The Quality Application of Deep Learning in Clinical Outcome Predictions Using Electronic Health Record Data: A Systematic Review

Khuder Alaboud[1]; Imad Eddine Toubal[2]; Butros M. Dahu[1]; Akram Abo Daken[3]; Ammar Ali Salman[4]; Nouha Alaji[5]; Wael Hamadeh[6]; Ahmad Aburayya[7]

[1]Institute for Data Science and Informatics (IDSI), University of Missouri, Columbia, MO, USA;
[2]Dept of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA;
[3]Emergency Medicine Specialist, Head of Emergency Department, Gargash Hospital, Dubai, UAE;
[4]Specialist Registrar, Rashid Hospital-SICU, Dubai Health Authority, Dubai, UAE;
[5]Specialist Emergency Medicine, Clemansue Hospital, Dubai, UAE;
[6]Specialist Emergency Medicine, Al-Garhoud Private Hospital, Dubai, UAE;
[7]Assistant Professor, College of Business, MBA Department, City University Ajman, Ajman, UAE;

KEYWORDS
Deep learning, Machine learning, Electronic Health Records, Clinical Outcome, Quality Prediction Method

ABSTRACT:

Introduction: Electronic Health Record (EHR) is a significant source of medical data that can be used to develop predictive modelling with therapeutically useful outcomes. Predictive modelling using EHR data has been increasingly utilized in healthcare, achieving outstanding performance and improving healthcare outcomes.

Objectives: The main goal of this review study is to examine different deep learning approaches and techniques used to EHR data processing.

Methods: To find possibly pertinent articles that have used deep learning on EHR data, the PubMed database was searched. Using EHR data, we assessed and summarized deep learning performance in a number of clinical applications that focus on making specific predictions about clinical outcomes, and we compared the outcomes with those of conventional machine learning models.

Results: For this study, a total of 57 papers were chosen. There have been five identified clinical outcome predictions: illness (n=33), intervention (n=6), mortality (n=5), Hospital readmission (n=7), and duration of stay (n=1). The majority of research (39 out of 57) used structured EHR data. RNNs were used as deep learning models the most frequently (LSTM: 17 studies, GRU: 6 research). The analysis shows that deep learning models have excelled when applied to a variety of clinical outcome predictions. While deep learning's application to EHR data has advanced rapidly, it's crucial that these models remain reliable, offering critical insights to assist clinicians in making informed decision.

Conclusions: The findings demonstrate that deep learning can outperform classic machine learning techniques since it has the advantage of utilizing extensive and sophisticated datasets, such as longitudinal data seen in EHR. We think that deep learning will keep expanding because it has been quite successful in enhancing healthcare outcomes utilizing EHR data.

1. Introduction

Electronic Health Records (EHRs) contain health and personal data routinely collected on patients. EHRs consist of heterogeneous structured and unstructured data elements, including demographic information, diagnoses, laboratory results, medication prescriptions, and free-text clinical notes. This information is extensively documented in EHR and provides an overview of a patient's health status. They provide opportunities for data-driven approaches to help understand population health and make better clinical decisions [1]. There is a huge volume of data in EHRs being generated from different sources that require computational power to process. Despite this volume of EHR data, its effective use remains an ongoing challenge to meet the promise of extracting the value of data to support and guide care providers' decision-making processes and bring in evidence-based practice for the care providers [2].

Predictive modeling using EHR data has been increasingly utilized in healthcare, achieving outstanding performance and improving healthcare outcomes. In recent years, deep learning has evolved as a powerful tool in many clinical tasks. As a result of deep learning, the data modeling paradigm has changed from expert-driven feature engineering to data-driven feature construction [3]. There is an increasing body of literature confirming the successful performance of deep learning methods over traditional machine learning and statistical techniques such as logistic regression, support vector machines (SVM), and random forest [4–6]. The growth of interest in deep learning in healthcare has two main reasons. First, deep learning models require less manual feature engineering. Second, deep learning enables large and complex dataset training, such as longitudinal data, with better performance. Despite the recent developments in deep learning models, however, the adaptation of these models in developing clinical decision-support has been very limited [3].

Research has focused on data sources and representation of different deep learning techniques, providing a guideline to apply deep learning using EHR data [3]. Others surveyed the research studies that applied deep learning techniques and frameworks and focused on technical details and general applications such as information extraction and representation learning [7]. In addition, other studies have reviewed deep learning in the field of health informatics for a broad range of applications, including genomics and medical imaging, speech recognition, computer vision, and natural language processing [2,8,9]. There is a lack of comprehensive review of the literature for summarizing and evaluating deep learning performance in clinical applications and outcomes using EHR data. The goal of this paper is to complete a systematic review focusing on evaluating and summarizing deep learning techniques in various clinical applications that target specific clinical outcome predictions using EHR data. We also plan to compare and evaluate the performances of deep learning with traditional machine learning methods only for the studies that applied both in any specific clinical problems.

