Computers in Biology and Medicine 101 (2018) 199–209

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology

Bhargava K Reddy[a], Dursun Delen[b,*]

[a] UCB Biosciences, Inc., 8010 Arco Corporate Drive, Suite 100, Raleigh, NC, 27617, USA
[b] Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, Tulsa, OK, 74106, USA

ARTICLE INFO

Keywords:
Readmission
Lupus
Machine learning
Predictive analytics
Deep learning
LSTM

ABSTRACT

Hospital readmission is one of the critical metrics used for measuring the performance of hospitals. The HITECH Act imposes penalties when patients are readmitted to hospitals if they are diagnosed with one of the six conditions mentioned in the Act. However, patients diagnosed with lupus are the sixth highest in terms of rehospitalization. The heterogeneity in the disease and patient characteristics makes it very hard to predict rehospitalization. This research utilizes deep learning methods to predict rehospitalization within 30 days by extracting the temporal relationships in the longitudinal EHR clinical data. Prediction results from deep learning methods such as LSTM are evaluated and compared with traditional classification methods such as penalized logistic regression and artificial neural networks. The simple recurrent neural network method and its variant, gated recurrent unit network, are also developed and validated to compare their performance against the proposed LSTM model. The results indicated that the deep learning method RNN-LSTM has a significantly better performance (with an AUC of .70) compared to traditional classification methods such as ANN (with an AUC of 0.66) and penalized logistic regression (with an AUC of 0.63). The rationale for the better performance of the deep learning method may be due to its ability to leverage the temporal relationships of the disease state in patients over time and to capture the progression of the disease—relevant clinical information from patients' prior visits is carried forward in the memory, which may have enabled the higher predictability for the deep learning methods.

1. Introduction

Digitization of electronic health records (EHRs) has created momentum for a positive change in the delivery of health care. However, adoption of EHRs is only a step forward. Mere adoption of EHR did not significantly improve the quality and delivery of care. Therefore, a push has arisen to not only collect data but also to make meaningful use of it. The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 allocates approximately $2 million to $10 million for each hospital that qualifies under "meaningful use" [23]. The goal is to improve quality and gain efficiencies in the delivery of health care. Financial incentives coupled with risk sharing models pave the way for a value-based model. In this value-based model, the value of care combines the measures of payment and patient outcomes such as rehospitalization, mortality, etc. to establish a quantitative metric of hospital performance. In this research, we focus on hospital readmission, which is defined as an event in which a patient is readmitted to the hospital for the same medical condition or for a different condition within a period of 30 days. This can be a planned or unplanned visit, and the admitting hospital can be the same hospital as the original admission or a different hospital.

Clinical decision support systems deployed on EHR databases enable real-time decision making by leveraging analytics models that can predict an event—such as rehospitalization —at the time of discharge. Systemic Lupus Erythematosus (SLE), also referred to as Lupus, is an autoimmune disease that affects multiple system organ classes. Therefore, diagnosis and disease management becomes a difficult task. Clinical manifestation of lupus is highly heterogeneous with most common organ manifestation being musculoskeletal, renal, and skin [34]. There is also a large variability in the disease severity and disease activity across patients with Lupus. Disease epidemiology shows that age-adjusted incidence rate of Lupus was 5.5 per 100,000 population

* Corresponding author. Management Science and Information Systems, Center for Health Systems Innovation, Spears and Patterson Endowed Chairs in Business Analytics, Spears School of Business, Oklahoma State University, 700 North Greenwood Avenue, North Hall 302, Tulsa, OK 74106, USA.
E-mail addresses: Bhargava.Reddy@ucb.com (B.K. Reddy), dursun.delen@okstate.edu (D. Delen).
URL: http://spears.okstate.edu/delen (D. Delen).

https://doi.org/10.1016/j.compbiomed.2018.08.029
Received 16 May 2018; Received in revised form 29 August 2018; Accepted 30 August 2018
0010-4825/

and the disease prevalence was 72.8 per 100, 000 population in the US [36]. The disease prevalence across countries and regions seem to be highly variable. Asia pacific had a prevalence of 4.3–45.3 per 100, 000 [21] while the disease is rarely diagnosed in African regions. A recent study conducted by Rees et al. [35] showed that North America had the highest incidence rate of 23.2 per 100,000 person-years while Africa and Ukraine had an incidence of 0.3 per 100,000 person-years and almost negligible, close to zero, in Australia. The same study also showed that women are more frequently diagnosed with the disease than men. There was also a significant variability in the prevalence across different ethnicity groups; the disease prevalence was very high in African Americans compared to the Caucasians. This variability in the disease status, across gender, ethnicity, and regions justifies and explains the complexity in building an analytical model that can predict the clinical outcome. Research shows that identifying subgroups of patients who are high-risk and allocating resources accordingly has resulted in an absolute reduction of five readmissions per 100 index admissions in the hospital [2]. Literature suggests that high-value drivers in health care are predicting readmissions, identifying high-cost patients or patients with chronic conditions, predicting an adverse events and disease progression, and treatment optimization for conditions such as lupus [3].

2. Motivation and background

Hospital readmissions accounts for $15 billion in health care spending annually as of 2007 [33], and approximately 20% of all hospital discharges with Medicare resulted in a readmission within 30 days [31]. The Medicare Payment Advisory Commission (MedPAC) estimates that at least 12% of those hospital readmissions are avoidable. These metrics show that a large amount of health care resources are not being utilized efficiently with the fee-for-service model [7]. Although lupus is not part of the six diseases under financial penalty, studies show that lupus ranks sixth among diseases with the highest hospital readmissions [16]. Therefore, research into understanding the cause of the disease, predictors for rehospitalization, and disease pathways is clearly needed to benefit health care providers, payers, and patients.

Previous work on rehospitalization for lupus primarily focused on identifying its predictors and understanding the characteristics of readmission. To the best of our knowledge, analytics models to predict hospital readmission for lupus patients have not been developed. This lack of analytic research may be attributed to (1) not having access to a high enough volume of data on this relatively rare chronic condition, and (2) the fact that lupus is not one of the six medical conditions under the hospital readmission reduction program (HRRP), which imposes financial penalties. HRRP lists the following six medical conditions that qualify under the financial penalty scheme: acute myocardial infarction (AMI), congestive heart failure (CHF), pneumonia, chronic obstructive pulmonary disease, coronary artery bypass graft, and hip and knee replacement [44]. One of the prior research studies on lupus performed by Yazdany et al. [43] developed a causal model leveraging a multilevel mixed effects logistic regression model to identify predictors. The predictors of rehospitalization identified from this study were age; race; public insurance (Medicaid, Medicare) versus private insurance; and clinical characteristics such as active lupus nephritis, serositis, and thrombocytopenia. Another study performed by Edwards et al. [14] also looked at the factors for admission and the characteristics associated with a poor outcome. Active lupus nephritis and multisystem disease make readmission more likely. A study in Malaysia on Asian patients diagnosed with SLE shows that the main causes for hospital admission were a flare of SLE, infection and renal biopsy [40].

A significant amount of research has been performed on understanding and predicting rehospitalization in general, but not much at the disease-specific level. Prior research has been performed in several disease areas, including predicting heart failure, AMI, pneumonia, and chronic obstructive pulmonary disease (COPD). Major risk factors found to be associated with hospital readmission are age, race, sex, drug use, number of previous hospitalizations, insurance type, and co-morbidities [2]. Literature shows that all the prior research has been performed by using baseline data from a single time-point and predicting the likelihood of readmission within 30 days. To the best of our knowledge, research using analytic methods that leverage large and feature-rich longitudinal EHR data on lupus patients has not been done to predict rehospitalization.

