ScriptChain Health White Paper

Background:

Healthcare presents numerous opportunities for improved preventive care. Hospitals have accumulated large patient databases that could be mined productively if only the data were accessible to clinicians in a familiar and easily interpretable format. Compared to the last 20-30 years, with the advancements in technology today, we can catch illnesses at an earlier stage. As a result, we can shift the care of patients to preventive based medicine rather than reactive management improving patient outcomes. With the emergence of telehealth, big data, artificial intelligence (AI), and federated learning, we are able to create efficiencies that were thought never to be possible. These efficiencies can occur in today's healthcare environment as long as access and adoption occurs. If healthcare executives were able to match the rate of innovation present in the technological space, the cost of materials, care, and readmission rates would decrease over the long-term increasing profitability, health equity, and improving patient outcomes at the same time.

In 2018 alone, over 18.5 million people died of chronic heart disease globally and that number has been growing annually. According to the U.S. Centers for Disease Control, heart disease alone will cost the United States healthcare system around $219 billion dollars annually. The cost of healthcare has been increasing steadily, but there are options on increasing efficiency and decreasing the costs using AI. ScriptChain Health offers providers a preventive care solution that can identify potential high-risk patients for readmission and cardiovascular disease by using artificial intelligence for early prevention. Preventive care is better than reactive based monitoring care. With the help of AI, we can process data lakes from decentralized data sources faster than a human being, thus resulting in time savings. With decentralized data sources including mobile devices, we can help assure precision medicine using the patients direct medical charts, lab tests, clinical notes and securely without the data having to leave the original location of where the data is stored.

Readmission has also been a very large problem in the United States, costing the U.S. healthcare system around $500 million dollars a week in losses and penalties. With AI becoming more popular in the healthcare space, knowing the type of technology that can be adopted into the work environment and used by physicians as a part of their daily responsibilities is imperative to get the healthcare system up to speed on improving operational cost efficiencies and overall quality of care.

Our business model is B2B operating as a SAAS company whose customers are primary care physicians and cardiologists. Our platform sits on top of electronic health record systems, and we utilize the patient medical data that sits within the database and request the critical endpoints that get the user through our algorithms and generate decision support for physicians when seeing a patient all directly within the EHR system so we fit the physician workflow. We follow ACC guidelines to build efficacy with our users and improve patient safety.

ScriptChain Intelligence System:

