International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI and Beyond: A Comparative Review of Health Forecasting Systems

MEENA M
Assistant Professor, Department of Information Technology, K.L.N College of Engineering, Madurai.
email- meena4neem@gmail.com

Abstract- Health forecasting is wellbeing circumstances or illness scenes and admonishing future occasions. It is also a sort of medicine or preventive care that engages public health planning and is aimed toward facilitating health care service provision in populations. The goal of predictive medicine is to predict the probability of future disease so as that health care professionals and thus the patient themselves are often proactive in instituting lifestyle modifications and increased physician surveillance. To understand the evaluation and evolution of health care prediction system various research based on machine and deep learning techniques was studied and various research is going on this way also. In this paper, we have discussed some of the popular techniques for health care prediction and we also present here about the performance of each technique.

Keywords: Healthcare, predictive analysis, machine learning, deep learning.

I.INTRODUCTION:

Being healthy ought to be a piece of your general way of life. Carrying on with a sound way of life can assist with forestalling ongoing infections and long haul sicknesses. An effective medical services framework can add to a major a piece of a nation's economy, improvement, and industrialization. Medical care is ordinarily viewed as a critical determinant in advancing the by and large physical and mental state and prosperity of people round the world. Prescient medication might be a field of medications that involves foreseeing the likelihood of sickness and establishing preventive measures in order to either forestall the infection inside and out or altogether decline its effect upon the patient. The eventual fate of medication's center may possibly move from treating existing infections, normally late in their movement, to forestalling illness before it sets in.

Health care prediction methods can be categorized as statistical methods, machine learning and data mining, network based approach and deep learning approach. Statistical methods focus on collecting, analyzing, interpreting, presenting and organizing numerical data. Machine learning involves the study of algorithms which will extract information automatically. Data Mining is an iterative cycle of finding different kinds of new and valuable examples that are intrinsic to the information. Data mining learning methods can be supervised or unsupervised based on the training dataset. In Network based approach, a network can be represented as a graph that consists of a set of nodes and edges. Nodes symbolize entities, whereas edges symbolize the relations between entities. Deep learning is a man-made insight (AI) work that impersonates the functions of the human cerebrum in handling information and making examples to be utilized in choosing Learning can be supervised, semi-supervised or unsupervised. This paper developed to discuss various methods used for health care prediction. The performance of health care prediction with various methods has been analyzed. The overall organization of paper is as follows: Section I describes Introduction. Section II explains with literature survey. Section III gives the conclusion.

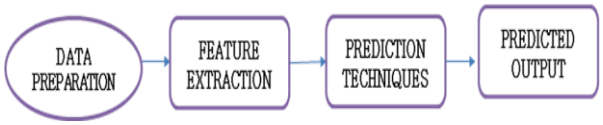


Figure1: General HealthCare Prediction Model

A) DATA PREPARATION: Data preprocessing is a process of preparing the raw data and making it suitable for a healthcare prediction model. It is the first and crucial step while creating a predictionmodel.

B) FEATURE EXTRACTION: Feature extraction is a piece of the dimensionality decrease measure, in which, an underlying arrangement of the crude information is isolated and diminished to more sensible gatherings

c) PREDICTION TECHNIQUES: Various prediction techniques such as statistical methods, data mining, machine learning etc can be used to predict the severity of disease.

D) PREDICTED OUTPUT: Based on the result of prediction techniques result can be categorized as low, medium and high risk.

II.LITERATURE REVIEW

Zahra Sadeghi et al [1], presented Decision trees or Rule based systems to provide explanations for specific diagnosis based on symptoms,results,and medical history.Also proposed predictive model for disease progression using anchors to identify features that influences prediction regarding diseases. However there is no specific collection of spatially well defined and consistently located features

Madhura Ingalhalikar et al. [2], presented Harmonization of multi-site neuron imaging data using the ComBat technique to address the curse of dimensionality caused by the larger sample size in the context of machine-learning based diagnostic classification. It is supported an empirical Bayes formulation to get rid of inter-site differences in data distributions, to enhance diagnostic classification accuracy. ABIDE (Autism Brain Imaging Data Exchange) multisite data is used for classifying individuals with Autism from healthy controls using resting state fMRI-based functional connectivity data. Its accuracy is 71.35%.