2. Objectives

This research aims to provide an understanding of the following: (1) Resources: What types of EHR data are available for clinical outcome modelling prediction? What data elements were used for deep learning models to facilitate the extraction of healthcare values from EHR data and derive insights to improve clinical outcomes? (2) Methods: What types of deep learning methods were used? What is the performance of these deep learning methods for specific clinical application tasks? (3) Applications: What are the targeted clinical problems and outcomes? What types of clinical application tasks are addressed? (4) Potential: How could deep learning potentially contribute to future research and healthcare? What is the impact of using deep learning and EHR data to improve overall clinical outcomes?

3. Methods

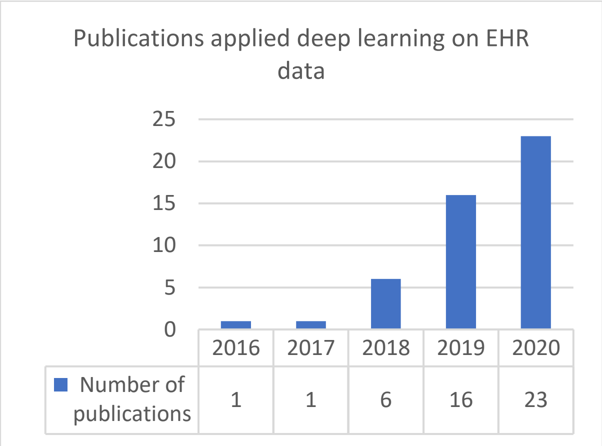
The search was conducted in PubMed database up until December 2020 to extract potentially relevant papers. The search we designed around 2 main concepts, deep learning, and EHRs. They were combined using AND operator. Table 1 presents the complete search strategy results. Figure 1 shows a distribution of the number of publications per year using deep learning models. It shows an overall yearly increase in the volume of publications relating to the application of deep learning on EHR-based data. In addition, the implementation of different deep learning models and the type of data used were analyzed. Finally, the different types of deep learning models and the number of clinical applications were presented in this study. The overall results show an increase in the volume of publications over the years in clinical applications. It started with a limited number of publications and then dramatically increased in recent years.

The inclusion criteria include: (1) studies that used any type of deep learning techniques, (2) studies that used EHR data consisting of unstructured data elements, including demographic information, diagnoses, laboratory results, medication prescriptions, free-text clinical notes (3) studies that applied deep learning techniques on specific clinical applications and targeted specific clinical outcomes. Since many studies reviewed deep learning techniques that have been applied to medical imaging and genomes data [6,7,10], we excluded the studies that applied deep learning to medical imaging and genomic data and focused on deep learning techniques that used EHR data for clinical problems and outcome predictions.

Table 1. Search Criteria in PubMed Database for Eligible Articles

No.	Search term in PubMed	Results
1	“deep learning” [TIAB]	12,808
2	“deep learning” [MeSH]	3,352
3	("deep learning" [TIAB] OR "deep learning"[Mesh])	13,340
4	"EMR"[TIAB]	7,529
5	"EHR"[TIAB]	6,946
6	"electronic health record"[TIAB]	9,568
7	“electronic medical record"[TIAB]	8,492
8	"electronic health records"[TIAB]	10,733
9	"electronic medical records"[TIAB]	10,247
10	"electronic health records"[Mesh]	21,131
11	("EMR"[TIAB] OR "EHR"[TIAB] OR "electronic health record"[TIAB] OR "electronic health records"[TIAB] OR "electronic medical record"[TIAB] OR "electronic medical records"[TIAB] OR "Electronic Health Records"[Mesh])	48,927
3 & 11	("deep learning" [TIAB] OR “deep Learning"[Mesh]) AND ("EMR"[TIAB] OR "EHR"[TIAB] OR "electronic health record"[TIAB] OR "electronic health records"[TIAB] OR "electronic medical record"[TIAB] OR "electronic medical records"[TIAB] OR "Electronic Health Records"[Mesh])	312

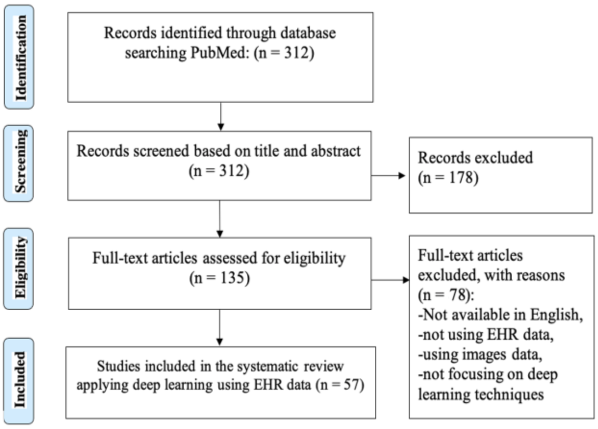
Figure 1. Publication counts over time



Two steps made up the selection procedure. To screen the publications based on the inclusion/exclusion criteria outlined above, we first examined the titles and abstracts of the citations returned by the search query. The second phase involves reading the complete texts of the citations that were chosen in the first. We decided whether the article was eligible for inclusion based on that. The earliest acceptable paper was published in 2016, despite the fact that the search criteria had no restrictions on publication dates due to the field's recent development.

