Computer (Long. Beach. Calif).*, vol. 50, no. 3, pp. 30–38, 2017.https://doi.org/10.1109/MC.2017.91.

[40] G. Valenza *et al.*, “Predicting Mood Changes in Bipolar Disorder Through Heartbeat Nonlinear Dynamics,” *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 4, pp. 1034–1043, 2016.https://doi.org/10.1109/JBHI.2016.2554546.

[41] John Devapriam, A. Rosenbach, and R. Alexander, “In-patient services for people with intellectual disability and mental health or behavioural difficulties,” *BJPscyh Adv.*, vol. 21, no. 2, pp. 116–123, 2015.https://doi.org/10.1192/apt.bp.113.012153.

[42] N. Satyanarayana, Y. Ramadevi, and K. Koteswara Chari, “High Blood Pressure Prediction based on AAA using J48 Classifier,” *2018 Conf. Signal Process. Commun. Eng. Syst.*, pp. 121–126, 2018.