Review of Deep Learning techniques for patient disease prediction using health data

[1]Atiya Khan, [2]Neha Purohit, [3]Mohammed Shahid
[1,2,3]G. H. Raisoni College of Engineering, Nagpur, India.
[1]atiya.khan@raisoni.net, [2]neha.purohit@raisoni.net, [3]*mohammad.shahid@raisoni.net*

Abstract - Chronic diseases like obesity, stroke, depression, and cancer are one of the biggest threats to human life. These Diseases are related with our lifestyle and can affect health on large scale. Disease prediction using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. Many works had applied deep learning techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of disease. The review will focus on the chronic diseases prediction based on historical data available into a multi-label classification problem using various techniques like convolutional neural network, Heterogeneous Convolutional Neural and recurrent neural networks. In Convolutional Neural Networks, the EHR data can be plotted as graph with heterogeneous attributes for risk prediction of multiple comorbid diseases. The paper discuss various deep learning techniques and through a comparative study on patient EHR data, the conclusion can be drawn that HCN achieves better performance than traditional convolutional neural networks on the risk prediction of comorbid diseases. This paper recommend that deep learning methodologies can be used as a vehicle for predicting disease based on huge biomedical data to enhanced human health.

Keywords—Deep Learning, HCN,CNN,RNN, Risk Prediction, Electronic Health Records, disease.

I. INTRODUCTION

Disease prediction using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. Many works have been applied data mining techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of disease. Also, some approaches try to do prediction on control and progression of disease. The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little preprocessing and produce more accurate results. Numbers of papers have been published on several data mining techniques for diagnosis of heart disease such as Decision Tree, Naive Bayes, neural network, kernel density, automatically defined groups, bagging algorithm and support vector machine showing different levels of accuracies in diseases prediction.

Electronic Health Record (EHRs) are systematic collections of longitudinal patient health information, such as diagnoses, medications, lab tests, procedures, and other information to characterize patient health, generated by one or many encounters across multiple healthcare providers or organizations. While EHR data provides a unique opportunity to characterize disease patterns and imminent risk of disease, it also presents many challenges due to its heterogeneity, high-dimensionality, sparsity, irregularity, and bias [11]. To overcome these challenges, previous studies have developed various methods for feature extraction and representation from EHR data.

Deep learning is a new name for large-scale multi-layer back-propagation neural networks, which is revolutionizing many fields, including computer vision, speech recognition, and natural language processing [2]. The emerging types of neural networks for deep learning models can be divided into two categories: unsupervised learning (deep Boltzmann machines and its varieties) and supervised learning (convolutional neural network (CNN), multilayer perception (MLP), and recurrent neural network (RNN)). CNNs have mainly attracted attention for modeling the associations between signal channels in various applications such as image recognition and captioning.

Convolutional Neural Networks (CNNs) have been proposed in to investigate symbolism information. The name of these systems originates from the convolution administrator that is a simple approach to perform complex operations utilizing convolution channel. CNN does not utilize predefined parts, but rather adapts privately associated neurons that speak to information particular portions. Since

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.

these channels are connected more than once to the whole picture, the subsequent network resembles a progression of covering responsive fields. The fundamental favorable position of a CNN is that amid back-propagation, the system needs to modify various parameters equivalent to a solitary case of the channel which definitely lessens the associations from the common NN design [10].

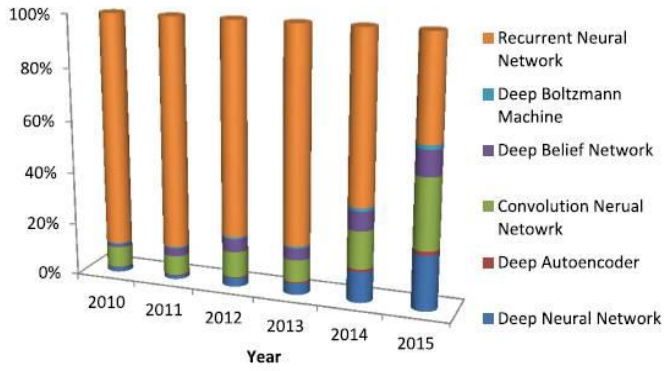

Figure 1: Percentage of most used deep learning methods in healthcare[7]

II. RELATED WORK

There are many existing approaches for feature representation of EHR data and several deep learning techniques targeted at exploring these new representations. In this section, we review some related works.