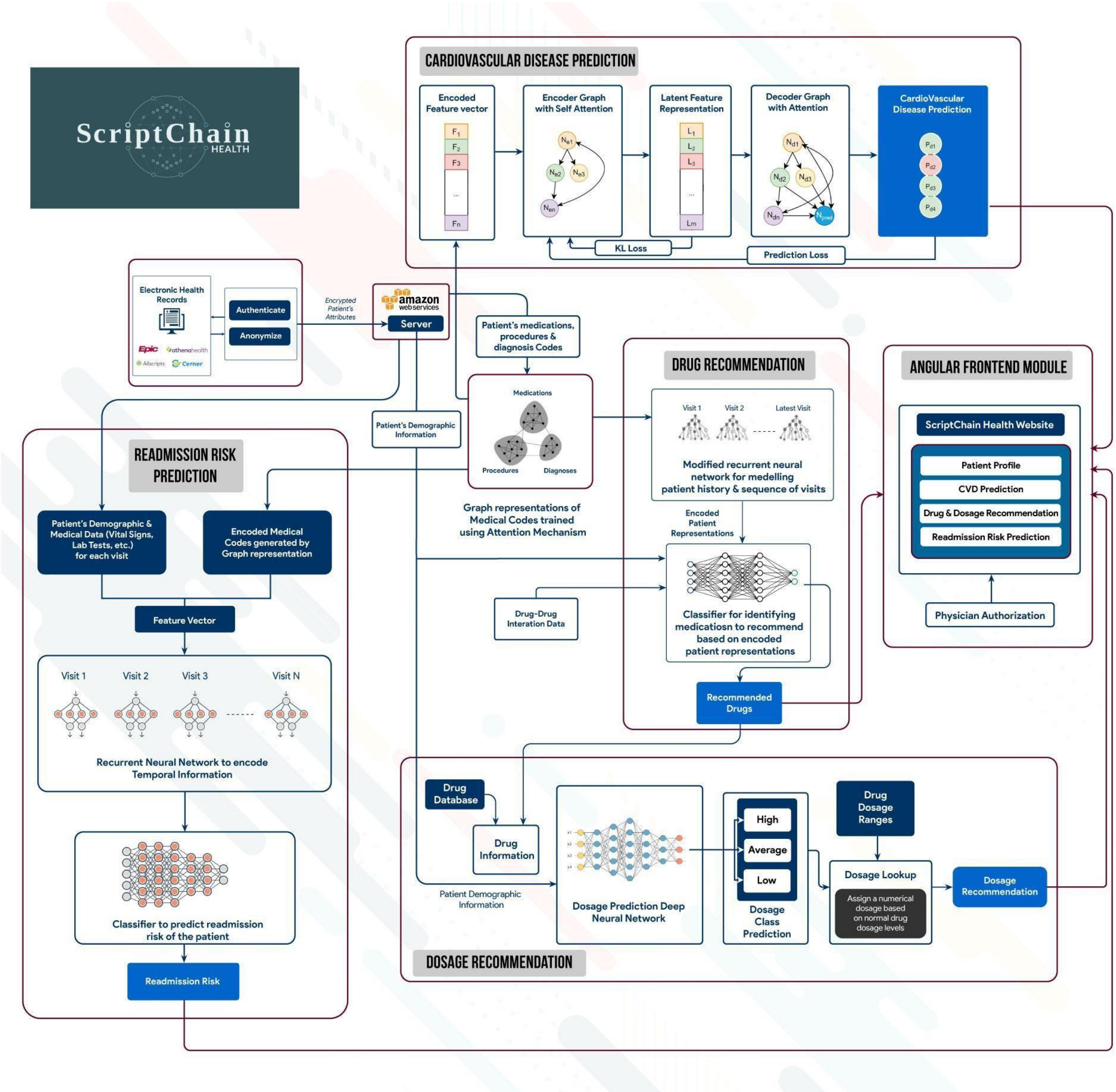

Figure 1: System Diagram of ScriptChain Intelligence System: Readmission Prediction (Left), Cardiovascular Disease Predictions (top), Drug Recommendation (middle), Dosage Recommendation (bottom)

Cardiovascular Disease Preventive Model:

Motivation:

Cardiovascular Disease (CVD) is the number one leading cause of death in the United States [14], associated with 23% percent of deaths in the United States in 2019. The costs as a direct result of CVD are estimated to be $103 million annually in the United

States alone [15]. Fortunately, CVD is not purely random, as it strongly correlates with factors such as genetics, weight, diet, exercise, blood pressure, etc. Many of these factors correlated with CVD are quantitatively documented in patient medical records. While providers are generally quite adept at noticing the early indications of these diseases, the schedule of most physicals does not allow them to systematically and thoroughly analyze each patient's records in many cases. This presents the opportunity to build an improved toolset that allows physicians to early detect CVDs and make the changes needed to avoid a diagnosis of these chronic Diseases in the Future.

Sciptchain Health Proposes a CVD prediction model to assist physicians to detect early signs of cardiovascular disease using AI integrated software. Our web application will give providers, physicians, and patients a probability that they will be diagnosed with several different cardiovascular diseases in the near future. Our software allows patients and physicians to make the medical and lifestyle changes needed to prevent these diseases.

Approach:

Using patient-level electronic health record (EHR) data, ScriptChain Health is building predictive models to detect early indications of heart disease to assist physicians in early detection and prevention of heart disease. Through research and experimental measures, ScriptChain Health is developing a state-of-the art machine learning framework that stores medical data into a graph network. The graph network retains the temporal and hierarchical relationships inherent in medical data that are often lost or unutilized in most deep learning approaches. Graph based deep learning on EHR data has been shown to outperform other ML approaches in diagnosis and prediction [3,4,5]. ScriptChain Health's current model is designed to predict the following 4 cardiovascular diseases: Hypertension, Atrial Fibrillation, Congestive Heart Failure and Coronary Artery Disease. The CVD prevention model returns the probability that a patient will be diagnosed with each of the 4 CVDs within 90 days.