Figure 2. shows the process of identification of articles for our study using the PRISMA guideline [61]. Through the use of the PubMed search engine, a total of 312 articles were located. We initially checked the eligibility of the titles and abstracts of those 312 papers. According to the title and abstract reviews, a total of 178 articles were eliminated because they did not match the eligibility requirements. The remaining 135 publications had their eligibility and data extraction evaluated in full text. A full-text review resulted in the exclusion of 78 publications. A total of 57 chosen publications were ultimately confirmed for data extraction.

Figure 2. PRISMA flow diagram



We distilled the essential details from the chosen papers. Three key factors were assessed for each paper: 1) The kind of deep learning models utilized and how well they perform, 2) The kind of EHR data used, and 3) The kinds of clinical applications and the clinical problems they are intended to address. For certain clinical application tasks, we assessed the deep learning model's performance and contrasted the outcomes with those from more conventional machine learning models. (for instance, Random Forest (RF), Support Vector Machines (SVM), and Logistic Regression (LR)).

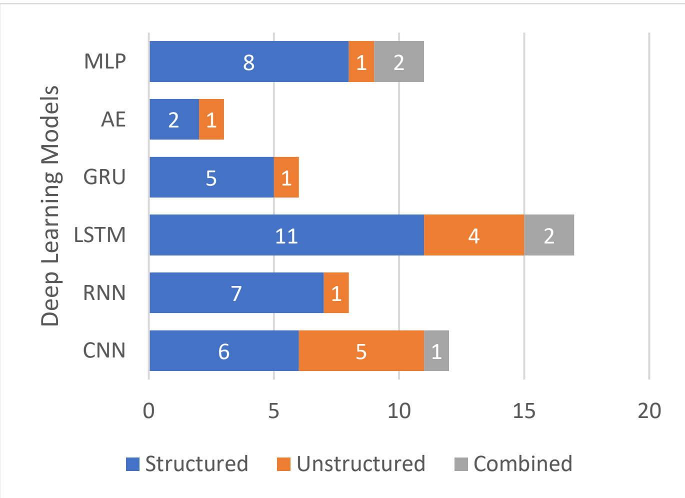
In addition, for each publication, we identified the precise data items used in the studies (such as diagnosis codes, procedure codes, medication codes, labs, vitals, and clinical notes) as well as the patient data types in EHRs (structured or unstructured). The following categories were used to categorize the identified target clinical application tasks: prediction of disease, prediction of intervention, prediction of mortality, prediction of readmission, and prediction of length of stay. Finally, we identified the type of deep learning models used in the papers [(Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long short-term memory (LSTM), Gated recurrent unit (GRU), Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), Autoencoder] and the corresponding performance results using the common performance metrics [(e.g., area under the receiver operating characteristic curve (AUC/AUROC), F1, MSE, Recall, Precision, Accuracy)].

4. Results

EHR Data Types and Elements

Of the 57 research papers, 39 (blue) publications applied deep learning models only to structured data, while 13 papers (orange) incorporated unstructured clinical notes. The remaining 5 papers (gray) combined the use of structured codes and free-form notes. Although merging heterogeneous resources is a promising strategy, we found that the raw data in these experiments was not entirely combined into the same model. Figure 3 depicts the use of various deep learning models and the kinds of data (structured and unstructured) that were employed. The findings demonstrate that deep learning models mostly need structured data. The most widely used deep learning models on unstructured EHR data are RNN and CNN.

Figure 1. Deep Learning models and type of EHR data



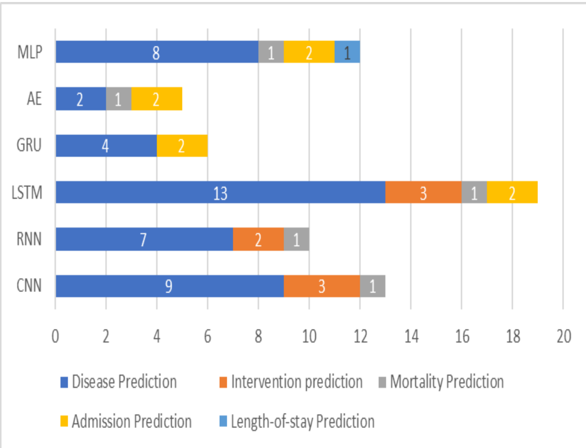
Datasets and Sample Size

EHR data contain protected health information (PHI); therefore, most papers described experiments conducted on non-publicly available datasets, and most papers were based on data from a single healthcare institution. However, most of the papers mentioned that the datasets are available upon reasonable request, and some required permission from the Healthcare Institutional Review Board (IRB). 12% of the papers (n=7) made use of publicly available datasets from the Medical Information Mart Intensive Care III (MIMIC-III), which is in a de-identified format. Some papers used Real-World EHR data from Cerner HealthFacts and Epic systems. Datasets and patient sample size used in the actual study, (n=33) papers had less than 60,000 samples, (n=7) papers had between 300,000 and 1 million samples, and only (n=6 papers) had over 1 million samples. Only (n=4) papers made the code to build and train the model openly available on GitHub.