Zhongzhi Xu et al. [3], proposed a comorbidity knowledge aware (CKA) model for the task of patient's risk of disease development. Risk propagation model Framework takes one patient and one target disease as the input, and outputs the predicted probability that the patient will develop the disease in the future. In Disease Risk Propagation, For each disease, use one-hot embedding to initialize its embedding vector. Given the one-hop propagation set of patient p and disease d, calculate the relevance probability for the ith disease-label-disease component is calculated and for each aggregation is feed into a corresponding LSTM unit to derive the 1-order response of patient p's history diagnoses to target disease d. Patient's response of last hop contains all the information from previous hops and finally patient embedding and disease embedding are combined to output the anticipated disease risk. Maximum posterior estimation problem ,maximize the posterior probability of model parameters , given the co morbidity network G and the historical disease records Y. However this method faces some limitation as Patients may develop quite different diseases even though they have same historical diagnoses. This is due to personal information except medical diagnoses cannot be accessed due to data sensitivity. And only the co-occurring frequency is considered as the external knowledge. Other relation types among diseases, such as shared genes and shared proteins are not considered. CKA reported its accuracy as 72.7%.

Ming Dong et al.[4], presented Long Short-term Memory Network based sequence learning model to predict long-term health indices for power asset classes. In Long Short-term Memory Network, the hidden layer nodes of the feed-forward neural network at the previous time step are connected with the hidden layer nodes of the feed-forward neural network at the following time step. The training of any neural network is driven by the gradient produced from the loss function at the output layer. The gradient propagates back from the output layer to the front layers of the neural network. When they are too small, the problem is called vanishing gradient and when they are too large, the problem is called exploding gradient and the training on the front layers cannot stabilize Softmax Layer is used for output layer. It is a function that converts a vector of numbers into a vector of probabilities, where the possibilities of every value are proportional to the relative scale of each value in the vector. Each value in the output of the softmax function is interpreted as the probability of membership for each class. It reported precision of 83% and recall of 82%.

Aniello De Santo et al.[5], presented LSTM (Long Short Term Memory) based model combining SMART attributes and temporal analysis to extend the Remaining Useful Life (RUL) of hard drives, and to minimize service shortage and data loss in hard disk drive health assessment. It consists of 3 steps. Hard drive health degree definition status (or health level) is defined for each hard drive according to its time to failure; Sequences extraction: in which sequences in a specific time window are extracted for each hard drive; Health Status assessment through LSTM: in which a health level is associated to each temporal sequence. It is designed with the purpose

of learning long-term dependencies. Input to each LSTM layer is a three dimensional data structure of size $z \times w \times n$, where: z - total number of sequences (or the batch size at each iteration); w - size of each sequence -that size of a time window, in terms of time steps; n - total number of features describing each time step. However an extensive, detailed investigation of different health degree settings, evaluating the trade-offs of incorporating constraints from real-world applications has to be done. The model reported efficiency as 98.45%, precision as 98.33% and recall as 98.34%

Gavin Tsang et al. [6], presented Entropy regularization with ensemble deep neural networks (ECNN) to perform feature decrease on a scanty, high-dimensional dataset of clinical occasions. Deep neural networks consist of multiple layers of preceptor, each aggregate the given input space through a weighted and biased sum. In Entropy weight regularization, probability mass function, P(X), the information entropy variable will approach zero and highest at the midpoint, P(X) = 0.5.It classifies as i) Disconnected ii) Partially connected iii)Fully Connected. Snapshot ensembles allow for multiple ensemble NNs to be generated through training a single model. In Feature ranking & selection, features can be categorized based upon the sparse weight matrix. However times-series based modeling methods should be used to acknowledge the continually changing health of the individual patient over time. And some of the greater statistical analysis of ranked features for improved ranking should be used. ECNN reported accuracy in term of mean ± standard deviation as 75.5%.

