# Applying Deep Learning to Individual and Community Health Monitoring Data: A Survey

Zhen-Jie Yao[1]        Jie Bi[1]        Yi-Xin Chen[2]

[1]Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China

[2]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

**Abstract:** In the recent years, deep learning models have addressed many problems in various fields. Meanwhile, technology development has spawned the big data in healthcare rapidly. Nowadays, application of deep learning to solve the problems in healthcare is a hot research direction. This paper introduces the application of deep learning in healthcare extensively. We focus on 7 application areas of deep learning, which are electronic health records (EHR), electrocardiography (ECG), electroencephalogram (EEG), community healthcare, data from wearable devices, drug analysis and genomics analysis. The scope of this paper does not cover medical image processing since other researchers have already substantially reviewed it. In addition, we analyze the merits and drawbacks of the existing works, analyze the existing challenges, and discuss future trends.

**Keywords:** Deep learning, healthcare, electronic health records (EHR), neural networks, survey.

## 1 Introduction

In the big data era, data has become more important than ever before. Meanwhile, rapid information technology improvement facilitates data processing. This trend is particularly obvious in the field of healthcare. Conventional data in heahthcare are all collected, standardized and stored in information system like hospital information system (HIS), laboratory information system (LIS), picture archiving and communication systems (PACS), etc., including demographic information, medical records, medical image, lab tests, medications, procedures and diagnosis, etc. Novel technologies such as wearable device, medical website, drug discovery and genetic testing enrich the categorization, scope and amount of the healthcare data. Intelligent analysis of these data is helpful for diagnosis, treatment, decision making support, prescription, disease prediction, curative effect evaluation, etc. It is an urgent task to extract valuable knowledge from the massive data in healthcare.

Deep learning (DL) can be traced back to around 2000. In recent years, deep learning outperforms traditional machine learning methods with a significant margin in image recognition, speech recognition and natural language processing[1]. The great success demonstrated its strong ability of modeling sophisticated data. Nowadays, deep learning is applied to more and more fields, and has

obtained many encouraging results. It has been shown that healthcare is one of the most promising directions.

Applying deep learning in healthcare results in many surprising works. The success of deep learning in the field of computer vision and image processing can be extended to the field of medical image processing directly. It is used for imaging, image segmentation, image recognition, lesion detection, etc. This inspired lots of works along the direction of medical image processing; hundreds of papers were published in the last 3–5 years. Litjens et al.[2] and Greenspan et al.[3] already gave substantial reviews on deep learning in medical image analysis. As mentioned above, health-care is more than medical image processing with a number of other aspects, therefore many researchers applied deep learning to analyze other types of data, and obtained significant improvements. This paper focuses on health-care areas other than medical image processing. By reviewing these works systematically, it is expected that this article could help the readers understand the general situation of this field. To our knowledge, the most similar review was given by [4]. Although their review covered many aspects, it was not extensive enough. For example, the most important improvement in recurrent neural networks (RNN), gated recurrent unit (GRU), is missing. Some latest important works are not included due to its publication time, such as electrocardiography (ECG) applications and electronic health records (EHR) analysis. Furthermore, we will share our own experiences with the application of deep learning on healthcare.

In this paper, all the works are categorized into 7 categories according to the data analyzed, which are EHR,

ECG, electroencephalogram (EEG), community health-care data, data from wearable devices, drug analysis and genomics analysis.

The remainder of this paper is organized as follows. Section 2 overviews commonly used deep learning algorithms, including their benefits and drawbacks. Section 3 reviews the relevant applications in healthcare in detail. Section 4 analyzes the limitations and challenges in this area. Finally, the conclusion of this study is drawn and the prospect of further development is discussed in Section 5.

## 2 Deep learning

Neural networks (NN) can be traced back to 1940 s, and was first implemented as a perceptron in 1950 s[5], which is a bionic inspired linear classifier. Perceptron started the first wave on NN research. However, early neural networks did not achieve very good performance, and it was found that perceptron has its own limitations. Research works on NN were revived when the multi-layer perceptron (MLP) was designed and trained with the back propagation (BP) algorithm in the 1980 s[6, 7]. These works gave rise to the second wave of NN research. Then, a NN usually had 3 layers, which were input layer, output layer and 1 hidden layer, it was referred to as the shallow model. Although these works did improve the model performance, NN has been shadowed by other machine learning methods represented by support vector machine (SVM) in 1990 s. Many factors kept NN from becoming deep architectures as we have today.