With the increased success of the new applications of deep learning methods in medicine, especially on the medical image and sound data, including electroencephalography (EEG) [1], CT scan [42], and ultrasound [5] and advancement in computing capabilities on hardware and software, deep learning methods have gained much attention. Deep learning has also produced tremendous success in natural language processing (NLP) tasks for sentiment analysis [11], language translation [38], topic classification, and obtaining answers for questions [6]. Another major field where deep learning has been extensively researched is voice/audio processing. Audio or speech recognition [19] and audio classification using convolutional deep belief networks are successful application areas for deep learning [28].

Recent successes in the application of deep learning in other industries have created an opportunity to explore deep learning methods in the health care domain. A recent study was done applying recurrent neural networks leveraging longitudinal EHR data for early detection of heart failure, which showed higher performance compared to traditional machine learning methods [10]. Another study utilized deep learning methodologies such as Restricted Boltzmann Machines (RBM) to differentiate disease states for patients with Parkinson's disease. The results from the deep learning methodology show a higher generalization performance compared to classical machine learning methods [18]. Similarly, some work has been done in the pharmaceutical industry for predicting the activity of molecules to perform quantitative structure activity relationships (QSAR), where deep neural nets show higher performance compared to random forest and support vector machines [30]. A recent study, conducted on patients in a critical care setting for various medical conditions, shows that the recurrent neural networks with long short-term memory (LSTM) architecture perform better than the multilayer perceptron neural network and the linear classification models in predicting the multiclass diagnosis model [29].

To the best of our knowledge, there is a dearth in the current literature in using deep learning methods that leverage longitudinal EHR data to predict hospital readmissions. Similarly, no research has been done on applying predictive analytics models to lupus. Therefore, combining robust deep learning methods and leveraging longitudinal data from patient EHR databases to predict hospital readmission in an autoimmune disease that is highly complex in terms of the disease state makes it a challenging research problem to solve. Traditional machine learning methods are not designed to benefit from the temporal relationships that exist in the longitudinal data. EHR data is longitudinal in nature and captures temporal relationships within clinical data for a patient, which is very critical in understanding the progression of the disease. Patient disease progression over time can predict hospital readmission more accurately than a traditional classification model, which cannot capture the temporal relationships well.

3. Methodology

Deep learning is an advancement of artificial neural networks with multiple levels of representation that are created by nonlinear transformations at each level. Deep learning discovers intricate structures and complex relationships in large datasets by using algorithms that allow for tuning and readjustment of features that are used to compute the representation in a layer based on the representation from the previous layers [27]. Good features in a classification model can be learned automatically without having to manually select features, which is a major advantage compared to other methods. Historically, deep learning methods performed well, but their biggest limitation is their inability to train fast due to lack of quick and efficient processing capabilities. This limitation is now being addressed with the advancements of the big data analytics technologies.

3.1. Data

The data for the current research was extracted from the Cerner HealthFacts EMR database using the data collected between the years 2000–2015 inclusive. The Cerner database consists of de-identified patient-level data collected across approximately 500 health care facilities in the United States with 63 million unique patients [13]. For the purpose of this study, we identified lupus patients by following the International Classification of Diseases (ICD) 9 standards, extracting patients that have a diagnosis code of 710.0.

After the data was extracted, we filtered for those patients who had more than two encounters. This was done so that we had at least two clinical encounters in the observation window and at least one observation in the prediction window. The dataset size reduced significantly to 30,291 unique patients. By filtering for patients who had more than one clinical encounter, we had 58,004 unique patients. Deep learning LSTM model requires longitudinal sequential patient level data that has been collected over time across several clinical encounters. The deep learning sequential model, LSTMdeveloped in this research study require a sequence of records in the observation window, unlike traditional machine learning models. Traditional classification models require one record per patient containing the predictors (independent variables) and a response variable (dependent variable) which is a binary variable.

The objective of the current study was to predict 30-day rehospitalization; therefore, we only included patients who were admitted to the health care facility as inpatients or under an emergency category. As a result, we were left with 11,007 patients. We then checked to see whether there was more than one hospital visit, indicating that the patient was rehospitalized. Now that we had subjects who were rehospitalized, we checked whether they were admitted within 30 days after the observation window or later. The observation window, as shown in Fig. 1, consisted of all the medical records for patients from day 1 to day 60. Information from these medical records was used as predictor information. Data from day 61 to day 91 was used for prediction. If a patient was admitted after the observation window, i.e., between day 61 and day 91, he/she was flagged as rehospitalized within 30 days. If a patient did not have a hospitalization record between day 61 and day 91, we flagged that record as the patient had not been rehospitalized within 30 days.

We filtered for patients who had at least two clinical visits in the observation window, creating sequential data for prediction. The number of patients were further reduced. The next step was to ensure the completeness of the data. We dropped the patients that had missing data values for the predictors and only kept the ones that had complete data. That way, we try to avoid inclusion of artificial values through the selection and implementation of any data imputation techniques on the missing values. In the end, we were left with 9457 unique patients in the final data set. Data for the modeling included information from the encounters dataset available in the Cerner Health Facts database. The final data set included demographics information, diagnosis information, diagnosis priority, payer information, cost of the medical encounter, and discharge information. The visual representation of the detailed steps involved in the process—starting with data extraction to preparing the analysis-ready dataset—is shown in Fig. 2.

The artificial neural networks (ANN) and regularized regression methods have similar data structures for the modeling-ready dataset but differ from the recurrent neural network method. For deep learning methods, a one-hot vector method was used to represent the data for modeling. A one-hot vector representation of the predictors or events is a representation of the data in which the unique values of each predictor P, is represented as a P-dimensional vector where one dimension is set to 1 and all the other dimensions are set to 0. The one-hot vector representation for each patient is performed on clinical events, diagnosis, etc. for each hospital visit from day 1 to day 60. However, for the ANN method and regularized regression methods, a two-step method was followed where data was also represented as a one-hot vector but aggregated across time points from day 1 to day 60. Not all the clinical variables were converted into the one-hot vector format due to the issue of sparseness. An equivalent higher-level variable was used in creating the one-hot vector.

The distribution of the response variable for the entire dataset, including training and testing subsamples, shows that approximately 82.8% of patients were not rehospitalized in the next 30 days, while 17.2% of them were rehospitalized (see Fig. 3). It is worth noting that there was a significant imbalance in the distribution of the response variable, which is often recognized as a hindrance for developing accurate classification type prediction models [39]. However, in this study, balancing techniques such as SMOTE were employed to ensure that the model building and interpretation of the results were done accurately.

Data utilized for model building can be broadly classified into five categories, as shown in Table 1. The Cerner HealthFacts database houses more features than those listed in the table. However, only the relevant features (as per the prior studies, expert opinion, and exploratory modeling) were utilized for model building and testing purposes. For the aggregation of the counts from each one-hot vector, sparse variables were discarded but their corresponding higher-level codes were utilized.