Ai-Min Yang et al. [7], presented improved TSVR algorithm to predict the recurrence time of cancer patients, and to improve the survival rate of patients by corresponding clinical interventions. Twin support regression vector machine (TSVR) is based on Dependent nearest neighbor (DNN) where the target sample is mapped to the center point of the 2D vector plane. Eplion-TSVR model may be a improved DNN weighted algorithm with local information mining function. Cuckoo algorithm is used to determine the optimal parameters of DNN to determine the optimal DR domain. However improved cuckoo algorithm should be used for more stable and faster search and model prediction accuracy must be improved. E-TVSR model reported prediction accuracy of 91%.

Walaa N. Ismail et al.[8], presented CNN based regular health data analysis model to address the issues of increase in computation load and huge memory requirement in Convolution neural network (CNN) model. Correlation coefficient method classifies the input data as the positively and negatively correlated health factors. CNN-based health knowledge model takes set of selected strong correlated factors as input and calculate the regularity of attribute e. If e satisfies the defined regularity threshold value given by the user, add to the subset of the regular candidate set. It produces the set of regular correlated attributes as output. However, different raw data preprocessing methods must be considered as data quantification methods will affect the CNN model learning accuracy and performance. It reported accuracy as 95%.

Hassan Harb et al.[9], proposed sensor-based data analytics for real-time patient monitoring and assessment to address problem of the limited sensor energies, and the prediction of the progress of patient situation for heath-based IoT applications. Model is divided into 3 phases such as Emergency detection, adapting sensing Frequency and real time prediction of patient situation using Long Short-term Memory (LSTM). In emergency detection, periodic patient monitoring is done by biosensor. Early Warning Score used to track the criticality, range 0 to 3,0-normal, 3-critical. Adapting sensing frequency phase classify patient as low, medium, high risk based on frequency. Real time prediction of patient situation phase uses Long Short-term Memory (LSTM), receive the training dataset, perform data normalization using Min-Max scaling [0,1] and determine the neural network parameters such as number of block, number of time steps(ε), number of features(F).Training the LSTM contain 2 function such as loss function to calculate variation between training data and Optimizer to optimize error in loss function using Mean Square Error. However, Sensor does not take account of the correlation between neighboring nodes when sending the data to the sink this lead to repeated collision and introduce a phase shift between the two transmission sequences.

Md. Ekramul Hossain et al.[10], discussed different risk prediction models based on electronic health data and its limitation. Methods include statistical methods, machine learning and data mining approach, network based approach. Statistical methods include Regression model and Cox Proportional Hazards. Data mining learning can be supervised or unsupervised. Data Mining supervised algorithm include Artificial Neural Network, Support

Vector Machine, Decision Tree and Random Forest, Rule-based Scoring. Data mining unsupervised algorithm include association analysis. Authors also discussed limitation of various methods.

Ying An, Nengjun Huang et al.[11], presented DeepRisk model based on attention mechanism and deep neural Networks to deal with the feature representation for sequential, high-dimensional heterogeneous EHR data. It takes 4 inputs such as "diagnosis sequences", "diagnosis and laboratory sequences", "laboratory sequences" and demographic data". Inputs are respectively loaded into their embedding parts to produce independent embedding vectors. Train four models based on attention-based deep neural networks by utilizing Bi-LSTM both forward and backward direction. Representation Learning produce representation vectors for patients. The connected portrayal is taken care of into softmax layer to anticipate the high-hazard of patients. However, the performance of Deeprisk depends on high quality temporal data. On the opposite hand, its effectiveness on prediction tasks for other diseases got to be further validated. A-LSTM reported precision as 0.7055 ± 0.0128 and recall as 0.6160 ± 0.0111.

Guanjin Wang et al. [12], presented additive LS-SVM based Classifier to tackle the issues of missing and noisy data in community health research. Input dataset X is associated with two separate classes with the class labels +1 and -1 stored in the output dataset Y. Additive Gaussian kernel calculate the corresponding values of the kernel depending on whether the features contain missing values or not. Fast Leave-One-Out Cross Validation used to determine the optimal value for choice of the parameter. Inf used to measure of the degree of influence. The greater the value of Inf, the more significant the influence of a feature with missing values on the classification performance. However the model developed using the data may not be representative for a broader population of elderly people. Additive LS-SVM based Classifier reported the mean accuracy of 74.38% and the maximum accuracy of 76.12%.