The current model uses tabular data combined from medical record sources. Future work is aimed at incorporating physician notes along with tabular data in an ensemble learning method as an all-encompassing solution. **Table 1** below details the features data currently used in the GNN based model.

Record collection	Features
Admissions	Patient ID, Admission ID, Admission date/time, Marital Status, Ethnicity
Patient data	Patient ID, Gender, Patient Age Group

Laboratory Tests	Total Cholesterol (TC), HDL, LDL, Troponin - T, Hemoglobin
Diagnoses	Patient ID, Admission ID, Diagnosis ICD code, Diagnosis Severity
Prescriptions	Patient ID, Admission ID, Start time, Stop time, Dosage value, Dosage Unit , Generic Sequence Number (GSN)

Table 1: Dataset Features used in CVD Model

Drug Recommendation System

Motivation:

In the United States, on average, patients younger than 65 purchased a mean of 10.8 prescription drugs and those 65 or older purchased a mean of 26.5 prescription drugs annually[1]. This level of prescribing opens the door for errors to occur. Medication errors occur in about 3.7% to 16.6% of hospital admissions, leading to an estimated 44,000 deaths per year. ScriptChain Health assists physicians in making better decisions by building models that take into account individualized long term patient histories, complex drug interactions, and demographic information among other factors. Therefore, by leveraging big data associated with EHR medical records along with advances in neural networks and machine learning technologies, ScriptChain Health is attempting to solve this issue by assisting physicians in predictive healthcare and support decision making.

An important aspect of improving the quality of drug therapy is to minimize negative drug-drug interactions. Given that adverse drug events were the leading cause (19.3%) of all non-operative adverse events, reducing these negative interactions by even 10% can have a significant impact on health outcomes for patients. However, the increasing complexity of polypharmacy medication can make it challenging for providers to always be able to predict and avoid such interactions individually, whereas leveraging big data and machine learning can scale effectively to model and avoid increasingly complex interactions and assist providers by providing better recommendations to minimize these adverse drug reactions.

Economic Impact:

Medication errors comprise a huge negative financial impact on both patients as well as providers in the current state. It is estimated that the annual cost of ADEs is about 19.5 billion in the US alone. Incorrect medication recommendations, cases of prescription errors where physicians may fail to take into account allergy / DDI interactions etc. can be especially financially costly since it can open providers and healthcare systems up to malpractice claims as well. In a 2017 study estimated the per-patient medications review savings cost as $218 per patient for likely ADEs. Providing an AI powered recommendation system to provide this review and analysis to physicians can help reduce negative outcomes and capture these cost savings. It also allows providers to claim improved quality of care outcomes. Given the Pay for Quality incentives mandated at the federal level in the US and implemented by insurance providers, having improved quality of care outcomes can help providers charge higher rate for services as well as potentially preventing claims. Therefore, having strong medication recommendation systems to assist physicians can have a significant positive financial impact for providers.

Data:

1. We use long term EHR data
2. Patient demographic information
3. Medical databases such as DrugBank / TwoSides.

Approach:

We combine different state-of-the-art machine learning techniques for leveraging available EHR data efficiently and effectively. Our model uses multi-headed attention mechanisms to learn embeddings for medical codes along with using graph-based neural networks to capture complex structural relationships between different conditions, procedures, and medications. This is then connected to a modified recurrent neural network architecture for modeling patient histories so we can ensure both efficiency and performance from our model. Moreover, we integrate the proposed graph-based neural networks into an advancing federated learning framework, via which we train the networks on end-users in a decentralized and distributed manner to guarantee the security of user information. We establish asynchronous communication between our server and end-users and distribute the initial networks to their devices and get the model trained locally. Then We harvest the well-trained model parameters rather than user data to avoid data leaks. After collecting locally-trained model parameters from a large group of users, federated average and federated proximity technologies are applied to merge local networks into a decentralized model with strong capability and robust performance.

