IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 12, 2019, accepted June 30, 2019, date of publication July 11, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928363

Deep Learning for Electronic Health Records Analytics

GASPARD HARERIMANA, (Student Member, IEEE), JONG WOOK KIM, (Member, IEEE), HOON YOO, AND BEAKCHEOL JANG, (Member, IEEE)

Department of Computer Science, Sangmyung University, Seoul 03016, South Korea

Corresponding author: Beakcheol Jang (bjang@smu.ac.kr)

This work was supported by the National Research Foundation of Korea Grant funded by the Korea Government under Grant NRF-2019R1F1A1058058 and Grant NRF-2018R1E1A2A02058292.

ABSTRACT Recent technological advancements have led to a deluge of medical data from various domains. However, the recorded data from divergent sources comes poorly annotated, noisy, and unstructured. Hence, the data is not fully leveraged to establish actionable insights that can be used in clinical applications. These data recorded in hospital's Electronic Health Records (EHR) consists of patient information, clinical notes, charted events, medications, procedures, laboratory test results, diagnosis codes, and so on. Traditional machine learning and statistical methods have failed to offer insights that can be used by physicians to treat patients as they need to obtain an expert opinion assisted features before building a benchmark task model. With the rise of deep learning methods, there is a need to understand how deep learning can save lives. The purpose of this study was to offer an intuitive explanation for possible use cases of deep learning with EHR. We reflect on techniques that can be applied by health informatics professionals by giving technical intuitions and blue prints on how each clinical task can be approached by a deep learning algorithm.

INDEX TERMS Electronic health records, convolutional neural networks, recurrent neural networks, adverse drug events, EHR raw features.

I. INTRODUCTION

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 raised an increase in the adoption of Electronic Health Records (EHR) [1] by hospitals. Hospitals and other points of care have diversified their efforts in constructing robust Electronic Health Records facilities to capture and leverage these data which are usually ill-understood. Currently there is a high ubiquity of health raw data mainly caused by the abundance of state-of-the-art clinical testing devices and medical Internet of Things (mIoT) [2]. This opportunity is a milestone to healthcare and there is undoubted belief that precision and personalized healthcare will be boosted. EHRs contains highly multidimensional, heterogeneous, multimodal, irregular, time series data like laboratory test results, doctor notes, medication prescriptions, demographic information, diagnoses, epidemiology, behavioral data, etc. With these vast data the clinical tasks can range from critical care to long term planning. Data in EHR can help into a choice of treatment, finding patient similarity, integrating genomics data for personalized treatment, predicting the hospital Length of Stay (LoS), and predicting patients risks of readmission. However, due to this high heterogeneity there is a high probability of missing or erroneous entries resulting into high reluctance by practitioners in using these usually expensive technologies, mainly because they still need to use abductive reasoning in getting clinical insights from them to perform effective diagnosis.

Though hospitals have effectively used the EHR for other administrative and corporate tasks like patients logging, assets management, transfers management, and mainly billing operations, there is a need to find ways to effectively use the EHR for patient's diagnosis. The only solution to this is the use of EHR analytic solutions that will support the physician's expertise. With the recent achievements of artificial intelligence, machine learning methods ranging from simple regression to complex Recurrent Neural Networks (RNN) can be used to bridge the inferential gap for

The associate editor coordinating the review of this manuscript and approving it for publication was Ziyan Wu.

TABLE 1. Studies that covered deep learning application to HER.

Study	year	approach
Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis [5]	2018	This study covered the aspects of EHR data, recent deep learning in healthcare projects. The paper describes some of deep learning algorithms used in health informatics, and some of the clinical tasks that leverage deep learning methods. The paper does not provide a technical detailed implementation.
Deep Learning for Health Informatics [6]	2017	This study covered different deep learning architectures used in health informatics. They also covered various software packages that are used to provide deep neural networks implementations. Mainly the study covered various clinical tasks that deep learning is used into like Medical imaging, Pervasive Sensing etc..
Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review [7]	2018	A literature review which covered clinical analytics tasks that leverage EHR data. It covers also various deep algorithms that are famously used in health informatics.
Deep learning in pharmacogenomics: from gene regulation to patient stratification [8]	2018	This study also covered opportunities and challenges of deep learning in healthcare. They focused mainly on issues of pharmacogenomics and drug targets.
current Study: Deep learning for Electronic Health Records Analytics.	2019	We describe the challenges of deep learning to EHR data. The focus of the paper is to provide technical intuitions on the utilization of deep learning algorithms to each task of EHR clinical knowledge discovery. We try to cover each task that clinicians perform on daily basis. This will help EHR based application developer to consider a task-oriented approach than a data oriented approach in analytics development.

various EHR tasks. However, various complex challenges to integrate them coupled with the limited availability of labeled data for training models as well as privacy issues associated with mistrusts between providers, hinder the effective use of these learning systems to achieve effective care. Though deep learning techniques are highly regarded as crosscutting novelty, there are still tasks in the EHR that can be efficiently be solved by classical machine learning techniques like regression, random forests and Bayesian techniques. Machine learning has empowered the newest methods like computational phenotyping in medical care as well as integrating genomics data into clinical procedures.

A. MOTIVATION FOR THIS WORK

Mining the EHR longitudinal data for clinical insights is a tiresome aspect of building health analytic solutions. Hospitals use customized EHRs which are comprised of heterogeneous mix of elements many of whom are voluminous and unstructured content. The noisiness and sparsity of the EHRs requires effective feature extraction and phenotyping before extracting insights from the data. Though there are various works done to explore methods used to mine data from EHR, there is a need to understand the EHR data mining from an aggregation point of view. For example, adverse event prediction, a process intended to find impending risk of a hospitalized patient can be performed by aggregating insights from doctor's notes (unstructured text data), MRI test (image data), ICD-10 nomenclature database (structured Text data) etc. Hence, this process needs analytic solution to aggregate insights from these diverging data. In this paper we intend to help EHR analytic designers to use deep learning technique for effective analytic techniques to be included in Clinical Decision Support System (CDSS) by tipping them with techniques and mechanisms to extract, transform, load, and leverage disparate EHR data.

B. ORGANIZATION OF THE PAPER

This paper is organized as follows: In Section II, we cover the related works to review approaches used by various authors in coming closer to providing a concise insight of using deep learning methods to EHR. In section III, we cover the anatomy and structure of EHR data using example from a real EHR database exploring its various aspect unraveling hidden patterns. In Section IV, we cover the challenges that an EHR analytics designer is likely to face. In Section V, we try to tip developers by give a glance of techniques per clinical task by covering a successful case study. In Section VI, we conclude by giving future directions.

C. METHODOLOGY

We considered hospital's workflow by covering clinical tasks that can be performed by clinicians. For each task we give insights about the type of EHR data that can be used. We show the challenges associated by each task, and we give a blue print of an appropriate deep learning model by either analyzing an already made model or proposing how it can be designed to produce the required insights. Our approach is to answer the question "how did they do it?" wherever there

TABLE 2. Anatomy of EHR data (With Intuitions from the MIMICIII data set).

Type of Data	Data source	purpose of data	Example entries	deep learning task contribution
Patient Demographic information	Hospital database	The background information Including name (de-identified for privacy), age, allergies, and any other demographic data.	Admission_type: Emergency Admission location: Emergency Room: Admit Discharge location: DISCTRAN CANCER/CHLDRN H Diagnosis: BENZODIAZEPINE OVERDOSE	▪ Adverse events Detection [15] ▪ Clinical trial recruitment
Clinical notes/manuscripts	Notes filled by physicians	describe patients' conditions during hospitalization or during discharge.it contains prescriptions, reactions and adverse conditions.	See Fig. 1	▪ Chronic Disease Prediction [16] [17] ▪ Medication information extraction [18] ▪ disease phenotyping [19]
Lab tests and mIoT Events	Hospital database	recording all laboratory measurements based on their LOINC codes defined as defined in another table	2160-0: Creatinine, blood 8248-7: Sperm test, urine 2950-4: Sodium body fluid, chemistry	▪ Adverse events detection [20] ▪ Adverse event prediction [21] [22] ▪ Disease phenotyping [23]
Drugs codes	Hospital dataset	Contains RXnorm codes for each drug in the hospital	00597016203: Dulcolax	▪ Medication information extraction [[24]
Diagnosis codes	Online standard dataset	data for ICD diagnoses where each code corresponds to a single diagnostic concept	0880: Bartonellosis 08881: Lyne disease 08882: Babesiosis 0914: Syphilis adenopathy	▪ Disease phenotyping [25]
EHR events	Hospital data records for each event per patient	the friendly/adverse events related tables records all events like ventilator settings, MRI readings, laboratory values, code status, mental status, etc..	27: Abdominal Assessment 41: Alarm activated 81: Brachial pulse	▪ Adverse Event detection [26] ▪ Adverse Event prediction [27] ▪ Patients phenotyping
Treatment procedures	Standard CPT coding	contains types of procedures as coded in the CPT standard. Each patient has to have specific procedures performed on him in a separate table	161: Removal of penetrating foreign body from eye. 1522: Shortening procedure on one extraocular muscle	▪ Adverse event detection ▪ Adverse event prediction

is an existing deep learning solution to an EHR task, and "how can we do it?" from our own perspective. Due to the high mathematical sophistications behind deep learning applications to EHR, we try to explain the concepts in simpler terms. However due to complexity of deep learning, it is impossible to thoroughly explain every concept, hence a modest understanding of machine learning is required to understand the content of this work.