Demographics and disposition are shown below in Table 2. The distribution of data is skewed for demographic parameters such as gender and race. However, the distribution of age is not as skewed as the other demographic parameters. Note that the average number of days that the patient stayed in the hospital is 16.39 whereas the median is eight days, showing some outlier patients with a larger duration of hospital stay. The number of days between hospital visits is also very skewed, but that skewness is explained by the patients who were not rehospitalized in the next 30 days. These patients were rehospitalized,

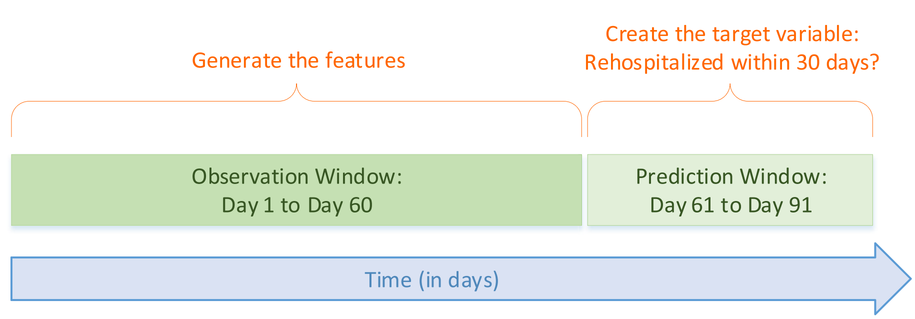

Fig. 1. Study design showing the observation and prediction window.

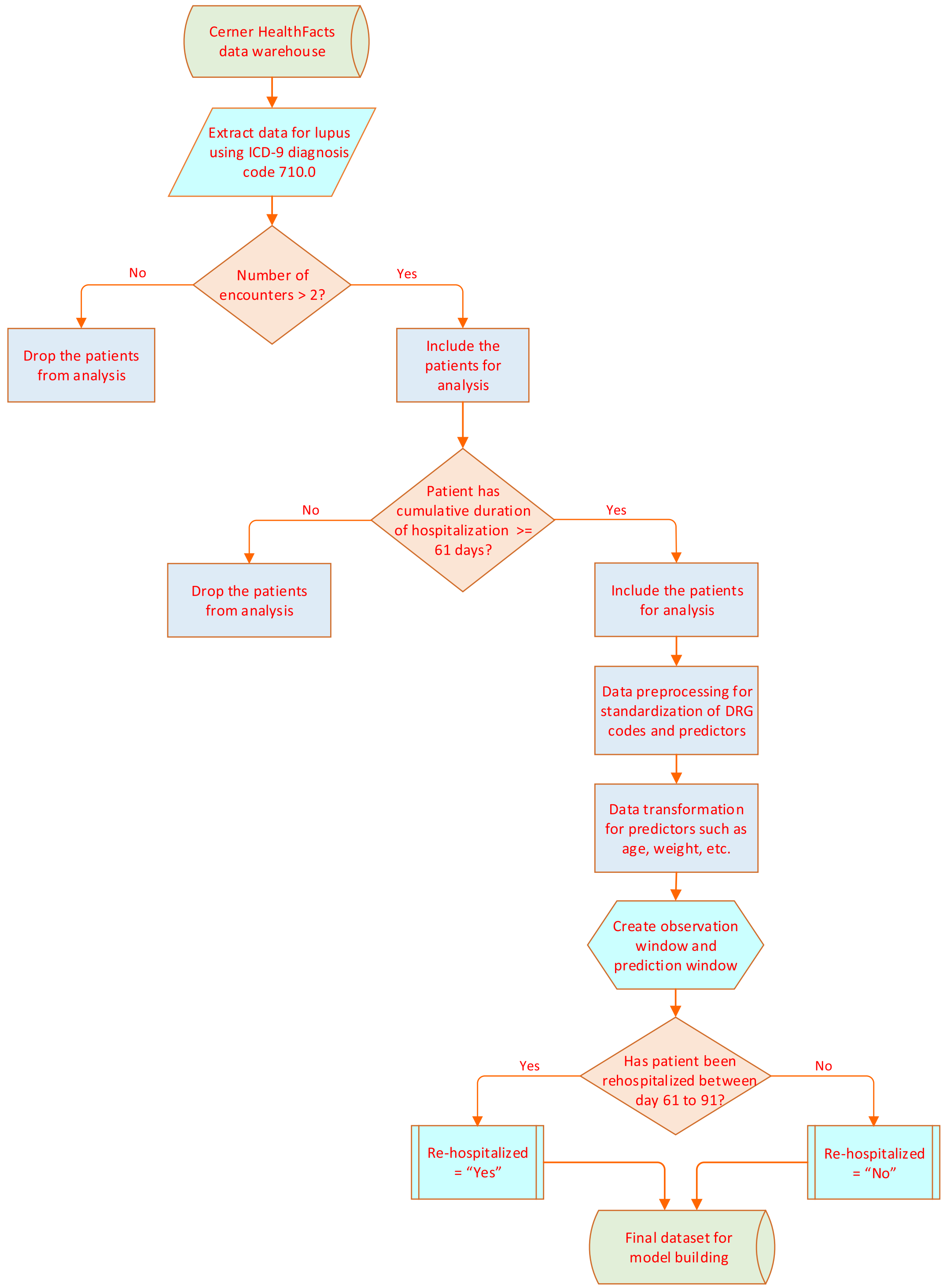

Fig. 2. The process of preparing the data for building the predictive models.

but not within the 30 days after completion of the observation period.

Data preprocessing such as data cleaning and data transformation to make an analysis-ready dataset for modeling was performed using SAS 9.4 software. The data exploration and data understanding steps were performed using SAS JMP 12.0 and data transformation for the ANN method was performed in R. The specific software tools and options used for the modeling are explained in the methods section.

3.2. Analytic methods

The current research utilizes regularized regression, artificial neural networks, and recurrent neural networks for predicting rehospitalization within 30 days. Regularized regression and ANNs are traditional methods that utilize a snapshot of data to predict a response. However, Recurrent Neural Networks (RNNs) utilize the longitudinal data for each patient and predict the response by leveraging the temporal relationships between the clinical data from each time point. Multiple software tools were utilized to help the development and testing of the machine learning methods. Penalized logistic regression model development and validation were performed in RStudio (version 1.0.136) utilizing the GLMNET package. ANN-multilayer perceptron model development and validation were performed using the MONMLP package

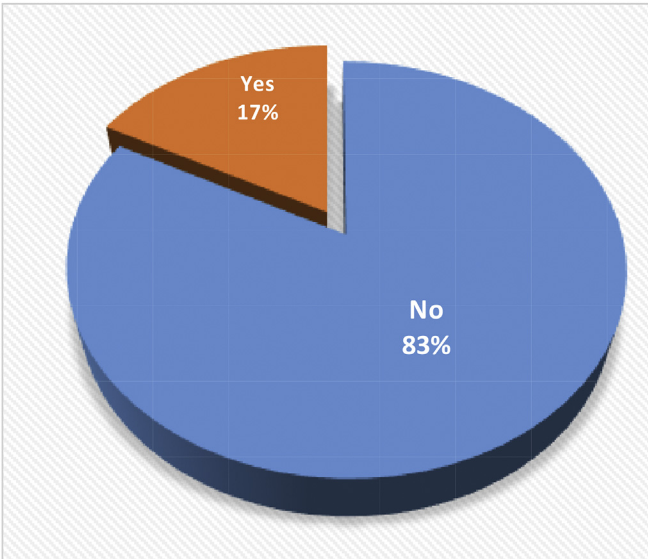

Fig. 3. Distribution of the response variable for patients across the full dataset.

Table 1
Features included in the model along with their higher-level categories.

Category	Feature Variables
Demographics	Age, Gender, Race, Marital Status
Clinical Diagnosis	DRG Code, MDC Code, Diagnosis Priority
Health Care Setting	Teaching Hospital, Bed Size, Region, Medical Specialty
Insurance	Type of Insurance
Disease Status	Acute, Emergency, Inpatient, Discharge Status, Duration of Stay in the Hospital
Derived	Duration of Hospital Stay, Duration Between Visits

within RStudio.

