Journal of Medical Systems (2019) 43: 111
https://doi.org/10.1007/s10916-019-1243-3

MOBILE & WIRELESS HEALTH

LSTM Model for Prediction of Heart Failure in Big Data

G. Maragatham[1] · Shobana Devi[1]

Received: 30 January 2019 / Accepted: 6 March 2019 / Published online: 19 March 2019

Abstract

The combination of big data and deep learning is a world-shattering technology that can make a great impact on any industry if used in a proper way. With the availability of large volume of health care datasets and progressions in deep learning techniques, systems are now well equipped in diagnosing many health problems. Utilizing the intensity of substantial historical information in electronic health record (EHR), we built up, a conventional predictive temporal model utilizing recurrent neural systems (RNN) like LSTM and connected to longitudinal time stepped EHR. Experience records were contribution to RNN to anticipate the analysis and prescription classes for a resulting visit during heart disappointment (e.g. diagnosis codes, drug codes or method codes). In this paper, we also investigated whether use of deep learning to model temporal relations among events in electronic health records (EHRs) would enhance the model performance in predicting initial diagnosis of heart failure (HF) compared to some of the traditional methods that disregard temporality. By examining these time stamped EHRs, we could recognize the associations between various diagnosis occasions and finally predicate when a patient is being analyzed for a disease. In any case, it is hard to access the current EHR data straightforwardly, since almost all data are sparse and not standardized. Along these lines, we proposed a robust model for prediction of heart failure. The fundamental commitment of this paper is to predict the failure of heart by means of a neural network model based on patient's electronic medicinal information. In order to, demonstrate the diagnosis events and prediction of heart failure, we used the medical concept vectors and the essential standards of a long short-term memory (LSTM) deep network model. The proposed LSTM model uses SiLU and tanh as activation function in the hidden layers and Softmax in output layer in the network. Bridgeout is used as a regularization technique for weight optimization throughout the network. Assessments subject to the real-time data exhibit the favorable effectiveness and feasibility of recommended model in the risk of heart failure prediction. The results showed improved accuracy in heart failure detection and the model performance is compared using the existing deep learning models. Enhanced prior detection could expose novel chances for deferring or anticipating movement to analysis of heart failure and diminish cost.

Keywords LSTM Model · Electronic health record · Recurrent neural systems · Long short-term memory

Introduction

A typical problem in healthcare nowadays is mostly doctors approach large volume of data on patients, however little time or devices. Smart clinical decision help imagines the data at the motivation behind thought that is specific to the patient and needs of supplier. Electronic wellbeing records (EHR), are now standard in U.S. medicinal services, address both the patients and masters longitudinal experience. To predict future events these data are used with expanding frequency. Even though, forecast models are created to envision needs, most of the existing works has focused on specific expectation models that foresee a compelled set of results. In any case, ordinary clinical practice incorporates an unprepared and diverse blend of circumstances and requirements distinctive predictive models in the 100's to 1000's. To create and deploy particular models one after another is unrealistic.

Prior to determination of a disease, a person's progression interceded by pathophysiologic changes recognizes the individuals who will in the long run get the disease from the individuals who won't. Discovery of temporal occasion successions that dependably use controls that recognize disease

This article is part of the Topical Collection on *Mobile & Wireless Health*

✉ G. Maragatham
maragathamhaarish@gmail.com

Shobana Devi
shobanak07@gmail.com

[1] Department of Information & Technology, SRM Institute of Science and Technology, Chennai, India

cases might be especially helpful in enhancing the performance of predictive models. For this reason, we examined whether models built using RNN could be adjusted thus, changing over clinical occasion successions and time stamped related information into trails essential to prior discovery of sickness. Electronic health record (EHR) data contains clinically rich and related time stamped data. Patient medicinal services experiences are all around archived and time stamped (eg, findings, drugs, and techniques). Be that as it may, EHR information are very unpredictable, given the structure also, broadness of data caught (spanning provider conduct, care use, pathways for treatment, and infection state of patient) and unpredictable frequency of sampling. So far, most prescient demonstrating work utilizing EHR information depends on aggregate values of features (e.g., count of event and average value of event). Utilizing these strategies, some of the features that are disaggregated in the temporal relations are most certainly not caught (e.g., at one time medication solution requested and at another time method performed).

Heart failure, likewise named to as congestive heart failure, happens when the heart can't sufficiently direct blood to address the needs of a body [1]. Most of the risk indicators for failure of heart incorporate hypertension, an earlier heart attack, stoutness, smoking, liquor misuse, vitamin lacks, and lack of sleep, overwhelming metal poisonous quality, a non-healthy eating like fats of animals, and being inactive [2]. Failure of heart is more typical in the middle of individuals beyond 65 years old, individuals with more weight, and individuals who had past heart failure. The diagnostic technique for failure of heart is fundamentally constructed using the patient's medicinal and past histories of family, a physical check-up and outcomes of test. Basic symptoms and signs of failure of heart are additionally mutual in different situations. In this way, doctors recognize any harm in patient's heart and diagnosis the pumping of blood in heart of the patient. These symptomatic techniques give enormous data that are sequential, and this is an unimportant job to implement most accurate diagnosis with such data of huge volume especially during the initial times.

In reality, a technique for prior heart failure diagnosis which gives a less error rate is fundamentally required for preliminaries in clinical and medications [3]. Through dissecting these datasets that are sequential, we need a chance to give initial conclusions and treatments to individuals who are probably going to suffer failure of heart and benefit them to have longer, energetic lives. The desired procedure to determine the issues of finding diagnosis with more accuracy and the conveyance of focused treatments gives the regular execution of thorough physical assessments [4]. Patients with heart failure and the society could profit on the off chance that we would give to the patients an accurate and efficient diagnostic service. Finally, this paper builds up another way to deal with this undertaking by means of an improved long short-term memory (LSTM) networks model and information driven structure.

In particular, we give every patient using a dynamic framework that is to be estimated by a time series data, for example, the consequences of various test results of lab, records of patients and markers of medical. Our crucial thought is to investigate these data that are time stamped. The value of a statistical indicator with respect of time that provides a sequence is termed as time series data [5]. A time series data demonstrates the pattern of the numerical estimation and examination protest over a specific period of the statistical index. The conventional expectation strategies dependent on time series fundamentally include the exponential smoothing strategy [6], (ARIMA) autoregressive integrated moving average model [7], (RNN) [8] recurrent neural network model, and (LSTM) long short-term memory model [9].

As of now, the analysts frequently gather the identification events by means of vectors using an unsupervised way. Interestingly, it was tremendously profitable to demonstrate the events of analysis with correspondence learning. We recommend an innovative technique, in this paper for diagnosis event modelling that are constructed using medical concept vectors by improving the one-hot vectors via adding time stamp to each input Xt and utilizes LSTM methodology for prediction of failure of heart using the diagnosis events are modelled as the input. Our experimental outcomes using a real-time data set exhibit the improved diagnosis prediction performance strategy.

The remaining part of the paper is structured as follows. We discussed the work background and related work of the deep learning and analysis of time series using EHR data, in Section 2. Section 3 we explained the projected LSTM model framework, and we experiment the proposed medical concept vector based LSTM model using real-time EHR data in Section 4. In Section 5, we discussed the result analysis of the model and comparison of the performance with the existing conventional deep learning models and In Section 6, we discussed our work conclusion and highlighted the directions for future research.