In fact, restrictions on the development of the neural network is not only the model itself, but also the limitations of hardware, such as memory capacity and computing power, and the volume of the training dataset. With rapid development of hardware, things changed gradually. Lecun et al.[8] developed the famous convolutional neural networks with multiple hidden layers. Hinton et al.[9] formally proposed the concept and method of deep learning. It was found that training a deep neural network is feasible. Meanwhile, rapid development of internet and mobile internet facilitated the collection of large dataset. The third wave of NN research has been set up, and its development has been rapid and enormous. Deep learning, which refers to NN with more than two hidden layers, got great success in computer vision and speech applications[1, 10, 11]. In the past 6 years, a variety of deep NN were developed and applied in various domains. In this paper, we group them into 5 categories, and introduce them one by one. Details about these deep NN could be found in [1, 4].

**Deep neural networks (DNN).** Generally, all the 5 kinds of deep learning models can be called deep neural networks. For clarity, in this paper, deep NN refers to all kinds of deep learning models. DNN specifically refers to the basic structure, which is the conventional NN with more than two hidden layers, as shown in Fig. 1. Training of DNN is difficult. Moreover, DNN is prone to suffer the problem of vanishing of the gradient, which means that in the BP training process, gradients (or errors) become negligible when it reaches the first several layers through many backward layers. Solving these problems led to the booming of the deep learning, and various kinds of deep NN.
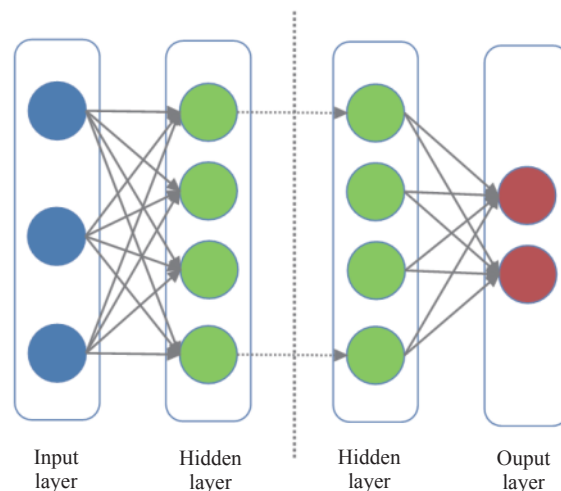


Fig. 1    Architecture of DNN

**AutoEncoder networks (AE).** AutoEncoder is an unsupervised model. The output layer of AE is not the label, but the input itself or its noisy version. Training of AE is to reconstruct the original input data. From the viewpoint of information theory, reconstructing original data means no loss of information. In other words, training of AE is searching the best coder to minimize information loss. The coder is the compressed data with least information loss. It is an optimal representation of original data. Because the coder is with low dimension and discriminative, AE is well suited for feature extraction. Some researchers used AE to extract features from various kinds of healthcare data[12–15].

**Restricted Boltzmann machine (RBM) related networks (DBN and DBM).** Restricted Boltzmann machine[16] is a variant of Boltzmann machine, which mimics the conception of statistical thermodynamics, the neural variables range from 0 to 1. Restriction on Boltzmann machine leads to a bipartite graph structure. Restriction simplifies the network structure, which makes the training procedure of RBM more tractable than Boltzmann machine. Two kinds of deep NN frameworks can be derived from RBM, deep belief networks (DBN)[9] and deep Boltzmann machines (DBM)[17]. Both DBN and DBM are initialized by layer-wise greedy training of RBM, and fine-tuned by target labels.

**Convolutional neural networks (CNN).** CNN consists of interleaved convolution layer and pooling (subsample) layer, which is robust to shifting, rotation and

scaling[8]. CNN is the most commonly used deep learning model. Its success on ImageNet[10] is the tipping point of deep learning research boom. Although CNN is designed for image recognition[8], it can be used on any fixed, ordered, and location related data. For example, it is convenient to apply CNN to 1-d time sequence with fixed length and 3-d video.