Deep learning sequential methods such as LSTM, RNN, and gated recurrent unit (GRU) were developed and validated using the Python programming language (version 3.6). The models were implemented using TensorFlow v1.0 as the backend and Keras API as the front end. TensorFlow is an open-source library developed for numerical computation using data flow graphs with nodes represented as tensors or multidimensional arrays. The software library was developed originally by engineers at Google [41]. Keras is an open-source, high-level neural network API written in Python running on top of TensorFlow that can run seamlessly on CPU and GPU [25].

***Regularized/Penalized Logistic Regression*:** The foundation for the development of regularized methods was the trade-off between bias and variance. A model is fit by including all the predictors and then regularizing the coefficient estimates to near zero for those that are less correlated to the response. This ensures the variance is minimized. The two best techniques that are part of the regularization or penalized techniques are ridge regression and the lasso method.

The objective function that is minimized in the penalized or regularized logistic regression method is as shown below in Equation (1), which uses the negative binomial log likelihood.

$$\frac{1}{N}\sum_{i=1}^{N} y_i * (\beta_0 + X_i^T\beta) - \log\left(1 + e^{(\beta_0 + X_i^T\beta)}\right) + \lambda\left[\frac{(1-\alpha)\,||\beta||_2^2}{2} + \alpha||\beta||_1\right] \tag{1}$$

Equation (1): Objective Function for Regularized Regression.

The coefficient $\hat{\beta}$ is calculated by applying the penalty λ as shown in Equation (2). The tuning parameter, λ, is the minimum value selected from cross-validation.

Table 2
Demographic and disposition of the parameters in the data.

Disposition or Demographic Parameters		(n = 9457)	(% of total)
Gender	Female	8493	(89.80)
	Male	962	(10.10)
	Not Recorded	2	(0.10)
Race	African American	4171	(44.10)
	Asian	91	(0.96)
	Asian Pacific Islander	23	(0.25)
	Biracial	20	(0.22)
	Caucasian	4193	(44.34)
	Hispanic	174	(1.84)
	Mid-Eastern Indian	2	(0.02)
	Native American	84	(0.89)
	Not Mapped	25	(0.26)
	Other	581	(6.14)
	Unknown	93	(0.98)
Insurance Type	Government	3998	(42.27)
	Private	5459	(57.73)
Census Region	Midwest	2004	(21.19)
	Northeast	2375	(25.11)
	South	4258	(45.03)
	West	820	(8.67)
Acute Status	Acute	9184	(97.12)
	Not-Acute	273	(2.88)
Urban-Rural Status	Rural	2053	(21.71)
	Urban	7404	(78.29)
Age	Mean	53.74	
	Median	43.00	
	Min	1.00	
	Max	> 100.00	
	Q1	30.00	
	Q3	55.00	
Duration of Hospital Stay	Mean	16.39	
	Median	8.00	
	Min	1.00	
	Max	817.00	
	Q1	5.00	
	Q3	17.00	
Duration Between Clinical Visits	Mean	138.10	
	Median	64.00	
	Min	0.00	
	Max	2250.00	
	Q1	19.00	
	Q3	120.00	

$$\beta\hat{} = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + e^{(\beta_0 + X_i^T\beta)}\right) + \lambda\left[\frac{(1-\alpha)\,||\beta||_2^2}{2} + \alpha||\beta||_1\right] \tag{2}$$

Equation (2): Estimating $\hat{\beta}$ Coefficients.
Ridge penalty: $||\beta||_2^2$
Lasso penalty: $||\beta||_1$
λ: Regularization parameter.

The ridge penalty shrinks the coefficients to near zero but not exactly to zero, whereas the lasso shrinks the coefficients to zero for those predictors that are not relevant to the model function. Therefore, the lasso method is primarily used for feature selection, making it an easier model to interpret. Although, the lasso method has high interpretability, the limitation is that when the pair-wise correlations between a group of relevant predictors are high, the lasso method selects only one of the predictor at random. The ridge penalty method keeps all the features in the model, making it harder to interpret the predictors for the model [22]. The current study employs a hybrid regularization method called elastic net which is also referred to as penalized logistic regression. This penalization technique is a combination of the ridge and the lasso methods.

Artificial Neural Networks (ANN): This methodology was developed by taking the inspiration from neuroscience and how a human brain works. ANNs are used to model complex nonlinear relationships in the data to predict a continuous or a categorical outcome. We utilized

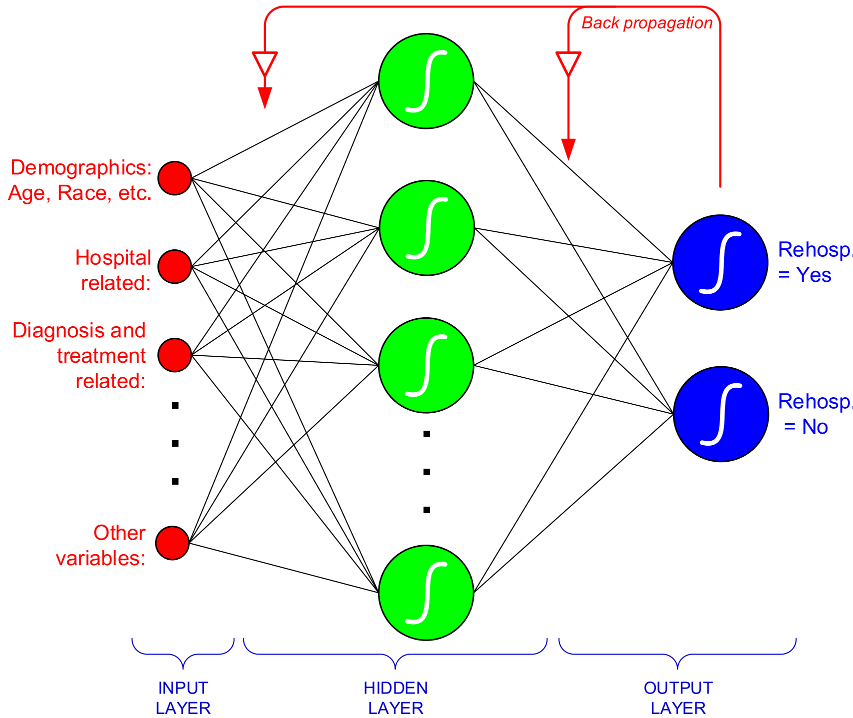

Fig. 4. Artificial neural network with feedforward connection.

the feedforward network, also called the multilayer perceptron (MLP), and the sigmoid activation function to predict the class variable of hospital readmission. A single hidden layer with no weight decay parameter was applied. There are three main components in the network structure and an input layer that contains the predictors that were transformed. The second layer, which is the hidden layer, is the vector values that connects the input and output layers and the output layer, which is the output of the activation function as shown in Fig. 4.

3.3. Deep learning methods

Deep learning methods are machine learning methods that are composed of multiple levels of nonlinear operations with many hidden levels within neural networks. Due to several transformations and multiple hidden layers that capture the relationships, complex functions can be learned to discriminate the response classes in classification problems. This is the primary difference between traditional ANNs and deep learning methods. Several deep learning methods were developed and enhanced recently due to the computing capabilities such as GPUs that can handle multiple layers of neural networks and build a complex function [17]. Some of the recent developments of deep learning methodologies and enhancement to the existing methods include convolutional neural networks, deep belief neural networks, restricted Boltzmann machines, Hopfield networks, autoencoders, and recurrent neural networks. For the purposes of this research, the recurrent neural networks and their variants, i.e., LSTM and GRU methods, are used to model rehospitalization and hence, are described in more detail herein.