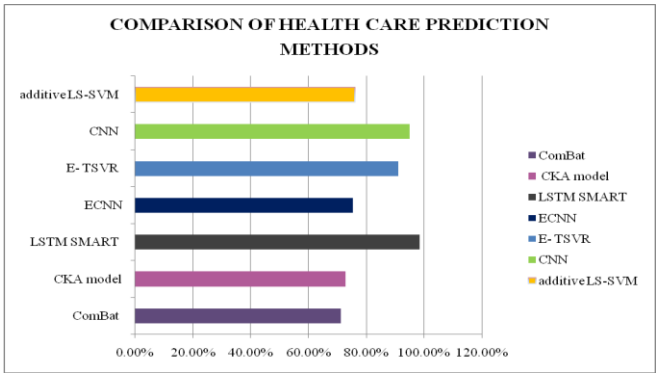
Olav Skrøvseth et al. [13], presented Feature Selection (FS) strategies based on the weights of the linear SVM to perform the task of early detection of anastomosis leakage (AL), a severe difficulty after elective medical procedure for colorectal malignant growth (CRC) medical procedure, utilizing free content extricated from EHRs. It was performed in 3 steps a simple statistical criterion (leave one-out test); an intensive-computation statistical criterion (Bootstrap re sampling test); and advanced statistical criterion (kernel entropy based threshold).But the results have to be enhanced to support early diagnosis decisions. This method reported sensitivity and specificity as 81% and 85% respectively.

TABLE 1: OVERVIEW OF HEALTH CARE PREDICTION APPROACHES

REF. NUMBER.	METHODOLOGY	PROS	CONS
[1]	Bagged tree-based classifier (BTBC), DT, RF, XGB, LR, NB	Patient-independent multimodel data using the proposed framework	No specific collection of spatially well defined and consistently located features
[2]	co morbidity knowledge aware (CKA) model	Knowledge-aware prognostic prediction model.	Historical diagnoses cannot be done due to data sensitivity
[4]	LSTM based model with SMART attributes and temporal analysis	Extend the RUL and minimize service shortage and data loss	Different health degree settings were not considered.
[5]	ECNN	Feature reduction of a sparse, high-dimensional dataset.	Continually changing health of the individual patient over time was not considered.
[6]	E-TSVR	Predict the recurrence time	Nutritional indicators does not improve the patient's survival time
[7]	CNN based data analysis model	Address the issues of increase in computation load and huge memory requirement in CNN	Data preprocessing methods affect the CNN model learning accuracy and performance .
[8]	Sensor-based data analytics	Address problem of the limited sensor energies, and the prediction progress of patient situation.	Repeated collision and phase shift between the two transmission sequences.
[10]	A-LSTM	Deal with the feature representation for sequential, high-dimensional heterogeneous EHR data	Performance depends on high quality temporal data.
[11]	Additive LS-SVM based Classifier	Tackle the issues of missing and noisy data	Not representative for a broader population

TABLE II: COMPARISON OF ACCURACY FOR THE TASK OF HEALTH CARE PREDICTION

AUTHOR	METHOD	ACCURACY
Zahra Sadeghi et al [1]	Bagged tree-based classifier (BTBC), DT, RF, XGB, LR, NB	99.5
Madhura Ingalhalikar et al. [2]	ComBat	71.35%
Zhongzhi Xu et al. [3]	CKA model	72.7%
Aniello De Santo et al.[4]	SMART LSTM	98.45%
Gavin Tsang et al. [5]	ECNN	75.5%
Ai-Min Yang et al. [6]	E- TSVR	91%
Walaa N. Ismail et al.[7]	CNN	95%
Guanjin Wang et al. [11]	additive LS-SVM	76.12%



III.CONCLUSION

A reliable health forecast significant for wellbeing administration conveyance, since it can upgrade preventive medical care/administrations; make alarms for the administration of patient floods (in circumstances of pinnacle interest for medical care administrations); and significantly reduce the associated costs in supplies and staff .If we are considering statistical method, it does not consider the linear relationship between variables and does not show great precision within the sight of complex connections among input factors. Data mining approach sets aside a ton of computational effort to prepare the organization for an unpredictable order issue. Whereas traditional network models do not have longitudinal or spatial aspects that are needed to predict disease risks. Various machine learning and deep learning methods and its performance have been studied. In this paper, an investigation of different medical care forecast method is proceeded just as different likeness estimations strategies are considered.

