



Risk Assessment and Diagnosis Code Prediction from Electronic Health Records (EHR) using deep learning

Priyanka S. Rane ^{1 *}, Dr. Uruj Jaleel ²

1. PhD Scholar, Kalinga University, Raipur, C.G. India, India
vimtpune@gmail.com ,

2. PhD Guide, Department of Management, Kalinga University, Raipur, C.G. India, India

Abstract: Knowledge in the medical field is often expressed by distinct and subjective norms. Much research and development efforts have focused on using deep learning algorithms to predict the likelihood of illnesses from EHR in recent years. When it comes to risk prediction, deep learning-based techniques outperform more conventional machine learning models. But nothing in the literature fully accounts for what doctors already know, such as the connections between illnesses and their risk factors. This research examines the use of Multi-layer Perceptron models for the categorisation of diagnoses in electronic health records. The raw data and a modified version of the EHR dataset are used to train two MLPs with distinct topologies. For comparative purposes, a Random Forest is used as a baseline. To phenotype patients using their electronic health records, we provide a deep learning method. Predictive modelling of chronic illnesses is the particular scenario used to verify the suggested model on a real-world EHR data warehouse. Many deep learning applications on EHRs have been effective, and there is still a lot of potential to be realised. It has been discovered that deep learning models can learn from the limited EHR dataset, but not to a level where they outperform the baseline model.

Keywords: Deep Learning, Human, Disease, Electronic Health Record, Diagnosis

----- X -----

INTRODUCTION

A patient's paper chart may now be accessed digitally with the use of an electronic health record (EHR). Electronic health records (EHRs) are accessible, secure, and real-time records that are centred on the patient. Although electronic health records (EHRs) do include patients' medical and treatment histories, EHR systems are designed to include more than just the typical clinical data obtained at a provider's office. By analysing the vast amounts of data stored in Electronic Health Records (EHR), researchers and healthcare practitioners may get closer to the goal of personalised treatment. But there are problems with raw EHR data as well, including high dimensionality, inconsistency, bias, sparsity, and irregularity in time. The key objective in the medical area, known as risk prediction, involves predicting patients' probable ailments. However, typical machine learning or statistical models [1,2] become much more difficult to directly apply due to these obstacles. Consequently, stronger models are required to address the difficulties brought about by the use of raw EHR data in risk prediction tasks. In several fields, such as computational phenotyping, risk prediction, and diagnostic prediction, deep learning models have recently shown the capacity to directly extract relevant characteristics from unstructured electronic health information. In order to forecast the occurrence of heart failure, attention-based recurrent neural networks (RNN) are used, particularly for risk prediction tasks. Also included are convolutional neural networks (CNNs), which boost

performance by capturing the local temporal features of patients' visits and using them to forecast illness risks.

The risk prediction tasks have been successfully completed by the aforementioned deep learning-based models; nevertheless, these models do not take into account the significance of previous medical information, such as the correlations between illnesses and their associated risk factors. Everyone knows that previous medical knowledge is crucial in the healthcare field. Before taking a patient's present symptoms into account, doctors thoroughly examine their medical history, which includes things like the patient's current medicines, smoking habits, alcohol use, and any illnesses that run in the family. The doctor may be able to make an initial diagnosis based on the patient's present symptoms and medical history. Patients may have symptoms such as a racing heart, shortness of breath, increased nighttime urination, chest discomfort, and fainting. For over eight years, he or she has battled hypertension and coronary artery disease. A rapid diagnosis of heart failure, rather than another illness, may be made by the doctor based on their experience and the patient's present symptoms. Reason being, heart failure may be triggered by hypertension and coronary artery disease. Thus, it is crucial for risk prediction tasks to take into account past medical information.

LITERATURE REVIEW

Raju, K & Vidyarthi, Ankit & Dara, Suresh & Gupta, V & Khan, Baseem. (2022) [3] This research proposes a methodology that combines Edge-Fog-Cloud computing to provide quick and accurate results. Data from several patients is collected by the hardware components. Important characteristics are retrieved by extracting cardiac features from signals. Also collected are the results of feature extraction for additional characteristics. In this case, Galactic Swarm Optimisation (GSO) is used to optimise the CCNN hyperparameters. According to the results of the performance study, the proposed GSO-CCNN outperforms PSO-CCNN, GWO-CCNN, WOA-CCNN, DHOA-CCNN, DNN, RNN, LSTM, CNN, and CCNN in terms of accuracy by 3.7%, 3.6%, 7.6%, 67.9%, 48.4%, 33%, 10.9%, and 7.6%, respectively. Therefore, the proposed system's efficacy above the traditional models is guaranteed by the comparison study.