II. RELATED WORKS

Using the vastly available EHR data for clinical analytics has recently gained a big deal of attention. However, few studies would come up with a complete set of methodologies and techniques that can be used to mine this unexplored big data. More of related researches have focused on applying data mining methods for an aspect of EHR data mining. Ching et al. [3] thoroughly discussed opportunities and challenges in using deep learning for biology and medicine, though this study was much exhaustive it did not elaborate more on the technical side of the processes involved. Reference [4] covered DeepEHR by surveying recent advances in deep learning techniques for EHR. This study focused on identifying key works done in deep learning for EHR. The description of these works done as well as their approaches are detailed in Table 1.

III. ANATOMY OF THE EHR DATA

In this section we cover the structure and anatomy of EHR data. The EHR is composed of huge longitudinal, time series data sourced from daily recordings by practitioners and hospital instruments. Each patient record is saved in a table in EHR warehouse. Though structure of database can vary depending on specific medical and computing requirements, examples and cases described in this section are retrieved from the MIMICIII dataset [109]. A more detailed description of EHR data is found in Table 2 with the following components being the big constituents.

A. PATIENT INFORMATION

Patient basic information is perhaps the simplest and the most structured data of the EHR data. It contains basic information of a patient like his hospital ID, an identifier which will identify the patients through his stay. It contains

his gender, date of birth, date of discharge (or death) and other demographic data. More data related to admission like admission time, discharge time, admission type (emergency or elective) insurance information and any other basic information of interest. Though the recording of these information looks straightforward many workers at hospital do not consider accurate recordings as one of their critical tasks and this becomes worse to physicians who are mostly preoccupied by saving lives than proper recording, hence errors in EHR records can be produced at any point. For instance, statistics at the English National Health Service (NHS), showed that about 20,000 adults were recorded in pediatric outpatient services, similarly 17,000 men were admitted to obstetrical services, and 8,000 men admitted to gynecology services [9].

B. CLINICAL NOTES/MANUSCRIPTS

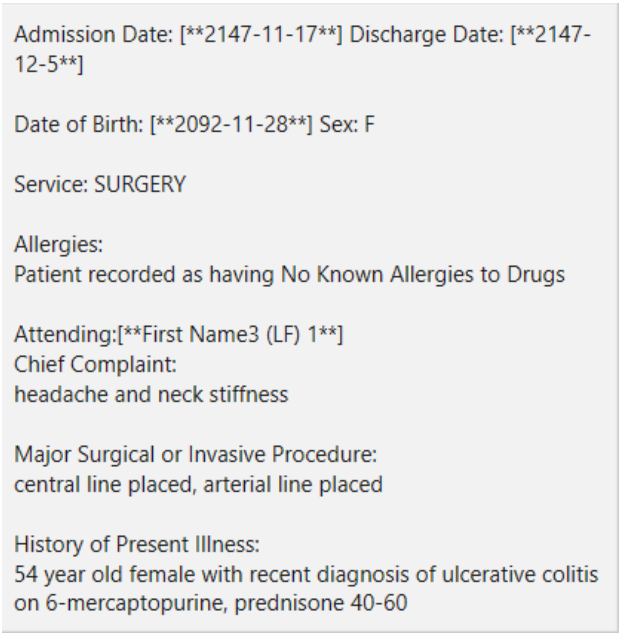
Perhaps the most rich but unstructured, vague and noisy of all EHR data are the physicians/nurses' clinical notes. The 2018 national physician poll [10] showed that though physicians view the EHR as necessary, they did not view it as a powerful clinical tool but as a mere data storage tool, and surprisingly only half of them agreed that using an EHR detracts from their clinical effectiveness. Moreover, the EHR does not provide a cognitive support design which is causing doctor to be reluctant in using the EHR interfaces and the continuation of relying on their manuscript-based documentation to reduce clinical burnout. Deep learning methods helps in transforming these manuscripts into database readable formats. Fig. 1 shows Example of a clinical note as extracted from the MIMIC III database. Clinical notes can be analyzed using deep learning models to predict adverse events like heart attack, death, hospitalization length etc. However, these notes must be treated by a vectorization and a feature representation algorithm before being fed to a deep learning model.

C. LABOLATORY MEASUREMENTS AND MEDICAL IOT READINGS

EHR has a lab events table that is associated with lab measurements for each patient. Each laboratory observation is linked to a lab item which is defined in another table containing all the definitions for laboratory measurements. The definitions contain the Logical Observation Identifiers Names and Codes (LOINC) for lab measurements. For instance, the MIMIC III hospital dataset's lab items table has 27,854,055 lab events associated to all 60,000 patients.

D. MEDICATION, DIAGNOSIS PROCEDURE AND DRUG CODES

This EHR sections contain standard codes for diseases and symptoms described by the International Classification of Diseases (ICD) [11] and diagnosis related groups codes DRG (used for identifying billable items that the patient received) [12]. Drugs are described by their RXnorm drugs


Admission Date: [**2147-11-17**] Discharge Date: [**2147-12-5**]

Date of Birth: [**2092-11-28**] Sex: F

Service: SURGERY

Allergies:
Patient recorded as having No Known Allergies to Drugs

Attending:[**First Name3 (LF) 1**]
Chief Complaint:
headache and neck stiffness

Major Surgical or Invasive Procedure:
central line placed, arterial line placed

History of Present Illness:
54 year old female with recent diagnosis of ulcerative colitis on 6-mercaptopurine, prednisone 40-60

FIGURE 1. Example of a clinical note as extracted from the MIMIC III database.

classification codes [13], treatment procedures are described by their Current Procedural Terminology (CPT) codes [14].

E. EHR EVENTS

EHR also contains 21,146,926 rows of input events (ex; Heart rate, Glucose levels etc...), 330,712,483 charted events, 4,349,218 Output Events, and many other events which records whatever happens to a patient.

IV. CHALLENGES FOR EHR MINING

EHR Feature engineering is moving away from the usual expert-driven feature engineering to data-driven paradigms or the combination of both [28] for sophisticated clinical tasks like feature construction, risk factors identification and diseases phenotyping. Hence analytic processes rely on the capability to find proper machine learning techniques for a distinct task. For example, while Natural Language Processing (NLP) [29] will help in dissecting clinical insights hidden in a million of clinical manuscripts it will be of little to no help in the understanding of an MRI brain scan. The following are key challenges that analytical solutions must address to provide actionable clinical insights.

A. COMPLEXITY OF EHR ANALYTICAL TASKS

Even the NLP which performs better in text-based sentiment analysis will hardly help in understanding clinical narrative and terms used in the clinical notes, recorded by medical expert care staff. The reason is that health care experts write these notes for individual or co-worker's reference with no machine learning applications to sight. Various toolkit that tailor the NLP for clinical texts have been invented like CLAMP (Clinical Language Annotation, Modeling, and

Processing) [30] which is a popular NLP tool that helps clinical applications developers to quickly build customized NLP pipelines. However, EHR tasks like prediction of clinical events need amalgamation of structurally diverging data like lab tests together with charted events and clinical notes. The more data we incorporate the more predicting accuracy is achieved. The structure of lab tests come as text flags with varying unit of measure hence combining them with clinical notes which are raw texts without a standard becomes a challenging task. Also, different EHR data do not contribute equally to illnesses that have to be predicted or detected. As an example, mental sickness might depend on narratives in a clinical note than on charted events as there might not be any events associated with the patient, hence coming up with a rationalized model that combines these data is a very complex task.

B. CONTEXTS OF EHR DATA

Even with tools that help people to design customized pipelines, challenges related to clinical data are hard to surmount. The big challenge comes from the nature of the data and the kind of insights we want to extract from it. Clinical experts are human beings who try to find solutions to intervention problems in a causal point of view. In his study about causality and machine learning, Pearl [31] argued that intervention questions cannot be answered from observational statistical information alone. He also argues that you cannot answer counterfactual question using intervention information. In a clinical example you cannot re-perform a trial on patients who were treated with a drug to inspect how they would have behaved had they not been given the drug. Machine learning algorithms which are observational algorithms that use statistical data exhibits these fundamental impediments that make their applications to clinical questions to require additional extra-statistical information.

C. SMALL LABELLED DATA

Perhaps, except the clinical notes and the patient's charted events used to perform certain deep learning tasks, most of EHR data lacks labelled ground truth data. Even the model that is built gets hardly into implementation due to lack of acceptance. The true outcome of a clinical event is a redundant operation that relies on abductive reasoning of a physician, hence deep learning gets stranded in this problem. As an example, you cannot find enough labelled cancer images that can be used to train a CNN for future predictions. Perhaps the most appropriate solution to the lack of labelled data seems to be the use of transfer learning. There are vastly available labelled data sets that have been trained for other tasks. these pre-trained models can be used to medical problems by only tailoring the last layer of the neural network to the EHR problem in question. Authors in [32] have used transfer learning on a pre-trained RNN model to establish phenotypes of various diseases. Another method is to use unsupervised CNN pre-training and perform a supervised fine-tuning.

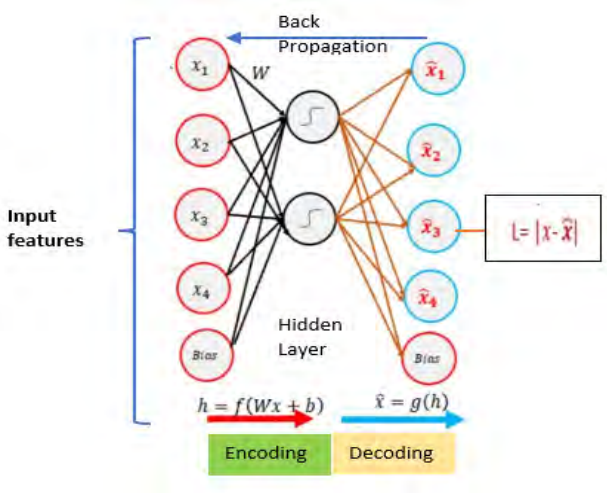


FIGURE 2. Basic architecture of auto encoders.