***Recurrent Neural Networks (RNNs)*:** These methodologies were primarily developed to process sequential data. Feedforward neural networks with cyclic connections are essentially the RNNs. An MLP or any feedforward network can only map from input to output vectors, whereas RNNs can map from the entire history of previous input data to each output, which is what makes RNNs good for processing sequential data. A simple RNN architecture is shown in Fig. 5. Each node in the diagram represents a layer of network units at each time point. There are three important connections: input to hidden layer, hidden to hidden layer, and hidden layer to output layer. The weighted connections from input to hidden layer are represented in the matrix U. The weighted connections from the hidden to hidden layer are represented as matrix W, and the weighted connections from hidden layer to output layer are represented as matrix V. The final weight matrix is passed through a sigmoid function to produce a scalar Y value, which is then classified as a binary variable referenced as $\hat{Y}$. The loss function is then applied to compare Y-actual and Y-predicted ($\hat{Y}$). It is important to note that the same weights are used at each time point. However, the current RNN has a vanishing gradient problem. The term "vanishing gradient" refers to the problem of the influence of a given input on hidden layer and subsequently on network output, which either decays exponentially or explodes exponentially over time as the data goes through transformations on the network. Although RNNs were built to learn long-term dependencies, empirical evidence shows that due to vanishing gradients, the network is unable to learn long-term dependencies [4].

Therefore, two popular enhancements to the current RNN method were developed: long short-term memory (LSTM) and, most recently, the gated recurrent unit (GRU). For the current research, we used RNN – LSTM architecture, which will be explained in the next section. A GRU method was also developed and tested to compare the performance of sequential models.

***Recurrent Neural Network – Long Short-Term Memory (RNN – LSTM)*:** The architecture consists of connected subnetworks, also called memory blocks. The natural function of memory blocks is to remember inputs for a long time [20]. Each memory block contains at least one self-connected accumulator cell and several multiplicative units such as input gate, forget gate, and output gate. These three gates allow us to store and access information by assigning a counter such as 0 and 1. The memory block architecture of LSTM is shown below in Fig. 6. The activation function is the logistic sigmoid function, which is similar to the MLP except that the activations come from the input at current time points and the activation from the hidden layers depending on the gates for LSTM.

Hyperparameters. The sequential models need regularization parameters and optimization methods. Table 3 lists the parameters and methods used in the modeling step of our effort.

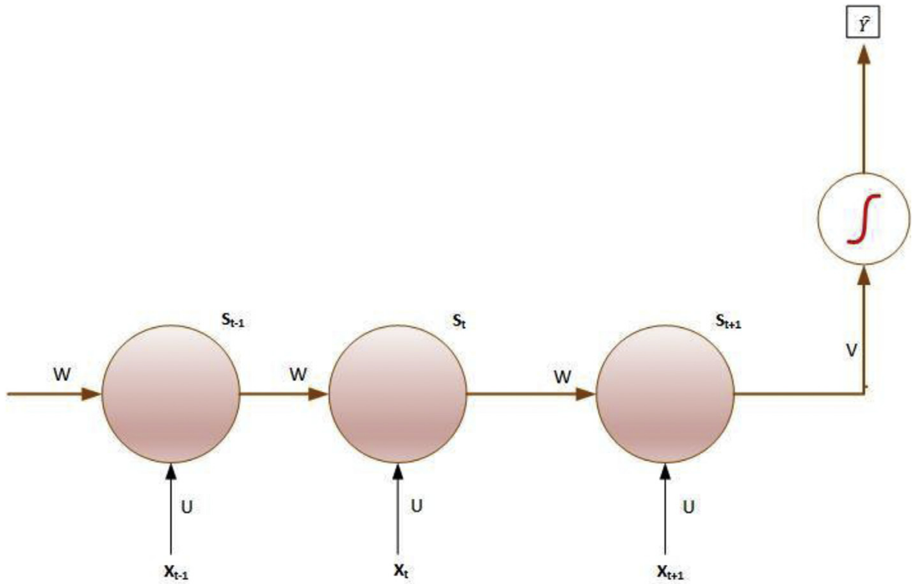

Fig. 5. Recurrent neural network unfolded in time.

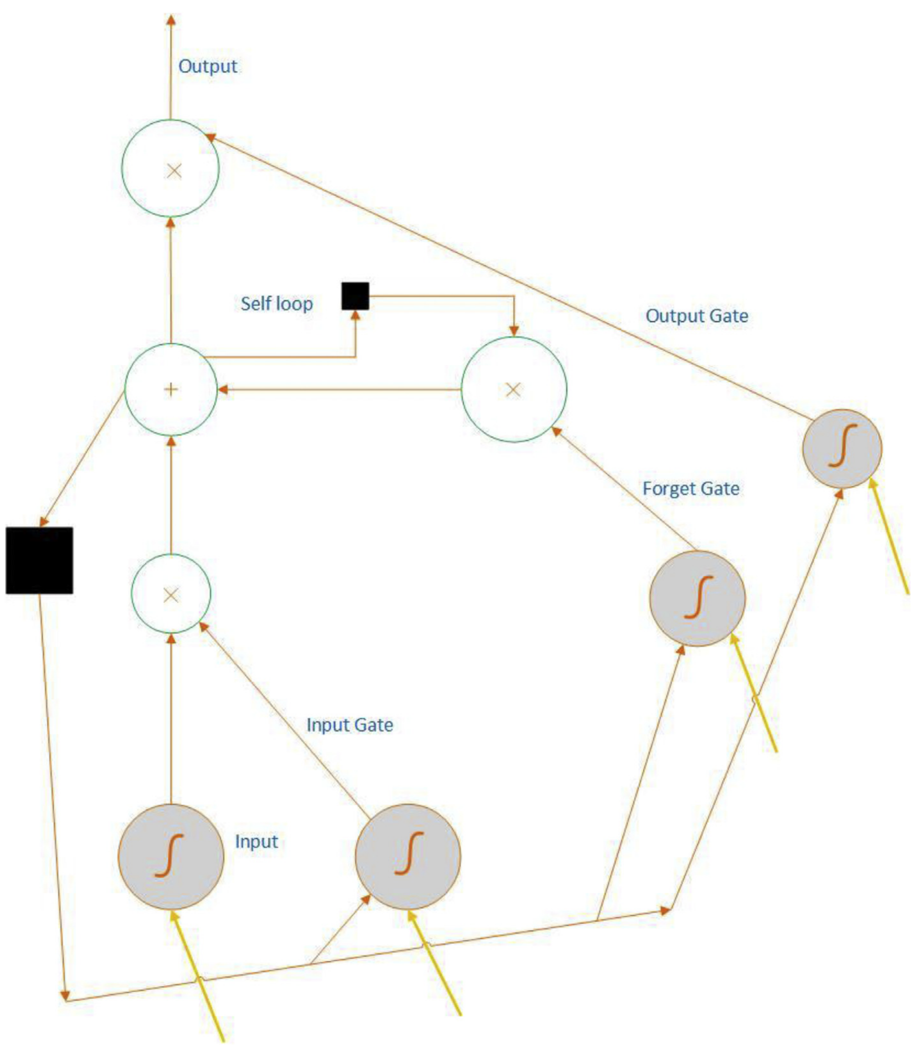

Fig. 6. LSTM block diagram showing the gates and the state function.

Table 3
Hyperparameters and learning algorithm.