We also model Drug-Drug interactions by using both NLP analysis of medical literature [3] as well as gold label databases in the workflow, to ensure we recommend drugs for minimizing side effects for patients.

<u>Dynamically Recommending Drug Dosages with Machine Learning</u>

Why This Matters:

Medication misdosing contributes about 7,000 to 9,000 deaths each year in the U.S. that is a direct result of medication errors. Medication misdosing is often a result of poor attention to detail or errors in communication. A quick analysis of patients' information could remedy these erroneous prescriptions by giving physicians an extra pair of eyes to look over the task at hand. The solution eliminates the complications of patient medical information through intelligent software that can read, assess, and make recommendations on medication dosing.

This system seeks to provide that solution. Trained on tens of thousands of past patient medication dosages, this system has learned to recommend the appropriate drug dosing for hundreds of cardiovascular-related medications. By analyzing several factors within the Electronic Medical Record (EMR), the system is able to provide a dosing recommendation that is specific to each patient. Demographic information about the patient is leveraged to make a statistical analysis for recommended dosage amounts.

The patient's medical history is used to gain further confidence in a dosage amount that is custom tailored for the case.

Economic Impact

The estimated cost of treatment for medication-related issues is valued at over 20 billion dollars annually[10]. An intangible but invaluable cost of medication errors is the public's loss of trust in the healthcare system. If we can prevent just 10 percent of medication misdosing errors, the mistreatment of thousands of patients would be avoided and 4 billion dollars could be saved annually. By reallocating additional critical resources and increasing hospital stays, medication errors cost hospitals an average of $3,000 per patient and increase length of stay by 3.1 days[16]. Additional costs can accrue from medical malpractice lawsuits, which can close at hundreds of thousands of dollars, depending on the severity[17].

Data Sources:

1. Offsides
2. Drugbank
3. Patient level medical data from Beth Israel Deaconess

Variables and Motivation:

We incorporate several factors to determine a specifically tailored dosage plan for each case. Starting with demographic information, age, weight, and height are the most significant factors when assigning drug dosage amounts. We also utilize ethnicity as a predictor as research shows that genomics play a role in the body's reaction to a drug. Patient medical history is another factor considered. This data can notify the system of the patient's past reaction to treatment in order to make an informed decision. Drug information from the patient's Electronic Medicine Administration Record (eMAR) is included to assess potential secondary reactions with the body and known interactions with other drugs. We aim to effectively treat patients with no overdosing and minimal negative reactions to treatment.

Algorithms:

With a patient's EHR data, along with demographic information, a specific drug dosage can be recommended to fit each patient's needs. Leveraging deep features learned by the drug recommendation network, a graph-based neural network with multi-headed attention mechanisms to learn embeddings for medical codes, we can also retrieve detailed drug dosage for each drug the network recommends. The algorithm is also integrated into a federated learning framework, in which the model is trained in a decentralizing manner to avoid collecting end-users information into a server, reduce biases and increase precision based health.

Readmission Risk Prediction

A major focus of current hospital quality efforts is reducing unplanned readmissions, in particular, for cardiovascular diseases patients. The estimated annual cost of readmissions for Medicare is estimated to $26 billion annually; $17 billion of it can be considered avoidable [2]. In 2015, one in five elderly patients was readmitted to the hospital within 30 days of being discharged, and one in four patients with some of the common diagnoses including Congestive Heart Failure are readmitted within 30 days. Of the readmitted, Medicare and Medicaid patients have a higher percentage of readmission rates, while the uninsured have the lowest.