Askar, Shavan & Jameel, Zhala & Kareem, Shahab. (2021) [4] In this review article, we'll look at few ways to summarise DL's roles in the FC industry. Advanced DL customers with top-tier services have emerged as a result of FC's DL deployment, paving the way for more in-depth analytics and smarter mission replies.

Nancy, A.A.; Ravindran, D.; Vincent, D.R.; Srinivasan, K.; Chang, C.-Y. (2023) [5] In order to detect cardiovascular illness, this study proposed a smart healthcare system that makes use of fog. For pre-processing and predictive analytics tasks, it integrated a fuzzy inference system (FIS) with the variation of the gated recurrent unit (GRU) from the recurrent neural network model. With a classification accuracy of 99.125%, the suggested approach demonstrates much better performance outcomes. The suggested method outperforms cloud computing in terms of latency, response time, and jitter, with most healthcare data analytics processing taking place at the fog layer. In particular, deep learning models excel in predictive analytics and other complex tasks. outcomes from the experiments show that time-critical healthcare applications may benefit from the decentralised fog model and deep learning's unique ability to provide

near-perfect outcomes.

Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023) [6] A wide range of ailments, including Alzheimer's, cancer, diabetes, chronic heart disease, TB, stroke and cerebrovascular, hypertension, skin, and liver disease, may be diagnosed using AI approaches, as covered in this article's complete survey. The medical imaging dataset, feature extraction method, and prediction methodology were all part of our comprehensive study. Prediction rate, accuracy, sensitivity, specificity, area under curve precision, recall, and F1-score are some of the quality metrics that are used to evaluate the outcomes based on the research of many publications on illness diagnosis.

Vu Khanh, Quy & Nguyen, Van-Hau & Anh, Dang & Ngoc, Le. (2021) [7] We compare many computer systems in this post. Then, for Fog-IoHT applications, we provide a shared architectural framework that is based on fog computing. In addition, we highlight potential uses and difficulties of fog computing in Internet of Things (IoT) healthcare applications. The investigation revealed that fog computing-based IoHT applications had enormous promise. Our research is founded on the premise that fog-based healthcare IoT applications might serve as a valuable roadmap for their future evolution.

RESEARCH METHODOLOGY

We test the proposed model on a real-world EHR data warehouse and see how well it performs in two clinical scenarios: early CHF diagnosis and COPD prediction. Our model-learned features not only improve prediction performance, but they also make clinical sense, according to the findings.

Approach

Two MLPs, one with a naïve model and the other with an architecture supplied by the principle, are used in the tests. Our model is constructed using the patient's electronic health record's temporal matrix form. A longitudinal event matrix is used to represent the electronic health record (EHR), with time stamps serving as the horizontal axis and event values as the vertical axis. In an electronic health record matrix, the i -th entry is set to 1 if the j -th time stamp for the relevant patient observes the i -th occurrence. Unfortunately, there are a number of reasons why this event matrix encoding is not amenable to the traditional CNN model, unlike photos and videos [8].

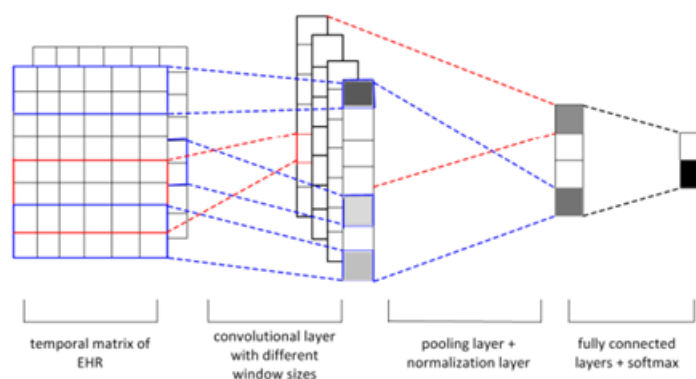


Figure 1: An electronic health record data example's fundamental model architecture

You can see a small variation of the CNN design in Figure 1, which represents the fundamental model architecture [9]. Represented as X , where $X \in \mathbb{R}^{d \times t}$, is every event matrix of length t . The i -th event item's d -dimensional event vector is denoted as $x_i \in \mathbb{R}^d$. Typically, the combination of elements x_i , x_{i+1} , ..., and x_{i+j} is denoted as $x_{i:i+j}$. In order to create a new feature, a filter $w \in \mathbb{R}^{d \times h}$ is used in a one-side convolution operation on a window of h event features. As an example, a non-linear function like rectification (ReLU) or tangent (Tanh) is used to create a feature c_i from a window of events $x_{i:i+h-1}$ by solving the equation $c_i = f(w \cdot x_{i:i+h-1} + b)$, where b is a bias component and f is defined in \mathbb{R} .