Data in medical fields, are generally represented in form of images and texts such as MRI, medical notes, and ECG. Various researchers have focused in working with different technologies to extract features from medical data to solve various problems, such as classification, regression, and retrieval. The recent years have witnessed a surge of interests in data analytics with patient Electronic Health Records ***(EHR). Electronic health records (EHRs) contain a wealth of detailed clinical information and provide several distinct advantages for clinical research, including cost efficiency, big data scalability, and the ability to analyze data over time.***

There are various research work which are carried out along with patient EHRs for data analytics, like for example [1] method was capable of Capturing features about the relationships between heterogeneous attributes of the graphs. Through a comparative study on patient EHR data, HCNN achieves better performance than traditional convolutional neural networks on the risk prediction of comorbid diseases**.** [2] The raw EHRs are feed into a CNN model which captures the consecutive sequential information to learn a vector representation, based on supervised classification method and triplet loss based distance metric learning method are used to learn the similarity of patient pairs. MOHD USAMA et al [3] proposed a model based on RCNN for disease risk assessment to extract fine-grain features of chronic infarction disease by utilizing structured and unstructured data and improve the accuracy of existing models.[4] have sketched out how deep learning has empowered the improvement of more data driven arrangements in healthcare by permitting programmed age of highlights that decrease the measure of human intercession in this procedure. Daniele Rav` et al [7] concluded that deep learning has provided a positive revival of NNs and connectionism from the genuine integration of the latest advances in parallel processing enabled by coprocessors as a computational resources. Jyoti Islam et al [9] proposed a deep convolutional neural network for Alzheimer's disease diagnosis using brain MRI data analysis. While most of the existing approaches perform binary classification, their model can identify different stages of Alzheimer's disease and obtains superior performance for early-stage diagnosis. [12] Proposed a deep learning framework for analyzing patient EHRs, where framework is composed of four layers: input layer, one-side convolution layer, max-pooling layer and softmax prediction layer for specific scenario of predictive modeling of chronic diseases.

III. DEEP LEARNING APPROACHES

Deep Learning methods are a modern update to Artificial Neural Networks that exploit abundant cheap computation.

They are concerned with building much larger and more complex neural networks and, many methods are concerned with very large datasets of labelled analog data, such as image, text. audio, and video.

The most popular deep learning algorithms are:

1. Convolutional Neural Network (CNN):- CNN is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used. The deep learning CNN models is used to train and test image, where input image will pass through a series of convolution layers with filters called Kernels, Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. The figure given below is a complete flow of CNN to process an input image and classifies the objects based on values.

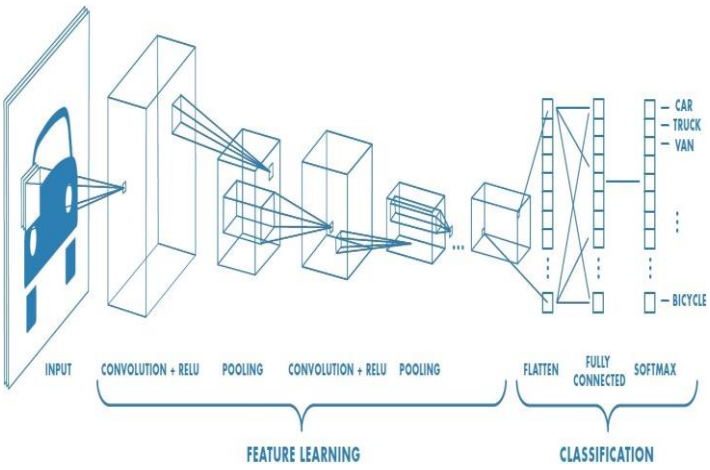

Figure 2 : Neural network with many convolutional layers

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.