Authors in [33] have been able to use this method to classify lung tissue in high resolution Computed Tomography (CT) data.

V. POPULAR DEEP EHR ALGORITHMS

Deep learning is a special branch of machine learning that utilize layered computational nodes with each node in each layer performing computation on inputs and its respective weight. A non-linearity function is applied to produce the node activation. The overall Artificial Neural Network is built on updating the weights of each node to minimize the final cost associated with the deviation of output predictions from the ground truth labels. The neural network first initializes the parameters (weights and bias) and use the forward propagation to calculate a cost then the chain rule is used to perform back propagation for weights updates. The process is referred to as gradient descent due to the process of finding an optimal path to a minimum cost. Various more advanced optimization algorithms that solves the basic deep learning problem have been discovered and used in practice. These are Stochastic Gradient descent [33], RMS Prop [33], Ada boost [34], Adam [35] etc. Deep learning is effective than other machine learning algorithms as there is no need to spend more efforts on feature engineering using a domain expert, rather using raw data as the features can be learned by the system. However, as we will see in later sections, due to complexity of EHR data and special intolerance to errors, feature representation and selection usually assisted by domain expert might be a key to the success of a deep learning model. In this section we are going to describe briefly popular deep learning algorithms used with EHR. A complete reference of these algorithms and their use with EHR can be found in Fig. 7.

A. SPARSE AUTO ENCODER

This is an unsupervised representation learning mostly used for the features engineering stage. They are used for

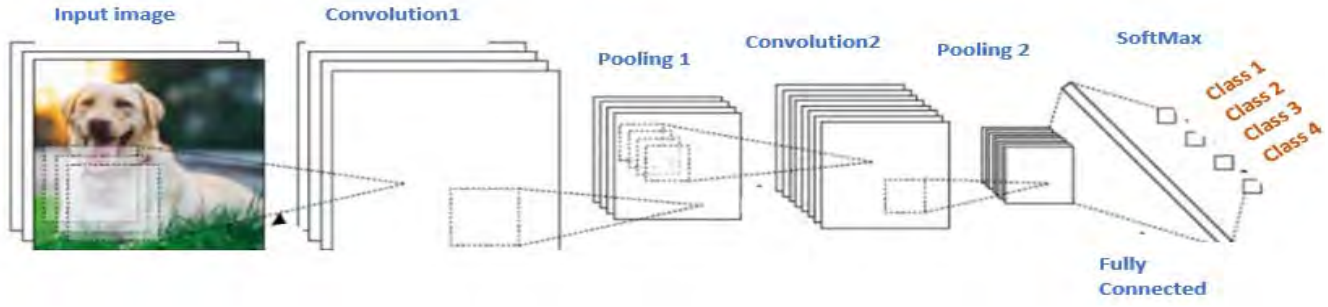


FIGURE 3. Basic convolutional neural network.

non-linear dimensionality reduction and comes as a better alternative to other traditional dimens ionality reduction techniques like principal component analysis (PCA) [36] and singular value decomposition (SVD) [37]. From Fig. 2, an auto encoder is used to transform(encode) a much bigger vector into much smaller data vector by taking the input x, encoding it to discover latent feature representation then decoding the latent feature representation to reconstruct the input. Auto encoders are used in applications that require features compression like finding document similarity, feature reduction etc. Many variations of auto encoders have been used extensively. Convolutional auto encoders are special types of autoencoders that do not use fully connected layers (each node in a layer connected with each node in next layer) rather using convolutional layers.

B. CONVOLUTIONAL NEURAL NETWORKS(CNN) [38]

Convolutional Neural Networks are special algorithms that perform extremely well in image classification problems. In the EHR context CNNs can yield good results in medical image analysis like mammography, MRI images, CT scans etc. They can be used to detect and differentiate malignant cancer cells with the benign cells from medical images. From Fig. 3, CNN is composed of layers where each layer is composed of a convolutional layer, a pooling layer and an activation to produce input to the next layer. CNN are special architecture where each node from the previous layer is not connected to each node of the next layer, rather each layer is composed of a filter(kernel) or several filters that are applied to the input to produce intermediate values. The resulting next layer input is a sum of products of each input feature value with the filter. We say that a filter is convolved with the input image. Each convolution stage defines certain attributes of the input such as lines curves and edges. As an example, if a 256×256 image is input to a CNN the input layer will be $256\times256\times3$ in size (with 3 representing RGB channels). the convolutional layer will perform a dot product between a receptive field and a kernel on all the dimensions of the input. To minimize the training time and avoiding overfitting, the pooling layer reduces the dimensionality in the network by taking a maximum or average of a certain number of inputs cells. At each layer an output is obtained by applying a non-linearity function usually Rectified linear unit(relu). A fully connected layer is added towards the end of the network followed by a SoftMax layer which produces the predictions. Various special types of CNNs have been produced and are being used with EHR like Resnets [39], VGGNet16 [40], Inception [41] etc.

C. RECURRENT NEURAL NETWORKS (RNN)

With some types of data in EHR like clinical notes, input data do not have the same length to be used with basic ANN. For instance, some medical applications can require processing vast amount of text (like clinical notes, web based medical queries platforms etc.) to find keywords that are relevant to standard clinical entities like ICD codes and CPT codes. This application requires performing a Named Entity Recognition (NER) [42] as a primordial step to the understanding of the bulk text. To understand RNN in medical context let's take a user who tweets about an Adverse Drug Event (ADE). Moreover, RNN can be used to identify drug names present in the tweet in the process of identifying the ADE from the tweets. With the RNN depicted in Fig. 4 taking the input tweets as a vector x we want to produce a vector y that contains 1 in a position that holds a drug name and 0 in a position that holds any other word. Using an NLP dictionary, we can build a one hot encoding of each word present in the document and feed the resulting vectors to the RNN.

D. DEEP TRANSFER LEARNING: SOLVING THE SMALL LABELLED DATASET ISSUE

One of the greatest impediments of machine learning to EHR data is to find enough labeled data for training. For instance, if we are analyzing CT scans to find a malignant tumor, we may not find enough recorded events that can be used for training a deep learning model. Transfer learning is a deep learning technique that takes intuition from the human learning which uses knowledge gained from one problem to another problem. In a deep learning world, we can use the weights learned while modeling one problem to another problem. As an example, in Fig. 5, we can use a model that was trained on the cat and dog dataset to MRI images that can classify if a brain tumor is malignant or benign.

VI. TECHNIQUES FOR HER TASKS

A. CLINICAL ADVERSE EVENT DETECTION

One of the primary tasks of hospitals is to detect a clinical event in real time. All the causes of clinical events

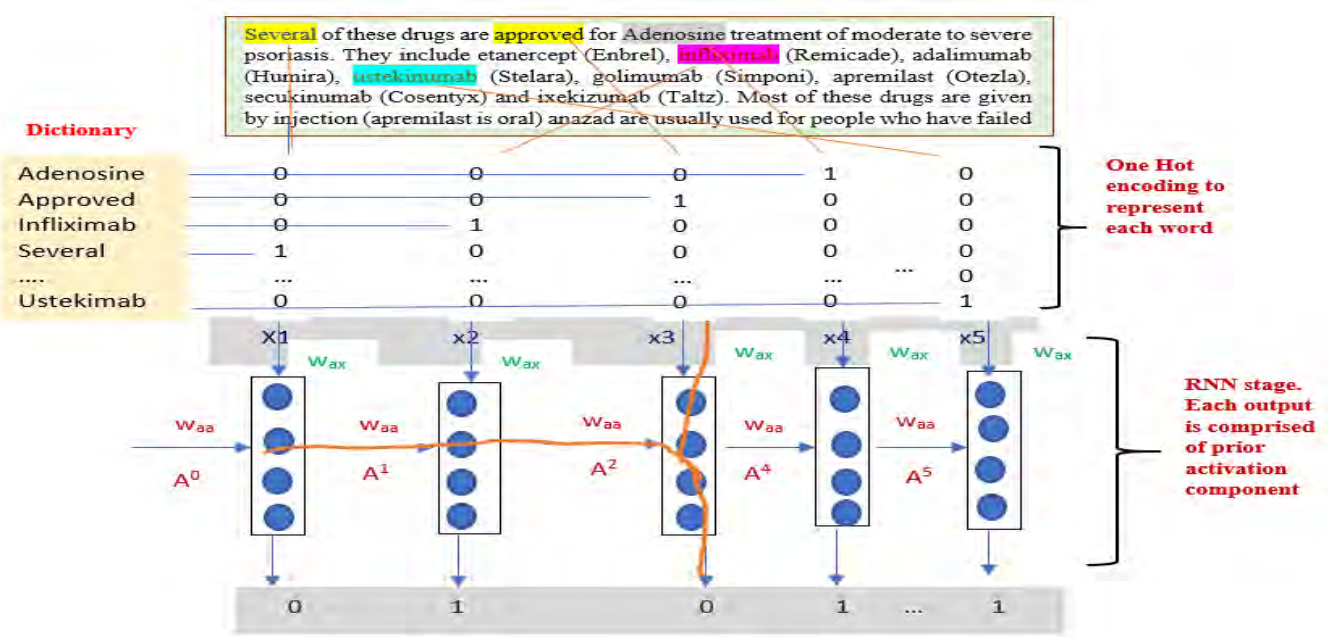


FIGURE 4. Basic RNN a 1 in the output vector represent a presence of a drug name in the input text.

