PREDICTION OF HEALTHCARE DEMAND USING DOCTOR AI ALGORITHM

DISSERTATION

submitted in partial fulfillment of the requirements

for the award of the dual degree of

Master of Science

in

MATHEMATICS AND COMPUTING

by

VANSHIKA

(302203014)

Under the guidance of

DR. KAVITA

July 2024

DEPARTMENT OF MATHEMATICS

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

PATIALA-147004

PUNJAB, INDIA

CERTIFICATE

This is to certify that the thesis entitled **"Prediction of Healthcare demand using Doctor AI algorithm"**, being presented in partial fulfillment of the requirements for the award of the degree of Masters of Science in Mathematics and Computing and submitted to **Department of Mathematics**, Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of **Dr. Kavita**.
The matter presented in this thesis has not been submitted for the award of any other degree from this or any other institute.

(Vanshika)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Dr. Kavita
Associate Professor
Department of Mathematics
Thapar Institute of Engineering and Technology,
Patiala

ACKNOWLEDGEMENT

This thesis represents the completion of a magnificent two-year journey to obtain my Master's degree. It entails spending many hours online studying machine learning techniques used in the medical industry. Throughout my trip, I have received support and guidance from various individuals. I would want to thank everyone.

Words cannot describe my heartfelt appreciation to **Dr. Kavita**, my supervisor, for her acceptance, patience, and insightful feedback. She helped me move from beginning to end by leading several conversations and teaching me the fundamentals of a research field in which I had no prior knowledge. I also want to express my gratitude to all of the professors who helped me over my two-year journey. My dissertation was easy to write because of their intuitive advice and instructional materials.

Next, I am grateful to my parents for their unwavering support and encouragement at home. What has kept me going thus far has been their prayer for me. Last but not least, I want to convey gratitude to God for helping me survive through all the obstacles. I've felt your support every day. You are the one who gave me the courage to finish my degree. I have always trusted you, and I will in the future too.

July 2024 — **Vanshika**

Place: Patiala — **Roll no.: 302203014**

ABSTRACT

Hospitals must estimate the need for healthcare a year in advance and negotiate with insurance providers to establish a budget that will allow them to treat everyone while also ensuring the highest possible income. The hospitals will have to bear the entire cost of treatment on their own if the forecasts prove to be off. The more accurate the prediction, the more compensation is earned.

The sequences and labels were created to accurately collect patients' historical electronic health records (EHRs) on a temporal basis. Deep learning, a subset of machine learning, is employed in this study because of its ability to detect hidden patterns in data that are frequently overlooked by humans. This dissertation attempts to use the Doctor AI algorithm (Choi, Bahadori, Schuetz, Stewart, and Sun (2016a)) to properly estimate future healthcare demand.

The results show that for the top 100 categories, the highest accuracy@top-30 of 94.375% was achieved with 500 nodes in the GRU layers when trained for 20 epochs. On increasing the number of output classes, there was a decrease in accuracy as well. Overall it can be concluded that machine learning can be used for prediction purposes in the future since it can handle EHR data quite efficiently. More research and the comparison between several other methods also need to be done in the future.

LIST OF FIGURES

CONTENTS

1. INTRODUCTION

1.1 PROBLEM UNDERSTANDING

It is widely known that the health business is often resistant to change, and in certain instances, this argument is valid. However, there have been notable transformations in recent decades, such as a shift away from the paternalistic medical culture, the adoption of more advanced electronic health records (EHRs), a substantial decrease in hospital stays, and an increased focus on high-level professional work. Even more rapid change was sparked by the epidemic, including a complete shift toward community-driven support models, digitally delivered healthcare, novel population health surveillance methods, and widespread public involvement in clinical trials.

There is a mutually advantageous relationship between insurance companies and hospitals. Hospitals work out a fair financial agreement that benefits both sides by negotiating prices with insurance companies. These agreed-upon prices aid in controlling healthcare expenses and guard against overcharging. Insurance encourages hospitals to prioritize early detection and preventative care since it lowers long-term medical costs associated with better patient health. This partnership ensures a consistent flow of patients as well. Despite the benefits, this connection is not free of obstacles. The challenges include claim processing, coverage denials, administrative complications, and escalating healthcare costs. A more effective and fair healthcare system may be achieved by implementing strategies for improving transparency, streamlining billing procedures, and increasing coverage alternatives to solve these issues.

The number of particular treatments required for each specialism is known as the hospital's healthcare demand. In this research, hospitals are obliged to forecast healthcare demand a year ahead of time in order to sell individual specialty treatments to multiple health insurance providers and halt the rise in healthcare costs to zero percent. Hospitals and insurers establish the maximum number of treatments that will be covered in an upcoming year by considering projected demand and national prices. By establishing a budget one year ahead, the hospital can guarantee that it will obtain the most feasible income for the year, as further expenses will not be covered by insurance companies. Therefore, any patient treatment that exceeds the estimated amount will be provided at the hospital's full cost. Insurers will also not pay for treatments that fall short of the anticipated amount, limiting the amount of compensation available for recovered therapy. Thus every time the prediction goes wrong, the hospital expenses are directly impacted.

To be more precise, the hospital is required to create a budget and develop a prediction one year in advance about the annual number of patients and the treatments that are believed to be administered to them in order to reach appropriate agreements with the various insurance companies. This indicates that there is a predetermined number of patients over the course of a year who are able to consume a predetermined number of treatments. In most cases, the budget for a particular specialization is predetermined, since it is possible to reallocate the budget within that specialization under very specific conditions during the year, which means that if one healthcare product does too well, the bad performance of another can make up for it. For instance, if there is a higher number of arthrosis treatments for the hips or pelvis in a given year, but a lower demand for arthrosis treatment of the knee, the budgets can be balanced. This is because these treatments fall under the orthopedic department and specialization. In the event that the forecast within a certain specialty is carried out precisely, the insurance companies will pay the hospital the total sum that was agreed upon for the year. Nevertheless, if there are fewer patients treated in a particular year, the hospital will only receive a portion of the compensation that is due to it. On the other hand, the hospital will be responsible for paying the whole cost of any treatment that exceeds the budget. This yearly continuous balance between budgets that are either too low or too high creates a scenario wherein the profitability of the hospital is reduced by every deviation in fault margin from the predicted healthcare demand. This condition is a result of the continuously occurring balance between budgets.

Accurate prediction in healthcare is very challenging, first because of the complexity of healthcare delivery itself. The treatment of a patient requires a wide range of resources, each with unique capabilities and expenses, including personnel, equipment, space, and supplies. These resources are utilized throughout a patient's care journey, beginning with their initial contact with the organization and progressing through clinical consultations, treatments, and administrative procedures. The patient's medical state determines the route they take inside the system. Treating a patient's illness involves multiple, mostly distinct organizational units thus making it more complex because of the highly fragmented way by which the services are delivered. (Kaplan, Porter et al. (2011a))

The data is the foundation for any future predictions. Data may come from several resources such as clinical notes, physiological waveforms, structured EHR data, radiological images, and even unstructured data from publications. Healthcare professionals are fully aware of the degree of errors, inconsistencies, and inaccuracies in health data, especially clinical notes. When choosing the outcome of interest, the important thing is to have access to accurate data regarding that target. Since laboratory tests are not 100% accurate, so it is impossible to obtain complete certainty, but ML techniques can handle the presence of uncertainty in the data. Also in the past, a significant portion of clinical data were disregarded or not gathered at all. This restriction resulted from the volume and complexity of the data as well as the lack of methods for gathering and preserving it. These data are often underutilized. New data gathering and storage systems, such as electronic health records, can address the issue of undervaluation and analysis. In particular, a general infiltration of machine learning (ML) into the clinical literature has long started (Wiens and Shenoy (2018)). The machine learning applications using electronic health record data demonstrate that algorithms using machine learning are not only capable of determining treatment demand, but also of revealing patterns in the emergence of various diseases and their interrelationships that were previously obscure (Lin, Zhang, Ivy, Capan, Arnold, Huddleston, and Chi (2018); Jensen, Jensen, and Brunak (2012)).

Five percent of all patients in the healthcare system account for fifty percent of the overall costs; in addition, the number of chronic illnesses requiring continuous care has steadily climbed nationwide. Machine learning has the potential to diagnose patients and identify those who may be more susceptible to recurrent ailments. Furthermore, about 90% of trips to the emergency room are avoidable. By avoiding costly, time-consuming emergency care facilities, machine learning can help detect conditions and refer patients to the right care while reducing expenses. In nutshell, machine learning has the potential to help both patients and providers in terms of better care and lower costs (Bhardwaj, Nambiar, and Dutta (2017)). In the event that healthcare providers are unable to effectively estimate the future demand for healthcare, they run the risk of having to treat patients at their own expense. Despite the fact that profit margins are not and should not be the primary motivators in the healthcare industry, Raghupathi and Raghupathi (2014) says, "it is extremely crucial for healthcare organizations to develop the ability to employ machine learning tools properly or else risk losing possibly millions of dollars in revenue and profits". Thus, the purpose of this research is to evaluate the predictive performance of machine learning in order to make a contribution to the validation of machine learning applications that are used in the real world and to improve forecasting performance.