2. Recurrent Neural Networks (RNNs)

Recurrent neural networks is a type of deep learning-oriented algorithm, which follows a sequential approach. In neural networks, we always assume that each input and output is independent of all other layers. These type of neural networks are called recurrent because they perform mathematical computations in sequential manner.

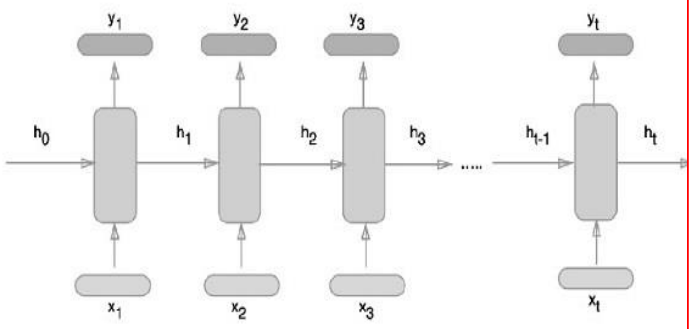

Figure 3:- Recurrent Neural Networks

3. Long Short-Term Memory Networks (LSTMs)

Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long term memory but can give more accurate predictions from the recent information. It is used for processing, predicting and classifying on the basis of time series data.

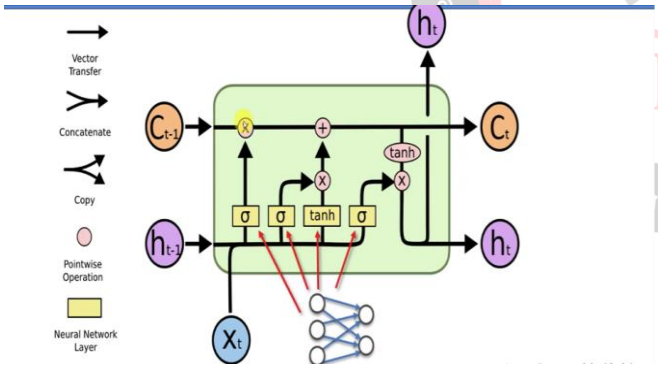

Figure 4:- Long Short-Term Memory Networks

4. Deep Boltzmann Machine (DBM) & Deep Belief Networks (DBN)

Deep Neural Networks (DNNs) are mainly based on stochastic gradient descent and back propagation training algorithms. Two main techniques are used for training DNNs. The first technique is based on a filtering strategy and the second one is based on unsupervised pre-training. The first filtering technique is used by Convolutional Neural Networks (CNNs) to locally filter inputs. Filtering is performed by convolving the input by weight matrices. In the second technique, information processing starts by using an unsupervised learning method. In this stage, unlabelled data can be used. Then, the DNN is fine-tuned by a supervised method using labelled data. Deep Belief Networks (DBN) and DBMs, are examples which use this semi-supervised technique.

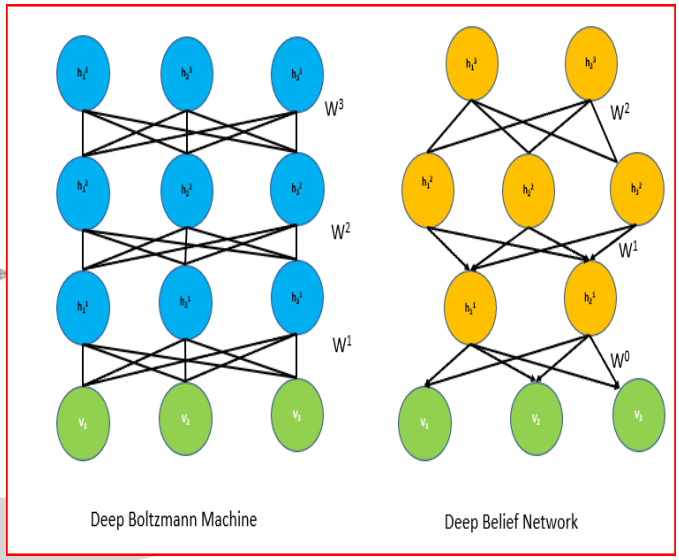

Figure 4:- DBM & DBN

IV. DEEP LEARNING IN HEALTHCARE

Deep learning is assisting medical professionals and researchers to discover the hidden opportunities in data and to serve the healthcare industry better. Deep learning in healthcare provides doctors the analysis of any disease accurately and helps them treat them better, thus resulting in better medical decisions

1. Drug discovery

Deep learning in healthcare helps in discovery of medicines and their development. The technology analyzes the patient's medical history and provides the best treatment for them.