Related work

Early location of heart disappointment Beginning of HF is related with an abnormal state of inability, health care costs, what's more, mortality (around half danger of five years of mortality analysis). [10, 11] There have been generally slight advancement in moderating the movement of heart failure disease, to a great extent since it is hard to recognize before real diagnosis. As a result, mediation has essentially been bound to the day and age after diagnosis, with next to zero effect on progression of disease. Prior heart failure detection could prompt enhanced results using patient commitment and even more confident treatment by means of angiotensin changing over inhibitors of enzyme or receptor blockers of angiotensin II, gentle work out, lessened intake of salt, and potentially other possibilities. [12–15] Earlier work on prior detection of heart failure has depended on traditional modelling strategies, for example, support vector machine (SVM) or logistic regression, that utilization of features demonstrating to the observation window events collection and reject time stamped relations of observation window between the events. [16–18] Interestingly, recurrent neural system (RNN) techniques catch temporal examples exist in the longitudinal EHR data. Models using RNN have demonstrated successful in numerous troublesome machine learning assignments, for example, picture captioning [19] and dialect translation. [20] Encompassing these strategies to health care data is practical.

Deep learning applications

Recently, Deep learning strategies have directed to the regeneration of models based on neural network. Revolutionary lessons presented loaded restricted Boltzmann machines [21] and loaded autoencoders, [22] that exhibited inspiring results in processing of image, which employs the layer-wise pretraining technique. In the meantime, variants of neural system applications have been investigated deep designs in the domain of computer vision, [23–25] processing of sound, [26, 27] and (NLP) natural language processing, [20, 28–30] among different fields. Models using RNN are normally appropriate to time series information, and a few variations need to be created on behalf of features that are sequenced. Schmidhuber and Hochreiter [31] suggested LSTM (long short-term memory), displaying noteworthy execution in various sequence centred assignments, for example, recognition of handwriting, [32] auditory demonstrating of speech, [33] dialect modelling, [34] and dialect translation. [35] Cho et al. recommended the GRU network model (gated recurrent unit), basically like however less complex than LSTM, and indicated practically identical, and gives better performance. [36] In the work using RNN defined in this, we utilized the LSTM structure in order to show the time series relations amongst health care information from EHRs of patient to anticipate the forthcoming prediction of heart failure.

Applications of deep learning in health care

Research Analysts have lately connected deep learning strategies to medical solicitations. Lasko et al. [37] utilized autoencoders to study phenotypic designs from estimations of serum uric acid. Che et al. [38] utilized deep neural systems on clinical time stamped data using incremental learning to find physiologic examples related by means of clinical known phenotypes. The works, [39, 40] be that as it may, concentrated on taking in patterns from clinical records as opposed to anticipating a clinical events. Hammerla et al. [41] connected restricted Boltzmann machines (RBM) on time stamped data gathered from sensors that are wearable to foresee the sickness province of patients who has Parkinson's disease. Lipton et al. utilized LSTM aimed at prediction of diagnosis with multilabel utilizing ICU time stamped data of paediatric (eg, pulse rate of heart, blood pressure value, level of glucose, and so on.). These two studies [42, 43] utilized multivariate with time stamped data starting patients, that concentrated on altogether diverse clinical situations, with persistent time series information. Our investigation centers on prior identification of heart failure aimed at the common patient population in view of broadly accessible EHR information, for example, codes that are time-stamped (diagnosis codes, drug codes, medication codes, procedure codes). We obtained since this earlier effort to use comparative illustration of clinical ideas over Skip-gram however centre on temporal modelling for predicting heart failure by means of RNN [44–47].

Time series data analysis of EHRs

Conventional time sequence strategies by means of direct prototypes for low dimensional information require to be generally connected towards EHRs: demonstrating the movement of chronic disease of kidney to disappointment of kidney utilizing the Cox relative hazard structure, [48] the progress of Alzheimer's illness by means of the hidden Markov model [49] and intertwined group Lasso, [50] glaucoma progress make use of a continuous time unseen Markov model with two dimensional, [51] the progress of disease in lung is identified by means of Gaussian procedure using graphical models, [52] the progress of obstructive pulmonary chronic illness uses Markov bounce process, [53] and progress of various infections by means of the Hawkes progression. [54] These past workings remained not ready to build model with high dimensional nonlinear associations and recurrent neural network. We concentrated on forecasting the beginning of heart failure uses longitudinal organized patient information (structured), for example, patient diagnosis, medication (drug) codes, and codes of procedure. We applied RNN that gives a non-linear development with model generalization and high scalability compared to a large number of conventional techniques, because of a more upgraded programming software packages and distributed/parallel systems named as a GPU (graphics processing unit).

Recurrent neural network model (RNN)

Through the present expanding computational power influence, deep learning have been utilized to fabricate numerous difficult neural systems, for example, (CNNs) convolutional neural systems [30, 55], (RNNs) recurrent neural systems [56], and deep neural systems (DNNs) [57]. These systems have empowered developments in (NLP) natural language processing, (IR) recognition of image, (SR) recognition of speech and different domains. RNN models are appropriate for managing time-sequence prediction issues. RNN models comprise of a layer for input, number of hidden layers, and one layer for output. The outcome of the hidden layer is identified with the present layer input and the output of previous layer. Utilizing this system, a RNN picks up the capacity to recall historical outcomes. Through exchange among the hidden layers, the data from previous layer is distributed to the following sequence that builds up the relationship over the time sequences.

Long short-term memory network model (LSTM)

In 1997, Hochreiter et al. [58] projected the LSTM display, or, in other words RNN special model. To accomplish long short-term memory, the model RNN desires to snare the condition of the present layer that is hidden to the past hidden layer's n-level state. This outcome gives an exponential increment in the measure of computation that expands the time budget of the model. Along these lines, RNN models are not straightforwardly utilized on behalf of long short memory computations. The LSTM layers are included towards the valve node the premise of the first network model RNN, or, in other words beating the issues of RNN through long short memory estimations. Also, this is the generally used methodology.

LSTM includes three basic gates in addition to essentials of the first network model RNN, first one is an input gate, second is a forget gate, and the third is an output gate. As of late numerous research scientists have rolled out minor improvements towards LSTM network model. The prevalent LSTM variation, presented in Gers et al. [59], includes expansion of "peephole connections" (i.e., we allow all the cell state to glance at the gate layers). Additional bigger variation is the limit unit cyclic network model (GRU) that suggested in Chung et al. [60]. At this time, the single update gate is used by combining the forget gate and input gate of LSTM model. Chung additionally rolls out some different improvements by blending hidden states and the cell states.

Materials and methods

As compared to the previously mentioned techniques, we build up a medical-concept-vector LSTM model that can mutually develop the LSTM framework with a one-hot vector strategy learned with accessible supervised requirements and combining at the end of each one-hot vector the time interval feature. Our work depends on the simple LSTM network model. Here, we had given the points of interest of our proposed methodology in the accompanying sections.