Hospital Readmission Reduction Program:

Health reforms in the U.S. identified hospital readmissions as a key area for improving care coordination and achieving potential healthcare savings, and in 2010, the Hospital Readmissions Reduction Program (HRRP) was enacted. The Centers for Medicare and Medicaid Services (CMS) started the implementation of the HRRP in 2012, financially penalizing hospitals with higher-than-expected risk-standardized 30-day readmission rates for six conditions or procedures, [3] including Acute myocardial infarction (AMI), and HF. Some of these conditions and procedures are among the most expensive conditions treated in US hospitals, including HF, and AMI, with HF having the highest readmission rate for patients aged 65 and above, with its prevalence expected to rise to over 8 million people by the year 2030 due to the aging population [4].

To ensure that a hospital is not unfairly penalized, excess readmissions are defined by measuring a hospital's readmission rates, adjusted for age, sex, and co-existing conditions, which are then compared to the national averages. The penalty is a percentage of total Medicare payments to the hospital; the maximum penalty has been set at 1% for 2013, 2% for 2014, and 3% for 2015. The penalties assessed to hospitals are CMS' savings [5].

Hospital Readmission rates are an important indicator of quality of care, as well as a significant factor of patients' emotional wellbeing. One study has shown that patients often believe that their readmissions were preventable, and they linked them with issues such as discharge timing, follow-up, and skilled services [6]. In order to decrease readmission rates, hospitals must focus, including other factors, on using data analysis to determine patients most at risk of readmissions, due to social determinants of health or their specific health conditions.

Multiple studies have shown that targeting patients at high-risk of readmission is significant to reduce overall readmission rates. Certain patient populations are at higher risk for hospital readmission, and research in the Journal for Healthcare Quality [7] found socioeconomic factors, such as race, income, and payer status, are correlated with rehospitalization rates, and patients with certain conditions, including heart failure, chronic obstructive pulmonary disease, and renal failure, also have higher rates of readmission.

Economic Impact:

As mentioned earlier, the annual cost incurred to Medicare due to readmissions is estimated around $26 billion annually, out of which $17 billion is attributed to avoidable readmissions. can be considered avoidable. To quantify the impact our proposed system can have financially, on the healthcare providers that provide services to patients with medicare and Medicaid, a 10% reduction in the avoidable costs due to readmission would result in saving $1.7 billion annually.

Reducing Readmission Rates:

ScriptChain provides a data-driven approach to generate 30-day readmission risk prediction for CVD patients. Our AI-based system provides hospitals with readmission risk of patients, enabling them to identify patients most likely to be readmitted. This enables caregivers to deliver focused interventions to vulnerable patients, preventing 30-day readmissions, and optimizing resource allocation. Our predictive model utilizes variables available from EHR systems, including patient's demographic information (age, marital status, race, etc.), as well as medications, procedures, diagnoses, insurance provider, residential zip code, lab tests, and doctor's notes.

The above listed features were selected after observing patterns in the data. We've trained a graph convolutional network on the medications, procedures, and diagnoses to capture structural correlations for all the patients. The final readmission risk is predicted after taking in the node representations from the graph network, and the demographic & medical features for each of a patients' admission, as our machine learning algorithm also captures temporal dependencies. The block diagram in Figure 1 shows an overview of the readmission risk prediction model as a part of the ScriptChain intelligence system.

Using advanced deep-learning algorithms that can model the spatiotemporal dependencies in the medical data, our approach has identified risk factors that were previously unrecognized. Using a Graph network based on attention-mechanism, the algorithm determines the importance of each feature during the training phase, without human intervention.
Our model is able to correctly predict 70% of the patients that are readmitted within 30 days, and we are working on improving the prediction accuracy. ScriptChain's portal will display the readmission risk as part of the EHR system, providing health-care providers additional insight, and the opportunity to improve clinical outcomes and

quality of care for high-risk patients. Our system, when combined with guided clinical interventions, has the potential to substantially reduce 30-day readmissions of CVD patients substantially.

Software Platform:

ScriptChain Health's web application is a robust, full-stack application. Being a healthcare-based company, the primary focus of ScriptChain Health is on security. ScriptChain Health is focused on patient privacy. Our software practices focus on protecting our users and remaining compliant with modern regulation, such as HIPAA and HITECH compliance. Using AWS, we are ready to scale our application to meet an increased user demand as well as keeping a centralized audit compliant environment.