Dropout	0.1
Recurrent dropout	0.1
Mini batch	512
Optimization method	Adam (Adaptive moment)
Loss	Binary Cross-Entropy
Output activation	Sigmoid

Dropout is a regularization technique used to minimize overfitting. Based on affine transformations, several units are randomly dropped from the network [37]. The probability of the number of units to be kept is a hyper parameter that is assigned. In the current research, we assigned a probability of 0.1. This indicates that we will drop 10% of the units at random from the network. By discarding several units at random, a much smaller network is created and the weights are shrunk and distributed across the predictors in the model. For the recurrent models, a recurrent dropout rate must be applied to ensure that the dropout mask is the same from time point to time point, in our case from visit to visit. By utilizing the same dropout mask at all the time points, the learning error is able to propagate to the end.

Mini batch size is the number of samples processed from the training set for each iteration. In the current research, we selected a mini batch size of 512 (2^9) to allow for faster processing of the stochastic gradient descent. A large mini batch size takes too long per iteration and a small batch size does not reach the local minima. The cost function on the mini batch trends downward but does not always decrease, unlike the batch gradient descent.

Adam, also referred to as adaptive moments is a learning rate optimization technique that is an enhancement of the combination of RMSProp and moments technique. The effect of gradient descent with RMSProp is combined with the effect of gradient descent with moments [26]. The rationale for selecting Adam as the learning algorithm is that it requires less memory and is best suited for sparse parameters. The learning rate parameter, α, needs to be tuned.

The weights and bias are updated as

$$W{:} = W - \propto \frac{Vdw(corrected)}{\sqrt{Sdw(corrected)} + \in}$$

$$b{:} = b - \propto \frac{Vdb(corrected)}{\sqrt{Sdb(corrected)} + \in}$$

Binary cross-entropy, also referred to as the cross-entropy, is the loss consisting of negative log-likelihood between the training set and the probability distribution of the model. It is the loss function between the actual and the predicted on the training set that is being minimized. The sigmoid function is the output activation function that is applied to the final layer after combining the output from the hidden states and the inputs from each time point. The sigmoid activation function finally converts the values into a probability in the range of 0–1.

3.4. Training and test strategy

Data were split randomly between training and test datasets in the proportion of 70 to 30, respectively. All three machine learning methodologies, including the regularized logistic regression, ANN, and LSTM methods, used the same seed number to split the data between the training and test sets. This approach was taken to ensure reproducibility and accurate comparison of the results. Although a 10-fold cross-validation method is frequently used for building and testing machine learning models' performance [12], herein, we chose to utilize a 70:30 split. The rationale for that decision was to take into consideration the computing needs of the RNN-LSTM method, which requires significant amounts of processing time for a relatively large dataset with a large number of hidden layers. The proportional split of data seems to be a common practice for deep learning applications when dealing with a large volume of data [9]. Due to huge imbalances in the data, a Synthetic Minority Oversampling Technique (SMOTE) was used to balance the training dataset. This is a widely accepted technique used for class-imbalance problems where the minority class is oversampled, taking each minority class and introducing synthetic examples by joining the *k*-nearest neighbor patients available in the dataset [8,39]. For the current research, *k* equal to 100 is utilized to create synthetic patients to oversample the training dataset for the minority class. The key is that the distribution of the response variable in the test set is very similar to the training set.

Table 4
Model performance metric for all the models.

Machine Learning Model	Area Under the Curve (AUC)	Sensitivity	Specificity	Accuracy
Artificial Neural Network (ANN)	0.66	50.80%	70.16%	66.79%
Regularized/Penalized Logistic Regression	0.63	69.63%	72.12%	71.69%
Long Short-Term Memory (LSTM)	0.70	74.49%	56.61%	70.54%

4. Results and discussion

Results from all the proposed models are shown in Table 4. Performance shows that the LSTM method had significantly higher performance compared to the traditional classification models based on the area under the curve (AUC) metric. Due to the significant imbalance in the test sample, prediction accuracy is not considered to be the best metric to gauge the model's performance, therefore AUC metric is utilized in assessing the models' performance. The AUC for both the traditional classification models seem to be within a similar range; however, the AUC for the deep learning method-LSTM is significantly higher compared to the rest of the traditional classification methods. This superior performance by the LSTM model can be attributed to its ability to capture the temporal relationship in the longitudinal data. Results of key metrics such as AUC, Sensitivity, Specificity, and Accuracy for the three proposed models are shown in Table 4.

The key metrics, such as sensitivity, were significantly higher for the LSTM model compared to the traditional classification models. This means that predicting patients who will be rehospitalized when they were actually rehospitalized was significantly higher with the LSTM model at 74.49% compared to regularized logistic regression, which was 69.63% followed by ANN at 50.80%. The most important value driver for health care delivery is being able to predict accurately those patients who will be rehospitalized as opposed to those that will not.

As noted earlier, the hypothesis of the study was to compare traditional machine learning methods and the LSTM model. However, variants of the LSTM model were also developed and tested. Simple RNN and GRU are the two additional sequential deep learning methods that were also developed. Comparison of the results from the three sequential models is shown in Table 5. Model performance based on

Table 5
Model performance metrics for deep learning sequential models.

Machine Learning Model	Area Under the Curve (AUC)	Sensitivity	Specificity	Accuracy
Long Short-Term Memory (LSTM)	0.70	74.49%	56.61%	70.54%
Gated Recurrent Unit (GRU)	0.70	55.46%	73.06%	70.00%
Simple Recurrent Neural Network (RNN)	0.69	32.28%	88.00%	78.32%

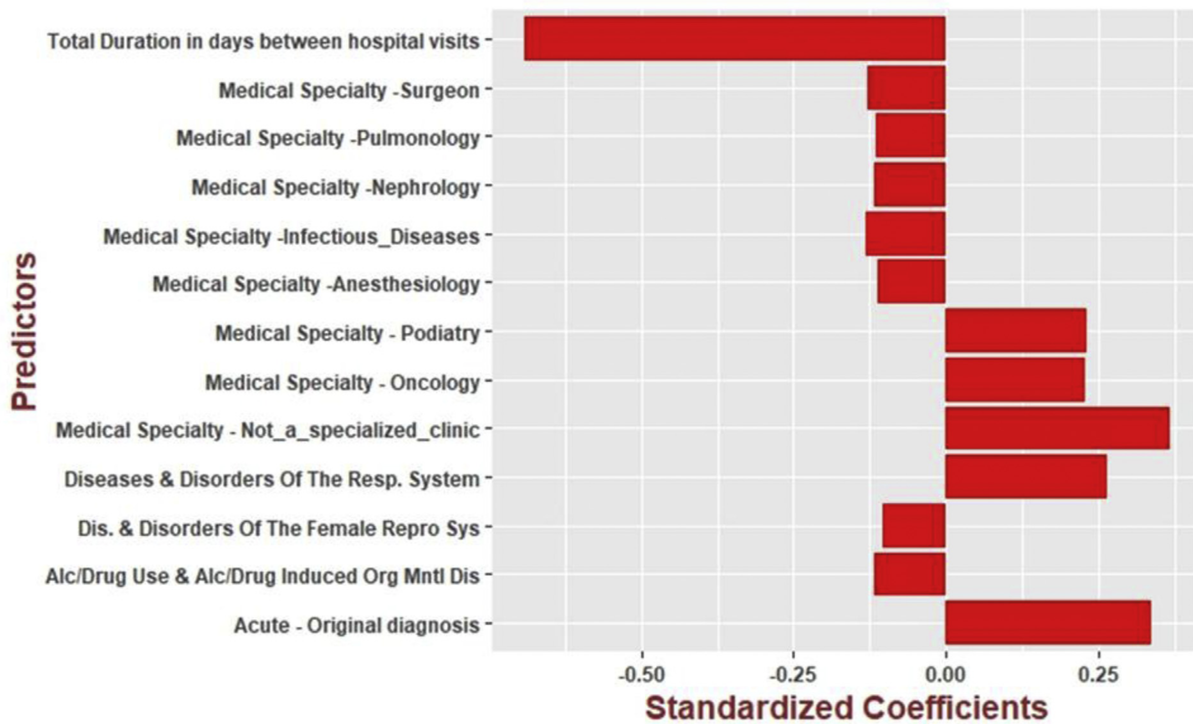

Fig. 7. A graphical depiction of the variable importance measures of the predictor variables calculated based on the standardized coefficients from the penalized logistic regression model.