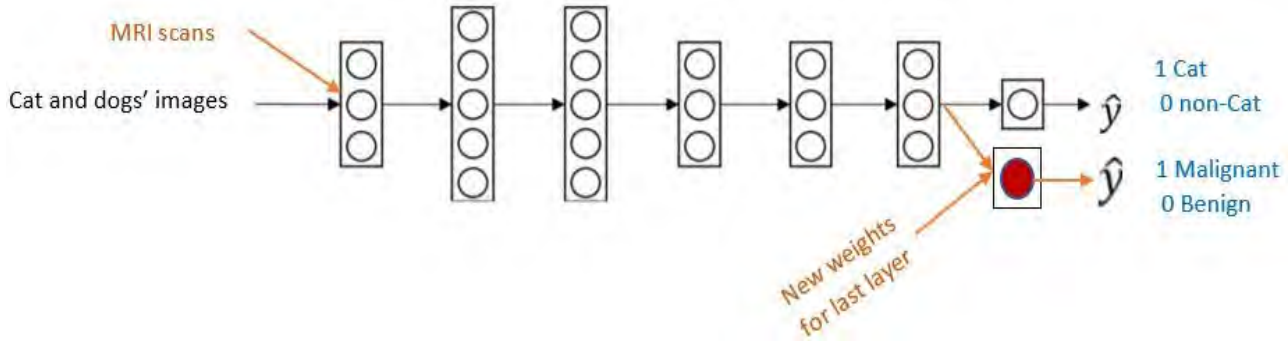


FIGURE 5. Basic Transfer learning. Weights learned from training a cat classifier are used to predict tumor malignancy from MRI images by only changing the last layer and introducing weights for the last layer.

including medication, diagnosis, and adverse drug events etc. can be found buried in longitudinal data in the EHR. Critical medical events can be conceived as negative changes in patient's medical status. Authors in [43] have applied bi-directional Recurrent Neural Networks (RNN) on EHR to predict medical events. The experiment used Sequence labeling techniques for extraction of medical events from unstructured text in EHR. The study in [44] tried to use EEC (electroencephalograms) signals from the EHR and Deep Convolutional Neural Network (DCNN) [45] to detect Epileptic seizure. First the EEG signal features were extracted using EMD algorithm [46] to decompose the EEG signals into oscillation instances with varying frequencies called the Intrinsic Mode Functions (IMFs). The next step was to feed the data to a Deep CNN for classifying the seizure into three classes of epilepsy; ictal (amid seizure), normal, and inter ictal (amongst seizures).

EHR Use Case: Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks [47]: One of the big challenges for health-related detection and classification is the absence of enough labeled data. In Fig. 6, researchers combined data from open-access dermatology repositories, which were annotated by dermatologists as well as data from the EHR. These skin lesion images were fed to an Inception V3 [48] Deep CNN which predicted if the subject in the image is having malignant melanocytic lesion or benign melanocytic lesion. this work leveraged the power of transfer learning by using the Inception V3 a special type of deep CNN with reduced number of learned parameters. It achieves this property by performing a factorization into smaller convolutions through replacing a 5×5 filter with two 3×3 filers. This technique helps in reducing the number of parameters to be learned hence shrinking the computational cost of the deep network.

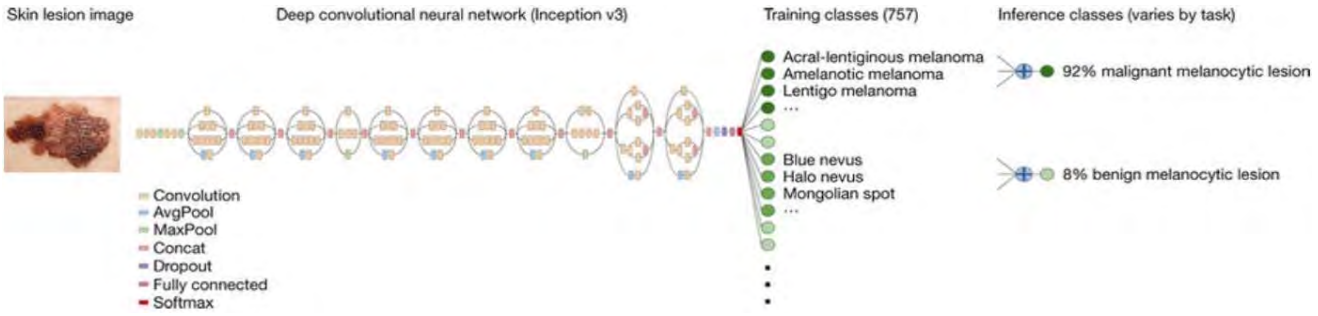


FIGURE 6. Architecture of the skin cancer detection by adapting the Deep CNN inception V3model.

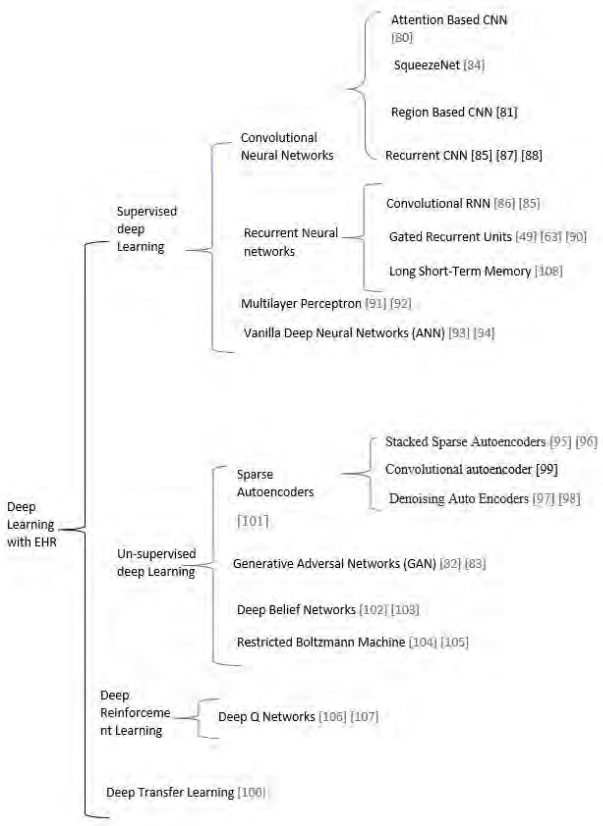


FIGURE 7. Popular deep learning algorithms used with Electronic health records.

B. CLINICAL ADVERSE EVENT PREDICTION

Clinical adverse Event Prediction sub task by a learning algorithm is to predict the onset of diseases a process that predicts the probability that patients might develop certain diseases given their current clinical status. Specific objective is to predict future events (hospitalization, suicide risk, heart failure risk etc...) from longitudinally diverse events. For intelligent support system to provide patient centered support each aspect type of data would need a support system. Choi et al. developed DoctorAI [49] a generic system that uses Recurrent Neural Networks (RNN) to predict clinical events via a system that performs multi label prediction using diagnoses, medication categories and visit time of a patient. They were able to use each patient visit to predict about diagnosis, medication order in the next visit as well as the time to next visit. Razavian et al. [50] were able to use longitudinal lab tests to perform early diagnosis of diseases for people who do not yet have the disease.

Miotto et al. proposed DeepPatient [51] a system that leverage raw patient data from EHR like medication, diagnoses, procedures, lab tests by applying them to unsupervised deep feature learning algorithm to produce patient representations that will be applied to perform more advanced clinical tasks like personalized prescription, drug targeting, clinical trial recruitment, detecting patient similarity etc. Prediction of future clinical events can be achieved by modeling the EHR record as longitudinal event matrix, with the horizontal dimension corresponding to the time stamps and vertical dimension corresponding to the event values and applying non-standard CNN [52]. Many prediction algorithms leverage various EHR data types to predict an outcome, however clinical notes contain rich amount of patient's data than other sources. Though unstructured they can be a source of a big number of clinical predictions. However raw text cannot be applied directly as meaningful features to deep learning models, hence to acquire vectorized inputs, before applying deep learning, a word embedding algorithm like Word2Vec and Doc2Vec must be applied to produce word vectors that can be understood by the learning algorithms.