Deep Learning Models and Architectures:

This systematic review's main goal is to identify the deep learning models that have been used to predict particular clinical outcomes. As shown in Figure 4, we depict the variety of clinical applications employed along with the number of papers that have applied each deep learning model as the foundation architecture. Overall, (n=17) research used LSTM for various therapeutic applications, with CNN coming in second with (n=12) studies. LSTM is effective for datasets that can predict health outcomes over time, which is crucial for EHR analysis. MLP was used in (n=11) trials, and RNN was used in (n=8) investigations. These are the most typical deep learning models used for clinical applications with EHR data. Additionally, only (n=2) investigations have used autoencoder. Our research indicates that RBM and DBN have not applied EHR data to clinical applications. This conclusion makes sense given that local or sequential information is extracted from large-scale data using both CNN and RNN architectures. RNN has also attracted a lot of attention for mimicking the sequential nature of EHR in time series data applications.

Figure 2. Deep learning models and Clinical Outcomes



Clinical Outcome Predictions and Applications:

We picked five categories to test clinical outcome predictions after reading the chosen articles: illness prediction (n=33), intervention prediction (n=6), death prediction (n=5), readmission prediction (n=7), and length-of-stay prediction (n=1). Five clinical applications and outcomes are listed in Table 2 along with a description of the deep learning model type. Our review's findings indicate that most studies have used deep learning to predict two or more clinical outcomes for certain clinical issues or illnesses, leading to a large amount of overlap in the outcome predictions across several studies. Disease prediction studies, which forecast whether a patient will develop a particular ailment, are the most popular clinical outcome prediction studies (n=33). The second most frequent task with studies (n=7) is readmission prediction, followed by intervention prediction with 6 studies. In addition, studies (n=5) used deep learning for predicting death and just one research for predicting duration of stay.

In addition, there were 16 studies that used EHR data to apply deep learning methods to cardiovascular diseases such as (heart failure, cardiac arrest, myocardial infarction, and ischemic stroke). Two studies used deep learning for Alzheimer and dementia, and three studies applied deep learning to predict asthma. There are three studies that used deep learning for kidney diseases, two studies for liver diseases, two studies for type-2 diabetes, and six studies for sepsis. Three studies for cancer (Prostate cancer, lung cancer, and Colorectal cancer).

Table 2. Application of Deep Learning and Clinical Outcome Predictions

Clinical Outcomes	Paper	Author (Year)	Applications	Deep Learning Model
Disease Prediction	11	(Akyea et al., 2020)	Detecting familial hypercholesterolaemia in primary care	MLP
	12	(Rank et al., 2020)	Real-time prediction of acute kidney injury	RNN
	13	(Ljubic, Roychoudhury, et al., 2020)	Alzheimer's disease prediction	LSTM
	14	(Mandair et al., 2020)	Prediction of incident myocardial infarction	MLP
	15	(Ioannou et al., 2020)	Predict Hepatocellular Carcinoma in Patients with	RNN

			Hepatitis-C Cirrhosis.	
	16	(Ljubic, Hai, et al., 2020)	Predicting complications of type-2 diabetes	RNN-GRU, LSTM
	17	(J. Lee et al., 2020)	Predict Post-Induction Hypotension	CNN
	18	(Xiang et al., 2020)	Asthma Exacerbation Prediction and Risk Factor	LSTM
	19	(Bedoya et al., 2020)	Early detection of sepsis	RNN
	20	(J. X. Wang et al., 2020)	Personalized clinical order set recommendations	MLP
	21	(Obeid et al., 2020)	Predicting Intentional Self-Harm in EHR	CNN, LSTM
	22	(Park et al., 2020a)	Detection of Bacteremia in Surgical In-Patients	RNN
	23	(Islam et al., 2020)	Predictive Model for Prostate Cancer	CNN
	24	(Lauritsen et al., 2020a)	Early detection of sepsis	CNN, LSTM
	25	(Miotto et al., 2020)	Acute Low Back Pain Episodes in Primary Care	CNN
	26	(Z. Xu et al., 2020)	Sub-phenotypes of acute kidney injury	LSTM
	27	(Alsaad et al., 2019)	Patient-Specific risk prediction	LSTM
	28	(Jin et al., 2019)	Detection of Hypoglycemic Events	CNN, RNN
	29	(R. Wang et al., 2019)	Predict outcomes after lung cancer radiotherapy	MLP
	30	(Ge et al., 2019)	Predicting post-stroke pneumonia	GRU
	31	(Nguyen et al., 2019)	Predicting the onset of type-2 diabetes	MLP
	32	(Y. H. Wang et al., 2019)	Detection of Colorectal Cancer	CNN
	33	(Tomašev et al., 2019)	Prediction of future acute kidney injury	RNN
	34	(Jang et al., 2020)	Early detection of cardiac arrest in ED	LSTM
	35	(Maragatham & Devi, 2019)	Prediction of Heart Failure in Big Data	LSTM
	36	(Norgeot et al., 2019)	Forecast Outcomes in Patients with Rheumatoid Arthritis	CNN, LSTM, GRU
	37	(Hung et al., 2019)	Prediction of ischemic stroke	MLP
	38	(Li et al., 2019)	Detection of Bleeding Events	CNN, LSTM, AE
	39	(Saqib et al., 2018)	Early Prediction of Sepsis	LSTM
	40	(Y. Lee et al., 2018)	Predicting Cardiac Arrest	MLP
	41	(Rasmy et al., 2018)	Prediction of heart failure onset risk	RNN
	42	(Huang et al., 2018)	Risk Prediction of Acute Coronary Syndrome	MLP, AE
	43	(Choi et al., 2017)	Early detection of heart failure onset	RNN-GRU
Intervention prediction	15	(Ioannou et al., 2020)	Predict Hepatocellular Carcinoma in Patients with Hepatitis-C Cirrhosis.	RNN