1.2 RESEARCH OBJECTIVE

The aim of this study is to ascertain whether machine learning is capable of successfully forecasting healthcare demand and producing more accurate estimates. For this purpose, a literature review is carried out in order to ascertain which of the currently available machine learning algorithms would be the most suitable, from a theoretical standpoint, for the application of healthcare product prediction in state progression/time series within the context of healthcare. Secondly, it is necessary to extract the appropriate variables that are required as input for the procedures that have been chosen. Thirdly, the machine learning approach that was discovered is trained and tested on the collected dataset of variables to determine its efficacy in forecasting the demand for healthcare services. After that, a discussion is held regarding the limitations and scope of the future work while also concluding the work.

1.3 THESIS FRAMEWORK

In the way it was described above, Chapter 1 provided an introduction to the topic that will be investigated in this study and outlined the questions that will be addressed in subsequent chapters. The purpose of Chapter 2 is to examine the existing literature on the subject of machine learning in the healthcare industry, discussing the characteristics of electronic health record (EHR) data, and the implementation of deep learning within this sector. It also provides a presentation of the methodology that is utilized in the process of conducting a systematic literature study in order to identify the machine learning algorithms that are the most appropriate for the situation. Because of this, the approach of the machine learning algorithm that was used for this research is explored in further detail in Chapter 3. This chapter also provides an introduction to the performance metrics, which are utilized in the process of evaluating the performance of the methods implemented. The outcomes of the various setups of the machine learning algorithm are analyzed and discussed in Chapter 4. This research is evaluated and discussed in Chapter 5, which provides an overview of the findings.

2. THEORETICAL CONTEXT

The complexity of Electronic Health Record(EHR) data and the particular difficulties it presents over other forms of data is discussed in the first section of this chapter. The topic of machine learning and its use in the healthcare industry are covered in the part that follows, along with some obvious differences from other application areas. Deep learning is discussed in more detail in the next section since it has applications in the areas of diagnostics and healthcare demand forecasting. The literature study helps identify and comprehend current machine learning techniques for diagnostic prediction. It does this by first explaining the process and then explaining the evaluation of the methods that were discovered. This chapter ends with the identification of three machine learning algorithms that are currently in use, based on the combination of newly gained information and the findings of the literature study. The current challenge of forecasting future healthcare demand can be resolved by using these three approaches.

2.1 CONVOLUTION OF EHR DATA

Electronic Health Record (EHR) is digitized details about a patient's medical history, treatments, diagnoses, medications, lab tests, and other relevant healthcare information of any hospital visit or clinical encounter. EHR data is stored electronically on an individual-level basis and can be accessed by authorized healthcare providers to aid patient care. There are several sources from which EHR data can be obtained, particularly from insurance claims, hospital warehouses, and pharmacy details. Since different records are spread across different sources, the data needs to be compiled together for further analysis. Due to the absence of a single criterion of EHR data types which includes diagnostic codes, procedure codes, insurance amount claimed, the amount reimbursed, stay period, lab reports, etc., Zhao, Ash, Ellis, Ayanian, Pope, Bowen, and Weyuker (2005) in their work proposed an array of different datasets for predictive modeling.

EHR data systems are intended to improve patient care, lower medical errors, and offer a thorough picture of a patient's past but the challenge of working on this data is its temporal nature. The data types can be numerical quantities such as body mass index, date-time objects (date of birth or time of admission), categorical values such as gender, drug codes, or codes from controlled vocabularies like ICD-9, natural language free-text such as clinical notes or discharge summaries (Shickel et al. (2017)). These records keep complete track of past information to predict future outcomes. Thus, temporal data must be converted into sequential data to generate useful results. To tackle the problems with heterogeneous data, techniques like one-hot encoding, label encoding, and categorization of data were applied but failed because of high dimensionality, computational complexity, missing/loss of information, and because they treat every dimension independently. The method that comes to the rescue is the Representation Learning method which captures the semantics in context.

Representation Learning methods are being used largely not only in the medical but also across various other fields as it ejects the need for feature engineering. It aims to recognize various complex relationships within the EHR data to meaningfully extract the patterns and features dependently. By advanced algorithms, this learning can uncover hidden structures in the data that may not be seen or understood by traditional analysis methods. In the healthcare sector, the medical information in EHRs is converted to vector forms (Deng, Seltzer, Yu, Acero, Mohamed, and Hinton (2010); Coates and Ng (2011); Yu, Lin, and Lafferty (2011); Bengio, Courville, and Vincent (2013)). By means of underlying linkages between visits, and diseases, or recognition of predictive diagnosis by different approaches, representation learning can tackle the problems with sparse high-dimensional medical codes and the challenge of capturing the temporal-hierarchical structure of EHR data (Choi, Bahadori, Searles, Coffey, Thompson, Bost, Tejedor-Sojo, and Sun (2016b); Zhou, Jia, Motani, and Chew (2017)).

Due to different policies in different countries, EHR data is not collected in a standardized manner. Even in hospitals, the data collected from different specializations can differ resulting in the complexity of data. Given that we utilize this data to draw generic conclusions, Morton, Nagpal, Sadanandan, and Bauhoff (2016) highlighted the importance of standardized codes by pointing out the insignificance of raw data entries from different hospitals in different formats. Despite significant efforts by the World Health Organization (WHO) to standardize medical classification codes (ICD-standard) for both procedures and diseases, different countries and institutions may nevertheless use different implementations of this framework.

As discussed, EHR can contain a large variety of datatypes i.e. due to its complexity, researchers are faced with difficulties in applying their methodologies. While additional validation of current methodologies is required to demonstrate results' general ability, machine learning, and deep learning, in particular, can find and use patterns in EHR data that were previously hidden, helping to predict future healthcare needs.

2.2 MACHINE LEARNING IN HEALTHCARE

Over the years, the idea of employing statistical models and algorithms to let computers learn from data and make predictions or judgments has been developing. With the availability of electronic health records and the need for more efficient and accurate diagnostic and treatment methods, machine learning gained attention. As Thomas H. Davenport writes in the Wall Street Journal, "Humans can typically create one or two good models a week (while) machine learning can create thousands of models a week". This quickly developing field has the potential to completely transform the provision of healthcare. In this field, machine learning can be employed for disease diagnosis, personalized treatment recommendations, patient outcome predictions, and medical image analysis.

Even though there are immense different ML algorithms for a given task, they can be generally grouped depending on how the data set is approached and what kind of data is handled. In healthcare, two main approaches are widely adopted, namely supervised and unsupervised learning. In supervised ML, the model is trained on a range of features associated with a known outcome. In the clinical field, examples include: identifying patients at risk of re-admission to a hospital (Billings, Blunt, Steventon, Georghiou, Lewis, and Bardsley (2012)), measuring the performance of a hospital through mortality rate (Pitocco and Sexton (2018a), predicting mortality rate (Pitocco and Sexton (2018b)), forecasting demand in the outpatient department (Jiang, Chin, Wang, Qu, and Tsui (2017)), predicting future medical codes (Shickel et al. (2017)), predicting healthcare costs (Morid, Kawamoto, Ault, Dorius, and Abdelrahman (2017)), hospitalization due to heart diseases (Dai, Brisimi, Adams, Mela, Saligrama, and Paschalidis (2015)) etc. In unsupervised ML, patterns within the dataset are identified by algorithms without a predefined outcome. The examples include: uncovering patients at risk from doctors' notes (Mikolov, Chen, Corrado, and Dean (2013a)), discovering patient clusters based on genomics (Lopez, Tucker, Salameh, and Tucker (2018)), detecting lung cancer (Rahane, Dalvi, Magar, Kalane, and Jondhale (2018)), etc.

While training any ML model, the quality of data must be the first consideration since the data ultimately determines the outcome. Thus in the medical field, developing efficient data management procedures at all levels becomes crucial. A more significant obstacle is that, unless models like decision trees are used, predictions made using machine learning typically do not include justifications. Using machine learning to correctly predict future healthcare demand is helping in future healthcare planning and optimizing future expenditures. Many researchers through their studies concluded that chronic disease, expensive treatments, and elderly people having multiple diseases consume the largest proportion of hospital expenditures (Wammes, Tanke, Jonkers, Westert, Van der Wees, and Jeurissen (2017); Kam, Sung, and Park (2010)). Thus these are the patients that appear to utilize the highest future consumption of healthcare and the expenses, so predicting the kind and quantity of these patients correctly is of great importance. Kaplan, Porter et al. (2011b) emphasized ways to address the healthcare expense challenge throughout the article.

The most commonly used traditional machine learning methods are random forest, decision trees, logistic regression, and support vector machine (SVM). These methods were employed on EHR data but didn't give the expected result due to their inability to handle raw data. Also capturing the underlying dependencies in each visit of a patient was a bit challenging. With the constant expansion of healthcare data, conventional machine-learning methods fail to keep up without specialized knowledge. Deep learning techniques have recently surpassed classical techniques as the most popular area of study in the healthcare field (Shickel et al. (2017)). Deep learning is more efficient since it doesn't require additional work on feature engineering by a domain expert; instead, it can learn the features from raw data. Thus tracking down hidden relationships in the data that may otherwise remain uncovered. Sometimes the model makers are unable to find enough recorded events to train their model, in such cases, transfer learning, a technique of deep learning is used that applies knowledge gained from one situation to another similar situation just as humans do.