REFERENCES

[1] Zahra Sadeghi a , Roohallah Alizadehsani b,* , Mehmet Akif CIFCI c,d,e "A review of Explainable Artificial Intelligence in healthcare" Computers and Electrical Engineering (2024)

[2] Madhura Ingalhalikar, Sumeet Shinde, Arnav Karmarkar, Archith Rajan, D. Rangaprakash, Gopikrishna Deshpande," Functional connectivity-based prediction of Autism on site harmonized ABIDE dataset", IEEE Transactions on Biomedical Engineering, 14 May 2021

[3] Zhongzhi Xu , Jian Zhang , Qingpeng Zhang , *Senior Member, IEEE*, Qi Xuan , *Member, IEEE*, and Paul Siu Fai Yip," A Comorbidity Knowledge-Aware Model for Disease Prognostic Prediction", IEEE Transactions on Cybernetics, 07 May 2021

[4] Ming Dong, *Senior Member, IEEE*, Wenyuan Li, *Life Fellow, IEEE*, Alexandre B. Nassif, *Senior Member, IEEE,"* Long-term Health Index Prediction for Power Asset Classes Based on Sequence Learning", IEEE Transactions on Power Delivery, 29 January 2021

[5] Aniello De Santo, Antonio Galli, Michela Gravina, Vincenzo Moscato, Giancarlo Sperl," Deep Learning for HDD health assessment: an application based on LSTM", IEEE Transactions on Computers, 02 December 2020

[6] Gavin Tsang; Shang-Ming Zhou; Xianghua Xie," Modeling Large Sparse Data for Feature Selection: Hospital Admission Predictions of the Dementia Patients Using Primary Care Electronic Health Records", IEEE Journal of Translational Engineering in Health and Medicine (Volume: 9), 24 November 2020

[7] Ai-Min Yang; Yang Han; Chen-Shuai Liu; Jian-Hui Wu; Dian-Bo Hua," D-TSVR Recurrence Prediction Driven by Medical Big Data in Cancer", IEEE Transactions on Industrial Informatics (Volume: 17, Issue: 5, May 2021), 24 July 2020

[8] Walaa N. Ismail; Mohammad Mehedi Hassan; Hessah A. Alsalamah;," CNN-Based Health Model for Regular Health Factors Analysis in Internet-of-Medical Things Environment", IEEE Access (Volume: 8),16 March 2020

[9] Hassan Harb; Ali Mansour; Abbass Nasser; Eduardo Motta Cruz; Isabel de la Torre Díez," A Sensor-Based Data Analytics for Patient Monitoring in Connected Healthcare Applications", IEEE Sensors Journal (Volume: 21, Issue: 2, Jan.15, 15 2021), 02 March 2020

[10] Md. Ekramul Hossain; Arif Khan; Mohammad Ali Moni; Shahadat Uddin," Use of electronic health data for disease prediction: A comprehensive literature review", IEEE/ACM Transactions on Computational Biology and Bioinformatics (Volume: 18, Issue: 2, March-April 1 2021), 27 August 2019

[11] Ying An, Nengjun Huang, Xianlai Chen, FangXiang Wu, *Senior Member, IEEE* and Jianxin Wang, *Senior Member, IEEE,"* High-risk Prediction of Cardiovascular Diseases via Attention-based Deep Neural Networks", IEEE/ACM transactions on computational biology and bioinformatics, january 2019

[12] Guanjin Wang; Zhaohong Deng; Kup-Sze Choi," Tackling Missing Data in Community Health Studies Using Additive LS-SVM Classifier", IEEE Journal of Biomedical and Health Informatics (Volume: 22, Issue: 2, March 2018), 01 December 2016

[13] Olav Skrøvseth, Fred Godtliebsen, Kim Mortensen, Arthur Revhaug, Rolv-Ole Lindsetmo, Knut Magne Augestad, Robert Jenssen *Member, IEEE,"* Support Vector Feature Selection for Early Detection of Anastomosis Leakage from Bag-of-Words in Electronic Health Records", IEEE Journal of Biomedical and Health Informatics (Volume: 20, Issue: 5, Sept. 2016), 08 October 2014.