Use Case1 (Using EHR Clinical Notes and Convolutional Neural networks (CNN) to Predict Death): This sub-section serves as an intuition and use case of clinical notes generated at the point of care into predicting adverse future event. An imminent patient's death is a result of various time series events manifested after admission into the hospital. The unexplored clinical notes produced by physicians or nurses contain a rich content in a form of text that requires critical analysis. The process of adverse event prediction is described in Fig. 8. The task of deep learning is to aggregate many data with or without known outcome(labels), and to train a model which can predict an outcome for new scenarios. As clinical notes cannot be directly analyzed by the deep learning model, they are vectorized by a Word2vec

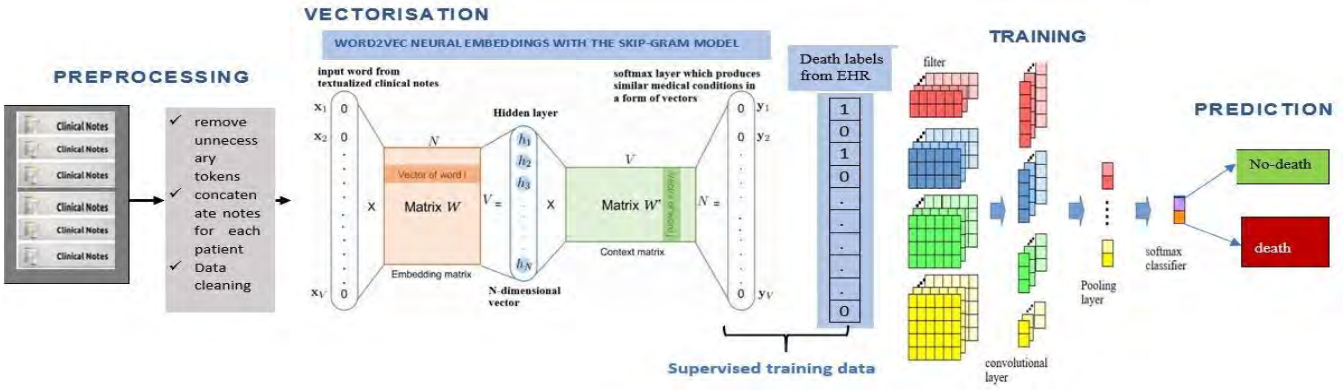


FIGURE 8. use case scenario clinical notes are vectorized using Word2Vec skip gram model, then using labels obtained from patient history weather he died or not, train a CNN model which can predict a near future death prediction using patient's hospital notes.

or Doc2Vec word embedding models that use skip gram to vectorize textual information. However, clinical notes contain ambiguous terms as well as important terms that are related to a certain disease phenotype (as an example we expect a clinical note written for a patient suffering a heart attack to contain terms like chest pain, discomfort, shortness of breath, lightheartedness etc.). Hence, before vectorization we must dissect the content of the clinical notes using standard ontologies for medical terminologies like the (SNOMED CT) [53] or the Unified Medical Language System (UMLS) [54], [60]. after extracting these words that are related to patient phenotype the notes can now be fed as input features to a Word2Vec model for vectorization. These resulting vectors can now constitute labeled training data for Convolutional Neural Network which can predict the death probability (labels are obtained in the end status of a patient whether he died, or he was discharged).

Use Case2 (Using Charted Clinical Events to Predict Medical Adverse Future Events): Predicting the length of stay and readmission probability helps in improving quality of care as well as the potential to decrease unnecessary healthcare costs. However, being able to aggregate all the patient's data and decide on which one that can have more weight in an intended prediction is highly an iterative process. Various machine learning, and statistical models have been deployed to predict death risks for hospitalized patients. Medical charted events like ventilator settings, mIoT device's alarms, laboratory values, heart rate, MRI readings, code status, mental status, and so on, can be used to predict patient's risk of imminent death or hospitalization period. For instance, a patient in the MIMIC III database who was admitted with hemorrhagic CVA (Cerebrovascular accident) hospitalized for 5 days recorded among others a total of 9172 charted events,68 prescriptions, and 12 microbiology events. These records contain a potential source of data for prediction. Because all the outcomes are known (Death or discharge), if we consider each patient and build a representative vector that accommodates all these events we can train a deep neural network that can predict the patient's outcome. Esteban et al. used Recurrent Neural Networks (RNNs) and static information like patient gender, blood type, etc. and dynamic information like clinical charted events to predict future adverse events [55].

C. EHR DRIVEN PHENOTYPING

Clinical phenotyping is a process of establishing diseases characteristics. This process is performed by expert opinions and many years of researches which have already established phenotypes of each disease. However, with the diversification and polymorphism of existing diseases coupled with individual genetic variations, there is high need to find other methods to establish disease phenotypes as well as individual patients' phenotype using huge data stored in EHRs. Many studies have used methods that include a mix of clinical expert opinions and automated methods. A. Neuraz et al. [56] have developed a method that used the frequency and TF-IDF [57] to establish the relationship between clinical phenotype and rare diseases. To access the performance of deep learning methods to phenotyping tasks, Gehrmann et al. [58] have thoroughly compared the results of CNNs with those obtained from concept extraction-based methods using clinical narratives and those from n-gram based models. Concept extraction is a popular method utilized extensively in phenotyping of many diseases. One popular project is the cTAKES (clinical Text Analysis and Knowledge Extraction System) developed at Mayo [59]. cTAKES is an openNLP toolkit that can be used to extract clinical meaning from many clinical notes. It produces named entities from each word in the clinical note and check its meaning from the UMLS through its concept unique identifier (CUI).

Use Case 1 (Creating Clinical Phenotypes Using Multi-Layer Perceptron Deep Neural Network (RNN) on EHR Data): Arguing with the precision of the International Classification of Diseases (ICD) codes that establish medical codes and associated phenotypes, Rashidian et al. *[61]* used lab results, patients' demographics, as well as medication data to establish a more trustworthy coding scheme using deep learning for ensuring the credibility of these codes, they partnered

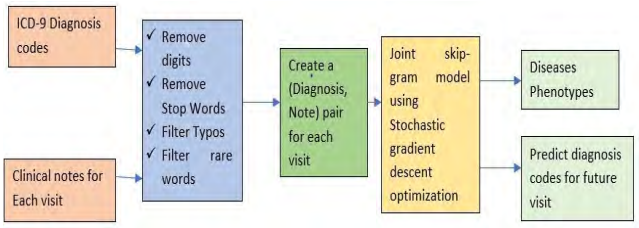


FIGURE 9. Jointly embedding ICD9 codes with clinical notes in a unified vector space to establish diseases phenotype and predicting future visits.

with medical experts who verified the trustworthiness of the model codes vis-a-vis the accepted ICD9 codes. Their model was found to provide extensive and precise phenotypes than those described in the ICD9 standard.

Use Case 2 (Embedding Medical Concepts and Words Into a Unified Vector Space [62]): Most of the studies who tried to leverage EHR data for patient's phenotyping used the embedding of medical codes like the ICD9 and fed the resulting vectors to a neural network to establish diseases phenotypes or to predict a clinical adverse event [63]–[66]. Other approaches have tried to embed the extracted medical codes and accompanying words separately. This approach can have its drawbacks as the words will lose their medical contexts. Rather than using the normal skip gram where the context words of the current word are established by calculating the probability of each neighboring word being a context word, Bai et al. used a Joint Skip-gram approach to jointly embed the medical codes and words from clinical notes. it is done by representing each patient visit by a pair made of diagnosis codes and words from clinical notes (D, N) where D = {C1, C2, C3...} and N = {N1, N2, N3.). With the MIMIC III data set 54,965 such pairs have been obtained. The Joint Skip gram was used to define the context of the diagnosis code in question with also other codes in the same visit, as well as all words in the clinical note. To aggregate data for the model, for each patient visit, all diagnosis codes and all clinical notes were extracted. As shown in Fig. 9 Stochastic gradient algorithm with negative sampling was used as an optimization algorithm to predict ICD-9 codes associated with future visit as well as establishing diseases phenotypes.

D. PATIENT'S FEATURES REPRESENTATION

To perform an adverse clinical event prediction or any other EHR task a precise patient representation and stratification is very paramount. It is highly erroneous to directly feed the EHR raw data to a deep network to perform clinical tasks like predictions, clinical trial recruitment or disease detection because of the high heterogeneity and sparsity. Hence, before performing these clinical support tasks with deep networks a feature learning framework which can represent the patient's features with less information overlap has to be constructed from the vastly heterogeneous EHR data. Various models have represented patients in a form of a 2D vector with patients on one dimension and amalgamation of each patient's records (ICD9 diagnosis, lab tests, clinical notes content...) in another dimension. A common approach is to have a clinical domain expert manually annotate the patterns to look for including the clinical features and the targets of the learning scheme. However, annotating features using a domain expert in an ad hoc manner is tiresome and imprecise. Recently unsupervised deep learning has revolutionized the process of feature learning and selection. Authors in [67] have use unsupervised learning for feature selection. First the EHR raw data was divided into continuous features and categorical features. Continuous features were first changed into representational features using stacked auto encoders and combined with categorical features then SVM was applied for features selection. The resulting features were fed to a model which can predict the amount of LVMI (Left Ventricular Mass Index) a common indicator of heart damage risk.

The most challenging hustle for deep leaning models is the small size of the input data set. This creates a natural incompatibility of EHR with deep learning models because when small data sets are directly fed to a deep network it leads to overfitting. One approach is to fuse deep features (obtained by using a deep network) with traditional features like texture feature, color moment obtained by traditional methods like Haarlick [68] method. The study in [69] used lung tumor images and transfer learning techniques using 3 existing CNN models that were pre-trained on ImageNet public data set [70] and combined obtained features with traditional features to predict Survival among Patients with lung adenocarcinoma. Authors in [71] used a CNN based Coding Network for medical image classification using deep features obtained with convolutional neural network and some selected traditional features obtained with a solid background knowledge of medical images like color histogram, color moment and texture features.

Choi et al. proposed Med2Vec [72] a patient representation that learns from medical codes associated with a clinical visit to predicts codes that are likely to characterize the next visits. The issues addressed in the study are that representations obtained from RNN are difficult to interpret and difficult to scale with high dimensional EHR data. Moreover, these representations fall short of critical information that is embedded in the patient's demographic information. The authors adapted the usual embedding skip-gram model to medical concepts. The first step of the solution is to represent a patient's visit as a unified vector consisting of codes(diagnosis, subscriptions, etc..).using these codes as inputs, ReLu activation was applied to obtain an intermediate vector which was then combined with patient's demographic information to produce Vt an intermediate visit vector and use it to train a SoftMax classifier that is able to predicts the medical codes of other visits within a context window.