	18	(Xiang et al., 2020)	Asthma Exacerbation Prediction and Risk Factor	LSTM
	44	(Park et al., 2020b)	Detection of Bacteremia in Surgical In-Patients	RNN
	45	(W. Chen et al., 2020)	Surgical Site Infection	CNN
	46	(Lauritsen et al., 2020b)	Early detection of sepsis	CNN, LSTM
	47	(Dandala et al., 2020)	Drug Safety Surveillance	CNN, RNN
Mortality Prediction	48	(D. Xu et al., 2020)	Improve Management of Cardiovascular Patients	AE
	49	(Maheshwari et al., 2020)	Prediction of mortality in intensive care units	LSTM
	50	(Payrovnaziri et al., 2019)	Predicting the risk of mortality for patients with acute myocardial infarction (AMI)	MLP
	51	(Mayampurath et al., 2019)	Predict in-hospital mortality.	CNN, RNN
	52	(L. Wang et al., 2019)	Mortality Prediction of Dementia for Earlier Palliative Care Interventions.	LSTM
Readmission Prediction	48	(D. Xu et al., 2020)	Improved Management of Cardiovascular Patients	AE
	53	(P. Chen et al., 2020)	Predict readmissions for heart failure patients	MLP, AE
	54	(Danilov et al., 2020)	Predicting Postoperative Hospital Stay in Neurosurgery	GRU
	55	(Ashfaq et al., 2019)	Readmission prediction with Congestive Heart Failure (CHF)	LSTM
	56	(Reddy & Delen, 2018)	Predicting hospital readmission for lupus patients	LSTM
	57	(Golas et al., 2018)	Readmission prediction in patients with heart failure	MLP
	58	(Xiao, Ma, et al., 2018)	Readmission prediction in patients with Congestive Heart Failure (CHF)	RNN-GRU
Length-of-stay Prediction	59	(Alsinglawi et al., 2020)	Predicting Length of Stay for Cardiovascular Hospitalizations in the ICU	MLP

Deep Learning Performance in Clinical Outcome Predictions:

Table 3 evaluates different deep learning models in clinical outcome predictions. Further, it provides a comparative review of some studies that applied deep learning and traditional machine learning techniques. All reviewed articles used EHR data to build their models. Also, the red color in the table indicates the model with the best performance based on the evaluation metrics used in the study. Blank space means not reported. Here are some examples, a study [12] developed a deep-learning-based algorithm that can predict postoperative Acute Kidney injury (AKI) prior to the onset of symptoms and complications. The study applied RNN for the real-time prediction of AKI after cardiothoracic surgery and compared the performance of deep learning models against that of experienced clinicians. The RNN significantly outperformed clinicians in predicting AKI after cardiothoracic surgery with AUC = 0.901. It could potentially be integrated into hospitals' electronic health records for real-time patient monitoring, and it might assist in early AKI detection and perioperative care treatment modification. Another study [51] implemented CNN model for predicting in-hospital mortality while enabling clinician interpretability using EHR data. CNN was able to predict in-hospital mortality with an AUC of 0.91, which is significantly better than the modified Early Warning Score with an AUC of 0.76, and better than the Sequential Organ Failure Assessment score of 0.89.

In addition, other studies compared deep learning performance with traditional machine learning to see how well deep learning performs in clinical outcome predictions. One study [15] implemented deep learning RNN models using EHR to predict the risk of developing hepatocellular carcinoma (HCC). With an AUC of 0.759 vs. 0.689, RNN models outperformed traditional linear regression models, indicating that they might be used to identify patients with HCV-related cirrhosis who were at a high risk of developing HCC. A study [16] developed deep learning models to predict complications of diabetes mellitus, which could help with more targeted measures that would prevent or slow down the development of diabetes mellites. RNN-LSTM, and RNN-GRU deep learning methods were designed and compared with RF and MLP traditional models. In this study, three settings and the number of hospitalizations between the diagnosis of diabetes and the health issues were studied along with the prediction accuracy of a few specific health difficulties. The RNN-GRU model, followed by the RNN-LSTM model, had the best results. The RNN-GRU model's prediction accuracy was between 73% and 83% for myocardial infarction and chronic ischemic heart disease, compared to 66% to 76% for classical models.