Processing of Data

Sixty-91,141 samples make up the data. The number of features in each sample is 1209. There are a total of 1209 elements in the questionnaire; 1208 of them are numerical and represent the various patient responses. The numerical values in this collection come from several sources and are organised as follows: a set of ternary columns where 0=No, 1=Not sure, and 2=Yes; and a set of columns with ranking numeric values, some of which are normalised and some of which are not.

Structure of Networks

You may implement the fundamental architecture of MLPs by selecting an optimisation technique, an activation function, and a cost function. Each model's parameters, including its number of layers and nodes, dropout probability, batch size, etc., are experimentally modified. The models are then evaluated and compared based on the parameters that provide the highest performance [10].

- **MLP-a:** MLP-a is a Upon finishing the parameter tweaking, MLP-a is selected as a three-layer network. The first hidden layer has thirty neurones, the second contains forty, and the final contains fifteen. With a rectified linear unit activation function, the output is zero in the case when the total of the neurone inputs is less than zero, and raw output in all other cases.

Table 1: Details on MLP-a

Activation function	ReLU
Cost function	Mean Squared Error
Layer-1, width	30
Layer-2, width	40
Layer-3, width	15
Learning rate	0.002
Dropout keep rate	0.7
Epochs	450
Batch size	300

- **MLP-b:** A three-layer network with ten neurones in the first hidden layer, twenty in the intermediate layer, and fifteen in the final layer is selected as MLP-b. Along with being a softmax function, the cost function also includes cross entropy, a measure of the dissimilarity between the actual and predicted class probability distributions that is sometimes employed as a substitute for squared error.

Table 2: Details on MLP-b

Activation function	Softmax
Cost function	Softmax with cross entropy
Layer-1, width	10
Layer-2, width	20
Layer-3, width	15
Learning rate	0.009
Dropout keep rate	0.8
Epochs	400
Batch size	300

DATA ANALYSIS

Comparison models

After making its own adjustments to the dataset, the Random Forest model was 61.8% accurate. On average, recall is 62.2% and precision is 63.4%. Thus, 62.8% is the F1 score. The accuracy drops to 38.8%, the precision to a pitiful 0.94%, the recall to 2.6%, and the F1 score to 1.4% when using a naïve model that guesses simply the majority class.

Matrix of MLP confusion on D_{min}

For the MLPs on D_{min} , the confusion matrix shows the relationship between y_{pred} and y_{real} ; the x-axis shows y_{pred} and the y-axis shows y_{real} .

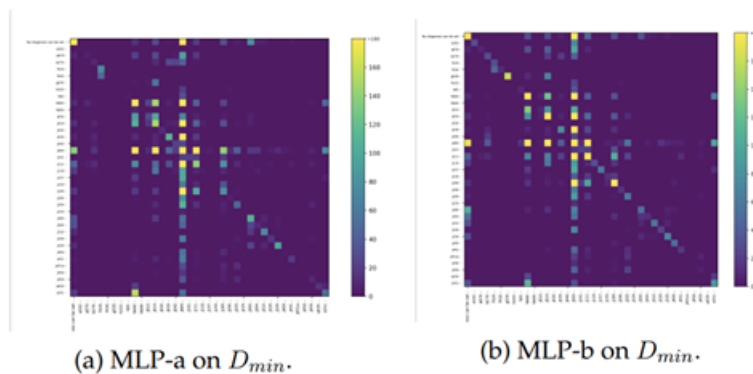


Figure 2: The MLPs' Operation on D_{min} Confusion Matrix

A matrix of confusions for MLPs on D_{alt}

The confusion matrix illustrates the relationship between y_{pred} and y_{real} for the MLPs on D_{alt} . On the x-axis, we see y_{real} , and on the y-axis, y_{pred} .