2. Medical imaging

Medical imaging techniques such as MRI scans, CT scans, ECG, are used to diagnose dreadful diseases such as heart disease, cancer, brain tumor. Hence, deep learning helps doctors to analyze the disease better and provide patients with the best treatment.

3. Insurance fraud

Deep learning is used to analyze the medical insurance fraud claims. With predictive analytics, it can predict fraud claims that are likely to happen in the future. Moreover, deep learning helps insurance industry to send out discounts and offers to their target patients.

4. Alzheimer's disease

Alzheimer is one of the significant challenges that medical industry is facing. Deep learning technique is used to detect Alzheimer's disease at an early stage.

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.

5. Genome

Deep learning technique is used to understand a genome and help patients get an idea about disease that might affect them. Deep learning has a promising future in genomics, and also insurance industry. Cell scope uses deep learning technique and helps parents to monitor the health of their children through a smart device in real time, thus minimizing frequent visits to the doctor. Deep learning in healthcare can provide doctors and patients with astonishing applications, which will help doctors to make better medical treatments.

V. Advantage of various Deep learning techniques in Healthcare section

1. Deep learning can use full-resolution mammogram images to accurately predict the likelihood of a woman developing breast cancer.
2. With the deep learning tool in hand, all radiologists can achieve expert-level performance on stroke diagnosis.
3. Machine learning can improve the response rate to questionnaires and improve the usefulness of the data through adaptive testing:
4. Deep learning provide more accurate and relevant data which make patient-reported data more actionable, combing through huge volumes of free-text responses or nuggets of information which are hidden in rambling clinical notes.
5. Convolutional neural networks (CNNs), is particularly well-suited to analyzing images, such as MRI results or x-rays.
6. CNNs are designed with the assumption that they will be processing images, allowing the networks to operate more efficiently and handle larger images.
7. Deep learning surpasses other machine learning methods for processing unstructured text, several significant challenges, including the quality of EHR data, are holding these tools back from true success.
8. In addition to skin cancer diagnosis, researchers are also using CNNs to develop tools for diagnosing tuberculosis, heart disease, Alzheimer's disease and other illnesses.

VII. Comparison of Deep learning techniques in Healthcare Sector

Sr.no	Title of paper	Author and date	Dataset	Deep Learning Technique Used	Application
1.	Deep Patient Similarity Learning for Personalized Healthcare [2]	Qiuling Suo , Fenglong Ma, Ye Yuan , Mengdi Huai, Weida Zhong, Jing Gao,Aidong Zhang (2018)	Electric health records (EHR)	CNN	Obesity, Diabetes, chronic Obstructive pulmonary disease (COPD).
2	Deep Feature Learning for Disease Risk Assessment Based on Convolutional Neural Network With Intra-Layer Recurrent Connection by Using Hospital Big Data[3]	MOHD USAMA, BELAL AHMAD, JIAFU WAN, M. SHAMIM HOSSAIN, MOHAMMED F. ALHAMID, M. ANWAR HOSSAIN (2018)	real-life medical big data obtained from a hospital in central China from 2013 to 2015	RCNN	risk assessment of cerebral infarction disease.
3	Deep Convolutional Neural Networks for Chest Diseases Detection[6]	Rahib H. Abiyev and Mohammad Khaleel Sallam Ma'aitah, Journal of healthcare Engineering, (2018)	National Institutes of Health Clinical Center	BPNN, CNN, CpNN	Diagnosis of chest diseases.
4	Detection of Alzheimer's Disease from MRI using Convolutional Neural Network with Tensorflow[8]	G. J. Awate, S. L. Bangare, Dr. G. Pradeepini, Dr. S. T. Patil.	Open Access Series of Imaging Studies dataset. (OASIS)	CNN	Alzheimer prediction
5	Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks[9]	Jyoti Islam,Yanqing Zhang, Springer, 2018	Open Access Series of Imaging Studies dataset. (OASIS)	CNN	Detection Alzheimer's disease
6	Deep Learning for Health Informatics[7]	Daniele Rav`ı, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo,and Guang-Zhong Yang (2017	EHR	CNN	health informatics, providing a critical analysis for various chronic disease.
7	A Convolutional Neural Network Model for Online Medical Guidance [5]	Cuili Yao, Yue Qu, Bo Jin,Li Guo, Chao Li, Wenjuan Cui, and Lin Feng (2016)	Real Life medical data	Named Entity Recognition (NER), CNN	automated medical consultation