**Recurrent neural networks (RNN).** Generally speaking, all the networks mentioned above are used as static models with fixed inputs. RNN[18] is a dynamic model, whose output is determined not only by the current states; but also by the previous state. RNN suffers the same problem as other NNs, which is vanishing of the gradients. Furthermore, besides vanishing layer by layer, gradients of RNN also vanish along time. To overcome this problem, Hochreiter and Schmidhuber[19] replaced the node of RNN by long short-term memory (LSTM). Chung et al.[20] proposed a simpler gated recurrent unit (GRU), which get similar performance as LSTM. RNN got success in many tasks for sequential data processing, such as language modeling for speech and text[11].

## 3 Healthcare applications

In this section, we introduce the application of deep learning in 7 healthcare areas, which are EHR, ECG, EEG, community healthcare, data from wearable devices, drug analysis and genomics analysis.

### 3.1 EHR

Electronic health records (EHR), or electronic medical records (EMR), is the systematized collection of patient-centered health information in a digital format. EHRs may include a range of data, including demographics, medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital sign, etc. It is the most extensive and important data source for healthcare.

Liang et al.[21] applied DBN for unsupervised feature extraction, and then performed supervised learning through a standard SVM for healthcare decision making. A dataset on Chinese medical diagnosis and treatment prescription and a dataset on hypertension retrieved from EHR were used to test the model. The experimental results indicate that the proposed deep model performed much better than the conventional shallow models, such as SVM and decision trees.

Che et al.[22] proposed a novel knowledge distillation methodology called interpretable mimic learning (IML), where they mimicked the performance of state-of-the-art deep learning models with gradient boosting trees (GBT). For mortality prediction and ventilator free days prediction tasks in intensive care units (ICUs), IML and deep learning give much better results than conventional methods such as SVM, logistic regression, decision tree and

GBT. It was shown that the IML can mimic deep learning architectures including stacked denoising autoencoders (SDAE) and LSTM well, and can give comparable or even better results. As we know, the GBT is much more interpretable than deep model.

Jagannatha and Yu[23] took bidirectional RNN for medical event detection in EHR. The dataset is 780 English EHR notes of 613 593 word tokens. The annotated events can be broadly divided into two groups, medication and disease. The medication group contains drugname, dosage, frequency, duration and route. The disease group contains events related to diseases (adverse drug events (ADE), indication, etc.) and their attributes (severity). The RNN could be composed of LSTM or GRU. Test results showed that RNN with GRU got the best performance. The F1-score is 0.803 1, compared to the baseline 0.723 0, which is got by conditional random field (CRF).

Miotto et al.[12] proposed an unsupervised representation of patient from the EHRs, named Deep Patient, which is a three-layer SDAE. This representation is capable to capture hierarchical regularities and dependencies in the aggregated EHRs. This model was evaluated by predicting health states by assessing the probability of patients to develop various diseases. Test results on a dataset composed of 76 214 test patients validated the effectiveness of this representation. The area under curve (AUC) of Deep Patient is 0.773, compared to that of the best conventional representation (independent component analysis, ICA), which is 0.695.

Lipton et al.[24] presented LSTMs for multi-label classification to classify 128 diagnoses given 13 frequently but irregularly sampled clinical measurements. This model was evaluated on a dataset consisting of 10 401 ICU episodes, where each episode consists of multivariate time series of 13 variables. Episodes vary in length from 12 hours to several months. Trained on raw time series, the proposed model gave comparable performance as a multilayer perceptron trained on expert hand-engineered features. The AUC is 0.807 5 for LSTM and 0.803 0 for multilayer perceptron, respectively.

Esteban et al.[25] presented an RNN based on GRU that is specifically designed for the clinical domain, which combines static and dynamic information in order to predict future events. This model was evaluated on a database collected in the Charit′e Hospital in Berlin, containing EHRs of patients that underwent a kidney transplantation. The model was adopted to predict whether any of three endpoints will occur within the next six or twelve months after each visit to the clinic. That is rejection of the kidney, loss of the kidney and death of the patient. The AUC of the proposed model is 0.833, while the AUC of the logistic regression is 0.808.