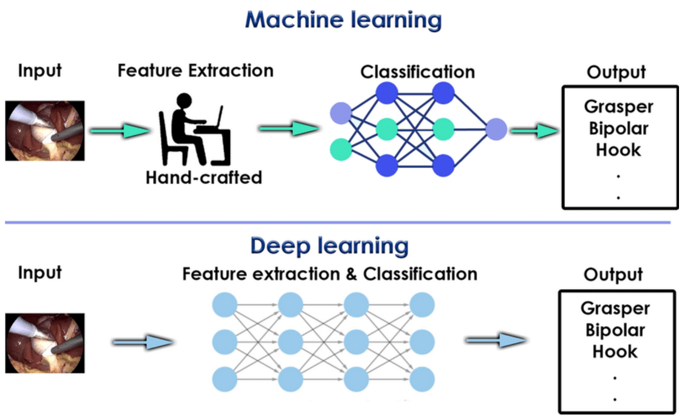

Fig. 2.1: ML vs DL approach (Jaafari et al. (2022)).

2.3 DEEP LEARNING IN HEALTHCARE

A specific subset of machine learning known as "deep learning" makes use of layered nodes, where each node calculates an input's weight. The node activation is generated by using a non-linearity function. The whole idea is to reduce the total cost by optimizing each node's weights. The neural network uses the chain rule to carry out backpropagation for weight updates after first initializing the parameters (weights and bias) and using forward propagation to determine a cost. This process of finding the optimal path to minimize the cost/error is called gradient descent.

The majority of deep learning algorithms are built upon the framework of artificial neural networks (ANN). ANNs are inspired by the biological neural networks that are the building blocks of the human brain. It consists of neurons(called nodes) that are linked to each other in various layers. Now we'll discuss some deep learning models used for deep EHR applications based on the architecture of ANN.

- **Multi-Layer Perceptron (MLP):** It is a feed-forward neural network with at least one input layer, an output layer, and several hidden layers where every node in the previous layer is completely linked to every other node in the subsequent layer. MLP uses a nonlinear activation function and the supervised learning method known as backpropagation for training. Sigmoid, tanh, and reLu are some activation functions used in MLPs. These neural networks are generally used for classification purposes in the clinical field. On applying four classifiers to Cleveland Heart Disease data, the findings demonstrate that the Deep Neural Network MLP classifier can more accurately detect a patient's condition (Krishna and Reddy (2019)).

- **Recurrent Neural Network:** When dealing with sequentially ordered data (as in this research), the Recurrent Neural Networks (RNNs) are the right choice. Whenever we feed data to RNNs, it is in the form (time step, input features). To function, RNNs update a hidden state h_t sequentially, taking into account not just the activation of the current input x_t at time t, but also the prior hidden state $h_{(t-1)}$, which was updated from $x_{(t-1)}$, and so on. In this way, information from each of the sequence's earlier pieces is present in the final concealed state that results from processing the complete sequence (as shown in Figure 2.2).

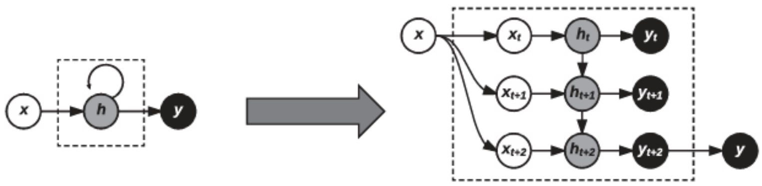

Fig. 2.2: Input sequence of length three, three hidden units, and a single output (Shickel et al. (2017) p. 5).

The gated recurrent unit (GRU) and long short-term memory (LSTM) models, sometimes known as gated RNNs, are two common RNN variations. In contrast to conventional RNNs, which are made up of interconnected hidden units, each unit in a gated RNN is substituted with a unique cell that has an internal recurrent loop and a gate mechanism that regulates the information flow. The example includes RNN with LSTM to predict the outbreak prediction of COVID-19 on big data (Natarajan, Kumar, Gadde, and Venugopal (2023)).

- **Autoencoders (AE):** An artificial neural network type called an autoencoder is used to automatically learn data encodings. Using a network trained to extract the salient features of an input picture, an autoencoder learns a lower-dimensional representation (encoding) for higher-dimensional data, usually for dimensionality reduction. The purpose of AEs is to convert the input data into a format that stores just the most crucial derived dimensions. They resemble conventional dimensionality reduction methods such as principal component analysis (PCA) and singular value decomposition (SVD) in this way, but still, they have a major advantage over them for complex problems because of the nonlinear transformations made possible by the activation functions of each hidden layer. For instance, in a research, clinical information of 714 elderly patients with Adverse Drug Events(ADE) preventability labels was provided as data by a hospital for the prediction of ADEs. Various methods were deployed to address the challenges of applying feature engineering to heterogeneous EHRs. Finally, a Dual Autoencoders model solved the problem of imbalance embedded in the training data (Liao, Derijks, Blencke, De Vries, Van Seyen, and Van Marum (2022)).

Physicians and other subject matter experts are specialized in a certain field of medicine and hence restricted to a specific anatomy. This results in a well-known knowledge gap regarding the various dependencies between illnesses from other specializations, or on the identification of patterns of visits when the doctor's knowledge is not the only factor considered. Deep learning is better at utilizing cross-domain existing patterns that were previously hidden since it is not constrained by or concerned with domain specialists' knowledge of a specific disease, specialization, or gland (Liang, Liu, Ou, Zhang, Li, and Huang (2019)). The two key aspects that deep learning is said to contribute uniquely to healthcare demand forecasting are (1) its automatic detection of patterns that are invisible to the human eye and (2) its ability to use these patterns to foresee scalable future outcomes. Given sufficient computational resources to execute the model end-to-end, deep learning approaches may speed up the process of accurate healthcare demand forecasting and support many other medical applications incorporating pattern recognition or probability prediction.

2.4 LITERATURE REVIEW

One of the most concerning areas when it comes to data collection and rectifying is the healthcare sector. A significant amount of multidimensional patient data, including clinical aspects, hospital resources, sickness diagnostic information, patient records, and medical equipment, are produced with the arrival of a patient. Extracting information for sound decision-making requires processing and evaluating massive, dense, and complicated data. This literature review is conducted to identify the machine learning, or rather deep learning methods that can satisfactorily make future predictions on this electronic health record data. Creating a predictive model that is reliable, accurate, and easy to understand is the main and most difficult task in diagnosis prediction. For the purpose of identifying algorithms that have been suggested in the past, the literature study makes use of the systematic studies that are currently available on the subject of deep learning applied to EHR data in order to build the appropriate search queries.

The purpose of creating a deep learning model for illness classification is to map the input EHR data to the output disease target using numerous layers of neural networks. One of the comparative systematic reviews that we'll consider is given by Xiao, Choi, and Sun (2018). From the examined publications, some used RNN on longitudinal outpatient data from Sutter Health to predict the future beginning of a new clinical condition, for example, heart failure (HF) (Choi, Schuetz, Stewart, and Sun (2017b)), some studies enable both binary (e.g., disease onset (Cheng, Wang, Zhang, and Hu (2016); Kam and Kim (2017)) and multi-class (e.g., Parkinson's disease stages (Che, Xiao, Liang, Jin, Zho, and Wang (2017)) classifications and use data from multiple modalities (e.g., cognitive tests, vital signs, medical imaging). In addition to diagnosing diseases, a number of studies developed the prescription of medications as a sequential prediction issue. For example, sequential medication prediction was performed (Bajor and Lasko (2022)) using 610076 patient information from Vanderbilt's Electronic Medical Record. Furthermore, a large number of publications used EHR data from a large number of patients to perform multilabel sequential prediction of clinical events. Multiple target labels may co-occur during a single visit for each patient according to multilabel prediction (e.g., multiple diagnoses in one visit). To anticipate the diagnosis categories for a follow-up appointment, for example, in Doctor AI, Choi et al. (2016a), 2,63,706 patients' encounter records (such as diagnosis codes, prescription codes, or procedure codes) were fed into an RNN model.

Most of the work in this domain relied on supervised learning where both labeled input and output were known. For instance, Dipole (Ma, Chitta, Zhou, You, Sun, and Gao (2017)), predicting clinical visits using RNN and Demographic Information (Wang, Li, Cui, Hong, and Yan (2018)), MSAM (Zeng, Feng, Moosavinasab, Lin, Lin, and Liu (2019)), learning to diagnose with LSTM RNN (Lipton, Kale, Elkan, and Wetzel (2015)), etc. Unsupervised learning showed another most common category of tasks. Clustering of unsupervised predecessors or neighboring diseases was commonly used to identify them. Graph-based Attention Model (GRAM) (Choi, Bahadori, Song, Stewart, and Sun (2017a)), Knowledge-based Attention Model (KAME) (Ma, You, Xiao, Chitta, Zhou, and Gao (2018)), Med2Vec (Choi, Bahadori, Searles, Coffey, Thompson, Bost, Tejedor-Sojo, and Sun (2016c)), and Deep Patient (Miotto, Li, Kidd, and Dudley (2016)) are some of the examples.

The majority of the methods listed above incorporated Recurrent Neural Networks with some using Gated Recurrent Unit (GRU) while others using the Long Short Term Memory Unit (LSTM) variant. In addition, GRAM, KAME, MSAM, Dipole used attention mechanisms in their methods while Med2Vec used binary vectorization. The Deep Patient method, which employs Stacked Denoising Autoencoders (SDA) is the only traditional method that does not require neural networks. The authors demonstrated that their representation learning technique outperformed shallow feature learning in predicting future illness. Since all these methods used different datasets with variable patient data to create different predicting models, it is likely impossible to make comparison amongst them.