LSTM model for prediction of heart failure

LSTM model on the way to decide if depiction of feature using the time among an occasion and the record date is combined by means of one-hot vectors enhanced the execution of model. In particular, the training of model is done by attaching to the input vector Xt an additional measurement that speaks to the logarithm time of the quantity of days among the occasion time t and T as the record date. In this, we connected logarithm change to limit skewed dispersion of the intervals and investigated the time among sequential visits of patients or without logarithm change, yet the above technique gave the best outcome. (Fig. 1)

The LSTM model formulation:

1. The Silu, tanh and softmax are used as an activation functions in layer network.
2. Gradient descent algorithm is used at each layer to adjust the weights of corresponding layers.
3. Backpropagation algorithm is used in the hidden layer to compute the derivatives of weights.
4. Input node:

   This unit, marked g_c, is a hub that takes actuation in the standard route from the layer of input X(t) at the current time step and using the hidden layer at the past time h(t-1) step. Normally, the total weighted information is gone through a SILU function, in spite of the fact that in the first LSTM paper, the sigmoid activation function is used.

   The initial segment of Fig. 2 (marked ``1″) figures out which data is disposed of after the cell state. Now, ht-1 speaks to the condition of a layer that is concealed at t - 1 min, and xt speaks to the yield at t minute. This choice is implemented by the information door. The entryway peruses the estimations of ht-1 and xt and yields the estimations of 0–1 to the ct-1 state of every cell over the SiLU work. State '1' implies all saved and state '0' implies all disposed of.

$$\boldsymbol{f}_t = SiLU\left(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f\right)$$

5. Input gate:

   Gates are a distinctive feature of the LSTM approach. A gate is a SiLU function unit that, like the input node, takes activation from the current data point X(t) as well as from the hidden layer at the previous time step. A gate is so-called because its value is used to multiply the value of another node. It is a gate in the sense that if its value is zero, then own from the other node is cut off. If the value of the gate is one, all low is passed through. The value of the input gate i_c multiplies the value of the input node.

   The second piece of Fig. 2 (marked ``2″) is utilized to refresh the cell status and incorporates the SiLU function layer and the tanh layer. The SiLU function layer figures

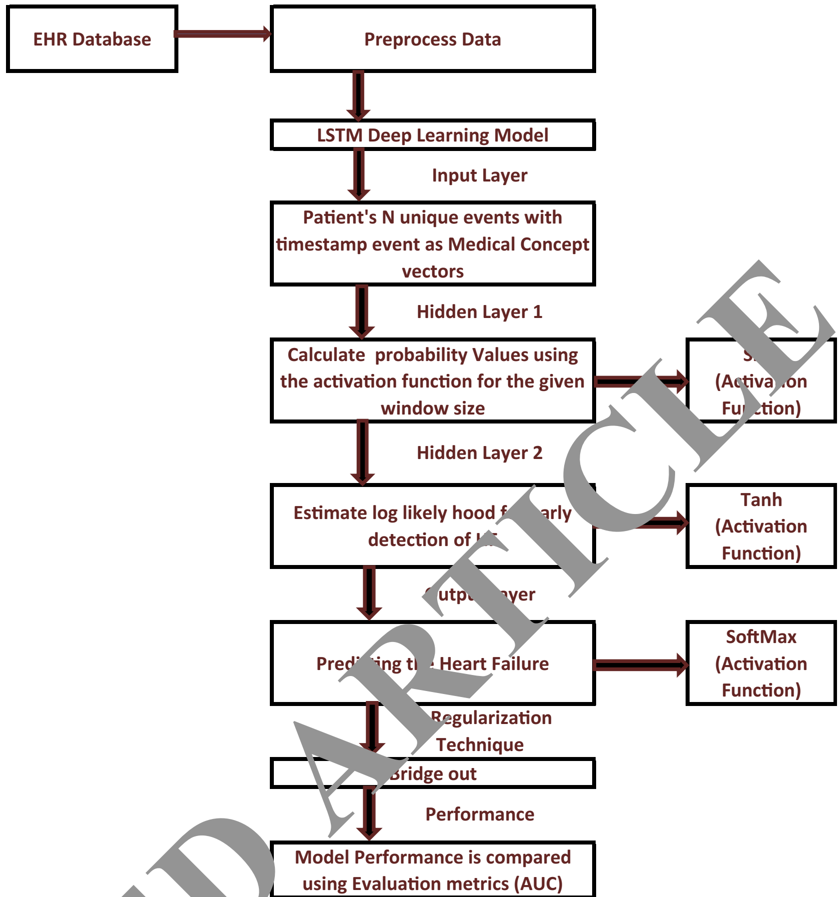

Fig. 1 Flow diagram of proposed model

out what output should be refreshed. The value is made through the tanh layer. It and Qct can be figured from the SiLU layer and the tanh layer.

$$
\begin{aligned}
i_t &= SiLU(\mathbf{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \\
\tilde{C}\sim_t &= \tanh(\mathbf{W}_C \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_C)
\end{aligned}
$$

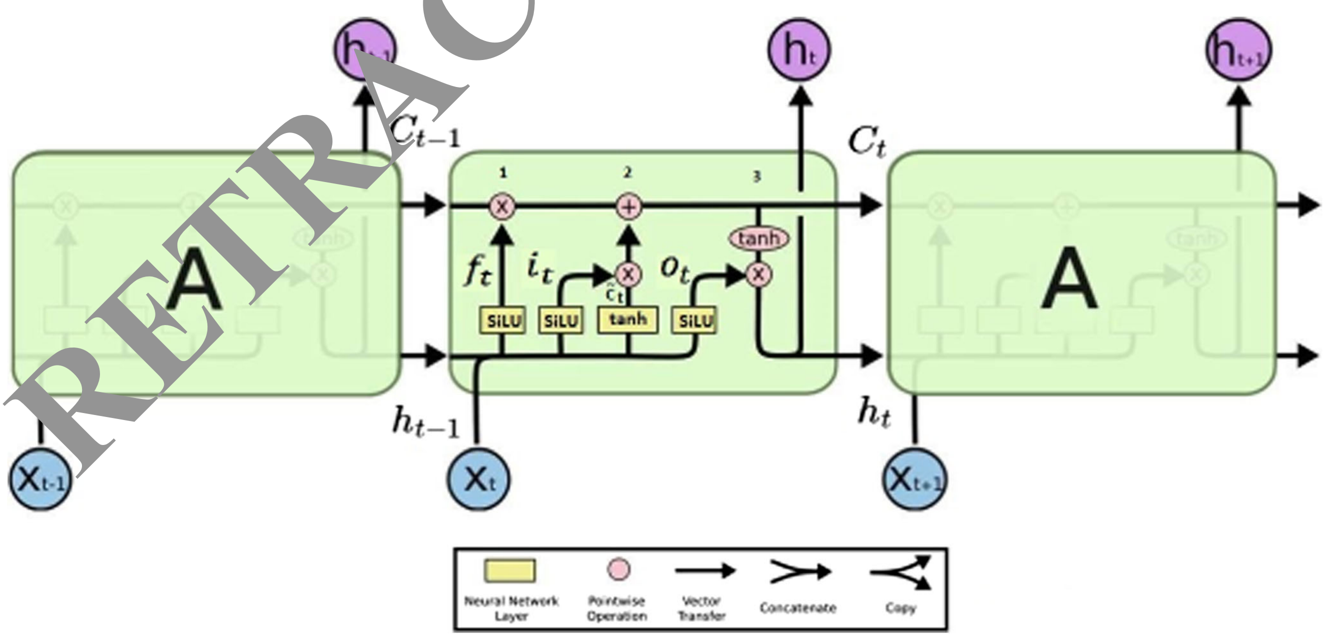

Fig. 2 The LSTM model architecture

6. Internal state:

At the core of every memory cell is a hub Sc with linear actuation, or, in other words in the first paper as the "internal state" of the cell. The internal state Sc has a self-associated repetitive edge with settled unit weight. Since this edge traverses contiguous time steps with consistent weight, error can stream crosswise over time ventures without vanishing. This edge is regularly called the consistent carousel error.