One of the studies [43] used recurrent neural network models for early detection of heart failure onset. Using a 12-month observation window, AUC for the RNN-GRU model was 0.777, compared to AUCs for logistic regression (0.747), (MLP) with 1 hidden layer (0.765), SVM (0.743), (KNN) (0.730). When using an 18-month observation window, the AUC for the RNN model increased to 0.883 and was significantly higher than the 0.834 AUC for the best of the baseline methods (MLP). GRU models also significantly outperformed traditional machine learning models that rely on aggregate features. Another study [59] aims at predicting the length of stay for cardiovascular hospitalizations in the Intensive Care Unit. The results of the three models (Gradient Boosting, Stacking Regression (GB), Random Forest) showed relatively close R-Squared (R2) and Mean Absolute Error (MAE) (R2:0.81, 0.81, and 0.80) and (MAE: 2.00, 1.92, and 1.98) respectively. The Gradient Boosting Regressor (GBR) outperformed all other models in this study with (R2 0.8±0.08). Noting that the deep learning model (Deep Learning Regression) did not report better results than regression models with R2 and MAE (0.77±0.06 2.30±0.18). The authors recommended examining DNN on a big EMR dataset with a large volume of heart failure hospitalizations and comparing it with regression models in a future study. Deep Learning underperformed due to a lack of enough data to train a high number of parameters, since DNN performs well in larger data size and high dimensional data due to its automatic learning feature benefit.

5. Discussion

Deep Learning Performance in Clinical Applications

EHR maintains patient information that can be used to enhance healthcare and offer individualized care. Healthcare providers find it tough and challenging to manage the enormous volumes of data in EHRs. The majority of EHR analysis approaches up until recently relied on traditional machine learning algorithms like LR, SVM, and random forest. In the past few years, deep learning approaches have shown considerable success in various clinical outcome predictions. Because deep learning has been so successful in so many other areas, there have been a lot more articles that use deep learning to analyze EHR data for therapeutic purposes.

Deep learning applications to simulate clinical data based on EHR data have significantly advanced recently. Table 3 presents a thorough summary of these accomplishments. According to our review, deep learning models have exhibited exceptional performance when applied to a variety of clinical outcome predictions, including those for disease, interventions, mortality, admission, and length of stay. 35 research have used RNNs as a deep learning model, which is the bulk of the studies.

Machine Learning and Deep Learning

Both machine learning and deep learning are data-driven (i.e., learn from data) to perform prediction (classification or regression.). While classical machine learning often requires domain knowledge and feature engineering, deep learning can automatically learn features from the data relevant to the task at hand. Since hand-crafting features require expertise, machine learning methods are very dependent on the quality of the features. This gives deep learning the advantage of being more autonomous. Deeper neural networks often produce higher-level features and perform better given enough data to train them.

Our findings from this review show that, when applied to EHR data, deep learning algorithms greatly outperformed conventional machine learning. The most common clinical issue was cardiovascular disease prediction. These findings demonstrate how applying deep learning to EHRs can result in patient representations that offer superior clinical predictions, as well as a deep learning framework for workflow and aid in clinical decision-making. Some studies found that deep learning performed worse than machine learning since there wasn't enough data to train models with a lot of complicated parameters. Due to the benefit of its automated learning characteristic, deep learning typically works well in larger data sizes and high dimensional data.

The optimal predictive model depends on the problem domain. There is no model that consistently outperforms other models for all problems. Consequently, it is often desirable to combine many diverse models to reduce the variability of models. This is known as ensemble learning which uses multiple learning algorithms to obtain better predictive performance. In addition, predictive model performance is not homogeneous across the whole patient population and understanding how it differs across different clinical subpopulations might assist in guiding decisions on the deployments of predictive models in the future.

Despite rapid advancement in deep learning, developing robust clinically applicable predictive models from routinely collected EHR data remains challenging. Clinically applicable models need to be reliable to deliver personalized insights on preventable conditions early enough to enable clinical intervention while also providing enough information to support decision-making. In addition, models need to be able to incorporate all the relevant medical data that is currently accessible and be evaluated on large representative datasets. The model evaluation needs to be performed with good sensitivity levels and achieved while maintaining clinically applicable levels of accuracy. These challenges can be a barrier to implementation. Deep learning has been used in numerous research to perform predictive modeling using EHR data. However, despite the success of deep learning on predictive tasks, it still lacks interpretability. Since deep neural networks are complex universal approximators, it is quite difficult to explain the architecture of the network and how it makes predictions. This can be a challenge when making informed decisions in a clinical application.

Conclusion

The outcomes reveal that deep learning techniques can outperform typical machine learning techniques thanks to the benefit of utilizing longitudinal EHR data. We predict that the expansion will continue, building on the current encouraging trends of utilizing deep learning methods with EHR data to enhance clinical outcomes. Deep learning on EHR data for targeted application tasks will accelerate the resolution of difficult clinical issues. The time has come to develop more sophisticated healthcare systems, where the optimum course of action is computationally learnt from EHR data by deep learning algorithms.