Typically, the approaches described depend on medical data from a patient's history that is organized in chronological order. RNNs are commonly used for such information as shown in Doctor AI (Choi et al. (2016a)), Dipole (Ma et al. (2017)), and RNN-INFO (Wang et al. (2018)). For the translation of medical/diagnoses codes, the Skip-Gram embedding method (Mikolov et al. (2013a)) was used for deep learning models by Choi et al. (2016c). Further to overcome the problem of insufficient data, Choi et al. (2017a) used their model to represent the patient's visit. KAME and GRAM both were unable to link the diagnoses codes with healthcare products and so they are unfit for our research. Since our study aims to forecast future healthcare by analyzing individual patient visit histories, thus the scope relevance is limited to machine learning approaches that include a temporal dimension. The most appropriate methods for this aim are outlined in the next section.

2.5 MOST SUITABLE METHODS

Our main target is to predict future healthcare demand correctly based on the previously recorded EHRs. The three methods that closely favored our target are Doctor AI(Choi et al. (2016a)), MSAM(Zeng et al. (2019)) and Dipole(Ma et al. (2017)).

Doctor AI is a predictive model that takes into account observed medical conditions as longitudinal time-stamped patient visit records, uses recurrent neural networks, and predicts the diagnosis and medication categories for the next visit. It also predicts the time of the patient's next visit. The use of a sizable patient database demonstrated that this approach addresses a wide range of illnesses, which therefore implies that diseases with less number of patients would result in performance degradation.

- Architecture: RNN,GRU(Supervised Deep Learning)
- Data: 2,63,706 patients
- Timespan: 8 years
- # of visits: 1,44,00,985
- # of unique medical codes: 38,594
- Train, Test split: 85%, 15% resp.
- Recall@30: 79.58%

Dipole(a diagnosis prediction model), the method proposed shortly after Doctor AI uses attention-based bidirectional RNNs to model EHR data and utilizes LSTM to remember the information of past visits to accurately provide meaningful interpretations. Since the significance of each patient's visit and the medical codes associated with each visit have varying importance, so this model assigns varying attention weights to each past visit. The training of the model was done on The Diabetes and Medicaid Datasets. The accuracy on both datasets was higher as compared to simple baselines and RNN variants.

- Architecture: BRNN, LSTM, Attention(Supervised DL)
- Data: 1,47,810 patients
- Timespan: 1 year
- # of visits: 10,55,011
- # of unique medical codes: 8,522
- Train, Test split: 90%,10% resp.
- Accuracy@30: 84.75%

MSAM(Multilevel Self Attention Model) captures the relationship between medical codes and medical visits to predict future disease as well as medical costs (similar to the one at hand). The PFK (Partner for Kids) and MIMIC-3 (Medical Information Mart for Intensive Care 3) datasets were used to train and assess the model's prediction performance. Mean absolute error (MAE) was defined as the absolute difference between the predicted cost and the true cost. The main advantage of the results above baseline models was their improved ability to handle irregular time periods.

- Architecture: FFNN, Attention(Supervised DL)
- Data: 1,46,287 patients
- Timespan: 2 year
- # of visits: 13,01,954
- # of unique medical codes: 12,334
- Train, Test split: 80%,20% resp.

- Recall@30: 79.48%

Overall the three methods discussed above used supervised deep learning models to enhance the prediction on EHR data by addressing issues like high dimensionality and temporality. However, practical validations can only be carried out on one of the approaches, as the research's nature limits its ability to model a method from scratch and in terms of time. Thus, Doctor AI is included in this research as it is similar to the most widely used approach in this field of study.

3. METHODOLOGY

3.1 THE CRISP-DM APPROACH

This research's methodology follows the CRISP-DM (Cross Industry Standard Process for Data Mining) structure (Chapman (2000)). Since its publication, this approach is by far the most popular technique for data mining, analytics, and data science initiatives. As shown in Figure 3.1, it has six main phases:

1. **Business understanding** - Understanding the project's goals and specifications is the main focus of the business understanding phase. This information is then turned into a description of the data mining challenge and a draft plan intended to accomplish the goals. A thorough discussion of this phase is done in the previous two chapters.

2. **Data understanding** - This phase deals with collecting, identifying the relationships, exploring, visualizing, and analyzing the data efficiently. The earlier chapters have already discussed this.

3. **Data preparation** - An important part of the CRISP-DM process is getting the data ready for analysis. During this stage, resolving missing values, cleansing data, and changing variables are crucial activities. The objective is to provide a well-organized, clean dataset that is prepared for further stages. The preparation of data will be done in this chapter.

4. **Modeling** - In this phase, various modeling techniques such as machine learning and statistics are applied to the dataset that we prepared earlier. Since we are dealing with EHR data to predict future healthcare demand,

so selecting the best model that accurately gives desired results becomes very crucial. This chapter covers modeling as well.

5. **Evaluation** - Models are subjected to thorough review once they are constructed. During this phase, performance, dependability, and capacity to achieve the business objectives are checked. The previous phases may be revisited to enhance and get better results. The evaluation metrics will be discussed in the current and next chapters.

6. **Deployment** - Successful data mining yields tangible results during the phase of deployment. Nevertheless, the client must have a clear understanding of the necessary steps to effectively utilize the generated model, even if he/she is not responsible for carrying out this phase practically. The next chapter deals with this phase.

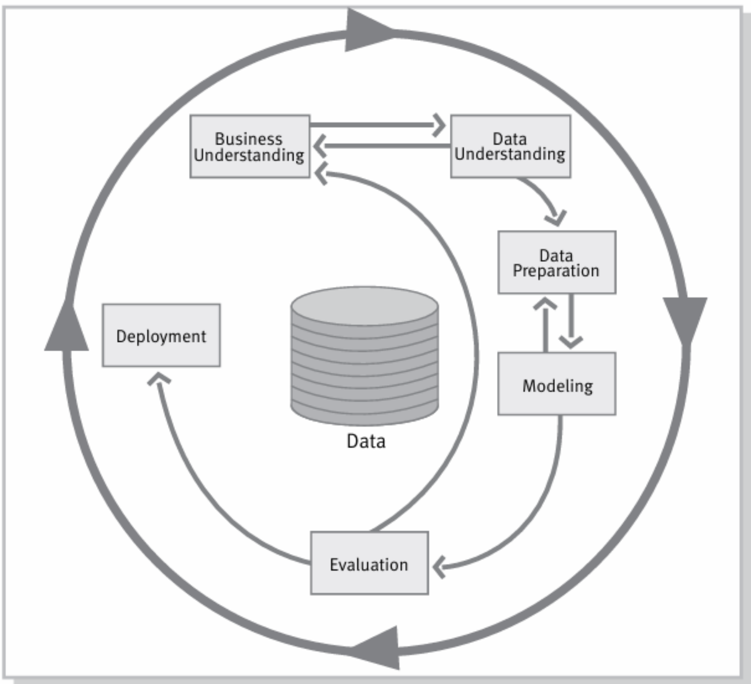

Fig. 3.1: The phases of CRISP-DM (Chapman (2000) p. 10)

3.2 DATA SCRAPING AND SELECTION

The data for this research is taken from the Medical Information Mart for Intensive Care (MIMIC-III) database (Johnson, Pollard, Shen, Lehman, Feng, Ghassemi, Moody, Szolovits, Celi, and Mark (2016)), which is the only freely accessible critical care database. It consists of clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. Information from 53,423 different adult patients (age 16 or older) hospitalized in critical care units between 2001 and 2012 is presented. Furthermore, it includes information on 7870 newborns hospitalized between 2001 and 2008. 49,785 hospital admissions and 38,597 unique adult patients are included in the data. There are 26 tables such as 'ICU_STAYS', 'ADMISSIONS', 'DIAGNOSES_ICD' 'DRG_CODES', etc. provided as CSV files. Tables are connected by identifiers, which often end in "ID".

Our task of predicting future diagnosis codes on the basis of the previous medical record of a patient required us to select 'ADMISSIONS', 'DIAGNOSES_ICD', 'PROCEDURES_ICD', and 'PRESCRIPTIONS' files from the complete dataset. The information regarding the patient's admission to the hospital is given by the 'ADMISSIONS' table and each unique hospital visit for a patient is assigned a unique 'HADM_ID'. The 'DIAGNOSES_ICD' table gives us the diagnosis codes, coded using the International Statistical Classification of Diseases and Related Health Problems(ICD-9), assigned by hospitals to each patient at each visit. The 'PROCEDURES_ICD' contains ICD-9 procedure codes for patients. 'PRESCRIPTIONS' table consists of medications ordered for a given patient.

3.2.1 Selected Variables

To start with, I extracted subject_id, hadm_id, admit_time, discharge_time columns from the 'ADMISSIONS' table and joined it with icd9_code (diagnoses) column from the 'DIAGNOSES_ICD' table using hadm_id as joining key via a SQL query, as shown in Source Code 3.1. Again these four columns from the 'ADMISSIONS' table are joined with icd9_code (procedures) from the 'PROCEDURES_ICD' table using hadm_id as the joining key. The step is repeated with the 'drug' column from the 'PRESCRIPTIONS' table again using hadm_id as the joining key. The length of common subject_id values after the intersection of these three resulting tables is 82 and that of hadm_id is 107. On merging the features (using Source Code 3.2), the extracted dataset includes a total of 7 columns described further in more detail. After removing null values, we have a total of 13,36,814 rows.

```
#Defining SQL query with JOIN
diagnoses_df = pd.read_sql_query(""" SELECT
    adm.subject_id, adm.hadm_id, adm.admittime, adm.dischtime
    , diag.icd9_code
FROM
    sql_admission adm
JOIN
    sql_diagnoses diag ON adm.hadm_id =diag.hadm_id;

""", conn)
```

Source code 3.1: Joining columns from two tables

```
merged_df = diagnoses_df.merge(medications_df, on=['
    subject_id', 'hadm_id', 'admittime', 'dischtime'], how='
    outer', suffixes=('_diag', '_med'))
merged_df = merged_df.merge(procedures_df, on=['subject_id',
    'hadm_id', 'admittime', 'dischtime'], how='outer',
    suffixes=('', '_proc'))
```

Source code 3.2: Merging the columns

Below is the description of each included variable.