Angular Frontend:

ScriptChain Health has a fully-responsive one-page web environment developed in Angular. Our web application uses the most modern web technology to build our software. The application navigates and loads content using JavaScript, allowing users to swiftly navigate around the website without waiting for the browser to load the content.

From our web application, we can employ ScriptChain Health's full suite of services. This includes user management/configuration, patient data management, and AI data analysis, with more to come. Interfacing with these services is done simply through the Angular frontend.

ScriptChain Heath seeks to give Providers a highly streamlined experience. Providers can view all of the patients they are treating, as well as a graphical representation of AI analysis done on the patient's historical data. This equips Healthcare Providers to actively manage many patients in a precise and efficient manner.

NodeJS Service:

Using NodeJS as our backend service, ScriptChain Health is able to have a more robust application, allowing us to integrate into Electronic Health Record systems through FHIR APIs. The Node service provides secure management of user data.

User security is our number one priority at ScriptChain Health. We have taken the following measures to ensure that our users remain safe:

1. The backend uses several forms of end-to-end encryption and JWT tokens to secure data and manage credentialing for the application's users.
2. All data coming into the NodeJS service is validated before being handled.

3. The full application uses an encrypted API key to further secure data processing within the application.

We do all this to allow our users to exchange information with their healthcare providers in a private manner.

Cloud-Based Architecture:

AWS architecture was designed and chosen for transparency and privacy, while taking advantage of several services offered by AWS. Hosting the full suite of our software services on AWS allows us to scale the application to meet user demand. AWS helps us build and maintain our services to be available to users worldwide.

The ScriptChain Health web application uses SSL encryption to secure data passing between AWS services. In-transit encryption is standard throughout a HIPAA compliant cloud-based architecture. ScriptChain Health maintains a CI/CD pipeline paired with an SDLC approach to software development practices. The SDLC inspires application progress, and the CI/CD pipeline prevents the mishandling of any secure user data. This model allows us to move our services forward in a secure and controllable manner.

Django Service:

Our Django backend service serves as our application's data analysis workhorse. Providers are able to integrate with several different EHR on FHIR systems to get insight into a patient's medical history. Using ScriptChain Health, providers will have immediate access to analysis of live patient data, and that's only part of what the service provides.

ScriptChain Health employs several AI models to form predictions about patients' cardiovascular health using EHR data. The Django service uses AI analysis to give providers insight into the patient's medical history. Analytics are reported to providers in a user friendly, graphical format.

References:

1. "PERFORMANCE of the MASSACHUSETTS HEALTH CARE SYSTEM SERIES: A FOCUS on PROVIDER QUALITY." n.d. Accessed August 23, 2021. https://www.chiamass.gov/assets/Uploads/A-Focus-on-Provider-Quality-Jan-2015.pdf.
2. Smeraglio A;Heidenreich PA;Krishnan G;Hopkins J;Chen J;Shieh L; "Patient vs Provider Perspectives of 30-Day Hospital Readmissions." BMJ open quality. U.S. National Library of Medicine. Accessed August 23, 2021. https://pubmed.ncbi.nlm.nih.gov/30687798/.
3. Liu, Zheng, Xiaohan Li, Hao Peng, Lifang He, and Philip S Yu. 2021.