Che et al.[26] developed a novel deep learning model, namely GRU-D, which is improved GRU. It takes two representations of missing patterns and effectively incor-

porates them into a deep model architecture so that it not only captures the long-term temporal dependencies in time series, but also utilizes the missing patterns to achieve better prediction results. Experimental results on real-world clinical datasets (MIMIC-III, PhysioNet) and synthetic datasets demonstrate that the models achieve state-of-the-art performance. For mortality prediction on MIMIC-III dataset, the AUC of GRU-D, GRU, and random forest are 0.852 7, 0.838 0 and 0.829 4, respectively.

Mehrabi et al.[27] attempted to discover the temporal pattern and association rule of the diagnosis codes. The authors modeled each patient′s records as a matrix of temporal clinical events with international classification of diseases-9 th version (ICD-9) diagnosis codes as rows and years of diagnosis as columns. A deep Boltzmann machine network with three hidden layers was constructed with each patient′s diagnosis matrix values as visible nodes.

Futoma et al.[28] compared predictive models for early hospital readmission. They focused this analysis on the five patient cohorts. That is chronic obstructive pulmonary disorder (COPD), heart failure (HF), pneumonia (PN), acute myocardial infarction (AMI) and total hip arthroplasty/total knee arthroplasty (THA/TKA). Penalized logistic regression (PLR) showed the best performance in traditional methods. DNN outperformed PLR on all the 5 diseases. Take HF for example, the AUC is 0.676 versus 0.654.

Putin et al.[29] designed a modular ensemble of 21 deep NN of varying depth, structure and optimization to predict human chronological age using a basic blood test. The input of the model is blood biochemistry records, including age, sex, and 46 standardized blood markers, the output is estimated age. The best performing DNN in the ensemble demonstrated MAE (mean absolute error) = 6.07 years in predicting chronological age within a 10 year frame, while the entire ensemble achieved MAE = 5.55 years. Furthermore, the ensemble also identified top 5 markers for predicting human chronological age: albumin, glucose, alkaline phosphatase, urea and erythrocytes.

Cheng et al.[30] predicted onset risk of congestive heart failure (CHF) and COPD half year later using all the historical patient records in EHR. First, the data was formatted as a matrix, with time as one dimension and event as the other dimension. Specifically, time dimension is number of days, and event dimension is onset of medical event categorized by ICD9. Then, a CNN with 4 layers was adopted to predict the risk. Moreover, the authors also designed 3 fusion strategies to handle the dynamic information in EHR, where a slow fusion method gave the best performance. Logistic regression is the baseline. For CHF, the AUC is 0.767 5, compared to that of baseline as 0.715 6. For COPD, the AUC is 0.738 8, compared to that of baseline as 0.662 4.

Choi et al.[31] modeled temporal relations among events in electronic health records (EHRs) to predict ini-

tial diagnosis of heart failure (HF) with deep learning. Each clinical event in EHR data was represented as a one-hot vector format, and then each EHR was a computable event sequence. RNN models using GRU were adapted to predict HF onset. When using an 18-month observation window, the AUC for the RNN model increased to 0.883 and was significantly higher than 0.834, the AUC for the best of the baseline methods (MLP).

Pham et al.[32] modeled the EHR data with a deep dynamic memory model named DeepCare, which is composed of 3 levels. The first level is a series of LSTM; the second level is multiscale pooling; the third level is conventional fully connected layer. The model was tested on two cohorts with heavy social and economic burden: diabetes and mental health for 3 tasks: disease progression modeling, intervention recommendation, and future risk prediction. Test results show significant improvement in modeling and risk prediction.

Avati et al.[33] predicted the mortality of a patient within 12 months by modeling the EHR data of one year with a DNN. Accurate prediction of mortality would help the medical care staffs improve palliative care. The dataset consists of 221 284 selected EHRs. The model achieved recall of 0.34 at 0.9 precision. The AUC was 0.93.