HADM_ID: HADM_ID is an integer data type. Each entry in this database contains a unique HADM_ID that corresponds to a particular patient's admittance to the hospital. HADM_ID runs from 1000000 to 1999999.

SUBJECT_ID: SUBJECT_ID is an integer data type. This table may include duplicate SUBJECT_IDs, suggesting that a single patient has been admitted to the hospital more than once.

ADMIT_TIME, DISCH_TIME: ADMIT_TIME indicates the date and time the patient was admitted to the hospital, whereas DISCH_TIME indicates when the patient was released. Their data type is TIMESTAMP.

ICD9_CODE: This feature's datatype is VARCHAR(10). The actual code for the diagnostic given to the patient for the specified row is contained in ICD9_CODE. All codes, as of MIMIC-III, are ICD-9 codes. ICD-9 is a medical classification list for the diagnosis of a disease given by the WHO and has been freely available online ever since. It is also the variable based on which the categorization of next diagnosis code will be done.

DRUG: This column contains the medication prescribed by a doctor to a patient at a particular visit. Its datatype is VARCHAR(100).

ICD9_CODE_PROC: This feature's datatype is VARCHAR(10). It provides the ICD-9 procedure codes for every patient that visits the hospital on the basis of the procedures list of ICD-9 codes released by the WHO.

Using Label Encoder, the three features, namely, icd9_code, drug, and icd9_code_proc having categorical data are converted to the numerical format.

```
label_enc_diag = LabelEncoder()
label_enc_med = LabelEncoder()
label_enc_proc = LabelEncoder()

merged_df['icd9_code'] = label_enc_diag.fit_transform(
    merged_df['icd9_code'].astype(str))
merged_df['drug'] = label_enc_med.fit_transform(merged_df['
    drug'].astype(str))
merged_df['icd9_code_proc'] = label_enc_proc.fit_transform(
    merged_df['icd9_code_proc'].astype(str))
```

Source code 3.3: Label Encoding

Upon examination, it was discovered that the output class has 523 unique values (as shown in Figure 3.2), with many classes having very few rows. Additionally, the findings of the application of Doctor AI indicate poor performance in terms of forecasting a large number of output classes. Hence, we determined the 100 most frequent values in the icd9_column for further analysis, and 9,67,204 is the resulting row count.

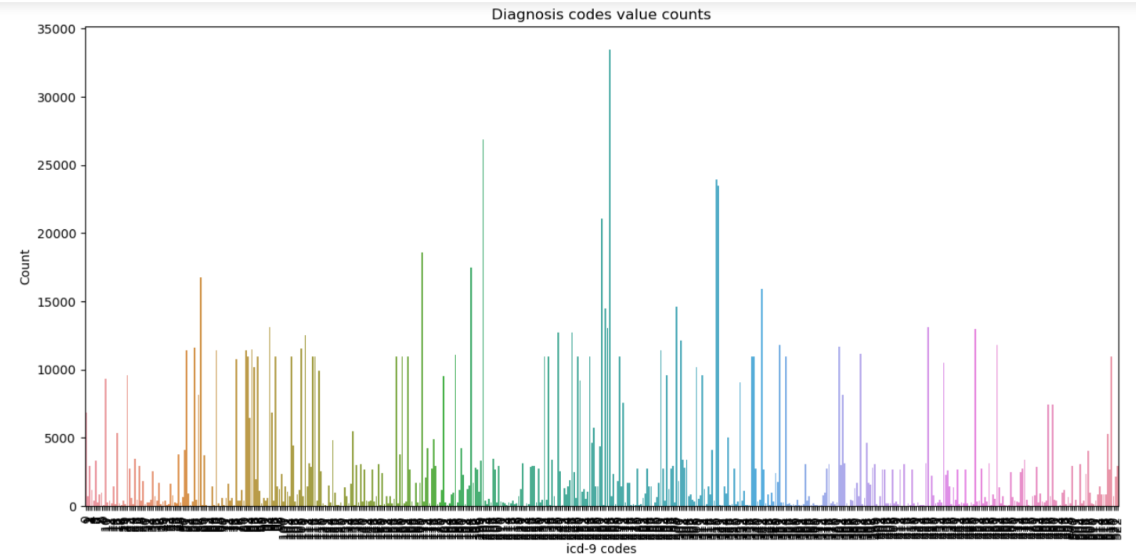

Fig. 3.2: Value counts for unique diagnosis codes

3.2.2 Sequences and Labels

Now, the selected variables cannot directly serve as input for the machine learning algorithm. We need to prepare sequences of medical events for each patient to put their medical history in a nutshell. After that, the corresponding labels will be generated for predictive modeling. To construct sequences, unique subject_IDs (patient IDs) are retrieved, the dataframe is filtered to extract just that patient's data, and sorted by admission time (admit_time). The sequences of diagnoses, drugs, and procedures are then extracted and combined into a single sequence for every patient. To efficiently handle data, machine learning algorithms require fixed-length inputs. To address this problem, the merged sequences are padded with zeros to ensure a constant length. To create labels for each sequence, the diagnosis(icd9_code) column is shifted to predict the next diagnosis. These labels are then one-hot encoded to be used for training the model. This method guarantees that all patient data is consistently structured and ready for forecasting.

3.3 DOCTOR AI ALGORITHM

Electronic health records, or EHRs, are being widely used in healthcare and serve as a repository for the medical history of both patients and physicians. Future events are being predicted using these data more and more frequently. Although demand has been anticipated through the development of predictive models, the majority of previous research is still specialized in forecasting a small range of outcomes. However, a daily clinical practice requires hundreds to thousands of different prediction models due to its unplanned and varied mix of events, so implementing the models one after the other is practically not feasible. Doctor AI (Choi et al. (2016a)) is a predictive model that uses recurrent neural networks with Gated Recurrent Units (GRU) to predict future healthcare demand for a patient. The hospital dataset of 260K patients and 2128 physicians over 8 years was used as longitudinal time-stamped EHR data. To create multilabel predictions, doctor AI evaluates the medical histories of its patients. Doctor AI was able to foresee clinical diagnoses with a significant accuracy of 79.58%recall@30, matching that of a human doctor, after 20 rounds of training. The complete framework of Doctor AI is explained in the following few paragraphs.

3.3.1 Vectorization

To start with, the observations are taken from a multilabel point process for each patient, where i = 1,2,.., n in the form of (t_i, $\mathbf{x}_i$). Each pair denotes a distinct visit to a hospital, containing numerous medical codes, such as 3-digit ICD-9 diagnostic codes, procedure codes (CPT i.e., Current Procedural

Terminology), and medication codes (Generic Product Identifier Drug Class). The vector $\mathbf{x}_i \in \{0,1\}^p$ is the multi hot-label representing medical codes given to a patient at any time t_i and p is the number of unique medical codes. Finally, for prediction purposes, higher level codes are extracted which are denoted by vector $\mathbf{y}_i$ (combination of diagnosis and medication codes). In each visit, a prediction about diagnosis and medication code in the next visit is made along with the time to next visit.

Numerous clinical applications may operate better if users learn effective representations of medical codes. Skip-Gram (Mikolov, Chen, Corrado, and Dean (2013b)) utilized real-valued multidimensional vectors to capture the latent representation of medical codes from EHRs. In this approach, a target vector is given while training and the related context words are received. Then the new predicted vector is compared with the context word vectors and similar medical codes are embedded close to one another. By using this variant of embedding, the medical codes were converted into lower-dimensional space. The Doctor AI can optimize the weights of W_{emb} (weight matrix between the input vector and embedding layer) during training rather than having to learn them from scratch thanks to this method of the Skip-Gram approach. In this manner, as the entire model is trained, the W_{emb} weights get better. The formula used for embedding is

$$h_i^{(1)} = [x_i^T \, W_{emb}, \, d_i]$$

These multi-hot vectors created by Skip-Gram are now the inputs for RNN and GRU setup.

3.3.2 The setup of RNN with GRU

Recurrent neural networks (RNNs) are a form of neural network designed to analyze sequential input. They are intended to retain information from earlier inputs. It enables them to employ context and interdependence across time steps, which are absolutely necessary for predicting future healthcare demand. RNNs come in several forms, such as gated recurrent units (GRUs) and long short-term memory (LSTM) networks. The issue of vanishing gradients in RNNs, which arises when the network's weight gradients are extremely tiny and the network struggles to train, is one that LSTMs and GRUs are both intended to solve. Tasks where the network must learn rapidly and adjust to new inputs are better suited for GRUs. GRUs consist of an update gate and a reset gate. The update gate decides which information from the prior concealed state and the current input to maintain, whereas the reset gate decides which information to discard. The final hidden state is a mix of the information stored by the update gate and the current input.

The Figure 3.3 depicts the structural setup of RNN to predict a patient's time to the next visit $d_{i+1} = t_{i+1} - t_i$ and the corresponding healthcare product demand in the next visit y_{i+1}. It takes input after a fixed time interval t_i. The first layer shows the conversion of higher dimensional input vectors to lower dimensions. The two layers above it are the recurrent units that understand the patient's condition at each timestamp of gathering data as a vector with actual values. On providing the status vector, these two dense layers help to generate the healthcare codes as well as the time to the next visit, given by the last layer.