The third part is utilized to refresh the cell state. These parts refresh $\tilde{C}\sim t$ to $\boldsymbol{Ct}$. Next, $\boldsymbol{Ct}$-$\boldsymbol{1}$ and $\boldsymbol{f}_t$ are duplicated, and the data that should be dropped is disposed of. At that point, add $\boldsymbol{i}_t * \tilde{Ct}\sim$, and get the esteem Ct of the new state:

$$C_t = \boldsymbol{f}_t * C_{t-1} + \boldsymbol{i}_t * \tilde{C}\sim_t$$

7. Forget gate:

They give a technique by which the system can figure out how to utilize the substance of the interior state. This is particularly helpful in persistently running systems.

The latter portion is utilized to acquire the value of output. The SiLU activation function is actualized to decide the portion, which should be the value of output. The tanh-treated cell state is (to get in the range of −1 to 1 value) and increased by SiLU output value. At that point, the output cell state ht. is acquired:

$$\begin{aligned} \boldsymbol{o}_t &= \quad SiLU(\mathbf{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \\ \boldsymbol{h}_t &= \quad \boldsymbol{o}_t * \tanh(C_t) \end{aligned}$$

8. Output gate:

The esteem Vc value at last delivered by a memory cell is the multiplication of internal state Sc value and the output gate Oc value. It is standard that the internal state initially be gone through a softmax function, as this gives the output value of every cell indistinguishable powerful range from a conventional softmax hidden unit. Be that as it may, in other neural system examine, ReLU units, which have a more noteworthy unique range, are less demanding to train. Along these lines it appears to be conceivable that the nonlinear activation function on the inner state may be discarded.

Medical concept vectors

In this section, we built medical concept vectors, which are used to enhance the input one-hot vectors Xt. We utilized a NLP inserting method, Skip-gram, to represent the training vectors of codes of procedure, medicine, and diagnosis [20]. The representation of resultant vector, is given in the form of medical concept vector, was used to identify the relations that are hidden of different codes.

The medical concept vectors are trained using sliding window with a size fixed on codes of sequence and the log likelihood maximization at every progression is given as:

$$\text{Maximize} \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \le j \le w, j \ne 0} \log p(c_{t+j} | c_t),$$

$$\text{where } p(c_{t+j} | c_t) = \frac{\exp\left(v(c_{t+j})^{\mathrm{T}} v(c_t)\right)}{\sum_{c=1}^{N} \exp\left(v(c)^{\mathrm{T}} v(c_t)\right)}$$

Where the value T is given as the quantity of codes in the majority of visits of patients, span of the window is given as w, the code at given time position t as ct, and v(ct) is the illustration of vector of the ct code. Basically, Skip-gram endeavours to expand the product of the representation of vector of proximal ideas that are temporal.

The span of the window and the vector dimensionality portrayal are the hyperparameters for most part set to 5 and 50–1000, separately. [28] The medical concept vectors are trained utilizing the experience, drug arrange, method request, and list of issues, using window estimate 5 value and coming about dimensionality value of 100. Another work introduced the application of Skip-gram for medicinal services through an emphasis on understanding of subsequent restorative idea vectors. [43] This examination, then again, centres on the successive idea of the longitudinal data of EHR via utilizing RNN for prior discovery of heart failure.

Table 1 Qualifying ICD-9 codes for heart failure

Description	ICD-9 Code
Heart failure, unspecified	428.9
Acute on chronic combined systolic and diastolic heart failure	428.43
Chronic combined systolic and diastolic heart failure	428.42
Acute combined systolic and diastolic heart failure	428.41
Combined systolic and diastolic heart failure, unspecified	428.4

Data set description

We utilized the EHR information commencing real-time datasets identified with congestive heart illness to conduct experiment. In the principal place, we removed the patient's records that had disease of heart failure disease for over 4 years. Records for the most part incorporate diagnostic and recording events. EHR information on essential consideration patients were removed from experiences happening between June 16, 2001, and June 23, 2014. The dataset reported delivered care in setting of outpatient and incorporated usage of tobacco, demographics and liquor consumption, clinical test and lab test values, The standard International Classification of Disease form 9 codes related by means of experiences, requests, and recommendations, technique data in (CPT) Current Procedural Terminology codes and pharmaceutical solution data in clinical terms. Dataset confined roughly 68,752,005 medicinal codes allotted to the patients. Supplier notes are not utilized in this proposed work, however rather might be incorporated into future exertion.

Definitions of cases and controls of patients

The density sampling has been applied with longitudinal archives of the total population of patients. Vijayakrishnan et al. [57] depicted that, 46 Cases from the population encountered the condition for instance beginning of heart failure and from Gurwitz et al. [58] at the time of HF analysis the patients are at the age of 40–85 years old are belong to Incident cases criteria. Diagnosis of heart failure (HFD) was functionally characterized as trails: (1) for heart failure the Qualifying ICD-9 codes showed up as symptom for medication order. (Sample of ICD-9 codes that are qualified is given in the accompanying Table 1.) (2) At least three clinical experiences using ICD-9 qualifying codes needed to happen within a year, where the diagnosis date was given to the 3 earliest dates. The time range among the first and second visit of the heart failure analysis code were > 12 months, the second appearance date was utilized by way of the most succeeding experience.

The fact upon every patient's course of events on which Heart Failure diagnosis were set up was signified as HFD. Up to ten qualified crucial care details of hospital, matched-age and sex (in five year ranges) codes were chosen for every occurrence of heart failure case, resulting a general proportion of nine controls for each case. Every control remained was additionally included a record date, which was matched with the case of HFD time point. Essential consideration patients were qualified to be controls in the event that they didn't encounter the functioning conditions for heart failure analysis preceding the time of 182 days plus the HFD time point of their comparing case. The records of 18-month time frame before the HFDx are extracted comprising a range that could be divided into two windows named prediction and observation window. For training models, the medical records of observation window were utilized as the dataset. Analysis, medicine, and codes of procedure given to every patient were ordered temporally. At a single visit, the multiple medications were indicated as random ordered multiple one-hot vectors.

Model evaluation

Performance comparison of Baseline models

In addition to our proposed long-short term memory (LSTM) model, we were trained four traditional classification models such as logistic regression, support vector machine (SVM), multilayer perceptron (MLP), and K-nearest neighbour (KNN).