Rajkomar et al.[34] analyzed the EHR extensively. First, the EHR data is converted to events recorded in containers based on FHIR (Fast Healthcare Interoperability Resources) and placed in temporal order, which served as the input of deep learning model for all the 4 tasks. The 4 tasks are inpatient mortality, 30-day unplanned readmission, length of Stay and discharge diagnosis. Three different network architectures were designed: one based on recurrent neural networks (LSTM), one based on an attention-based time-aware neural network (TANN) model, and one based on a neural network with boosted time-based decision stumps. Each model was trained separately, and the final model is an ensemble of predictions. The baselines are enhanced version of conventional scores, which are commonly used in hospitals. AUC of all the 4 tasks are improved consistently. For inpatient mortality prediction, the AUC of deep learning is 0.95, while that of baseline is 0.85.

Dernoncourt et al.[35] designed the first de-identification system based on a deep artificial neural network with LSTM as element. The model was tested on two largest public available de-identification datasets. This model outperforms the state-of-the-art systems. It yielded an F1-score of 97.85% on the i2b2 2014 dataset, with a recall of 97.38% and a precision of 97.32%, and an F1-score of 99.23% on the MIMIC de-identification dataset, with a recall of 99.25% and a precision of 99.06%.

**Discussion.** EHR is often regarded as the main body of healthcare data. In a sense, all kinds of healthcare data are components of EHR. Therefore, there are many researches on EHR. Table 1 lists the application fields, input data, and deep models of the above mentioned pa-

pers. As shown in Table 1, DL models have been applied to nearly all the fields in healthcare, including conventional applications such as intelligent diagnosis, disease risk prediction, and novel applications such as chronological age prediction and EHR de-identification. We can also find in Table 1 that most researchers modeled the EHR data as a sequence of events. RNN is the most effective and popular DL model for EHR analysis. Other DL models such as CNN, DNN, SDAE, DBN and DBM also work on some specific tasks for EHR analysis.

Table 1   Summary of applications in EHR

| Application | Input | Model | References |
| --- | --- | --- | --- |
| Intervention recommendation/ Decision making | EHR | DBN+SVM | [21] |
| | | RNN | [32] |
| Mortality prediction | | IML(GBT mimics SDAE&RNN) | [22] |
| | | RNN | [26, 33] |
| | | Ensemble of deep NN | [34] |
| Medical event prediction | | IML(GBT mimics SDAE&RNN) | [22] |
| | | RNN | [23, 25, 26] |
| | | Ensemble of deep NN | [34] |
| Intelligent diagnosis | | RNN | [24, 32] |
| | | Ensemble of deep NN | [34] |
| Disease risk prediction | | CNN | [30] |
| | | SDAE | [12] |
| | | RNN | [31, 32] |
| Hospital readmission prediction | | DNN | [28] |
| | | Ensemble of deep NN | [34] |
| Pattern and association rule discovery | | DBM | [27] |
| EHR De-identification | | RNN | [35] |
| Chronological age prediction | Blood test | Ensemble of deep NN | [29] |

## 3.2 ECG

Electrocardiography (ECG) is a noninvasive recording of the electrical activity of heart, with electrodes being placed on skin. Arrhythmia would cause various types of irregular heart rhythm or pattern on ECG.

Chauhan and Vig[36] utilized stacked LSTM to format a deep recurrent neural network. The deep recurrent neural network estimated the probability of the input ECG is normal. The output distribution was modeled by a multivariate Gaussian distribution, and a maximum likelihood estimator checked whether it is a normal or abnormal ECG. It is worth mentioning that, the deep model was trained by normal ECG only. Four different types

of arrhythmias could be detected, namely premature ventricular contraction (PVC), atrial premature contraction (APC), paced beats (PB) and ventricular couplet (VC). The dataset is extracted from MIT-BIH arrhythmia database. The overall precision and recall are 0.975 0 and 0.464 7, respectively.

Yan et al.[37] constructed a deep belief network and the restricted Boltzmann machine (RBM) based algorithm was used in the ECG classification problem. The algorithm was evaluated on the two-lead ECG dataset of MIT-BIH dataset and got the performance with accuracy of 98.829%. However, the dataset is not strictly person independent. Although there is no overlap between the training set and testing set, they may be from the same person. Person independent is a more practical scenario.

Al Rahhal et al.[14] extracted a suitable feature representation from the raw ECG data in an unsupervised way using SDAEs with sparsity constraint. A softmax regression layer was added on the top of the resulting hidden representation layer. During the interaction phase, the expert labeled the most relevant and uncertain ECG beats of the test record during each iteration. The labeled data were used for updating the model weights. Active learning like this is a promising direction for ECG analysis. The only drawback is that the active learning system needs experts for interaction, so it cannot run automatically. Results usually depend on how much data the expert could label.