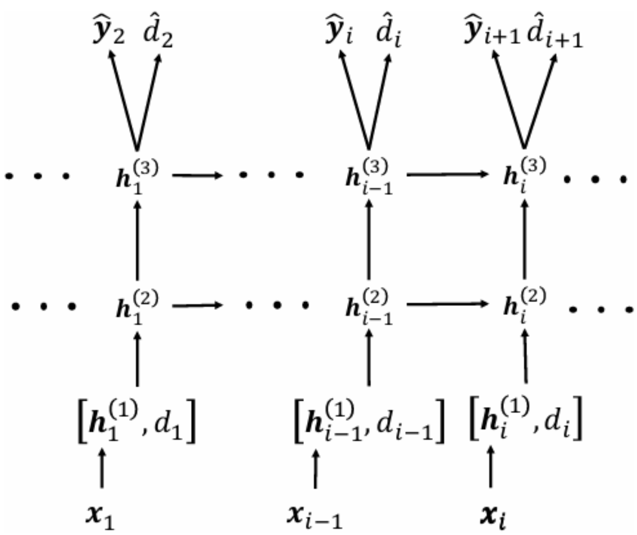

Fig. 3.3: The RNN architecture (Choi et al. (2016a) p. 5)

The formulae used in RNN-GRU architecture are given below:

$$z_i = \sigma(W_z x_i + U_z h_{i-1} + b_z)$$

$$r_i = \sigma(W_r x_i + U_r h_{i-1} + b_r)$$

$$\hat{h}_i = tanh(W_h x_i + r_i \circ U_h h_{i-1} + b_h)$$

$$h_i = z_i \circ h_{i-1} + (1 - z_i) \circ \hat{h}_i$$

At a particular timestamp t_i, x_i is the input, z_i is update gate, r_i is reset gate, $\hat{h}_i$ is intermediate memory unit, and h_i is the hidden layer. The values of all ***W's*** and ***U's*** were initialized to orthonormal matrices using singular value decomposition of matrices generated from the normal distribution (Saxe, McClelland, and Ganguli (2013a)). Using the sigmoid activation function, the output between 0 and 1 is obtained. The statement suggests that when the reset gate value is nearly zero, the input value x_i will be utilized, and the intermediate memory unit will discard the value from the previous hidden layer. Conversely, when the update gate value is close to one, the current hidden layer will retain the value from the previous timestep h_{i-1} and ignore the current input value.

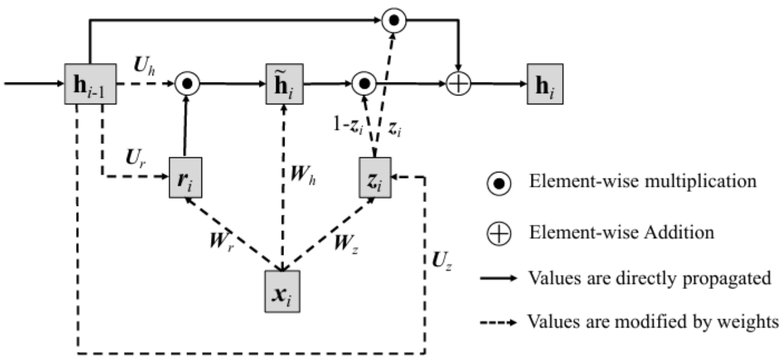

Fig. 3.4: The GRU architecture (Choi et al. (2016a) p. 14)

In Figure 3.4, the previous hidden layer h_{i-1} and the current input x_i do not directly affect the value of the current hidden layer h_i. Instead, they modify the values of both the gates and the intermediate memory unit $\hat{h}_i$. Then $\hat{h}_i$ and z_i update the current hidden layer h_i. To sum up, the previous hidden layer works as the memory of the system i.e., the previous record of the patient, the reset gate makes the hidden layer drop any information that is not required in making the prediction, and the update gate decides the amount of information to be propagated from previous to the current hidden layer.

3.3.3 Softmax Loss Function

Softmax Loss is simply the combination of a Softmax Activation and a Cross Entropy Loss. The probability for each class is produced by the activation function Softmax to normalize the output of the network, and these probabilities add up to one. The total of the probabilities' negative logarithms is the cross entropy loss. Using h_i, a softmax layer is added on top of GRU architecture for the prediction of medical and diagnosis codes, and a Rectified Linear Unit(ReLU) is added to predict the time duration until the next visit. The goal is to learn the values of ***W's***, ***U's***, and ***b's***. The initial value of $\mathbf{W}_{time}$ was chosen from the uniform distribution between -0.1 and 0.1 (Saxe, McClelland, and Ganguli (2013b)) and all ***b's*** were initialized to zeros. The cross-entropy loss function is used to find the difference between prediction and real values for a single patient.

$$\hat{y_{i+1}} = softmax(W_{code}^{T} h_i + b_{code})$$

$$\hat{d_{i+1}} = \max(\mathrm{W}_{time}^{T} \mathrm{h}_i + \mathrm{b}_{time}, 0)$$

$$L(\boldsymbol{W},\boldsymbol{U},\boldsymbol{b}) = \sum_{i=1}^{n-1} \{y_{i+1} log(\hat{y_{i+1}}) + (1 - y_{i+1}) log(1 - \hat{y_{i+1}})\}$$

3.4 TRAINING THE MODEL

The section 3.2 explained what Choi et al. (2016a) did in his own implementation of Doctor AI. After making a few tweaks and implementing the code using our own way, we describe the code that we used to train the model to predict the future diagnosis for each patient, while also including strategies to prevent overfitting and dynamically alter the learning rate.

After the preprocessing, now we split our data as 80% for training and 20% for testing. Since we are making a recurrent neural network, we'll separately elaborate upon the different layers that we used while training. The first one is the **Input Layers** of diagnosis, medications, and procedures for the model (as shown in Source Code 3.4). The maximum sequence length in our case is 50.

```
#Input Layers
input_diagnoses = Input(shape=(max_seq_length,), name='diagnoses_input')
input_medications = Input(shape=(max_seq_length,), name='medications_input')
input_procedures = Input(shape=(max_seq_length,), name='procedures_input')
```

Source code 3.4: Defining Input Layers

Next are the **Embedding Layers** for diagnosis, procedures, and medications. The Source Code 3.5 shows the embedded diagnosis layer. Here, input_dim is the number of unique diagnosis codes. Similarly, the other two features are embedded with an embedding size of 128. Then we concatenate the embeddings from different input types for the model to learn from the combined information.

```
embed_diagnoses = Embedding(input_dim=len(encoders['icd9_code'].classes_), output_dim=embedding_size, input_length=max_seq_length)(input_diagnoses)
```

Source code 3.5: Defining Embedding Layers

We will now define the **GRU layers** as we are utilizing the GRU variation of RNN. In the following chapter, we will examine the model's performance with 200 and 500 nodes. To disregard the zero-padded numbers, masking is used. Utilizing a bidirectional GRU layer, process the patient's prior medical history while taking the current input into account. A 0.9 dropout rate is used to avoid overfitting.

```
# GRU layers
x = Masking(mask_value=0)(merged_input)
x = Bidirectional(GRU(500, return_sequences=True))(x)
x = Dropout(0.9)(x)
x = BatchNormalization()(x)
x = Bidirectional(GRU(500, return_sequences=True))(x)
x = Dropout(0.9)(x)
x = BatchNormalization()(x)
```

Source code 3.6: GRU Layers

Next, we'll apply two **dense layers** one with activation function 'leaky reLu' and 256 units and the second that outputs the probabilities for each class using the 'softmax' activation function.

Now, the model is defined using diagnoses, medications, and procedures as inputs and it outputs a sequence of class probabilities. We used categorical cross-entropy loss, which is a preferred choice for multi-class classification. Finally, the model is trained on 10, 20, and 30 epochs for comparison. 20% of training data is used for validation. Callbacks are defined to stop the training to check the validation loss improvement.

```
# Train the model
history = model.fit(
    [X_train_diagnoses, X_train_medications,
   X_train_procedures],
    y_train, epochs=30, batch_size=120, validation_split=0.2,
    callbacks=[early_stopping, reduce_lr])
```

Source code 3.7: Model Training

3.5 PERFORMANCE METRICS

Examining the performance of supervised multiclass classification models requires determining the appropriate metric to use. Usually, this is dependent on the problem context. For instance, in the multiclass classification situation, the problem of class ambiguity may emerge when an algorithm struggles to distinguish between extremely similar classes, in contrast to the binary case where accuracy and recall alone may be the optimum choice for performance evaluation. The healthcare sector faces this problem very often where one gets confused between the diagnoses and treatment of the illness related to the same body part. This issue appears when misclassifying a tiny class might have significant negative effects, for instance, coping with an uncommon but potentially fatal illness. The expense of incorrectly predicting a sickness and testing a healthy individual on additional metrics is significantly less than the cost of failing to identify illness in an unwell person. As a result, it is critical to apply assessment metrics in healthcare prediction with greater prudence and from a much wider viewpoint.

