Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 10, October 2021: 568-574

A Review of Constructive Deep Learning Approaches for a Health Care System

[1]S.T.Shenbagavalli , [2]Dr.D.Shanthi
[1]Assistant Professor, Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Kothandaraman Nagar, Tamil Nadu 624622
Email : mailatshenbagavalli@gmail.com
[2]Professor & Head, Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Kothandaraman Nagar, Tamil Nadu 624622
Email: dshan71@gmail.com

Abstract

Human beings face various diseases due to ecological condition and their living behaviours. So, early prediction of diseases is needed for timely treatment. Data mining based prediction of diseases is considered a promising solution for early prediction. Acquiring knowledge from high-dimensional and heterogeneous medical data is considered a difficult task in data mining. To overcome the difficulties in medical data, deep learning (DL) algorithms are used to extract hidden information accurate from the data set and produce efficient prediction results. In this paper, a complete review of DL algorithms proposed by various researchers in medical data are explained. Then, difficulties in DL algorithms performance and parameters used for evaluating DL algorithms are explained. Finally, preceding suggestions for further research.

Keywords : Data mining, Deep Learning ,Medical Data, Data Set

1. Introduction

Good health of the human contributes a major role in the all-around growth of the world where biomedical health-related data are playing a pivotal role. The proper medical treatment for a particular disease needs correct and timely prediction or diagnosis of diseases. The prediction or diagnosis of diseases should be considered patients data like electronic health records (EHRs) , environment, and variability in molecular behaviours.

The huge amount of data available in the medical field adds more challenges and difficulties to health care research. The different pieces of medical data combined for predictive analysis and detection. Recently, the use of data mining algorithms in health care data is an interesting field for early prediction and treatment. Data mining algorithms include both machine learning(ML) and deep learning (DL) algorithms to process structured or unstructured medical data. The structured data includes diagnosis, medicines, laboratory tests and unstructured includes clinical notes. The example for data, mining algorithms are Decision trees, Artificial Neural Networks(ANN's) , Naïve Bayes Classifier, Support Vector Machines and K-Nearest Neighbour etc. ML algorithms may suffer from overfitting and are often relatively erroneous. It may lead to complex decision-making condition

DL is also a subset of ML where computations or processing of data's are motivated by the human brain. DL network consists of neurons to process the data feature based on weights and priority. DL is completed different from ML which process the data by multiple layers. The main difference between conventional ANN to DL is the number of layers count to train and learn the data. Every layer of the DL model denotes the observed patterns or features of input data which are performed in a nonlinear manner.

The difficulties associated with the DL algorithm are Data volume, Data quality, Temporality and Interpretability. Data volume issues related to the amount of data is needed to train the model. the complexity of the model depends on the amount of data trained for feature learning. Data used for training may be heterogeneous, noisy and imbalanced. The missing of data in training leads to reduce prediction quality of the model. Due to the time-varying nature of health data, trained models are to be updated continuously to tackle the accuracy issues. DL techniques in medical data are more complicated when compared to image and speech processing, applications. Further, the limitation of patients and heterogeneity of available data leads to the design of a model more complex.Section 2 discussed a detailed survey of proposed s DL models. Section 3, describes performance parameters used for prediction evaluation Finally, we conclude this work by summarizing the advantages of DL algorithms with future work.

2. Related work

Chen, J et al 2021 have proposed a hybrid deep learning-based stroke risk prediction algorithm by processing collected data. The attributes of age, Number of lymphocytes, Number of basophils, Potassium, Calcium and Fibrinogen has been used for forecasting. Generative Instance learning is applied for external stroke data processing and active instance transfer is applied for informative instances. Active instance learning applied for extremely correlated diseases. The proposed hybrid learning model outperforms all the performance parameters for both increasing and decreasing transferring layers.

Shahinda Mohamed et al 2021 kidney disease classification and prediction model using deep learning. The improved Deep Belief Network model is applied to extract relevant information from a data set. The deep Belief Network model includes stacked restricted Boltzmann machines to accurately predict and classify the data. Compared to another traditional model, the proposed model achieves a higher accuracy of 98.5% and sensitivity of 87.5 %.