Training strategy

With the dataset created using the records of 18-month time span prior to the HFD time point, all the developed models were trained. Iteratively the data is separated into the train, test, and valid data sets by a proportion of 6:1:1 ratio and the performance of proposed model is reported with the test data set by means of the region beneath the ROC curve (AUC). To calculate the AUC of baseline SVM, the confidence score of SVM is used since, the value of probability can't be computed explicitly. Detailed strategy of training model is given below. (Fig. 3)

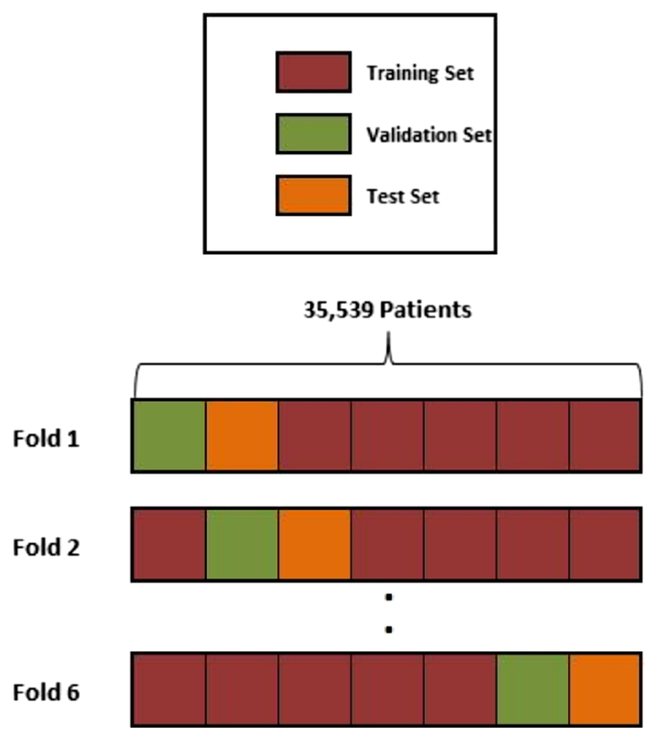

Fig. 3 Diagram of 6-fold cross validation

Input features

In this, we observed the condition of one-hot coded vectors such as sparseness and high-dimensionality could be diminished by means of 2 elective methodologies: The conventional methodology was to utilize regular medical idea groupers. Analysis codes of ICD-9 could be gathered by means of (CCS) Clinical Classification Software grouper determination [59] into the gatherings of 283, pharmaceutical codes might be assembled using Generic Product Identifier [60] into the gatherings of 96, and strategy codes of CPT could be assembled via CCS grouper strategy [61] into the gatherings of 244 nos. On the other hand, medicinal idea vectors dependent on the Skip-gram could be utilized to catch relations between procedures, medications, and diagnoses. Training information for the two LSTM models with window time and without window time duration data are built in a similar style for each of the three kinds of vector input (grouped codes, medical concept vectors, and one-hot encoding dependent using Skip-gram).

Using theses 3 kinds of input vectors, the other baseline models such as MLP, Logistic regression, KNN and SVM were also trained. In the observation window, each dimension of the collected one-hot input vector denotes the aggregate number of events for a particular code. Medical concept vectors and grouped code vectors are produced in a similar manner, with the exception of the one-hot vectors. Normalization of all aggregated three input vectors are done with unit variance and zero mean.

Evaluation strategy

The developed model's utility is connected, towards some extent for application how much data are required and for future how far an accurate forecast can be done. We directed experiments to inspect performance of model for prediction window of varying lengths (i.e., time prior HFD) and the observation window (i.e., before the start of the prediction window what is the length of time), where using the already defined observation window features were extracted (Figs. 4a and b). Note that for every observation window size we trained separate LSTM models and baseline models with the goal, by which the models might acquire in ideal features using the records of patients of various sizes.

Implementation details

The LSTM network, MLP, and logistic regression models are implemented using Theano 0.7. For training the model Adam [62] was used, since it was not based on the setting of the learning rate strongly. Using the Python Scikit-Learn version 0.16.1 the techniques KNN and SVM were implemented. To train all models a machine with Ubuntu Xeon E5-2697, memory of 1 TB, and NVIDIA K80 Tesla also used. Hyper parameters required for training every model were defined in the following Table 2.

Experiment results and discussion

Data processing

From arbitrary samples of 365,446 patients, incident 4289 cases of heart failure and 30,249 patients controls was recognized. The normal clinical codes number given out to every patient were around 82, and from that 19,192 are distinctive clinical codes (6950 determination, 6997 drug, and 4484 technique codes) altogether. The complete example of 365,446 were utilized for medical idea vectors training, and heart failure occurrence cases and the patient's controls was utilized for remaining all models training and evaluations of task.

Execution performance of diagnosis of HF prediction models

In Fig. 5, for all models the average values of AUC of 6-fold cross-validation are shown. In which with the dataset generated with time span of 12-month observation window all LSTM models and baseline models are trained and tested

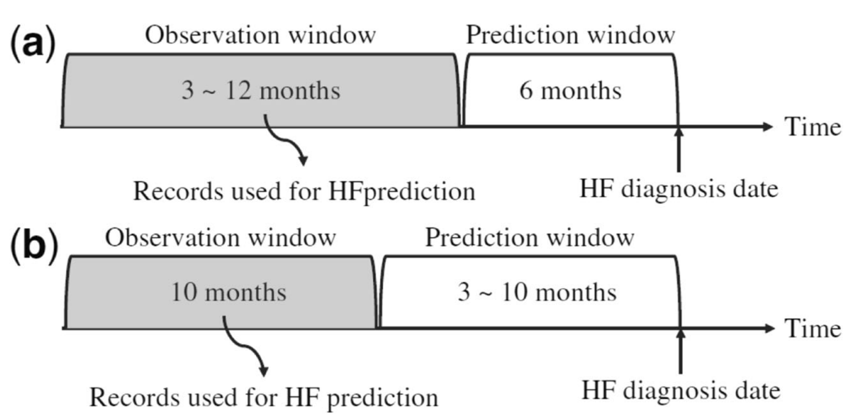

Fig. 4 Using the varied length of the both observation window and prediction window two experiments were conducted. In **a**, the 6 months fixed observation and prediction window of varied length. In **b**, the 9 months fixed observation window length the observation length was fixed and the prediction window of varied length

Table 2 Hyper-parameter settings for training the models

Model	Hyper-Parameter
KNN, One-hot vectors	Number of neighbors: 15
KNN, Grouped code vectors & Medical concept vectors	Number of neighbors: 100
Logistic Regression, One-hot vectors	Bridgeout: 0.1, Max epoch: 100
Logistic Regression, Grouped code vectors & Medical concept vectors	Bridgeout: 0.01, Max epoch: 100
SVM, One-hot vectors	Bridgeout: 0.000001, Dual: False
SVM, Grouped code vectors & Medical concept vectors	Bridgeout: 0.001, Dual: False
MLP, One-hot vectors	Bridgeout: 0.01, Hidden layer size: 15, Max epoch: 100
MLP, Grouped code vectors & Medical concept vectors	Bridgeout: 0.001, Hidden layer size: 100, Max epoch: 100
All LSTM models	Bridgeout: 0.001, Hidden layer size: 100, Max epoch: 100

separately. The colours used in Fig. 5 indicate to various error bars specify values of standard deviation got from the cross validation and the values of training input vectors. As shown in Fig. 5, LSTM models outperformed well when compared to other models.

The LSTM model show results in Fig. 5 characterize to cutting edge execution performance in prediction attained by implementing deep learning techniques to find complex associations inside EHR data. When compared to all the assessed models, the better execution was accomplished with LSTM that utilized time stamped information and trained using medical concept vectors. LSTM models additionally consistently performed well with conventional machine learning models which depend on total set of features and they are trained using grouped code vectors and one-hot vectors (Fig. 5).