AUC shows that all three sequential models were within the same range. The sensitivity is quite high for the LSTM model compared to the RNN and GRU methods. This shows that the LSTM model can predict accurately when patients are actually rehospitalized, whereas the RNN method has very low sensitivity. The LSTM and GRU methods are extensions of the simple RNN method that can store in memory relevant prior information from the features extracted.

Literature review on hospital readmissions showed that the prediction models developed over time were highly variable in their findings. Nine studies conducted in the US on a large population across multiple centers provided a predictive or discrimative ability of 0.55–0.65 based on the c-statistic metric [24]. The c-statistic which is an equivalent of AUC, is defined as the proportion of times the model accurately discriminates a pair of high and low risk individuals or groups. Although these prior models are being referred to as predictive models, they are actually inferential models developed on retrospective data. This means that the goal of the study was not to predict hospital readmission based on the prior EHR data at the point of care, but to be able to separate the patients that are likely to be readmitted restrospectively using a c-statistic. A single center study on heart failure patients utilizing predicitive models and manual intervention in US showed a 0.72 for the c-statistic [2]. The results are comparable to the current study but the underlying predictive models were not explicitly stated in the research and there was a manual intervention to first classify high-risk and low-risk patients. The lack of explanation of the methods in the research [2] used and the manual intervention makes the results less comparable to the current study.

Prior research in hospital readmission using deep learning methodologies was primarily done on the five disease conditions that were identified as part of HRRP act. Five separate predictive models were developed for each disease condition. Total hip and knee replacement (THKA) disease condition had the lowest AUC of 0.63 while patients readmitted for pneumonia had the highest AUC of 0.73. Patients admitted for heart failure had a AUC of 0.67, acute myocordial infraction (AMI) 0.64 and chronic obstructive pulmonary disorder (COPD) had an AUC of 0.71 [15]. Based on the AUC metric, we can clearly note that the current deep learning models developed by our study have superior performance than models developed by Ref. [15] for three disease conditions such as heart failure, AMI and THKA. The models developed for pneumonia and COPD are slighlty better ($< 3\%$) than the current models developed by us. This slightly lower performance for lupus compared to pneumonia and COPD can be attributed to several reasons, complexity of the disease condition which impacts several system organ classes, lack of timely diagnosis and insufficient specialists such as rheumatologists. All the aforementioned factors contribute to late clinical intervention which impacts the predictability of the rehospitalization. Further, the models developed by Futoma et al. [15] are not referred to as sequential models, which means those models (RBM) did not have the capability of utilizing rich longitudinal patient level data that spanned across several time points.

Yazdany et al. [43] was the only study focused on hospital readmissions in lupus patients which was a causal model intended to identify the predictors of hospital readmission and their influence on the readmission for a patient. Although a large number of predictors were used in the current modeling, only a few of them were found to be statistically significant and important based on the regularized logistic regression model. Predictors in Fig. 7 show that total duration between visits was found to negatively influence whether a patient would be rehospitalized. That means the longer the duration between clinical visits, the less likely the patient will be rehospitalized. Note that patients admitted for the first time in the medical specialties surgery, pulmonology, and nephrology are less likely to be readmitted within 30 days. However, it is also worth noting that patients admitted for podiatry conditions are more likely to be rehospitalized within 30 days. Several studies have already shown that patients with lupus have more biomechanical problems in the feet than the control population [32]. It is also worth noting that patients admitted initially with acute disease severity are more likely to be rehospitalized within 30 days than those patients who are admitted with nonacute severity.

Lupus patients admitted originally for alcohol and drug use are less likely to be rehospitalized within 30 days. Patients with reproductive disorders are likely to be readmitted within 30 days. These less-severe and nonlife-threatening conditions may not inform the scientific community related to the disease state but can be very useful in distinguishing the high-risk from the low-risk patients who are more likely to be readmitted. Variable importance for deep learning methods is found to be challenging to interpret and less useful for inferential reasoning. These deep learning methods are considered more of a black box in terms of interpretation; however, their higher predictive performance has attracted much attention.

5. Summary and conclusion

Rehospitalization is a key driver that is part of improving health care delivery, i.e., to transition from a fee-for-service model to an outcome-based model. Predicting rehospitalization helps ensure optimal utilization of the limited health care resources that are currently available, thereby improving quality as well. Results from this research show that the deep learning methodology, which can utilize longitudinal EHR data as sequential data, shows significant promise in predicting rehospitalization in lupus patients. The deep learning methods RNN with LSTM architecture can outperform all the other comparable models mentioned here due to the advantage of leveraging the longitudinal EHR patient data. Clinical manifestation of the disease state and progression of the disease state are among the primary indicators of future hospitalization. Therefore, deep learning methods may show superior performance over traditional machine learning methods.

The limitation of the RNN method with the further development of this research is the issue of sparsity. Traditional classification methods encountered a sparsity issue, but due to aggregation of the frequency of occurrence of clinical events for the observation window, the sparsity issue was resolved. However, for the deep learning method, sparsity seems to pose an issue. This issue was resolved by utilizing the diagnosis-related group (DRG) codes and major diagnostic codes. Higher level features were created for variables such as insurance type and admission types to reduce sparsity. The second biggest limitation with the deep learning methods is the inability to interpret the importance of the predictors. Although deep learning models show superior predictive performance compared to traditional machine learning models, the interpretation of the influence of a predictor on rehospitalization has been a black box.

EHR data that is captured and stored in a database has inherent issues such as missing, incorrect, and skewed data. Current LSTM architecture cannot handle missing data unless the data is imputed by some mechanism. A future area of work can be to build an LSTM model that can handle missing data without having to apply any imputation rule.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This study was conducted with the data provided by, and the support from, the Center for Health Systems Innovation (CHSI) at Oklahoma State University (OSU) and the Cerner Corporation. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of CHSI, OSU or the Cerner Corporation.

Conflict of interest statement

None Declared.