Use Case 3 (Deep Features Learning From EHR Raw Data): The most effective method is adopted by authors in [73], [23]. In these researches, authors argue that supervised feature learning lacks an ability to fully grasp novel

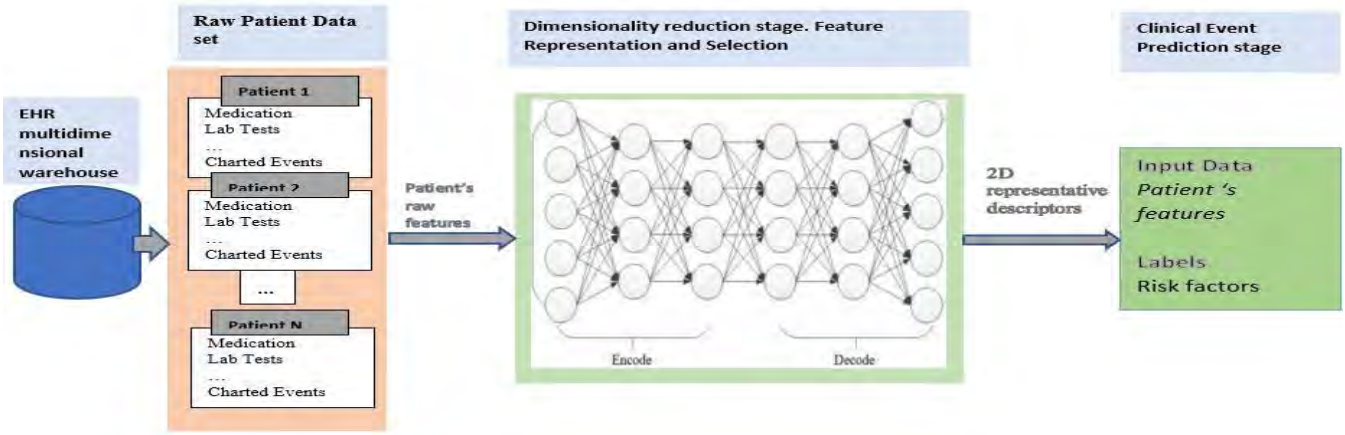
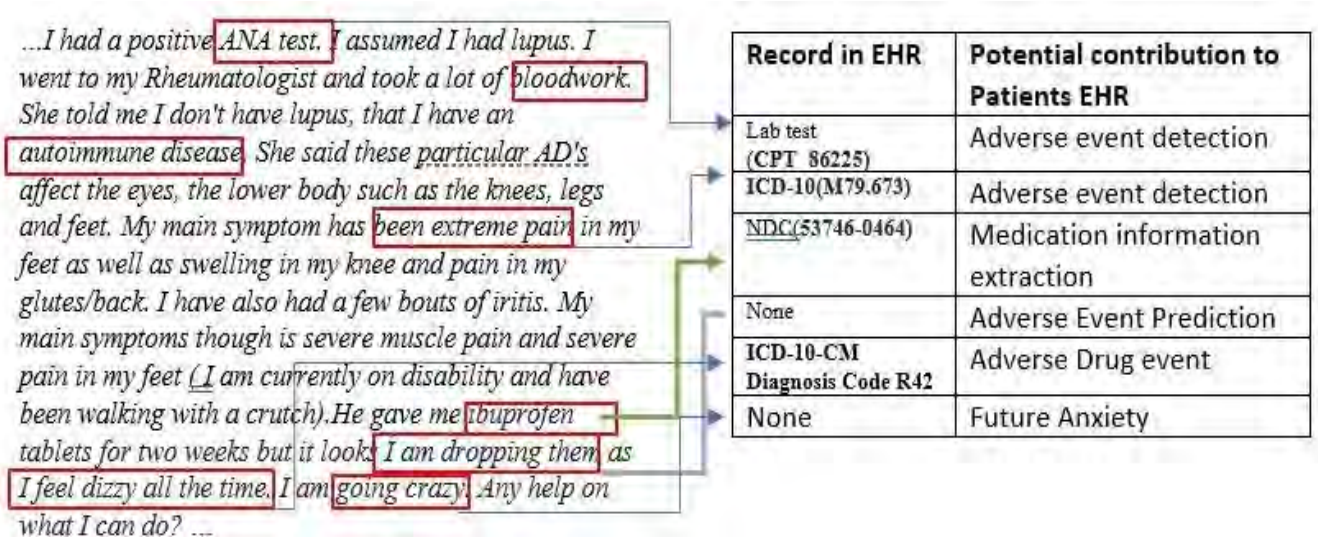


FIGURE 10. Patient's clinical features representation and selection overview.



Record in EHR	Potential contribution to Patients EHR
Lab test (CPT 86225)	Adverse event detection
ICD-10(M79.673)	Adverse event detection
NDC(53746-0464)	Medication information extraction
None	Adverse Event Prediction
ICD-10-CM Diagnosis Code R42	Adverse Drug event
None	Future Anxiety

FIGURE 11. Potential medical information hidden from social acquaintances and their possible contribution to HER.

patterns and features. They propose a data driven approach to automatically identify patterns and dependencies in the data without the need of a domain expert to annotate the features. Fig. 10 is a blueprint of this novel approach. The first step is to extract patient data (medication, diagnoses, lab tests, clinical notes ...). The data is pre-processed including appropriate embedding of clinical notes, then each patient is represented as a single vector. The next step is the dimensionality reduction stage which consists of feature representation and selection using stacked Denoising Autoencoders (DAE) which are unsupervised learning neural networks that can generate their own labels from the training data. The SDAEs are used to transform these patient vectors into more representative descriptors which can be input of another deep learning prediction model. The last stage is the use of supervised learning to perform various clinical support tasks like diagnosis proposition, adverse event prediction, clinical trial recruitment etc. As an example, these features can be used together with risk factors (like death or ECG readings) to train a supervised model which can predict adverse events.

E. MEDICATION INFORMATION EXTRACTION

Medication information is an important area of biomedical research as it contributes greatly to pharmacovigilance, adverse events' detection, bio curation assistance, integrative biology etc. Through much of the information can be extracted from social fabric like social networks, EHR contains also much of the immediate medication information. However, the process of mining this information from the EHR can be a lengthy tiresome process as the data is hidden deep in EHR's clinical narratives, patient's encounters, ICU discharges, and charted events. The task of a computerized Adverse drug event recognition involves 3 main tasks which are the Named Entity Recognition (NER) a process of detecting key drug mentions, identifying these named events a process of identifying the context of these mentions, and finding

relationship between them. The medication information extraction system aims to establish the medications names, and their signatures like dosages, duration, prescription reasons, complications, frequencies, route of administering, and any other information deemed necessary by the prescribing entity. Early use cases include MedEx [74] a system that automatically extract medication names and their signatures from clinical narratives using NLP. Authors of MedEx argue that usual text parsing methods like regular expression cannot apply in medication information extraction as they fall short of contextual information out of clinical narratives. MedEx uses a semantic-based approach with a much finer granularity.

F. INTEGRATING EHR SOCIAL NETWORK AND WEB DATA

It is most likely that a patient shares clinical insights like adverse drug event within social acquaintances than with his physician. With the explosion of social networks, there is huge, untapped medical insights which can be used together with hospital's EHR for clinical support systems. Though the medical research community agrees that social networks should be part of the EHR, the modalities of how to go about it remain a highly debated subject. The concerns of this reluctance are high noise due to spelling errors, imprecise descriptions, and ambiguous or casual use of medical terms. Some clinical tasks may even depend on social data than more formal EHR data. For example, recent researches have shown that these social network services can hold data related to pharmacovigilance and medication adherence than EHR because a big number of patients might not return to hospital to narrate the drug reactions unless there is an acute condition that resulted into taking the drug. Recently deep learning models have been applied to SNS data to contribute to various clinical tasks [75]–[79]. Integration of social media in the clinical care pipeline helps patients to participate in self-care, health promotion, and disease prevention efforts by the public. Ideas on how to integrate the SNS into EHR argue that these data should be supplemental not overriding other EHR data like charted events, lab events, lab tests, etc. Fig. 11 shows a patient message to his acquaintances and possible EHR tasks that can leverage these types of messages.

VII. CONCLUSION

We have given insights and technical intuitions of how to leverage the EHR data using deep learning approaches. We unraveled the technical side of various efforts that have been invested to apply deep learning models for clinical knowledge discovery using electronic Health Records vast data sets. Despite clear success of deep learning for other hospital's tasks like billing and patient management there is still much to do in the application of EHR data with deep learning methods. Available successes in this domain still depend on a supervision of a medical domain expert. More research needs to be done to bring AI and deep leaning on the patient's bedside. Unlike other deep learning applications, the medical field is challenged by the structure of the data itself and the acceptance of the models by the medical community. Even if the model might be working from a computing point of view its adoption will be hindered by the reluctance by clinicians who still exercise their professions using abductive reasoning. Though deep learning algorithms perform better even with little or no feature engineering, considering high risk factors associated with EHR tasks, coupled with high longitudinality, sparsity, and noisiness of EHR data there is a requirement to perform a thorough patient representation that consists of appropriate patient's feature selection and representation before a predictive deep learning model.