Acharya et al.[38] designed a deep CNN with 11 layers to detect myocardial infarction. Each beat of the ECG was segmented as a sample, which is input of the CNN. The CNN checked whether this beat is normal or myocardial infarction. The model was tested on data with and without noise. The average accuracy was 93.53% and 95.2% using ECG beats with and without noise respectively. However, the dataset was not person-independent.

Yao et al.[39] designed a multi-scale convolutional neural networks (MCNN). First, R wave was detected from ECG, and RR (R wave to R wave) interval sequence was the input of the MCNN for atrial fibrillation (AF) detection. The algorithm was tested on both public and private datasets. They got the best detection performance on the public dataset. The accuracy is 98.18%, compared to that of conventional method 97.99%. Test result on private dataset showed that the deep learning is more sensitive to the difference between training and testing datasets. Transfer learning is promising for solving this problem.

Rajpurkar et al.[40] collected the largest ECG dataset. 64 121 ECGs were recorded by a single lead wearable monitor from 29 163 patients. A special CNN named ResNet with 34 layers was built. 14 kinds of rhythms (13 arrhythmia and 1 normal) were recognized. They got cardiologist-level arrhythmia detection performance. The aggregated F1-score is 0.809, compared to the averaged F1-score of 6 certified cardiologists is 0.751.

🖄 Springer

**Discussion.** These works on ECG are listed in Table 2. From these papers, we can find that ECG can be used for heart related disease diagnosis and detection. RNN, SDAE, RBM and CNN have been tried for these tasks. The input data could be ECG data itself, or its derivatives (such as RR-intervals). According to the results reported by the works, CNN is the most effective DL model for ECG analysis.

Table 2    Summary of applications in time series

| Application | Input | Model | References |
|---|---|---|---|
| Intelligent diagnosis | ECG | DBN | [37] |
| | | SDAE | [14] |
| | | CNN | [40] |
| Disease detection | | RNN | [36] |
| | | CNN | [38, 39] |
| | EEG | DBN | [41, 42] |
| Affective state recognition | | RBM | [43] |
| Mind reading | | DNN | [44] |
| | | CNN | [45] |

## 3.3 EEG

Electroencephalogram (EEG) is a noninvasive recording the electrical activity of the brain, with electrodes placed on scalp. Anomalous electroencephalography indicates brain function problem. EEG could be used to diagnose conditions such as seizures, epilepsy, head injuries, dizziness, headaches, brain tumors and sleeping problems. It can also be used for simple thought reading.

Wulsin et al.[41] proposed a DBN approach to detect anomalies in EEG waveforms. This model was tested by a large set of training data. DBNs outperform traditional one class SVM. The F1-score of DBN is 0.475 2, compared to that of one class SVM 0.439 0. They also presented how the outputs of a DBN-based detector can be used to aid visualization of anomalies in large EEG data set.

Page et al.[42] explored the use of a variety of representations and machine learning algorithms to detect seizure. They compared conventional methods such as $k$-nearest neighbor (KNN), SVM, Logistic regression with DBN, and DBN gave the best performance.

Jia et al.[43] proposed a novel semi-supervised deep learning framework for affective state recognition. First, supervised label information and unsupervised structure information jointly made decision on channel selection. A generative restricted Boltzmann machine (RBM) model was adopted for the classification task. An active learning sketch was taken to solve the costly labeling problem. Two kinds of affective states were recognized. Although the sample sets were small (32 participants), the proposed method achieved better accuracy.

Sturm et al.[44] applied DNNs with layerwise relevance propagation (LRP) for EEG data analysis. Through LRP, DNN decisions were transformed into heatmaps indicating relevance of data for the outcome of the decision. The single-trial LRP heatmaps reveal neurophysiologically plausible patterns, resembling conventional common spatial pattern (CSP)-derived scalp maps.