Zhao, S et al 2020 have proposed a deep learning model for mortality prediction in the Intensive Care Unit . the data set collected from the Chinese hospitals includes the attributes of body temperature, pulse rate, respiration level, sugar level, mean arterial pressure and haemoglobin rate for risk prediction. The proposed model includes ten layers for training and testing . Results show that the proposed model achieves higher region of convergence values and lower brier scores compared to other models.Tsang, G et al 2021 have proposed an ensemble deep neural network with entropy regularization for processing electronic health records. The proposed model includes four stages: training using entropy weight regularization,backwards-stepwise feature selection, prediction and performance analysis by cross-validation. The proposed method shows higher accuracy than random forest and logic regression-based prediction algorithms

Sevi, M et al 2020 have proposed a modified Long short-term memory (LSTM) based COVID 19 prediction algorithm to make an aware of spreading rate. LSTM is an improved version of recurrent neural network that combines the advantages of both neural network and recurrent neural network. Data set collected from a different place in a country used to train a model.

Attiga, Y et al 2018 have developed a tensor flow-based neural network model for forecasting thyroid disease. The proposed model tested for 747,301 samples collected from different places. To balance missing data in data sets, down-sampling up-and sampling methods have been used. The performance of the proposed model was analyzed for varying activation functions like ReLu, sigmoid and softplus etc.Shamout, F. E et al 2019 have proposed a deep early warning model for predicting cardiac arrest. The features of mean and variance are extracted using the Gaussian Process Regression model. A total of 45314 vital signs of patients features were used for training the model. Results show that the proposed model has produce higher accuracy and true positive rates than other prediction models. Che, Z et al 2017 have analysed the difficulties of data prediction in the field of Computational Phenotyping. The difficulties of phenotyping are: how to handle lost data, how to develop a scalable model and how to differentiate features. Developed a learning mode using Gated Recurrent Units (GRU) [14], to solve the mentioned difficulties and to achieve higher prediction accuracy.

Jeremic, A et al 2020 have proposed an age-dependent multinomial regression model to process the health record of patients. The data set was separated into three portions:health status , pain remark and accident data. The proposed regression model includes 10 layers and 6 neurons for processing features. The optimal fusion rule is integrated to get better scalability based on data sets.Lopez-Martinez, D et al 2019 have proposed a sequential decision-making algorithm for calculating opioid dosing levels in ICU patients. The level of opioids dosage varied based on pain level. The proposed deep learning model utilized reinforcement learning to classify the pain level and quantify the amount of opioid dosage level automatically. Results show that the proposed model achieves an accuracy of about 96% compared to other machine learning models.

Cao, Y. et al 2020 have presented a hybrid model for processing EEG and EMG data of health records. The proposed model uses Stacked Auto-Encoder (SAE) to extract useful features from the biological data. The hybrid model combines a neural network with AdaBoost classifier for prediction. Compared to another model, the proposed model increases the accuracy by 12% for various data sets.

Huang, Z et al have introduced a stacked denoising auto-encoder (SDAE) prediction model for electronic health record mining. It is based on the symmetrical neural network used to handle the heterogeneity of health data. By considering patient health risks, the discriminating features of data sets are extracted carefully. The classification results of the model include low risk, medium risk and high risk.

Farhadi, A et al 2019 have proposed deep transfer learning (DTL) approach for breast cancer data classification. The proposed DTL has three layers of the input layer, an output layer and three hidden layers with a total of 8880 parameters. The hyperparameters of the model tuned by using search algorithm and backpropagation concept are used to adjust the weights.

Gao, C et al 2019 have proposed a regularized logistic regression model for Neonatal Encephalopathy(NE) prediction in health data. NE is a disorder that includes depressed neurological function caused by a deficiency of oxygen to the baby during birth. Qanita Bani Baker et al 2021 have proposed a deep learning classification model for differentiating severity level in lung cancer. The LSTM model combined with neural networks to classify the level into moderate or higher levels of infections. The residual errors of LSTM are optimized by neural networks to increase accuracy. Results on data sets show that the proposed optimized LSTM model achieves higher accuracy when compared to other methods.