REFERENCES

[1] A. Hoerbst and E. Ammenwerth, "Electronic health records," *Methods Inf. Med.*, vol. 49, no. 4, pp. 320–336, 2010.
[2] D. V. Dimitrov, "Medical Internet of Things and big data in healthcare," *Healthcare Inform. Res.*, vol. 22, no. 3, pp. 156–163, 2016.
[3] T. Ching *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *J. Roy. Soc. Interface*, vol. 15, no. 141, 2018, Art. no. 20170387.
[4] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, 2017.
[5] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.
[6] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.
[7] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, 2018.
[8] A. A. Kalinin, G. A. Higgins, N. Reamaroon, S. Soroushmehr, A. Allyn-Feuer, I. D. Dinov, K. Najarian, and B. D. Athey, "Deep learning in pharmacogenomics: From gene regulation to patient stratification," *Pharmacogenomics*, vol. 19, no. 7, pp. 629–650, 2018.
[9] W. R. Hersh, M. G. Weiner, P. J. Embi, J. R. Logan, P. R. Payne, E. V. Bernstam, and J. H. Saltz, "Caveats for the use of operational electronic health record data in comparative effectiveness research," *Med. care*, vol. 51, no. 8, pp. S30–S37, 2013.
[10] (2018). *How Doctors Feel About Electronic Health Records*. [Online]. Available: https://med.stanford.edu/content/dam/sm/ehr/documents/EHR-Poll-Presentation.pdf
[11] V. N. Slee, "The international classification of diseases: Ninth revision (ICD-9)," *Ann. Internal Med.*, vol. 88, no. 3, pp. 424–426, 1978.
[12] R. B. Fetter and J. L. Freeman, "Diagnosis related groups: Product line management within hospitals," *Acad. Manage. Rev.*, vol. 11, no. 1, pp. 41–54, 1986.
[13] S. Liu, W. Ma, R. Moore, V. Ganesan, and S. Nelson, "RxNorm: Prescription for electronic drug information exchange," *IT Prof.*, vol. 7, no. 5, pp. 17–23, 2005.
[14] J. A. Hirsch, T. M. Leslie-Mazwi, G. N. Nicola, R. M. Barr, J. A. Bello, W. D. Donovan, R. Tu, M. D. Alson, and L. Manchikanti "Current procedural terminology; a primer," *J. Neurointerventional Surg.*, vol. 7, no. 4, pp. 309–312, 2015.
[15] K. E. McBride, K. E. McBride, M. J. Solomon, J. M. Young, D. Steffens, T. J. Lambert, N. Glozier, and P. G. Bannon, "Impact of serious mental illness on surgical patient outcomes," *ANZ J. Surg.*, vol. 88, nos. 7–8, pp. 673–677, 2018.
[16] D. J. Feller, J. Zucker, M. T. Yin, P. Gordon, and N. Elhadad, "Using clinical notes and natural language processing for automated HIV risk assessment," *JAIDS J. Acquired Immune Deficiency Syndromes*, vol. 77, no. 2, pp. 160–166, 2018.
[17] J. Liu, Z. Zhang, and N. Razavian, "Deep EHR: Chronic disease prediction using medical notes," 2018, *arXiv:1808.04928*. [Online]. Available: https://arxiv.org/abs/1808.04928

[18] A. Jagannatha, F. Liu, W. Liu, and H. Yu, "Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0)," *Drug Saf.*, vol. 42, no. 1, pp. 99–111, 2019.

[19] B. S. Glicksberg, R. Miotto, K. W. Johnson, K. Shameer, L. Li, R. Chen, and J. T. Dudley, "Automated disease cohort selection using word embeddings from electronic health records," in *Proc. Pacific Symp. Biocomput.*, vol. 23, 2018, pp. 145–156.

[20] S. Liu, L. Wang, D. Ihrke, V. Chaudhary, C. Tao, C. Weng, and H. Liu, "Correlating lab test results in clinical notes with structured lab data: A case study in hba1c and glucose," *AMIA Summits Transl. Sci. Proc.*, vol. 2017, pp. 221–228, Jul. 2017.

[21] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, p. 1419, Sep. 2016.

[22] A. Passantino, F. Monitillo, M. Iacoviello, and D. Scrutinio, "Predicting mortality in patients with acute heart failure: Role of risk scores," *World J. Cardiol.*, vol. 7, no. 12, pp. 902–911, 2015.

[23] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data," *PLoS ONE*, vol. 8, no. 6, 2013, Art. no. e66341.

[24] S. Lim, K. Lee, and J. Kang, "Drug drug interaction extraction from the literature using a recursive neural network," *PLoS ONE*, vol. 13, no. 1, 2018, Art. no. e0190926.

[25] J. A. Sinnott, F. Cai, S. Yu, B. P. Hejblum, C. Hong, I. S. Kohane, and K. P. Liao, "PheProb: Probabilistic phenotyping using diagnosis codes to improve power for genetic association studies," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1359–1365, 2018.

[26] R. C. Zink, "Detecting safety signals among adverse events in clinical trials," in *Proc. Biopharmaceutical Appl. Statist. Symp.*, 2018, pp. 107–125.

[27] J. C. Lauffenburger, J. M. Franklin, A. A. Krumme, W. H. Shrank, O. S. Matlin, C. M. Spettell, G. Brill, and N. K. Choudhry, "Predicting adherence to chronic disease medications in patients with long-term initial medication fills using indicators of clinical events and health behaviors," *J. Managed Care Specialty Pharmacy*, vol. 24, no. 5, pp. 469–477, 2018.

[28] J. Sun, J. Hu, D. Luo, M. Markatou, F. Wang, and S. Edabollahi, "Combining knowledge and data driven insights for identifying risk factors using electronic health records," in *Proc. AMIA Annu. Symp. Proc.*, 2012, pp. 901–910.

[29] G. G. Chowdhury, "Natural language processing," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2003.

[30] E. Soysal, J. Wang, M. Jiang, Y. Wu, S. Pakhomov, H. Liu, and H. Xu, "CLAMP—A toolkit for efficiently building customized clinical natural language processing pipelines," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 3, pp. 331–336, 2017.

[31] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," 2018, *arXiv:1801.04016*. [Online]. Available: https://arxiv.org/abs/1801.04016

[32] P. Gupta, P. Malhotra, L. Vig, and G. Shroff, "Transfer learning for clinical time series analysis using recurrent neural networks," 2018, *arXiv:1807.01705*. [Online]. Available: https://arxiv.org/abs/1807.01705

[33] T. Schlegl, J. Ofner, and G. Langs, "Unsupervised pre-training across image domains improves lung tissue classification," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, 2014, pp. 82–93.

[34] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*. Berlin, Germany: Springer, 2013, pp. 37–52.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[36] I.T Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science*, M. Lovric Ed. Berlin, Germany: Springer, 2011.

[37] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," in *Linear Algebra*. Springer, 1971, pp. 134–151.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[42] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," 2016, *arXiv:1603.01360*. [Online]. Available: https://arxiv.org/abs/1603.01360

[43] A. N. Jagannatha and H. Yu, "Bidirectional RNN for medical event detection in electronic health records," in *Proc. Conf. Assoc. Comput. Linguistics. North Amer. Meeting*, 2016, pp. 473–482.

[44] H. G. Daoud, A. M. Abdelhameed, and M. Bayoumi, "Automatic epileptic seizure detection based on empirical mode decomposition and deep neural network," in *Proc. IEEE 14th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2018, pp. 182–186.

[45] T. N. Sainath, A.-R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8614–8618.

[46] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[47] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[49] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," 2015, *arXiv:1511.05942*. [Online]. Available: https://arxiv.org/abs/1511.05942

[50] N. Razavian, J. Marcus, and D. Sontag, "Multi-task prediction of disease onsets from longitudinal laboratory tests," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 73–100.

[51] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, May 2016, Art. no. 26094.

[52] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *Proc. SIAM Int. Conf. Data Mining*, 2016, pp. 432–440.

[53] K. A. Spackman, K. E. Campbell, and R. A. Côté, "SNOMED RT: A reference terminology for health care," in *Proc. AMIA Annu. Fall Symp.*, 1997, pp. 640–644.

[54] O. Bodenreider, "The unified medical language system (UMLS): Integrating biomedical terminology," *Nucleic Acids Res.*, vol. 32, pp. D267–D270, Jan. 2004.

[55] C. Esteban, O. Staeck, S. Baier, Y. Yang, and V. Tresp, "Predicting clinical events by combining static and dynamic information using recurrent neural networks," in *Proc. IEEE Int. Conf. Healthcare Inform. (ICHI)*, Oct. 2016, pp. 93–101.

[56] N. Garcelon, A. Neuraz, R. Salomon, N. Bahi-Buisson, J. Amiel, C. Picard, N. Mahlaoui, V. Benoit, A. Burgun, and B. Rance, "Next generation phenotyping using narrative reports in a rare disease clinical data warehouse," *Orphanet J. Rare Diseases*, vol. 13, no. 1, p. 85, 2018.

[57] A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[58] S. Gehrmann, F. Dernoncourt, Y. Li, E. T. Carlson, J. T. Wu, J. Welt, J. Foote, Jr., E. T. Moseley, D. W. Grant, P. D. Tyler, and L. A. Celi, "Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives," *PLoS ONE*, vol. 13, no. 2, 2018, Art. no. e0192360.

[59] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 5, pp. 507–513, 2010.

[60] B. L. Humphreys and D. A. Lindberg, "The UMLS project: Making the conceptual connection between users and the information they need," *Bull. Med. Library Assoc.*, vol. 81, no. 2, pp. 170–177, 1993.

[61] S. Rashidian, J. Hajagos, R. Moffitt, F. Wang, X. Dong, K. Abell-Hart, K. Noel, R. Gupta, M. Tharakan, V. Lingam, J. Saltz, and M. Saltz, "Disease phenotyping using deep learning: A diabetes case study," 2018, *arXiv:1811.11818*. [Online]. Available: https://arxiv.org/abs/1811.11818

[62] T. Bai, A. K. Chanda, B. L. Egleston, and S. Vucetic, "EHR phenotyping via jointly embedding medical concepts and words into a unified vector space," *BMC Med. Inform. Decis. Making*, vol. 18, no. 4, p. 123, 2018.