Schirrmeister et al.[45] studied deep convolutional neural networks with a range of different architectures, designed for decoding imagined or executed movements from raw EEG. The test results showed that the CNN methods reach or surpass that of the widely-used filter bank common spatial patterns (FBCSP) decoding algorithm. The two methods were compared on recognizing 4 movements, which are of the left hand, the right hand, both feet, and rest. The accuracy of deep CNN is 92.40%, while that of FBCSP is 91.15%. Moreover, they highlighted the potential of deep CNN combined with advanced visualization techniques for EEG-based brain mapping.

**Discussion.** The works on EEG analysis are listed in Table 2. From these papers, we can find that EEG can be used for brain related disease detection, affective recognition and mind reading. The input data is the original EEG. DBN, DNN, RBM and CNN have been tried for these tasks. According to the results, CNN is the most effective DL model for EEG analysis. Combined with the results of ECG analysis, we can find that although DBN is applied to ECG and EEG analysis very early (2010–2015), the performance is unsatisfactory. CNN is good at handling uniformly sampled data, such as image and time series (ECG and EEG).

## 3.4 Community healthcare

Social media data record the activity on internet, which is a novel field for healthcare. It is a powerful supplement of traditional healthcare data. Social media is helpful for monitoring of the mental health status and the spread of infectious diseases.

Nie et al.[46] inferred diseases from Q&A on health-related web. There scheme builds a sparsely connected deep architecture with three hidden layers with sparse constraint. Disease inference is helpful for health seeker online. The inference result was tested on dataset collected from medical Q&A website. Result showed that the method outperforms SVM, KNN, Decision tree, Naive Bayes and deep NN composed of stacked autoencoder and softmax.

Zhao et al.[47] designed a social media nested epidemic simulation model. Twitter data were used to continuously track health states from the public. DNN was used to mine epidemic features that are combined into a simulated environment to model the progression and diffusion of disease.

Zou et al.[48] employed a deep learning approach for creating a topical vocabulary, and then applied a linear

elastic net as well as a nonlinear Gaussian process for inference. Test results indicated that Twitter data contain a signal that could be strong enough to complement conventional methods for infectious intestinal disease surveillance. Such method could be extended to other infectious diseases.

Benton et al.[49] developed a deep neural multi-task learning (MTL) model for 10 prediction tasks (suicide, seven mental health conditions, neurotypicality, and gender), which map the Twitter data to each kind of problem. The MTL model is compared with single task learning (STL) models. Results showed that an MTL model performs significantly better than other models for all the tasks. For example, the AUC for suicide is 0.848. According to the result, we infer that MTL with related tasks is helpful for improving model performance, compared to independent STL.

**Discussion.** The works on community healthcare are listed in Table 3. From these papers, we have seen that the activities and texts on the internet, especially on social networks, can be used for physical disease inference, infectious disease surveillance and mental health monitoring. DNN is the most frequently used model for these tasks. Novel models such as RNN should be attempted for these tasks. Moreover, it is found that MTL model outperforms STL model. We can also see that training with more tasks typically improves the performance of the DL models.

Table 3   Summary of applications in novel healthcare data

| Application | Input | Model | References |
| --- | --- | --- | --- |
| Disease inference | Health Q&A | Stacked AE+softmax | [46] |
| Infectious disease surveillance | Social media | DNN | [47, 48] |
| Mental health monitoring | | DNN | [49] |
| Motion analysis | Accelerometer (W) | RBM | [50] |
| | | CNN | [51] |
| Disease detection | PPG(W) | RNN | [52] |
| Emotion recognition | Respiration (W) | SAE | [53] |
| Drug activity analysis | Structure | DNN | [55, 56] |
| | | AE with GRU | [15] |
| Adverse drug effects classification | ADE related texts | CNN & deviations | [57] |
| Cancer & survival analysis | Genomics | DNN | [59] |
| | | SDAE | [58, 13] |

## 3.5   Wearable devices

Wearable devices are smart electronic devices worn on the body, which can capture data consecutively. Data captured from wearable devices are valuable for they are captured in daily life, not in a special environment like

hospital. It is helpful for task such as disease monitoring and activity recognition.