References

1. Birkhead, G. S., Klompas, M. & Shah, N. R. Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health. *Annual Review of Public Health* **36**, 345–359 (2015).
2. Xiao, C., Choi, E. & Sun, J. Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association* vol. 25 1419–1428 (2018).
3. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* vol. 521 436–444 (2015).
4. Carin, L. & Pencina, M. J. On deep learning for medical image analysis. *JAMA - Journal of the American Medical Association* vol. 320 1192–1193 (2018).
5. Chan, H. P., Samala, R. K., Hadjiiski, L. M. & Zhou, C. Deep Learning in Medical Image Analysis. in *Advances in Experimental Medicine and Biology* vol. 1213 3–21 (Springer, 2020).
6. Shen, D., Wu, G. & Suk, H.-I. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering* **19**, 221–248 (2017).
7. Ravi, D. *et al.* Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics* **21**, 4–21 (2017).
8. Ayala Solares, J. R. *et al.* Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics* vol. 101 103337 (2020).
9. Si, Y. *et al.* Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. *arXiv* 103671 (2020) doi:10.1016/j.jbi.2020.103671.

10. Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Molecular Systems Biology* **12**, 878 (2016).
11. Akyea, R. K., Qureshi, N., Kai, J. & Weng, S. F. Performance and clinical utility of supervised machine-learning approaches in detecting familial hypercholesterolaemia in primary care. *NPJ digital medicine* **3**, 142 (2020).
12. Rank, N. *et al.* Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *NPJ digital medicine* **3**, 139 (2020).
13. Ljubic, B. *et al.* Influence of medical domain knowledge on deep learning for Alzheimer's disease prediction. *Computer Methods and Programs in Biomedicine* **197**, (2020).
14. Mandair, D., Tiwari, P., Simon, S., Colborn, K. L. & Rosenberg, M. A. Prediction of incident myocardial infarction using machine learning applied to harmonized electronic health record data. *BMC Medical Informatics and Decision Making* **20**, (2020).
15. Ioannou, G. N. *et al.* Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients With Hepatitis C Cirrhosis. *JAMA network open* **3**, e2015626 (2020).
16. Ljubic, B. *et al.* Predicting complications of diabetes mellitus using advanced machine learning algorithms. *Journal of the American Medical Informatics Association* **27**, 1343–1351 (2020).
17. Lee, J. *et al.* Comparative analysis on machine learning and deep learning to predict post-induction hypotension. *Sensors (Switzerland)* **20**, 1–21 (2020).
18. Xiang, Y. *et al.* Asthma exacerbation prediction and risk factor analysis based on a time-sensitive, attentive neural network: Retrospective cohort study. *Journal of Medical Internet Research* **22**, (2020).
19. Bedoya, A. D. *et al.* Machine learning for early detection of sepsis: an internal and temporal validation study. *JAMIA open* **3**, 252–260 (2020).
20. Wang, J. X., Sullivan, D. K., Wells, A. C. & Chen, J. H. ClinicNet: machine learning for personalized clinical order set recommendations. *JAMIA open* **3**, 216–224 (2020).
21. Obeid, J. S. *et al.* Identifying and Predicting Intentional Self-Harm in Electronic Health Record Clinical Notes: Deep Learning Approach. *JMIR medical informatics* **8**, e17784 (2020).
22. Park, H. J., Jung, D. Y., Ji, W. & Choi, C. M. Detection of bacteremia in surgical in-patients using recurrent neural network based on time series records: Development and validation study. *Journal of Medical Internet Research* **22**, (2020).
23. Islam, M. *et al.* Deep learning approach for the development of a novel predictive model for prostate cancer. in *Studies in Health Technology and Informatics* vol. 270 1241–1242 (IOS Press, 2020).
24. Lauritsen, S. M. *et al.* Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artificial Intelligence in Medicine* **104**, (2020).
25. Miotto, R. *et al.* Identifying Acute Low Back Pain Episodes in Primary Care Practice From Clinical Notes: Observational Study. *JMIR medical informatics* **8**, e16878 (2020).
26. Xu, Z. *et al.* Identifying sub-phenotypes of acute kidney injury using structured and unstructured electronic health record data with memory networks. *Journal of Biomedical Informatics* **102**, (2020).

27. Alsaad, R., Malluhi, Q., Janahi, I. & Boughorbel, S. Interpreting patient-Specific risk prediction using contextual decomposition of BiLSTMs: Application to children with asthma. *BMC Medical Informatics and Decision Making* **19**, (2019).
28. Jin, Y., Li, F., Vimalananda, V. G. & Yu, H. Automatic Detection of Hypoglycemic Events From the Electronic Health Record Notes of Diabetes Patients: Empirical Study. *JMIR medical informatics* **7**, e14340 (2019).
29. Wang, R. *et al.* Multi-objective ensemble deep learning using electronic health records to predict outcomes after lung cancer radiotherapy. *Physics in Medicine and Biology* **64**, (2019).
30. Ge, Y. *et al.* Predicting post-stroke pneumonia using deep neural network approaches. *International Journal of Medical Informatics* **132**, (2019).
31. Nguyen, B. P. *et al.* Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Computer Methods and Programs in Biomedicine* **182**, (2019).
32. Wang, Y. H., Nguyen, P. A., Mohaimenul Islam, M., Li, Y. C. & Yang, H. C. Development of deep learning algorithm for detection of colorectal cancer in EHR data. in *Studies in Health Technology and Informatics* vol. 264 438–441 (IOS Press, 2019).
33. Tomašev, N. *et al.* A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).
34. Jang, D. H. *et al.* Developing neural network models for early detection of cardiac arrest in emergency department. *American Journal of Emergency Medicine* **38**, 43–49 (2020).
35. Maragatham, G. & Devi, S. LSTM Model for Prediction of Heart Failure in Big Data. *Journal of Medical Systems* **43**, (2019).
36. Norgeot, B. *et al.* Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA network open* **2**, e190606 (2019).
37. Hung, C. Y., Lin, C. H., Lan, T. H., Peng, G. S. & Lee, C. C. Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database. *PLoS ONE* **14**, (2019).
38. Li, R. *et al.* Detection of Bleeding Events in Electronic Health Record Notes Using Convolutional Neural Network Models Enhanced With Recurrent Neural Network Autoencoders: Deep Learning Approach. *JMIR Medical Informatics* **7**, e10788 (2019).
39. Saqib, M., Sha, Y. & Wang, M. D. Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* vols. 2018-July 4038–4041 (Institute of Electrical and Electronics Engineers Inc., 2018).
40. Lee, Y. *et al.* Deep Learning in the Medical Domain: Predicting Cardiac Arrest Using Deep Learning. *Acute and critical care* **33**, 117–120 (2018).
41. Rasmy, L. *et al.* A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *Journal of Biomedical Informatics* **84**, 11–16 (2018).
42. Huang, Z., Dong, W., Duan, H. & Liu, J. A Regularized Deep Learning Approach for

Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records. *IEEE Transactions on Biomedical Engineering* **65**, 956–968 (2018).

43. Choi, E., Schuetz, A., Stewart, W. F. & Sun, J. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association* **24**, 361–370 (2017).
44. Park, H. J., Jung, D. Y., Ji, W. & Choi, C. M. Detection of bacteremia in surgical in-patients using recurrent neural network based on time series records: Development and validation study. *Journal of Medical Internet Research* **22**, (2020).
45. Chen, W. *et al.* Artificial Intelligence-Based Multimodal Risk Assessment Model for Surgical Site Infection (AMRAMS): Development and Validation Study. *JMIR medical informatics* **8**, e18186 (2020).
46. Lauritsen, S. M. *et al.* Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artificial Intelligence in Medicine* **104**, (2020).
47. Dandala, B., Joopudi, V., Tsou, C.-H., Liang, J. J. & Suryanarayanan, P. Extraction of Information Related to Drug Safety Surveillance From Electronic Health Record Notes: Joint Modeling of Entities and Relations Using Knowledge-Aware Neural Attentive Models. *JMIR medical informatics* **8**, e18417 (2020).
48. Xu, D., Sheng, J. Q., Hu, P. J.-H., Huang, T.-S. & Hsu, C.-C. A Deep Learning Based Unsupervised Method to Impute Missing Values in Patient Records for Improved Management of Cardiovascular Patients. *IEEE journal of biomedical and health informatics* **PP**, 1–1 (2020).
49. Maheshwari, S., Agarwal, A., Shukla, A. & Tiwari, R. A comprehensive evaluation for the prediction of mortality in intensive care units with LSTM networks: Patients with cardiovascular disease. *Biomedizinische Technik* **65**, 435–446 (2020).
50. Payrovnaziri, S. N., Barrett, L. A., Bis, D., Bian, J. & He, Z. Enhancing prediction models for one-year mortality in patients with acute myocardial infarction and post myocardial infarction syndrome. in *Studies in Health Technology and Informatics* vol. 264 273–277 (IOS Press, 2019).
51. Mayampurath, A., Sanchez-Pinto, L. N., Carey, K. A., Venable, L. R. & Churpek, M. Combining patient visual timelines with deep learning to predict mortality. *PLoS ONE* **14**, (2019).
52. Wang, L. *et al.* Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients with Dementia for Earlier Palliative Care Interventions. *JAMA Network Open* (2019) doi:10.1001/jamanetworkopen.2019.6972.
53. Chen, P. *et al.* Interpretable clinical prediction via attention-based neural network. *BMC Medical Informatics and Decision Making* **20**, (2020).
54. Danilov, G. *et al.* Predicting postoperative hospital stay in neurosurgery with recurrent neural networks based on operative reports. in *Studies in Health Technology and Informatics* vol. 270 382–386 (IOS Press, 2020).
55. Ashfaq, A., Sant'Anna, A., Lingman, M. & Nowaczyk, S. Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics* **97**, (2019).
56. Reddy, B. K. & Delen, D. Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. *Computers in Biology and Medicine* **101**, 199–209 (2018).

57. Golas, S. B. *et al.* A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. *BMC Medical Informatics and Decision Making* **18**, (2018).
58. Xiao, C., Ma, T., Dieng, A. B., Blei, D. M. & Wang, F. Readmission prediction via deep contextual embedding of clinical concepts. *PLoS ONE* **13**, (2018).
59. Alsinglawi, B. *et al.* Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* **2020**, 5442–5445 (2020).
60. Nori, V. S., Hane, C. A., Sun, Y., Crown, W. H. & Bleicher, P. A. Deep neural network models for identifying incident dementia using claims and EHR datasets. *PLoS ONE* **15**, (2020).
61. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. Journal of clinical epidemiology, 134, 103-112.