E Macias et al 2019 have proposed hybrid imputing and deep learning combined accurate sepsis prediction method ICU patients. The proposed algorithm is divided into three steps: Data preparation using autoencoder, Exploiting temporal dependencies using LSTM network and utility score calculation for classification. The attributes of vital signs, demographics, and data from the autoencoder are given as input for LSTM. LSTM predicts sepsis conditions by processing series of data. Hartmann, M et al 2019 have presented a distilled knowledge learning based abnormal classification in ECG data. Three types of models proposed on distilled knowledge learning , teacher model, student model and standalone model. The proposed model was applied in both 14 beat and 3 beat data types and achieved a maximum accuracy level of 93.2%. Recently proposed DL algorithms and their implantation results are given in Table 1.

Table 1: Survey summary

Author	Modality	Method	Performance metric and results
Liu, L et al	IoT Platform	MASK R-CNN	Mean diagnosis time reduced by 37.5%
Klenk, J et al	fall events in older people	DNN	Accuracy rate 81.27
Brandon Ballinger et al	Cardiovascular Risk Prediction	multi-task LSTM	ROC of 0.701
Chang, W.-J et al	Chronic Patients	a deep learning training server	Recognition accuracy reaches 96.6%
W.-C. Wang et al	a hairy scalp detection system	Hybrid learning	Accuracy of 89.77%
S. Sarraf et al	Alzheimer's Disease	LeNet-5,	The accuracy of test data on trained data reached 96.85%.

3. Performance metric for valuation

To evaluate the efficiency of DL algorithms in health data prediction, the following metric has been used. Evaluation Metrics were used to validate the proposed method with existing methods.

Confusion Matrix (CM)

CM is a matrix table used to define the performance of a DL model on a set of test data for which the true values are known as shown in Figure 1. It correlates the true and false real values to obtained values (TP,TN,FN,FP)

		Actual Value	
		Yes (1)	No (0)
Predicted Value	Yes (1)	TP	FP
	No (0)	FN	TN

TP-True Positive
TN-True Negative
FP-False Positive
FN-False Negative

Accuracy

Accuracy is a measure of results to the true value. It defines the ability of the DL model to forecast values correctly.

$$\text{Accuracy} \ = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Specificity

It is also named true negative rates used to measure the ratio of actual negatives that are correctly recognized as such.

$$\text{Specificity} \ = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

F1 Score

F1 score is a measure of a test's accuracy. It reflects both the precision and the recall rate to calculate the score

$$\text{F1Score} \ = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

Sensitivity, precision and recall

These measures were used to find the number of TP and FN. It finds the ratio of real positives that are properly recognized

$$\text{Sensitivity} \ = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} \ = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

ROC Curve

It is an important metric for data classification. It is a measure of the DL model that can accurately distinguish between the two things. a poor model will have complications in differentiating between the two.

3. Conclusion

DL algorithms are used for the system to learn from the data in an effective way, so that they create a new way for the inventions of smart applications like wearable devices. DL supports various clinical activities to modernize the medical field like health risk prediction, treatment suggestions and data analysis. This paper reviewed the various DL based prediction algorithms for medical data. The challenges and open issues in DL algorithms are discussed. . we also study the performance of existing DL methods in terms of performance parameters. Future, the accuracy and performance of DL techniques can be enhanced by adjusting the hyperparameters of models using optimization algorithms.