The metrics that are most commonly used for multiclass classification in the medical field for evaluating ML methods are the top-k metrics. Extending the learning criterion from just allowing accurate (top-1) predictions to admitting k predictions, among which the correct class must be, is the aim of top-k learning. In other words, the method considers a class to be accurate if it is in the top k in terms of probability. For instance, in a recommendation system, we are in the mood to listen to songs of a particular genre i.e., our goal is to identify a group of intriguing recommendations rather than choose the best one. In top-k learning, k can vary depending on the model, category, or a mix of these in its generic version. The top-k metrics behave similarly to a clinician making differential diagnoses for patients. Using this system, doctors prioritize the most likely diagnosis and treat patients accordingly. The doctor and top-k metrics can apply the best therapy to a patient's diagnosis by following the most likely solution.

The evaluation metric that we are using for evaluating our model in this work is the accuracy@top-k. Accuracy@top-k measures the method's ability to accurately categorize output among the top-k results (i.e., no. of correct predictions in top-k divided by total no. of predictions). The higher the value of accuracy@top-k, the better our model is at predicting the next healthcare product.

4. ASSESSMENT OF RESULTS

In this chapter, we check the performance of our model. Since this study is conducted in a supervised environment, only fully labeled data is used for model validation, testing, and training. The following table compares the results of assessing the constructed model for 200 and 500 nodes using the accuracy@top-k performance metric for varying values of k and setting different epoch values.

		ACCURACY @ TOP - K		
NODES	EPOCH	Acc@10	Acc@20	Acc@30
200	10	79.125%	90.875%	84.750%
	20	90.125%	82.875%	92.875%
	30	83.249%	90.249%	90.875%
500	10	83.749%	84.500%	90.375%
	20	82.499%	90.375%	**94.375%**
	30	82.749%	94.125%	84.625%

Tab. 4.1: Performance metrics

The comparison above demonstrates that our model achieves its maximum accuracy of 94.375% accuracy@30 when we have 500 nodes in each of our GRU layers and we train for 20 epochs. This indicates that the algorithms' ability to correctly forecast the proper class is increased when the number of output classes is reduced while maintaining the healthcare products that occur most frequently.

The best-performing hyperparameter settings, which provided accuracy@30 of 94.375%, will now be used to further visualize our model's performance and improve our understanding of it. As can be seen from the graph of epochs versus categorical cross-entropy loss (in Figure 4.1), on increasing the number of epochs, our validation loss is decreasing. But if we keep on increasing the number of epochs further, our model will go towards overfitting.

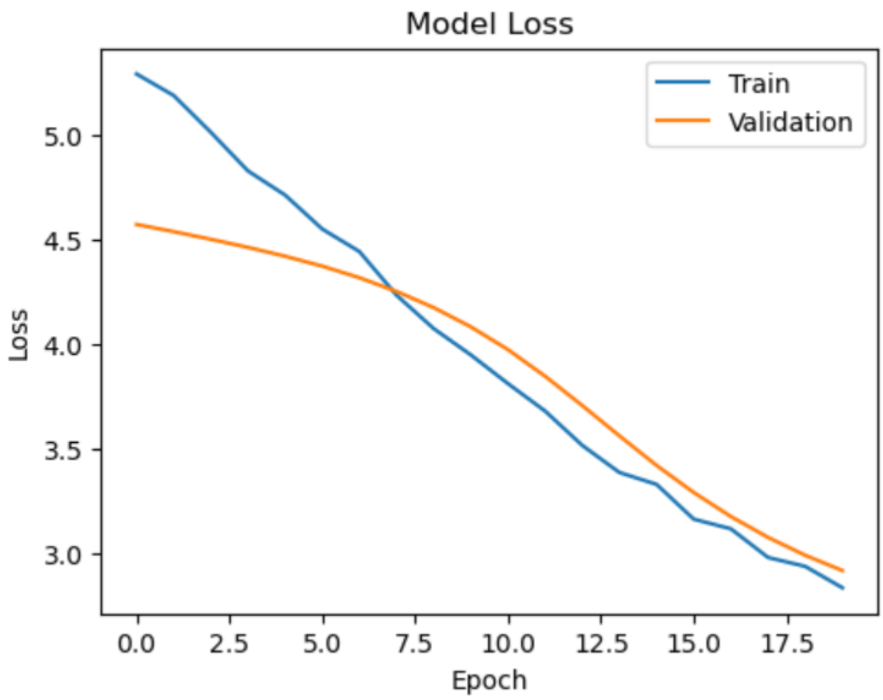

Fig. 4.1: Relation between loss and epoch values

Since there are 100 classes, so for a better visualization, a smaller version heatmap of the confusion matrix is drawn using Seaborn to check the model's performance for each class separately for test data (Figure 4.2). The plot is labeled with 'Predicted' on the x-axis and 'True' on the y-axis. The diagonal cell entries are filled with certain integers which represents the number of times the class was correctly predicted. For instance, 104 in the cell (8,8) indicates that the class number 8 was correctly predicted as class 8 104 times. Off-diagonal values give the misclassified values. For example, 1 in the cell

(0,7) indicates that class 0 was incorrectly predicted as class 7 once. Cell (14,11) has the highest non-diagonal entry of 50 showing that these two classes are highly getting confused with each other.

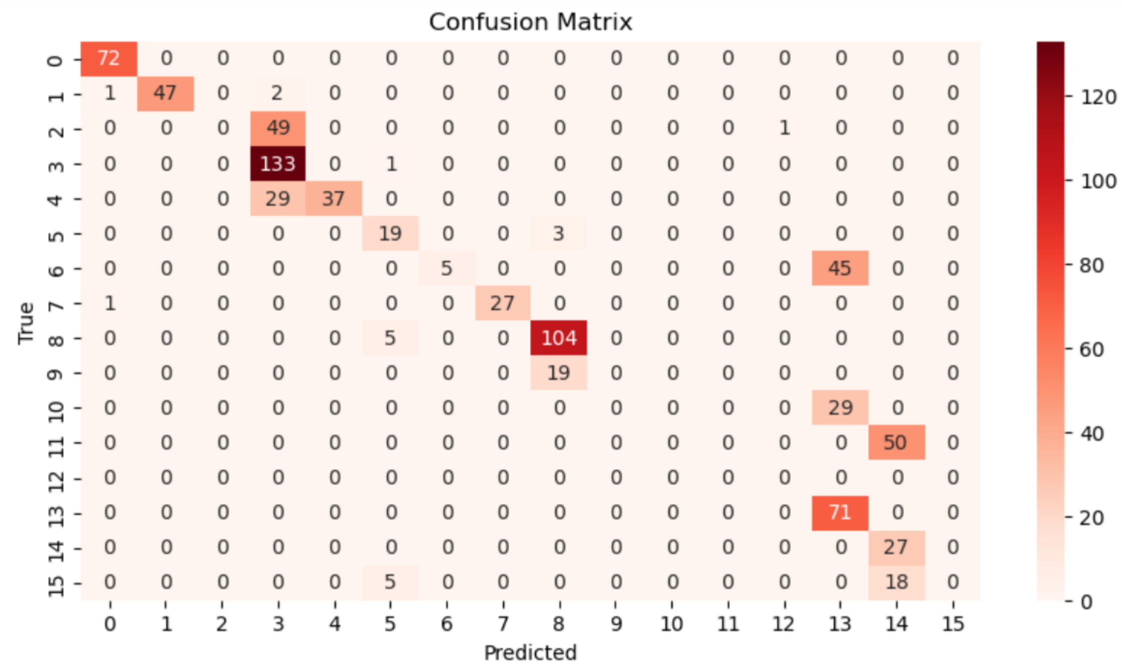

Fig. 4.2: True value versus predicted value

5. CONCLUSION AND FUTURE DIRECTIONS

5.1 CONCLUSION

The goal of this chapter is to gather all of the work completed in earlier chapters. The main focus of this research was to know whether hospitals might use machine learning on EHR data to correctly predict future healthcare demand. Hospitals are therefore able to forecast and effectively manage their budgets a year early. Based on a discussion of the difficulties of EHR, machine learning, and deep learning in the medical industry, as well as a literature review, the Doctor AI algorithm was chosen as the most prominent technique for the challenge. After downloading the MIMIC-III dataset from Kaggle, it was trained and evaluated in accuracy@top-k for three values of k and 200, 500 nodes. To evaluate overall performance and simplify the model, the number of output classes was reduced to the top 100 categories with the maximum amount of rows.

Hospitals may be able to collect more insurance payments if they successfully use machine learning capabilities to estimate healthcare demand. While establishing annual positive profit margins is widely thought to be important to maintain sustainable high healthcare quality, more profit is not necessarily a clear correlate of better healthcare (Beauvais and Wells (2006)). Hospitals may spend in critical areas such as staff training, capacity growth, funding of research, and maintaining affordable healthcare if they are able to predict insurance compensation. By predicting patient demand hospitals may more efficiently distribute their few assets— such as treatment facilities, nurses, and doctors—across their several departments.

The results of the conducted literature review indicate that deep learning is the predominant technique for predicting healthcare demand, despite the recent development of several machine learning variants and applications. Large volumes of EHR data must be processed, and while each deep learning approach has its own special characteristics, the fundamental power of finding hidden patterns in data and using these to forecast scalable end-to-end algorithms is evident. Doctor AI's RNN/GRU architecture has a track record of wide and generalizable prediction capabilities, supported by literature. Because we took fewer output classes into consideration, the model we used produced satisfactory results. Consequently, it illustrates the possible financial effects of using machine learning to anticipate healthcare demand. In conclusion, machine learning can be used for prediction purposes in the future since it can handle EHR data quite efficiently. Hence, this model might directly add value and act as the foundation for further research into the specific healthcare need.