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.

8	Classification using Convolutional Neural Network for Heart and Diabetics Datasets, [13]	Tharani.S, Dr. C. Yamini, (2016)	Dataset	CNN	Heart disease
9	Heterogeneous Convolutional Neural Networks for Comorbid Risk Prediction with Electronic Health Records[1]	Jinghe Zhang, Jiaqi Gong, Laura Barnes (2017)	EHR	HCNN	Study type 2 diabetes and chronic kidney disease leading congestive heart failure.
10	Predicting Disease By Using Data Mining Based on Healthcare Information System[14].	Feixiang Huang, Shengyong Wang, and Chien-Chung Chan , IEEE, 2012.	real world Healthcare Information System database	Decision Trees, Naive Bayes and Neural Network	Prediction of hypertension along with other disease like diabetes, and heart diseases.

VIII. CONCLUSION

Deep learning has gained a central position in recent years in machine learning and pattern recognition. Deep Learning (DL) has the potential to change the future of healthcare. The use of Artificial Intelligence (AI) has become increasingly popular and is widely used in cancer diagnosis and treatment. Deep learning for computer vision enables an more precise medical imaging and diagnosis. In this paper, we have outlined how deep learning has enabled the development of more data-driven solutions in health informatics by allowing automatic generation of features that reduce the amount of human intervention in this process.

To perform classification task of medical data, the neural network is trained using Convolutions algorithm(CNN). [6] Paper shows that outperformance is mainly due to the deep structure of CNN that uses the power of extracting different level features, which resulted in a better generalization capability.[10] Paper has conducted experiment with heart disease dataset by considering the single and multilayer neural network modes. Applications of deep learning algorithms in clinical settings provide the potential of consistently delivering high quality results.

REFERENCES

[1] Jinghe Zhang, Jiaqi Gong, and Laura Barnes, Member, IEEE, "HCNN: Heterogeneous Convolutional Neural Networks for Comorbid Risk Prediction with Electronic Health Records", IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, 2017.

[2] Qiuling Suo , Fenglong Ma, Ye Yuan , Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang , Fellow, IEEE, "Deep Patient Similarity Learning for Personalized Healthcare", IEEE TRANSACTIONS ON NANOBIOSCIENCE, VOL. 17, NO. 3, JULY 2018.

[3] MOHD USAMA, BELAL AHMAD1, JIAFU WAN, M. SHAMIM HOSSAIN, MOHAMMED F. ALHAMID3, AND M. ANWAR HOSSAIN 3, (Senior Member, IEEE), Deep Feature Learning for Disease Risk Assessment Based on Convolutional Neural Network With Intra-Layer Recurrent Connection by Using Hospital Big Data",IEEE ACCESS November 1, 2018.

[4] Amit S. Wale, Prof. Shilpa S. Sonawani, "A Survey on Deep Learning Approaches in Healthcare", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 5, May 2018.

[5] Cuili Yao, Yue Qu, Bo Jin, Senior Member, IEEE, Li Guo, Chao Li, Wenjuan Cui, and Lin Feng, "A Convolutional Neural Network Model for Online Medical Guidance", IEEE ACCESS, 10 January 2019.