References

[1] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. (2017), https://doi.org/10.1016/j.compbiomed.2017.09.017.
[2] R. Amarasingham, P.C. Patel, K. Toto, L.L. Nelson, T.S. Swanson, B.J. Moore, M.H. Drazner, Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study, BMJ Qual. Saf. 22 (12) (2013) 998–1005, https://doi.org/10.1136/bmjqs-2013-002407.
[3] D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, Health Aff. 33 (7) (2014) 1123–1131.
[4] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Network. 5 (2) (1994) 157–166, https://doi.org/10.1109/72.279181.
[5] M. Biswas, V. Kuppili, T. Araki, D.R. Edla, E.C. Godia, L. Saba, H.S. Suri, T. Omerzu, J.R. Laird, N.N. Khanna, A. Nicolaides, J.S. Suri, Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on Japanese diabetic cohort, Comput. Biol. Med. 98 (2018) 100–117, https://doi.org/10.1016/j.compbiomed.2018.05.014.
[6] A. Bordes, S. Chopra, J. Weston Question Answering with Subgraph Embeddings, Cornell University Press, Ithaca NY, 2014.
[7] Centers for Medicare and Medicaid Services, Readmissions Reduction Program, (2015) 6 https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html , Accessed date: 6 April 2017.
[8] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[9] E. Choi, M.T. Bahadori, A. Schuetz, W.F. Stewart, J. Sun, Doctor Ai: predicting clinical events via recurrent neural networks, Proceedings of the 1st Machine Learning for Healthcare Conference, vol 56, 2016, pp. 301–318.
[10] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inf. Assoc. 24 (2) (2016) 361–370, https://doi.org/10.1093/jamai/ocw112.
[11] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural language processing (almost) from scratch, J. Mach. Learn. Res. 12 (Aug) (2011) 2493–2537.
[12] D. Delen, Real-world Data Mining: Applied Business Analytics and Decision Making, FT Press, Upper Saddle River, New Jersey, 2015.
[13] J.P. DeShazo, M.A. Hoffman, A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample, BMC Health Serv. Res. 15 (2015) 384, https://doi.org/10.1186/s12913-015-1025-7.
[14] C.J. Edwards, T.Y. Lian, H. Badsha, C.L. Teh, N. Arden, H.H. Chng, Hospitalization of individuals with systemic lupus erythematosus: characteristics and predictors of outcome, Lupus 12 (9) (2003) 672–676, https://doi.org/10.1191/0961203303lu452oa.
[15] J. Futoma, J. Morris, J. Lucas, A comparison of models for predicting early hospital readmissions, J. Biomed. Inf. 56 (2015) 229–238, https://doi.org/10.1016/j.jbi.2015.05.016.
[16] A. Elixhauser, C. Steiner, Readmissions to U.S. Hospitals by Diagnosis, 2010: Statistical Brief #153. United Kingdom: Europe PubMed Central, (2006).
[17] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge MA, 2016.
[18] N. Hammerla, P. Andras, L. Rochester, T. Ploetz, PD disease State Assessment in Naturalistic Environments Using Deep Learning. Proceedings of the 29th AAAI Conference on Artificial Intelligence, (2015).
[19] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.R. Mohamed, N. Jaitly, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups, IEEE Signal Process. Mag. 29 (6) (2012) 82–97, https://doi.org/10.1109/MSP.2012.2205597.
[20] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[21] R.W. Jakes, S.C. Bae, W. Louthrenoo, C.C. Mok, S.V. Navarra, N. Kwon, Systematic review of the epidemiology of systemic lupus erythematosus in the Asia-Pacific region: prevalence, incidence, clinical features, and mortality, Arthritis Care & Research 64 (2) (2012) 159–168, https://doi.org/10.1002/acr.20683.
[22] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications, Springer Texts in Statistics, New York, 2013.
[23] A.K. Jha, Meaningful use of electronic health records: the road ahead, JAMA, J. Am. Med. Assoc. 304 (15) (2010) 1709–1710, https://doi.org/10.1001/jama.2010.1497.
[24] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, Risk prediction models for hospital readmission: a systematic review, Jama 306 (15) (2011) 1688–1698, https://doi.org/10.1001/jama.2011.1515.
[25] Keras Guiding, Principles, September 3, (2017) https://keras.io.
[26] D.P. Kingma, J.B. Adam, A Method for Stochastic Optimization, (2014) arXiv preprint arXiv:1412.6980 (2015).
[27] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539.
[28] H. Lee, P. Pham, Y. Largman, A.Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, Advances in Neural Information Processing Systems Proceedings, vol 22, 2009, pp. 1096–1104.
[29] Z.C. Lipton, D.C. Kale, C. Elkan, R. Wetzell, Learning to Diagnose with LSTM Recurrent Neural Networks, Cornell University Press, Ithaca NY, 2015.
[30] J. Ma, R.P. Sheridan, A. Liaw, G.E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure-activity relationships, J. Chem. Inf. Model. 55 (2) (2015) 263–274, https://doi.org/10.1021/ci500747n.
[31] C.K. McIlvennan, Z.J. Eapen, L.A. Allen, Hospital readmissions reduction program,

Circulation 131 (20) (2015) 1796–1803, https://doi.org/10.1161/CIRCULATIONAHA.114.010270.
[32] R. Morales-Lozano, J. Martínez-Barrio, M.L. González-Fernández, F.J. López-Longo, J.G. Ovalles-Bonilla, L. Valor, I. Monteagudo, The feet in systemic lupus erythematosus: are we underestimating their involvement and functional impact? Clin. Exp. Rheumatol. 34 (4) (2016) 609–617.
[33] M.P.A. Committee, Report to Congress: Promoting Greater Efficiency in Medicare, (2007).
[34] A.L. Nightingale, R.D. Farmer, C.S. de Vries, Incidence of clinically diagnosed systemic lupus erythematosus 1992–1998 using the UK General Practice Research Database, Pharmacoepidemiol. Drug Saf. 15 (9) (2006) 656–661, https://doi.org/10.1002/pds.1199.
[35] F. Rees, M. Doherty, M.J. Grainge, P. Lanyon, W. Zhang, The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies, Rheumatology 56 (11) (2017) 1945–1961, https://doi.org/10.1093/rheumatology/kex260.
[36] E.C. Somers, W. Marder, P. Cagnoli, E.E. Lewis, P. DeGuire, C. Gordon, W.J. McCune, Population-based incidence and prevalence of systemic lupus erythematosus: the Michigan lupus epidemiology and surveillance program, Arthritis & Rheumatology (Hoboken, N.J.) 66 (2) (2014) 369–378, https://doi.org/10.1002/art.38238.
[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov Dropout, A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[38] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, Proceedings of the 2014 Advances in Neural Information Processing Systems Conference, 2014, pp. 3104–3112.
[39] D. Thammasiri, D. Delen, P. Meesad, N. Kasap, A critical assessment of imbalanced class distribution problem: the case of predicting freshmen student attrition, Expert Syst. Appl. 41 (2) (2014) 321–330, https://doi.org/10.1016/j.eswa.2013.07.046.
[40] C.L. Teh, G.R. Ling, Causes and predictors of mortality in hospitalized lupus patient in Sarawak General Hospital, Malaysia. Lupus 22 (1) (2013) 106–111, https://doi.org/10.1177/0961203312465780.
[41] TensorFlow, About TensorFlow, (2017) https://www.tensorflow.org/.
[42] N. Tomita, Y.Y. Cheung, S. Hassanpour, Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans, Comput. Biol. Med. 98 (2018) 8–15, https://doi.org/10.1016/j.compbiomed.2018.05.011.
[43] J. Yazdany, B.J. Marafino, M.L. Dean, N.S. Bardach, R. Duseja, M.M. Ward, R.A. Dudley, Thirty day hospital readmissions in systemic lupus erythematosus: predictors and hospital- and state-level variation, Arthritis & Rheumatology 66 (10) (2014) 2828–2836, https://doi.org/10.1002/art.38768.
[44] R.B. Zuckerman, S.H. Sheingold, E.J. Orav, J. Ruhter, A.M. Epstein Readmissions, observation, and the hospital readmissions reduction program, N. Engl. J. Med. 374 (16) (2016) 1543–1551, https://doi.org/10.1056/NEJMsa1513024.