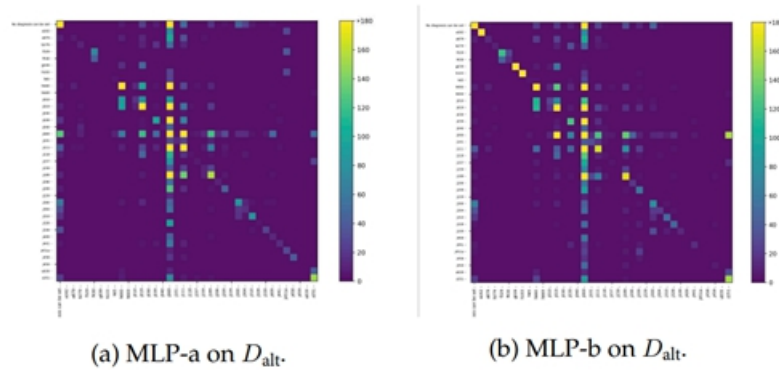


Figure 3: Matrix of Perplexities for MLPs Running on D_{alt}

Two variants of the dataset are used to assess the models. D_{min} is the one with little changes, while D_{alt} is the one with many changes that are often employed on machine learning datasets. When tested on D_{alt} , both models outperform D_{min} . The possibility that the dataset size is insufficient to allow the deep models to train adequately is one of the project's concerns, as mentioned in the thesis's limits section. This is supported by the fact that D_{alt} outperforms D_{min} in terms of performance. Data simplification (by, for example, narrowing the feature space) seems to aid the model in D_{alt} . This may also help to explain why the baseline model outperformed the MLPs. Compared to the MLPs, the baseline model is simpler, which allows it to generalise more effectively from a smaller dataset. The confusion matrices on D_{min} show that both models give the class j069 an exaggerated probability. This is because the model became overfit to the class that was more prevalent in the training data. In order to prevent overfitting, dropout is used. However, there is a trade-off that has to be made when setting the chance of a node being closed down during training iterations.

Evaluation

We tested our methods on a real-world EHR data warehouse with 319,650 patient records spanning 4 years to see how well they worked. To create the EHR sequences, we employ the diagnostic data according to the first three digits of the ICD-9 form. The challenge of predicting the likelihood of chronic illness emergence at an early stage will be our focus. Figure 4 shows the entire setup. We begin by obtaining a collection of case patients who have been medically diagnosed with a condition with the assistance of our domain expert. Then, we use patient demographics and clinical features to create a set of group matched controls.

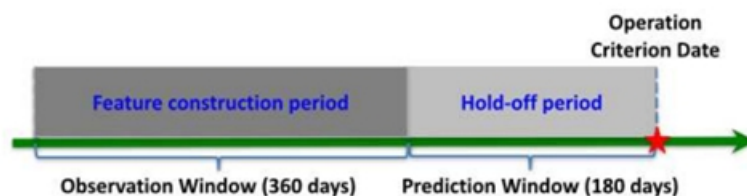


Figure 4: Research environment for the early prediction of the risk of chronic illness development

Chronic Obstructive Pulmonary Disease (COPD) and Congestive Heart Failure (CHF) are the two illnesses that our investigations focus on. A total of 3,850 healthy controls and 1,127 cases makes up the CHF patient group. The COPD patient group consists of 2,385 healthy controls and 477 actual cases. We use all the records in our database that were accessible before the prediction window to train our proposed model for both illnesses. We set the forecast window to 180 days. Put another way, we can anticipate the likelihood of a patient developing CHF or COPD six months down the road by analysing all of their medical information.

CHF:

Using area under the curve (AUC) across 10-fold cross validation, Table 3 summarises the findings for CHF prediction. The table clearly shows that our approaches substantially and repeatedly beat the feature-based baseline. In particular, out of all the approaches, the SF-CNN performs the best. Using AUC as a metric, SF-CNN improves prediction accuracy by 1.5 percentage points when 60% training data is used, and by 5.2 percentage points when 90%. The similarities in their designs explain why EF-CNN and the basic CNN model (BS-CNN) function similarly. By a hair, LF-CNN outperforms BS-CNN and EF-CNN in terms of prediction accuracy. The ability of LF-CNN to detect discriminative local temporal patterns for categorisation is one probable explanation. Convolutional neural network (CNN) based models would benefit from huge training data sets, as the performance gains of all suggested models grow in direct proportion to the amount of training data.

Table 3: Standard deviation and area under the curve for predictions on the CHF cohort using varying training data ratios

Method	60%	70%	80%	90%
Baseline	0.5317 ± 0.091	0.5992 ± 0.077	0.6593 ± 0.048	0.7156 ± 0.043
BS-CNN	0.5346 ± 0.106	0.6133 ± 0.091	0.6754 ± 0.070	0.7388 ± 0.047
EF-CNN	0.5389 ± 0.101	0.6195 ± 0.093	0.6797 ± 0.070	0.7402 ± 0.046
LF-CNN	0.5414 ± 0.112	0.6232 ± 0.094	0.6815 ± 0.067	0.7569 ± 0.048
SF-CNN	0.5462 ± 0.100	0.6309 ± 0.087	0.6963 ± 0.060	0.7675 ± 0.044
Combined	0.5405 ± 0.102	0.6038 ± 0.083	0.6779 ± 0.060	0.7355 ± 0.047