Nonetheless, the trained benchmark models by means of medical concept vectors demonstrated practically identical execution performance with the trained LSTM models using one-hot vectors. The outcome proposes a medical concept vectors potential advantage, or, in other words when decent domain ontology is missing. Strangely, the performance execution improvement for LSTM models accomplished by utilizing interval data was uncertain. This might be because of the unpredictable visits of patients' pattern to hospital. In spite of the fact that applying logarithm transformation enhanced the execution slightly, the inborn inconsistency of the pattern visit appears to create it troublesome for the LSTM to study interval features that are predictive.

As presented in the Fig. 6, the observation window size is increased to 18 months of full history of the patient and size zero of prediction window. For this new window size, all models were tested and trained once again. The LSTM model outperformed consistently when compare to the other all methods, with 0.894 AUC. As compared to the trained models using one-hot vectors and trained models via means of grouped code vectors, the trained models by medical concept vectors were performed well significantly.

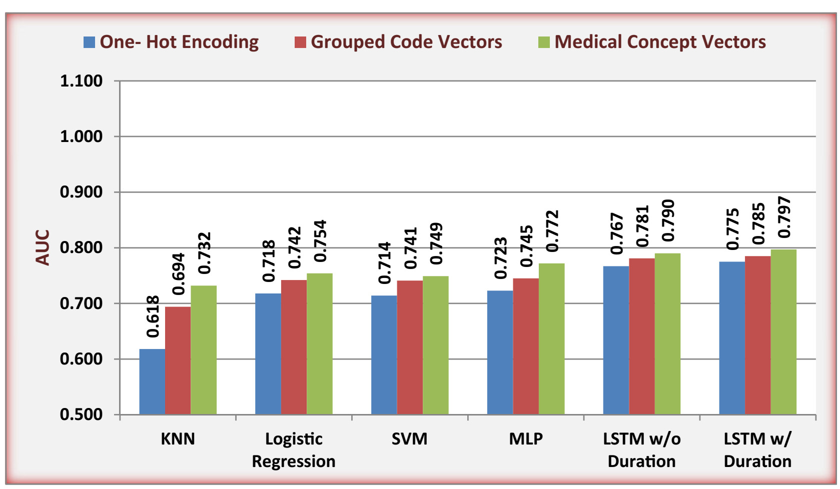

Fig. 5 The performance comparison of baseline and LSTM models for prediction of Heart failure. Using observation window of 12-months and prediction window of 6 months the dataset created and the trained and tested results of all models are given

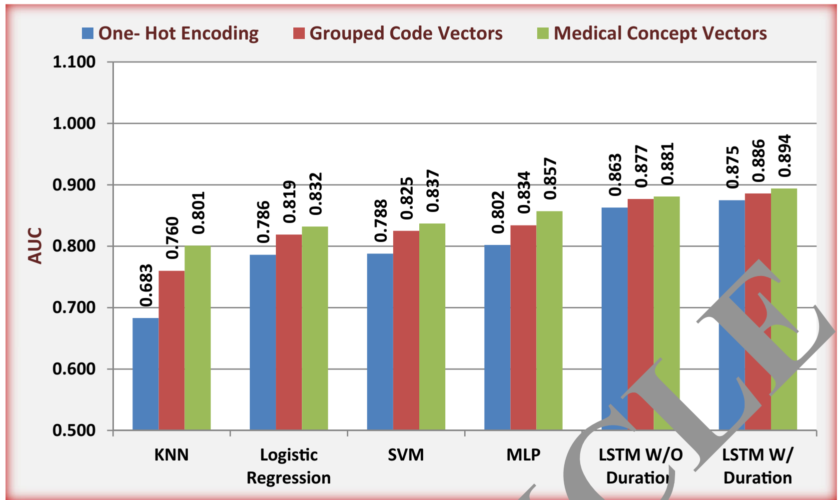

Fig. 6 The performance comparison of baseline and LSTM models for prediction of Heart failure. Using observation window of 18-months and prediction window of 0-month the dataset created and all models were trained and tested

Prediction power and observation/prediction length

The following Tables 3 and 4 demonstrate the AUCs cross-validation values coming after experiments depicted in the Fig. 4a and b. Both the tables demonstrate the model LSTM performed well with other traditional models in all prediction and observation sizes of window.

From Table 3, we understand the model LSTM performs well when compared to other standard models irrespective of the measure of observation window. It appears that long period value of the observation window, the all models performed more effectively. The LSTM model delivered the peak AUC value when the prediction windows were diminished to size of zero, and entire data was utilized on behalf of the observation

Table 3 AUC values of all models by varying the length of observation window

Observation Window (months)	KNN	Logistic Regression	SVM	MLP	LSTMw/duration
3	[illegible]	0.7210	0.7192	0.7312	0.7596
4	0.7084	0.7272	0.7256	0.7427	0.7617
5	0.7154	0.7314	0.7297	0.7455	0.7669
6	0.7165	0.7344	0.7327	0.7486	0.7722
7	0.7175	0.7388	0.7353	0.7530	0.7749
8	0.7213	0.7422	0.7398	0.7587	0.7823
9	0.7234	0.7441	0.7420	0.7603	0.7856
10	0.7261	0.7481	0.7434	0.7631	0.7902
11	0.7292	0.7522	0.7432	0.7679	0.7943
12	0.7319	0.7539	0.7489	0.7719	0.7969

Table 4 AUC values of all models by varying the length of prediction window of 3–9 months

Prediction Window (months)	KNN	Logistic Regression	SVM	MLP	LSTM w/ duration
3	0.7373	0.7501	0.7479	0.7650	0.7916
4	0.7313	0.7473	0.7459	0.7632	0.7886
5	0.7302	0.7458	0.7434	0.7620	0.7845
6	0.7284	0.7441	0.7420	0.7603	0.7821
7	0.7269	0.7426	0.7405	0.7587	0.7785
8	0.7247	0.7396	0.7376	0.7569	0.7759
9	0.7206	0.7364	0.7354	0.7528	0.7738

Table 5 For a single patient the Prediction times of all models

Performance Metric	KNN	Logistic Regression	SVM	MLP	LSTM
Prediction Time (seconds)	36.66	0.000002	0.000034	0.000259	0.024750

window. Although this outcome is fascinating in the methodological viewpoint, such a show has constrained clinical effectiveness since it doesn't forecast the heart failure prior to doctor.

We could realise using Table 4, all the models implemented are performed somewhat better while predicting cases of heart failure that are near-future. However, the size of prediction window was not appearing to influence the model performance prediction as compared to the size of observation window. It recommends access to interval time of history of patient is essential for accurate heart failure prediction, maybe since symptoms of incident sickness apparent over a period of time frame.

Prediction time

The values in Table 5 demonstrate the time necessary for single patient to create a prediction for every model. The trained models by means of medical concept vectors are used in this table. From the 6-folds cross-validation using only the CPU on the test data sets the time took to make predictions was computed using the other models averaging times.

Aside after KNN that needs computation of distance among all other points of data focuses on time of prediction LSTM, actuality a consecutive model, needs the lengthiest prediction time for a single patient, as can be realized in Table 5. In any case, it is as yet ready to calculate the risk of HF in 200 s for one million patients utilizing 100 CPUs, so price would not represent an issue for real-time medical application.

Scalability of our approach

The Fig. 7 delineates the time taken for training of two diverse LSTM models on behalf of fluctuating patient's count, with most extreme gradient descent value of epoch set to 10.