"Heterogeneous Similarity Graph Neural Network on Electronic Health Records." ArXiv.org. 2021. https://arxiv.org/abs/2101.06800.
4. Li, Yang, Buyue Qian, Xianli Zhang, and Hui Liu. 2020. "Graph NeuralNetwork-Based Diagnosis Prediction." *Big Data*, August. https://doi.org/10.1089/big.2020.0070.
5. Ahmedt-Aristizabal, David, Mohammad Ali Armin, Simon Denman, ClintonFookes, and Lars Petersson. 2021. "Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future." *Sensors* 21 (14): 4758. https://doi.org/10.3390/s21144758.
6. Pollock, Madelyn, Oralia V Bazaldua, and Alison E Dobbie. 2007. "AppropriatePrescribing of Medications: An Eight-Step Approach." *American Family Physician* 75 (2): 231–36. https://www.aafp.org/afp/2007/0115/p231.html.
7. Su, Chenhao, Sheng Gao, and Si Li. 2020. "GATE: Graph-Attention Augmented Temporal Neural Network for Medication Recommendation." ResearchGate. unknown. July 7, 2020. https://www.researchgate.net/publication/342762237_GATE_Graph-Attention_Augmented_Temporal_Neural_Network_for_Medication_Recommendation.
8. Allahgholi, Milad, Hossein Rahmani, Delaram Javdani, Gerhard Weiss, and Dezső Módos. 2020. "ADDI: Recommending Alternatives for Drug–Drug Interactions with Negative Health Effects." *Computers in Biology and Medicine* 125 (October): 103969. https://doi.org/10.1016/j.compbiomed.2020.103969.
9. Shang, Junyuan, Tengfei Ma, Cao Xiao, and Jimeng Sun. 2019. "Pre-Training ofGraph Augmented Transformers for Medication Recommendation." ArXiv.org. 2019. https://arxiv.org/abs/1906.00346.

10. Tariq RA, Vashisht R, Sinha A, et al. Medication Dispensing Errors AndPrevention. [Updated 2021 May 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK519065/
11. "Offsides and Twosides (NSIDES V0.1) – TLab." 2019. Tatonettilab.org.November 15, 2019. http://tatonettilab.org/offsides/.
12. "DrugBank Online | Database for Drug and Drug Target Info." 2019. Drugbank.com. DrugBank. 2019. https://go.drugbank.com/.
13. Johnson, Alistair, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo AnthonyCeli, and Roger Mark. 2020. "MIMIC-IV." Physionet.org. August 13, 2020. https://physionet.org/content/mimiciv/0.4/.
14. Kochanek, Kenneth, Jiaquan Xu, and Elizabeth Arias. 2020. "Mortality in theUnited States, 2019 Key Findings Data from the National Vital Statistics System." https://www.cdc.gov/nchs/data/databriefs/db395-H.pdf.
15. Virani, Salim S., Alvaro Alonso, Hugo J. Aparicio, Emelia J. Benjamin, Marcio S.Bittencourt, Clifton W. Callaway, April P. Carson, et al. 2021. "Heart Disease and Stroke Statistics—2021 Update." *Circulation* 143 (8).

https://doi.org/10.1161/cir.0000000000000950.

16. Oyebode, Femi. "Clinical Errors and Medical Negligence." Medical Principles andPractice 22, no. 4 (June 18, 2013): 323–33. https://doi.org/10.1159/000346296.
17. Wittich, Christopher M., Christopher M. Burkle, and William L. Lanier. "Medication Errors: An Overview for Clinicians." Mayo Clinic Proceedings 89, no. 8 (June 27, 2014): 1116–25. https://doi.org/10.1016/j.mayocp.2014.05.007.
18. Readmission References: (all numbering needs to be reorganized once thedocument is completed)
19. [2] 2015. Performance of the Massachusetts Health Care System Series: A Focus on Provider Quality. Center for Health Information & Analysis, Massachusetts .
20. [3] 2015. McIlvennan CK, Eapen ZJ, Allen LA. "Hospital readmissions reduction program. Circulation." PubMed Central 1796-1803.
21. [4] 2018. Zohrabian, Armineh, Julie M. Kapp, and Eduardo J. Simoes. "The economic case for US hospitals to revise their approach to heart failure readmission reduction." Annals of translational medicine 6 14.
22. [5] 2018. "Hospital Readmissions Reduction Program (HRRP)." NEJM Catalyst
23. [6] 2019. Smeraglio A, Heidenreich PA, Krishnan G, Hopkins J, Chen J, Shieh L. "Patient vs provider perspectives of 30-day hospital readmissions." BMJ Open Quality, 8 e000264.
24. [7] 2016. Fiscella K, Sanders MR. "Racial and Ethnic Disparities in the Quality of Health Care." Annual Review of Public Health