[6] Rahib H. Abiyev and Mohammad Khaleel Sallam Ma'aitah, "Deep Convolutional Neural Networks for Chest Diseases Detection", Hindawi Journal of Healthcare Engineering Volume 2018, Article ID 4168538, 11 pages.

[7] Daniele Rav`i, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo,and Guang-Zhong Yang, Fellow, IEEE, "Deep Learning for Health Informatics", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 21, NO. 1, JANUARY 2017.

[8] G. J. Awate, S. L. Bangare, Dr. G. Pradeepini, Dr. S. T. Patil, " Detection of Alzheimer's Disease from MRI using Convolutional Neural Network with Tensorflow", IEEE Xplore 2018.

[9] Jyoti Islam* and Yanqing Zhang ," Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks", Springer Open Access, 2018.

[10] Tharani.S, Dr. C. Yamini , " Classification using Convolutional Neural Network for Heart and Diabetics Datasets", IJARCCE, Vol. 5, Issue 12, December 2016.

[11] Ping Xuan, Yihua Dong, Yahong Guo, Tiangang Zhang and Yong Liu, "Dual Convolutional Neural Network Based Method for Predicting Disease-Related miRNAs", International Journal of Molecular Sciences, 23 November 2018.

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.

[12] Yu Cheng,Fei Wang,Ping Zhang,Jianying Hu, "Risk Prediction with Electronic Health Records: A Deep Learning Approach", Computer Science, Published in SDM 2016.

[13] Feixiang Huang, Shengyong Wang, and Chien-Chung Chan," Predicting Disease By Using Data Mining Based on Healthcare Information System", 2012 IEEE International Conference on Granular Computing.

[14] A. Rios and R. Kavuluru, "Convolutional neural networks for biomedical text classification: application in indexing biomedical articles," in Proceedings of the 6th ACM Conference on Bioinformatics, omputational Biology and Health Informatics, pp. 258–267, ACM, Atlanta, GA, USA, 2015.

[15] S. Hussain, S. Muhammad Anwar, and M. Majid, "Segmentation of glioma tumors in brain using deep convolutional neural network," Neurocomputing, vol. 282, pp. 248–261, 2018.

[16] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang and Joel T. Dudley, "Deep learning for healthcare: review, opportunities and challenges", Briefings in Bioinformatics, 2017.

[17] X. Shi et al., ``Multiple disease risk assessment with uniform model based on medical clinical notes," IEEE Access, vol. 4, pp. 7074_7083, 2016.

[18] Liu, Y.; Zeng, X.; He, Z.; Zou, Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. IEEE/ACM Trans. Comput. Boil. Bioinform. 2017,14, 905–915.

[19] J. Gu, Z. Wang, J. Kuen, L. Ma et al., "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354–377, 2018.

[20] S. Khobragade, A. Tiwari, C. Y. Pati, and V. Narke, "Automatic detection of major lung diseases u sing chest radiographs and classification by feed-forward artificial neural network," in Proceedings of 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016) 2016 IEEE, pp. 1–5, Delhi, India, 2016.

[21] C.-Y. Chen and K.-M. Liao, "Chronic obstructive pulmonary disease is associated with risk of chronic kidney disease: A nationwide case-cohort study," Scientific reports, vol. 6, 2016.

[22] G. S. Birkhead, M. Klompas, and N. R. Shah, "Uses of electronic health records for public health surveillance to advance public health," Annual review of public health, vol. 36, pp. 345–359, 2015.

[23] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," arXiv preprint arXiv:1506.05163, 2015.

[24] R. T. Gansevoort, R. Correa-Rotter, B. R. Hemmelgarn, T. H. Jafar, H. J. L. Heerspink, J. F. Mann, K. Matsushita, and C. P. Wen, "Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention," The Lancet, vol. 382, no. 9889, pp. 339–352, 2013.

[25] M. Edwards and X. Xie, "Graph based convolutional neural network," arXiv preprint arXiv:1609.08965, 2016.

 DOI : 10.35291/2454-9150.2019.0544 © 2019, IJREAM All Rights Reserved.