The LSTM model utilizing one-hot vectors showed a high linear relationship among the number of patients and training time, as appeared in Fig. 7. Then again, the LSTM model by medical concept vectors demonstrated increment in linear fashion. We ascribe this to two reasons. The main reason is, medical concept vectors are with dimensions of fixed size, when compared to one-hot vectors in which size of dimension increases if fresh patients' data are included with training data set. Meanwhile new coming patients may require new procedure, medication and diagnoses codes. Second reason is, one-hot vectors were basically with high dimension, and in this manner consumes a lot of VRAM that prompts to repeated collection of garbage. In spite of the fact that the one-hot vectors training time of the LSTM models carries on close linear relationship after the patients count reached 20,000, it is as yet negative with medical concept vector scenario. Hence, we suggest usage of medical concept vector for LSTM model training, as it improves together the performance of model and training time is reduced significantly.

The LSTM model that incorporated temporal relations that are trained using medical concept vectors gave the better performance when compared to all models using real-time situation, notwithstanding, the execution with LSTM model might be distinctive relying upon the idea of the real cohort. In spite of the fact that current work concentrated on heart failure, our proposed methodology is common and might be connected to an extensive cluster of health – connected prediction issues. Further, to encode medicinal data we used the medical concept vectors that were appeared to commonly enhance the execution of together deep learning and traditional models and might has usefulness in various health care related applications anywhere rich data should be represented concisely.

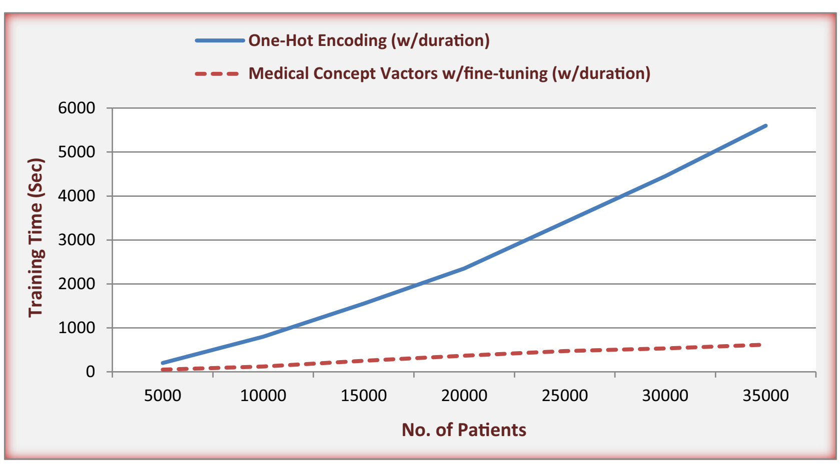

Fig. 7 Number of Patients vs. Training Time (sec)

Conclusion and future work

We have been proposed a novel LSTM basis predictive model framework to early diagnosis of heart failure by using the methods of deep learning. In predicting HF diagnosis, LSTM models (with and without window duration) showed a superior performance when compared to common methods as like KNN, logistic regression, SVM and MLP. By examining the outcomes, we have defined about the significance of the concerning the sequential nature of medical records. The future works would be comprised about the integrating all the expert knowledge's into my proposed model as well as escalating my methodology to other application of health care domains. It will be focused on calculating performance model for diseases prediction windows after 10 months, which may be producing good models for medical inference with better performance, and with added features of hi-level as well as better medical ontologies. Alternative future enhancement will have to be considering using distinct models for HF predictions to the various groups of diseases, such as diabetes and hypertension, which could be possibly can be added discriminative as part of the cohort is used to discover inter-relations among the diseases. Before, start developing distinct RNN models for every disease group, we should ensure an adequate size of sample is obtainable. Another possible research direction, the visualization of RNN models using temporal dynamics, in which presently limited work has been attempted. [63]

Compliance with ethical standards

Conflict of interests The authors declare that this article content has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Huang, H. et al., Uric acid and risk of heart failure: A systematic review and meta-analysis. Eur. J. Heart Fail. 16(1):15–24, 2014.
2. Ford, I. et al., ``Top ten risk factors for morbidity and mortality in patients with chronic systolic heart failure and elevated heart rate: The SHIFT risk model. Int. J. Cardiol. 184:163–169, 2015.
3. Choi, E., Schuetz, A., Stewart, W. F., and Sun, J., ``Using recurrent neural network models for early detection of heart failure onset. J. Amer. Med.Inform. Assoc. 24(2):361–370, 2016.
4. Hripcsak, G., and Albers, D. J., ``Next-generation phenotyping of electronic health records. J. Amer. Med. Inform. Assoc. 20(1):117–121, 2012.
5. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M., Time Series Analysis: Forecasting and Control. Hoboken: Wiley, 2015.
6. Bianchi, F. M., De Santis, E., Rizzi, A., and Sadeghian, A., ``Short-term electric load forecasting using echo state networks and PCA decomposition. IEEE Access 3:1931–1943, 2015.
7. Pati, J., Kumar, B., Manjhi, D., and Shukla, K. K., ``A comparison among ARIMA, BP-NN, and MOGA-NN for software clone evolution prediction. IEEE Access 5:11841–11851, 2017.
8. Su, Y.-T., Lu, Y., Chen, M., and Liu, A.-A., Spatiotemporal joint mitosis detection using CNN-LSTM network in time-lapse phase contrast microscopy images. IEEE Access 5:18033–18041, 2017.
9. Zhu, G., Zhang, L., Shen, P., and Song, J., ``Multimodal gesture recognition using 3-D convolution and convolutional LSTM. IEEE Access 5:4517–4524, 2017.
10. Roger, V. L., Weston, S. A., Redfield, M. M. et al., Trends in heart failure incidence and survival in a community-based population. JAMA. 292(3):344–350, 2004.
11. Murphy, S. L., Xu, J., and Kochanek, K. D., Deaths: final data for 2010. Natl Vital Stat Rep. 61(4):1–117, 2013.
12. Arnold, J., Yusuf, S., Young, J. et al., Prevention of heart failure in patients in the Heart Outcomes Prevention Evaluation (HOPE) study. Circulation. 107(9):1284–1290, 2003.
13. Sciarretta, S., Palano, F., Tocci, G., Baldini, R., and Volpe, M., Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. Arch. Intern. Med. 171(5):384–394, 2011.
14. Wang, C.-H., Weisel, R., Liu, P., Fedak, P., and Verma, S., Glitazones and heart failure critical appraisal for the clinician. Circulation. 107(10):1350–1354, 2003.
15. Wang Y, Ng K, Byrd R, et al., Early detection of heart failure with varying prediction windows by structured and unstructured data in electronic health records. In IEEE Engineering in Medicine and Biology Society. 2530–2533, 2015.
16. Sun J, Hu J, Luo D, et al., Combining knowledge and data driven insights for identifying risk factors using electronic health records. In American Medical Informatics Association. 901–910, 2012.
17. Wu, J., Roy, J., and Stewart, W., Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Med. Care 48(6):S106–S113, 2010.
18. Karpathy A, and Li, F., Deep visual-semantic alignments for generating image descriptions. Computer Vision and Pattern Recognition (CVPR), pp. 3128–3137. Boston, 2015.
19. Cho K, Van Merrienboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734. Doha, 2014.
20. Hinton, G., Osindero, S., and Teh, Y.-W., A fast learning algorithm for deep belief nets. Neural Comput. 18(7):1527–1554, 2006.
21. Bengio, Y., Learning deep architectures for AI. Foundations Trends Machine Learning. 2(1):1–127, 2009.
22. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NIPS), pp. 1106–1114. Lake Tahoe, 2012.
23. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In International Conference on Machine learning (ICML), pp. 1096–1103. Helsinki, 2008.
24. Le Q, Ranzato M, Monga R, et al. Building high-level features using large scale unsupervised learning. In International Conference on Machine Learning (ICML), Edinburgh, 2012.
25. Lee, H., Pham, P., Largman, Y., and Ng, A., Unsupervised feature learning for audioclassification using convolutional deep belief networks. In Advances in Neural Information Processing Systems (NIPS), pp. 1096–1104.Vancouver, 2009.
26. Hinton, G., Deng, L., Yu, D. et al., Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. Signal Process Mag. 29(6):82–97, 2012.