[63] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 2, pp. 361–370, Mar. 2017.

[64] Y. Choi, C. Y.-I. Chiu, and D. Sontag, "Learning low-dimensional representations of medical concepts," *AMIA Summits Transl. Sci. Proc.*, vol. 2016, pp. 41–50, Jul. 2016.

[65] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *J. Biomed. Inform.*, vol. 69, pp. 218–229, May 2017.

[66] A. Perotte, R. Ranganath, J. S. Hirsch, D. Blei, and N. Elhadad, "Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 4, pp. 872–880, 2015.

[67] M. Z. Nezhad, D. Zhu, X. Li, K. Yang, and P. Levy, "SAFS: A deep feature selection approach for precision medicine," in *Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, Dec. 2016, pp. 501–506.

[68] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[69] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma," *Tomography*, vol. 2, no. 4, pp. 388–395, 2016.

[70] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[71] Z. Lai and H. Deng, "Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron," *Comput. Intell. Neurosci.*, vol. 2018, Sep. 2018, Art. no. 2061516.

[72] E. Choi, M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, and J. Sun, "Multi-layer representation learning for medical concepts," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1495–1504.

[73] M. Z. Nezhad, D. Zhu, N. Sadati, and K. Yang, "A predictive approach using deep feature learning for electronic medical records: A comparative study," 2018, *arXiv:1801.02961*. [Online]. Available: https://arxiv.org/abs/1801.02961

[74] H. Xu, S. P. Stenner, S. Doan, K. B. Johnson, L. R. Waitman, and J. C. Denny, "MedEx: A medication information extraction system for clinical narratives," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 1, pp. 19–24, 2010.

[75] A. Cocos, A. G. Fiks, and A. J. Masino, "Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 4, pp. 813–821, Jul. 2017.

[76] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 3, pp. 671–681, 2015.

[77] L. Xia, G. A. Wang, and W. Fan, "A deep learning based named entity recognition approach for adverse drug events identification and extraction in health social media," in *Proc. Int. Conf. Smart Health*, 2017, pp. 237–248.

[78] J. Xie, D. D. Zeng, and Z. A. Marcum, *Using Deep Learning to Improve Medication Safety: The Untapped Potential of Social Media*. London, U.K.: Sage, 2017.

[79] S. Chowdhury, C. Zhang, and P. S. Yu, "Multi-task pharmacovigilance mining from social media posts," 2018, *arXiv:1801.06294*. [Online]. Available: https://arxiv.org/abs/1801.06294

[80] T. Huynh, Y. He, A. Willis, S. Rüger, "Adverse drug reaction classification with deep neural networks," in *Proc. COLING Tech. Paper*, 2016, pp. 877–887.

[81] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, and E. Barkan, "A region based convolutional network for tumor detection and classification in breast mammography," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 197–205.

[82] V. K. Singh, S. Romani, H. A. Rashwan, F. Akram, N. Pandey, M. K. Sarker, S. Abdulwahab, J. Torrents-Barrena, A. Saleh, M. Arquez, M. Arenas, and D. Puig, "Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 833–840.

[83] W. Zhu, X. Xiang, T. D. Tran, G. D. Hager, and X. Xie, "Adversarial deep structured nets for mass segmentation from mammograms," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 847–850.

[84] B. B. Ahn, "The compact 3D convolutional neural network for medical images," Standford Univ., Stanford, CA, USA, Tech. Rep., 2017.

[85] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: https://arxiv.org/abs/1802.06955

[86] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 280–290, Jan. 2019.

[87] R. P. K. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," in *Reconstruction, Segmentation, and Analysis of Medical Images*. Cham, Switzerland: Springer, 2016, pp. 83–94.

[88] M. Turan, Y. Almalioglu, H. Araujo, E. Konukoglu, and M. Sitti, "Deep EndoVO: A recurrent convolutional neural network (RCNN) based visual odometry approach for endoscopic capsule robots," *Neurocomputing*, vol. 275, pp. 1861–1870, Jan. 2018.

[89] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: https://arxiv.org/abs/1802.06955

[90] M. Golmohammadi, S. Ziyabari, V. Shah, E. Von Weltin, C. Campbell, I. Obeid, and J. Picone, "Gated recurrent networks for seizure detection," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2017, pp. 1–5.

[91] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Medical concept representation learning from electronic health records and its application on heart failure prediction," 2016, *arXiv:1602.03686*. [Online]. Available: https://arxiv.org/abs/1602.03686

[92] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Distilling knowledge from deep networks with applications to healthcare domain," 2015, *arXiv:1512.03542*. [Online]. Available: https://arxiv.org/abs/1512.03542

[93] M. Puri, Y. Pathak, V. Sutariya, S. Tipparaju, and W. Moreno, *Artificial Neural Network for Drug Design, Delivery and Disposition*. New York, NY, USA: Academic, 2015.

[94] T. T. Erguzel, S. Ozekes, O. Tan, and S. Gultekin, "Feature selection and classification of electroencephalographic signals: An artificial neural network and genetic algorithm based approach," *Clin. EEG Neurosci.*, vol. 46, no. 4, pp. 321–326, 2015.

[95] J. Guo, K. Yang, H. Liu, C. Yin, J. Xiang, H. Li, R. Ji, and Y. Gao, "A stacked sparse autoencoder-based detector for automatic identification of neuromagnetic high frequency oscillations in epilepsy," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2474–2482, Nov. 2018.

[96] P. Cerveri, A. Belfatto, G. Baroni, and A. Manzotti, "Stacked sparse autoencoder networks and statistical shape models for automatic staging of distal femur trochlear dysplasia," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 6, 2018, Art. no. e1947.

[97] Y. Qiu, W. Zhou, N. Yu, and P. Du, "Denoising sparse autoencoder-based ictal EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1717–1726, Sep. 2018.

[98] Z. Alhassan, D. Budgen, R. Alshammari, T. Daghstani, A. S. McGough, and N. Al Moubayed, "Stacked denoising autoencoders for mortality risk prediction using imbalanced clinical data," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 541–546.

[99] T. Katsuki, M. Ono, A. Koseki, M. Kudo, K. Haida, J. Kuroda, M. Makino, R. Yanagiya, and A. Suzuki, "Feature extraction from electronic health records of diabetic nephropathy patients with convolutioinal autoencoder," in *Proc. Workshops 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 451–454.

[100] S. Dubois, N. Romano, K. Jung, N. Shah, and D. C. Kale, "The effectiveness of transfer learning in electronic health records data," in *Proc. ICLR Workshop*, Toulon, France, Tech. Rep., 2017.

[101] H. Suresh, P. Szolovits, and M. Ghassemi, "The use of autoencoders for discovering patient phenotypes," 2017, *arXiv:1703.07004*. [Online]. Available: https://arxiv.org/abs/1703.07004
[102] M. Khademi and N. S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 727–732.
[103] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Syst. Appl.*, vol. 46, pp. 139–144, Mar. 2016.
[104] T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, "Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM)," *J. Biomed. Inform.*, vol. 54, pp. 96–105, Apr. 2015.
[105] K. H. Hoang and T. B. Ho, "Learning treatment regimens from electronic medical records," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2018, pp. 411–422.
[106] P. M. Shakeel, S. Baskar, V. R. S. Dhulipala, S. Mishra, and M. M. Jaber, "Maintaining security and privacy in health care system using learning based deep-Q-networks," *J. Med. Syst.*, vol. 42, no. 10, p. 186, 2018.
[107] Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang, and Y. Wang, "Deep reinforcement learning for dynamic treatment regimes on medical registry data," in *Proc. IEEE Int. Conf. Healthcare Inform. (ICHI)*, Aug. 2017, pp. 380–385.
[108] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," 2015, *arXiv:1511.03677*. [Online]. Available: https://arxiv.org/abs/1511.03677
[109] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, May 2016, Art. no. 160035.

GASPARD HARERIMANA (S'17) received the B.S. degree in computer engineering from the Ethiopian Defense University, in 2008, and the M.S. degree in information technology from Carnegie Mellon University, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Sangmyung University, Seoul, South Korea, where he is also a Research Assistant. He was a Visiting Lecturer with the Adventist University of Central Africa (AUCA), Kigali, Rwanda. His research interests include big data, machine learning, with the emphasis on deep learning.

JONG WOOK KIM (M'17) received the Ph.D. degree from the Computer Science Department, Arizona State University, in 2009. He was a Software Engineer with the Query Optimization Group, Teradata, from 2010 to 2013. He is currently an Assistant Professor of computer science with Sangmyung University. His primary research interests are in the areas of data privacy, distributed databases, and query optimization. He is a member of the ACM.

HOON YOO received the B.S., M.S., and Ph.D. degrees in electronic communications engineering from Hanyang University, Seoul, South Korea, in 1997, 1999, and 2003, respectively. From 2003 to 2005, he was with Samsung Electronics, South Korea, where he was involved in the world's first development of the satellite digital multimedia broadcasting (DMB) and the terrestrial DMB on mobile phones. He is currently a Full Professor with the Department of Electronics, Sangmyung University. His research interests are in the areas of 3-D integral imaging, computer vision, and digital signal processing.

BEAKCHEOL JANG (M'17) received the B.S. degree from Yonsei University, in 2001, the M.S. degree from the Korea Advanced Institute of Science and Technology, in 2002, and the Ph.D. degree from North Carolina State University, in 2009, all in computer science. He is currently an Associate Professor with the Department of Computer Science, Sangmyung University. His primary research interests include wireless networking, big data, the Internet of Things, and artificial intelligence. He is a member of the ACM.