5.2 LIMITATIONS AND FUTURE SCOPE

One significant limitation of this study is the unavailability of the implementation of two additional methods, MSAM and Dipole, in addition to Doctor AI, and the inability to compare them to decide which works better. To compare several ways, we would have had to develop the algorithm from start to finish ourselves, which would have required significant coding skills. Furthermore, completing the Master's thesis in the allotted time would not have permitted so. Future studies should not only concentrate on creating novel approaches to address real-world issues but also expand the field of machine learning applications by contrasting current approaches with other real-world issues to determine the best course of action. Another limitation might be the reduction of output categories, which resulted in the elimination of a huge number of diagnosis codes that the hospital cannot overlook. The next study should focus on this issue as well and provide a solution.

BIBLIOGRAPHY

Bajor, J. M. and T. A. Lasko (2022): "Predicting medications from diagnostic codes with recurrent neural networks," in *International conference on learning representations*.

Beauvais, B. and R. Wells (2006): "Does money really matter? A review of the literature on the relationships between healthcare organization finances and quality," *Hospital Topics*, 84, 20–29.

Bengio, Y., A. Courville, and P. Vincent (2013): "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, 35, 1798–1828.

Bhardwaj, R., A. R. Nambiar, and D. Dutta (2017): "A study of machine learning in healthcare," in *2017 IEEE 41st annual computer software and applications conference (COMPSAC)*, IEEE, vol. 2, 236–241.

Billings, J., I. Blunt, A. Steventon, T. Georghiou, G. Lewis, and M. Bardsley (2012): "Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30)," *BMJ open*, 2, e001667.

Chapman, P. (2000): "CRISP-DM 1.0: Step-by-step data mining guide," .

Che, C., C. Xiao, J. Liang, B. Jin, J. Zho, and F. Wang (2017): "An rnn architecture with dynamic temporal matching for personalized predictions of parkinson's disease," in *Proceedings of the 2017 SIAM international conference on data mining*, SIAM, 198–206.

Cheng, Y., F. Wang, P. Zhang, and J. Hu (2016): "Risk prediction with electronic health records: A deep learning approach," in *Proceedings of the 2016 SIAM international conference on data mining*, SIAM, 432–440.

Choi, E., M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun (2016a): "Doctor ai: Predicting clinical events via recurrent neural networks," in *Machine learning for healthcare conference*, PMLR, 301–318.

Choi, E., M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, and J. Sun (2016b): "Multi-layer representation learning for medical concepts," in *proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1495–1504.

——— (2016c): "Multi-layer representation learning for medical concepts," in *proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1495–1504.

Choi, E., M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun (2017a): "GRAM: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 787–795.

Choi, E., A. Schuetz, W. F. Stewart, and J. Sun (2017b): "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, 24, 361–370.

Coates, A. and A. Y. Ng (2011): "The importance of encoding versus training with sparse coding and vector quantization," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 921–928.

Dai, W., T. S. Brisimi, W. G. Adams, T. Mela, V. Saligrama, and I. C. Paschalidis (2015): "Prediction of hospitalization due to heart diseases by supervised learning methods," *International journal of medical informatics*, 84, 189–197.

Deng, L., M. L. Seltzer, D. Yu, A. Acero, A.-r. Mohamed, and G. Hinton (2010): "Binary coding of speech spectrograms using a deep auto-encoder," in *Eleventh annual conference of the international speech communication association*.

Jaafari, J., S. Douzi, K. Douzi, and B. Hssina (2022): "The impact of ensemble learning on surgical tools classification during laparoscopic cholecystectomy," *Journal of Big Data*, 9, 49.

Jensen, P. B., L. J. Jensen, and S. Brunak (2012): "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, 13, 395–405.

Jiang, S., K.-S. Chin, L. Wang, G. Qu, and K. L. Tsui (2017): "Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department," *Expert systems with applications*, 82, 216–230.

Johnson, A. E. W., T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark (2016): "MIMIC-III, a freely accessible critical care database," *Scientific data*, 3, 1–9.

Kam, H. J. and H. Y. Kim (2017): "Learning representations for the early detection of sepsis with deep neural networks," *Computers in biology and medicine*, 89, 248–255.

Kam, H. J., J. O. Sung, and R. W. Park (2010): "Prediction of daily patient numbers for a regional emergency medical center using time series analysis," *Healthcare informatics research*, 16, 158.

Kaplan, R. S., M. E. Porter, et al. (2011a): "How to solve the cost crisis in health care," *Harv Bus Rev*, 89, 46–52.

——— (2011b): "How to solve the cost crisis in health care," *Harv Bus Rev*, 89, 46–52.

Krishna, C. L. and P. V. S. Reddy (2019): "An efficient deep neural network multilayer perceptron based classifier in healthcare system," in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, IEEE, 1–6.

Liang, Z., J. Liu, A. Ou, H. Zhang, Z. Li, and J. X. Huang (2019): "Deep generative learning for automated EHR diagnosis of traditional Chinese medicine," *Computer methods and programs in biomedicine*, 174, 17–23.

Liao, W., H. J. Derijks, A. A. Blencke, E. De Vries, M. Van Seyen, and R. J. Van Marum (2022): "Dual autoencoders modeling of electronic health records for adverse drug event preventability prediction," *Intelligence-Based Medicine*, 6, 100077.

Lin, C., Y. Zhang, J. Ivy, M. Capan, R. Arnold, J. M. Huddleston, and M. Chi (2018): "Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM," in *2018 IEEE international conference on healthcare informatics (ICHI)*, IEEE, 219–228.

Lipton, Z. C., D. C. Kale, C. Elkan, and R. Wetzel (2015): "Learning to diagnose with LSTM recurrent neural networks," *arXiv preprint arXiv:1511.03677*.

Lopez, C., S. Tucker, T. Salameh, and C. Tucker (2018): "An unsupervised machine learning method for discovering patient clusters based on genetic signatures," *Journal of biomedical informatics*, 85, 30–39.

Ma, F., R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao (2017): "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1903–1911.

Ma, F., Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao (2018): "Kame: Knowledge-based attention model for diagnosis prediction in healthcare," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 743–752.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013a): "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*.

——— (2013b): "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*.

Miotto, R., L. Li, B. A. Kidd, and J. T. Dudley (2016): "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records," *Scientific reports*, 6, 1–10.

Morid, M. A., K. Kawamoto, T. Ault, J. Dorius, and S. Abdelrahman (2017): "Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation," in *AMIA annual symposium proceedings*, American Medical Informatics Association, vol. 2017, 1312.

Morton, M., S. Nagpal, R. Sadanandan, and S. Bauhoff (2016): "India's largest hospital insurance program faces challenges in using claims data to measure quality," *Health Affairs*, 35, 1792–1799.

Natarajan, S., M. Kumar, S. K. K. Gadde, and V. Venugopal (2023): "Outbreak prediction of COVID-19 using Recurrent neural network with Gated Recurrent Units," *Materials Today: Proceedings*, 80, 3433–3437.

Pitocco, C. and T. R. Sexton (2018a): "Measuring hospital performance using mortality rates: an alternative to the RAMR," *International Journal of Health Policy and Management*, 7, 308.

——— (2018b): "Measuring hospital performance using mortality rates: an alternative to the RAMR," *International Journal of Health Policy and Management*, 7, 308.

Raghupathi, W. and V. Raghupathi (2014): "Big data analytics in healthcare: promise and potential," *Health information science and systems*, 2, 1–10.

Rahane, W., H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale (2018): "Lung cancer detection using image processing and machine learning healthcare," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, IEEE, 1–5.

Saxe, A. M., J. L. McClelland, and S. Ganguli (2013a): "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," *arXiv preprint arXiv:1312.6120.*

——— (2013b): "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," *arXiv preprint arXiv:1312.6120.*

Shickel, B., P. J. Tighe, A. Bihorac, and P. Rashidi (2017): "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics*, 22, 1589–1604.

Wammes, J. J. G., M. Tanke, W. Jonkers, G. P. Westert, P. Van der Wees, and P. P. Jeurissen (2017): "Characteristics and healthcare utilisation patterns of high-cost beneficiaries in the Netherlands: a cross-sectional claims database study," *BMJ open*, 7, e017775.

Wang, W. W., H. Li, L. Cui, X. Hong, and Z. Yan (2018): "Predicting clinical visits using recurrent neural networks and demographic information," in *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*, IEEE, 353–358.

Wiens, J. and E. S. Shenoy (2018): "Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology," *Clinical infectious diseases*, 66, 149–153.

Xiao, C., E. Choi, and J. Sun (2018): "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," *Journal of the American Medical Informatics Association*, 25, 1419–1428.

YU, K., Y. LIN, AND J. LAFFERTY (2011): "Learning image representations from the pixel level via hierarchical sparse coding," in *CVPR 2011*, IEEE, 1713–1720.

ZENG, X., Y. FENG, S. MOOSAVINASAB, D. LIN, S. LIN, AND C. LIU (2019): "Multilevel self-attention model and its use on medical risk prediction," in *Pacific Symposium On Biocomputing 2020*, World Scientific, 115–126.

ZHAO, Y., A. S. ASH, R. P. ELLIS, J. Z. AYANIAN, G. C. POPE, B. BOWEN, AND L. WEYUKER (2005): "Predicting pharmacy costs and other medical costs using diagnoses and drug claims," *Medical care*, 43, 34–43.

ZHOU, C., Y. JIA, M. MOTANI, AND J. CHEW (2017): "Learning deep representations from heterogeneous patient data for predictive diagnosis," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 115–123.