27. Mikolov, T., Chen, K., Corrado, G., and Dean, J., Efficient estimation of word representations in vector space. In arXiv preprint arXiv:1301.3781. 2013.
28. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J., Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems (NIPS), pp. 3111–3119. Lake Tahoe, 2013.
29. Socher R, Pennington J, Huang E, Ng A, Manning C. Semi-supervised recursive autoencoders for predicting sentiment distributions. In Empirical Methods in Natural Language Processing (EMNLP), pp. 151–161. Edinburgh, 2011.
30. Hochreiter, S., and Schmidhuber, J., Long short-term memory. Neural Comput. 9(8):1735–1780, 1997.
31. Grosicki, E., El Abed, H., ICDAR 2009 handwriting recognition competition. In International Conference on Document Analysis and Recognition, pp. 1398–1402. Barcelona, 2009.
32. Sak, H., Senior, A., and Beaufays, F., Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In International Speech Communication Association, pp. 338–342. Singapore, 2014.
33. Zaremba, W., Sutskever, I., and Vinyals, O., Recurrent neural network regularization. In arXiv preprint arXiv:1409.2329, 2014.
34. Luong, M.-T., Sutskever, I., Le, Q., Vinyals, O., Zaremba, W., Addressing the rare word problem in neural machine translation. In Association for Computational Linguistics (ACL), pp. 11–19. Beijing, 2015.
35. Jozefowicz, R., Zaremba, W., and Sutskever, I., An empirical exploration of recurrent network architectures. In International Conference on Machine Learning (ICML), pp. 2342–2350. Lille, 2015.
36. Lasko, T., Denny, J., and Levy, M., Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PLoS One 8(6):e66341, 2013.
37. Che Z, Kale D, Li W, Bahadori M, Liu Y. Deep computational phenotyping. In Knowledge Discovery and Data Mining (KDD), pp. 507–516. Sydney, 2015.
38. Hammerla N, Fisher J, Andras P, Rochester L, Walker R, Plotz T. PD disease state assessment in naturalistic environments using deep learning. In AAAI, pp. 1742–1748. Austin, 2015.
39. Lipton, Z., Kale, D., Elkan, C., Wetzell, R., Learning to diagnose with LSTM recurrent neural networks. In arXiv preprint arXiv:1511.03677, 2016.
40. Minarro-Gimenez, J., Marin-Alonso, O., and Samwald, M., Exploring the application of deep learning techniques on medical text corpora. Stud Health Technol Inform. 205:584–588, 2013.
41. De Vine, L., Zuccon, G., Koopman, B., Sitbon, L., and Bruza, P., Medical semantic similarity with a neural language model. In International Conference on Information and Knowledge Management (CIKM), pp. 1819–1822. Shanghai, 2014.
42. Choi, Y., Chiu, C., and Sontag, D., Learning low-dimensional representations of medical concepts. San Francisco: American Medical Informatics Association on Clinical Research Informatics, 2016.
43. Choi, E., Schuetz, A., Stewart, W., Sun, J., Medical concept representation learning from electronic health records and its application on heart failure prediction. In arXiv preprint arXiv:1602.03686, 2016.
44. Tangri, N., Stevens, L., Griffith, J. et al., A predictive model for progression of chronic kidney disease to kidney failure. JAMA. 305(15):1553–1559, 2011.
45. Sukkar, R., Katz, E., Zhang, Y., Raunig, D., and Wyman, B., Disease progression modeling using hidden Markov models. Engineering in Medicine and Biology Society.:2845–2848, 2012, 2012.
46. Zhou, J., Liu, J., Narayan, V., and Ye, J., Modeling disease progression via multi-task learning. NeuroImage. 78:233–248, 2013.
47. Liu, Y.-Y., Ishikawa, H., Chen, M., Wollstein, G., Schuman, J., Rehg, J., Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden Markov model. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 444–451. Nagoya, 2013.
48. Schulam, P., Saria, S., A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. In Advances in Neural Information Processing Systems (NIPS), pp. 748–756. Montreal, 2015.
49. Wang, X., Sontag, D., and Wang, F., Unsupervised learning of disease progression models. In Knowledge Discovery and Data Mining (KDD), pp. 85–94. New York, 2014.
50. Choi, E., Du, N., Chen, R., Song, L., and Sun, J., Constructing disease network and temporal progression model via context-sensitive Hawkes process. In International Conference on Data Mining (ICDM), pp. 721–726. Atlantic City, 2015.
51. Zhang, G. P., ``Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175, 2003.
52. Jenkins, G. M., and Alavi, A. S., Some aspects of modelling and forecasting multivariate time series. J. Time Ser. Anal. 2(1):1–47, 1981.
53. Brown, R. G., ``Exponential smoothing for predicting demand. Oper. Res. 5:145, 1957.
54. Box, G. E. P., Jenkins, G. M., and Reinsel, G. C., *Linear Nonstationary Models* Time Series Analysis. 4th edition. Hoboken: Wiley, 1976, 93–136.
55. Gers, F. A., and Schmidhuber, J., Recurrent nets that time and count. Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN) 3:189–194, 2000.
56. Chung, J., Gulcehre, C., Cho, K., and Bengio, Y., Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. Available: https://arxiv.org/abs/1412.3555
57. Vijayakrishnan, R., Steinhubl, S., Ng, K. et al., Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. J. Card. Fail. 20(7):459–464, 2014.
58. Gurwitz, J., Magid, D., Smith, D. et al., Contemporary prevalence and correlates of incident heart failure with preserved ejection fraction. Am. J. Med. 126(5):393–400, 2013.
59. Clinical Classifications Software (CCS) for ICD-9-CM. Agency for Healthcare Research and Quality. https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed April 2016.
60. Medi-Span Electronic Drug File (MED-File) v2. Wolters Kluwer Clinical Drug Information. http://www.wolterskluwercdi.com/drug-data/medi-span-electronic-drug-file/. Accessed April 2016.
61. Clinical Classifications Software for Services and Procedures. Agency for Healthcare Research and Quality. https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp. Accessed April 2016.
62. Zeiler, M., ADADELTA: An adaptive learning rate method. In arXiv preprint arXiv:1212.5701, 2012.
63. Karpathy, A., Johnson, J., and Li, F., Visualizing and understanding recurrent networks. In arXiv preprint arXiv:1506.02078, 2015.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.