References

1. Shamout, F. E., Zhu, T., Sharma, P., Watkinson, P. J., & Clifton, D. A. (2019). Deep Interpretable Early Warning System for the Detection of Clinical Deterioration. IEEE Journal of Biomedical and Health Informatics, 1–1.
2. Attiga, Y., Chen, S.-Y., LaGue, J., Ovalle, A., Stott, N., Brander, T., … Francis-Lyon, P. (2018). Applying Deep Learning to Public Health: Using Unbalanced Demographic Data to Predict Thyroid Disorder. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).
3. Sevi, M., & AYDIN, I. (2020). COVID-19 prediction Using Deep Learning Methods. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI).
4. Tsang, G., Zhou, S.-M., & Xie, X. (2021). 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, 9, 1–13.
5. Chen, J., Chen, Y., Li, J., Wang, J., Lin, Z., & Nandi, A. K. (2021). Stroke Risk Prediction with Hybrid Deep Transfer Learning Framework. IEEE Journal of Biomedical and Health Informatics, 1–1.
6. Shahinda Mohamed Mostafa Elkholy, Amira Rezk & Ahmed Abo El Fetoh Saleh, 2021.Early Prediction of Chronic Kidney Disease using Deep Belief Network, IEEE Access (Early Access),pp. 1 – 1.
7. Zhao, S., Liu, P., Tang, G., Guo, Y., & Li, G. (2020). Development and Application of an Intensive Care Medical Data Set for Deep Learning. 2020 IEEE International Conference on Big Data (Big Data).
8. Che, Z., & Liu, Y. (2017). Deep Learning Solutions to Computational Phenotyping in Health Care. 2017 IEEE International Conference on Data Mining Workshops (ICDMW).
9. Jeremic, A., Nikolic, D., Kostadinovic, M., & Milicevic, M. S. (2020). Predicting the Assisted Living Care Needs Using Machine Learning and Health State Survey Data. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).
10. Huang, Z., Dong, W., Duan, H., & Liu, J. (2017). A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records. IEEE Transactions on Biomedical Engineering, 1–1.
11. Cao, Y., Zhang, H., Choi, Y.-B., Wang, H., & Xiao, S. (2020). Hybrid Deep Learning Model Assisted Data Compression and Classification for Efficient Data Delivery in Mobile Health Applications. IEEE Access, 8, 94757–94766.

12. Lopez-Martinez, D., Eschenfeldt, P., Ostvar, S., Ingram, M., Hur, C., & Picard, R. (2019). Deep Reinforcement Learning for Optimal Critical Care Pain Management with Morphine using Dueling Double-Deep Q Networks. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
13. Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S.-C., & Zheng, L. (2019). A Smart Dental Health-IoT Platform Based on Intelligent Hardware, Deep Learning and Mobile Terminal. IEEE Journal of Biomedical and Health Informatics, 1–1.
14. Klenk, J., Schwickert, L., Palmerini, L., Mellone, S., Bourke, A., … Becker, C. (2016). The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls. European Review of Aging and Physical Activity, 13(1).
15. Brandon Ballinger, Johnson Hsieh, Avesh Singh,,DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction, e Advancement of Artificial Intelligence , 1-1.
16. Chang, W.-J., Chen, L.-B., Hsu, C.-H., Lin, C.-P., & Yang, T.-C. (2019). A Deep Learning-based Intelligent Medicine Recognition System for Chronic Patients. IEEE Access, 1–1.
17. W.-C. Wang, L.-B. Chen, W.-J. Chang, “Development and experimental evaluation of machine-learning techniques for an intelligent hairy scalp detection system,” Applied Sciences, vol. 8, no. 5, article 853, pp. 1-28, May 2018.
18. S. Sarraf, Ghassem Tofighi, Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks,(2016),arXiv:1603.08631
19. Hartmann, M., Farooq, H., & Imran, A. (2019). Distilled Deep Learning based Classification of Abnormal Heartbeat Using ECG Data through a Low Cost Edge Device. 2019 IEEE Symposium on Computers and Communications (ISCC).
20. Farhadi, A., Chen, D., McCoy, R., Scott, C., Miller, J. A., Vachon, C. M., & Ngufor, C. (2019). Breast Cancer Classification using Deep Transfer Learning on Structured Healthcare Data. 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA).
21. Gao, C., Yan, C., Osmundson, S., Malin, B. A., & Chen, Y. (2019). A Deep Learning Approach to Predict Neonatal Encephalopathy from Electronic Health Records. 2019 IEEE International Conference on Healthcare Informatics (ICHI).
22. E Macias; G Boquet; J Serrano,Novel Imputing Method and Deep Learning Techniques for Early Prediction of Sepsis in Intensive Care Units, 2019 Computing in Cardiology (CinC), pp.8-11.
23. Qanita Bani Baker; Maram Gharaibeh; Yara Al-Harahsheh,Predicting Lung Cancer Survival Time Using Deep Learning Techniques, 2021 12th International Conference on Information and Communication Systems (ICICS), 24-26 May 2021.