COPD:

You may see the results of the COPD prediction in Table 4. The CHF cohort shows a similar pattern. Even with 90% training data, SF-CNN improves the prediction AUC by 5.3% compared to the baseline, making it the best performer. Nevertheless, LF-CNN's performance on this dataset falls short of its CHF Cohort counterpart. Reasons for this include the fact that ESRD Cohort is a smaller dataset and the fact that LFCNN is more prone to over-fitting due to its larger number of parameters. In fact, LF-CNN outperforms EF-CNN and BS-CNN when the training set is decreased to 60%.

Table 4: Standard deviation and area under the curve for COPD cohort predictions using various

training data ratios

Method	60%	70%	80%	90%
Baseline	0.4536 ± 0.103	0.5738 ± 0.086	0.6324 ± 0.061	0.6624 ± 0.052
BS-CNN	0.4643 ± 0.100	0.5814 ± 0.084	0.6512 ± 0.064	0.7072 ± 0.057
EF-CNN	0.4625 ± 0.094	0.5854 ± 0.080	0.6533 ± 0.058	0.7083 ± 0.047
LF-CNN	0.4517 ± 0.111	0.5865 ± 0.090	0.6583 ± 0.072	0.7109 ± 0.066
SF-CNN	0.4749 ± 0.105	0.6086 ± 0.077	0.6735 ± 0.063	0.7388 ± 0.054
Combined	0.4572 ± 0.099	0.5815 ± 0.086	0.6523 ± 0.061	0.6924 ± 0.050

CONCLUSION

In conclusion, we provide a novel approach to analysing patient EHRs using a deep learning architecture. The input, one-side convolution, max-pooling, and softmax prediction layers make up our system. To further analyse the temporal smoothness of patient EHRs in the suggested framework, other temporal fusion strategies are also explored. Lastly, we statistically and subjectively test the suggested model's efficacy using synthetic and real-world data. Many deep learning applications on EHRs have been effective, and there is still a lot of potential to be realised. Results from this experiment show that deep learning models can learn from the limited EHR dataset, but they still can't compete with the baseline model's performance. It seems like a better idea to use a simpler approach like Random Forest and put more effort into feature engineering in a situation when there is less data. But further study on deep learning categorisation would be really interesting as the dataset keeps getting bigger. Using a framework to reduce parameters might be an avenue for future research into improving the CNN model and avoiding over-fitting. It is also encouraging to apply the present approach to a different subject.

References

1. Joyce C Ho, Joydeep Ghosh, Steve R Steinhubl, Walter F Stewart, Joshua C Denny, Bradley A Malin, and Jimeng Sun. 2014. Limestone: High-throughput Candidate Phenotype Generation via Tensor Factorization. *Journal of Biomedical Informatics* 52 (2014), 199–211.
2. Joyce C Ho, Joydeep Ghosh, and Jimeng Sun. 2014. Marble: High-throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. 115–124.
3. Raju, K & Vidyarthi, Ankit & Dara, Suresh & Gupta, V & Khan, Baseem. (2022). Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model. *Computational Intelligence and Neuroscience*. 2022. 10.1155/2022/1070697.
4. Askar, Shavan & Jameel, Zhala & Kareem, Shahab. (2021) Deep Learning and Fog Computing: A Review. 10.5281/zenodo.5222647.
5. Nancy, A.A.; Ravindran, D.; Vincent, D.R.; Srinivasan, K.; Chang, C.-Y. (2023). Fog-Based Smart

Cardiovascular Disease Prediction System Powered by Modified Gated Recurrent Unit. *Diagnostics*, 13, 2071. <https://doi.org/10.3390/diagnostics13122071>.

6. Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, 14(7), 8459–8486. <https://doi.org/10.1007/s12652-021-03612-z>
7. Vu Khanh, Quy & Nguyen, Van-Hau & Anh, Dang & Ngoc, Le. (2021). Smart healthcare IoT applications based on fog computing: architecture, applications and challenges. *Complex & Intelligent Systems*. 8. 10.1007/s40747-021-00582-9.
8. F. Wang, N. Lee, J. Hu, J. Sun, and S. Ebadollahi, “Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach,” in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 453–461.
9. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Nov. 2011.
10. Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: (2014), pp. 1–15. ISSN: 09252312. DOI: <http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>. arXiv: 1412.6980. URL: <http://arxiv.org/abs/1412.6980>.