Hammerla et al.[50] assessed the states of Parkinson's disease by movement data collected in naturalistic settings. 2 wearable sensing devices, with a tri-axial accelerometer inside, were used to capture movement data. The sensors were worn on each wrist of the participant. 91 hand-craft features were extracted from each minute of the sensor recordings. These features were used to train a sequence of RBM, with a softmax top-layer to assess the state of Parkinson's disease. The mean F1-score is about 0.55, compared to F1-score of decision tree which is about 0.4. The authors also mentioned that this handcrafted feature extraction would be substituted with a convolutional architecture in the future.

Ravi et al.[51] presented a human activity recognition technique based on deep learning methodology, which is designed to enable accurate and real-time classification for low-power wearable devices. It is worth noting that all the inertial data was collected without any constraints. To obtain invariance against changes in sensor orientation, sensor placement, and in sensor acquisition rates, a feature generation process was applied to the spectral domain of the inertial data. To reduce the computation demands, a CNN with constraints was adopted to analyze the spectrum for activity recognition. The proposed method outperformed traditional methods on 2 out of 4 datasets. The performance is not very extraordinary, the main point is that the deep learning method is implemented on a very resource constrained device.

Aliamiri and shen[52] utilized the wearable device with build-in photo-plethysmography (PPG) sensor to provide a portable, non-intrusive and low-cost solution for AF monitoring and detection. An end-to-end deep learning AF detection system was built based on CNN, which can filter out poor quality signals and make reliable AF detection. The models achieved over 95% AUC in quality assessment task and over 99% AUC in AF detection task.

Zhang et al.[53] proposed a DL framework adopting sparse auto-encoder (SAE) to extract emotion-related features, and logistic regression for emotion recognition. One task was arousal classification and the other was valence classification. Only respiration data collected from wearable devices was used for recognizing human emotions. The accuracy was about 80%.

**Discussion.** The works on analyzing data from wearable devices are listed in Table 3. Data captured by wearable devices can be used for disease detection, disease assessment, emotion recognition and activity recognition. RBM, SAE and CNN did not get very impressive performance on these tasks. A key limitation is that DL model is computation-intensive, which can hardly work for wearable devices with limited battery, memory and computational capacity. Tradeoff between resources and performance is necessary. Some hardware specifically designed for DL (such as [54]) may be helpful to overcome

this problem.

## 3.6 Drug & compound analysis

Conventional drug discovery is an extended process that takes years to finish. Drugs should not only cure the disease, but also restrict toxicity and adverse drug reaction. This can be inspected by molecular structure analysis, and drug related records mining.

Unterthiner et al.[55] built a system named DeepTox. DeepTox normalized the chemical representations of the compounds. Then, it computed a large number of chemical descriptors that are used as input to machine learning methods. DNN with 5 layers was used to predict the toxicity of the compounds. Multi-task learning was taken to enhance the performance. A dataset of 12 000 environmental chemicals and drugs was measured for 12 different toxic effects. This system was compared with traditional machine learning method such as naive Bayes, support vector machines, and random forests. Deeptox outperformed all the other methods, the average AUC of 12 toxic effects was 0.846.

Ma et al.[56] took DNN to model the quantitative structure activity relationships (QSAR). QSAR is a commonly used technique in the pharmaceutical industry for predicting on-target and off-target activities. Such predictions will reduce the experimental work that needs to be done during the drug discovery process. The metric to evaluate prediction performance is $R^2$, the squared Pearson correlation coefficient between predicted and observed activities in the test set. The mean $R^2$ of DNN is 0.496, compared to $R^2$, of random forest as 0.423. The authors also found that the DNN for QSAR is generalized. DNN with recommended parameters can achieve better performance than random forest for most of the datasets.

Xu et al.[15] proposed a deep auto-encoder (DAE) network for molecular representation. A multi-layered gated recurrent unit (GRU) network was used to map the input molecule into a continuous feature vector of fixed dimensionality, and then another deep GRU network was employed to decode the continuous vector back to the original molecule. Such an auto-encoder framework was expected to get the continuous encoding vector, which contains enough information to recover the original molecule (no information loss) and predict its chemical properties. The resulting continuous feature vector was fed into Adaboost, gradient boost and random forest, respectively, for chemical properties prediction. In the wateroctanol partition coefficient experiment, the accuracy is 0.766 4, compared to that of traditional method 0.608 0.

Huynh et al.[57] investigated different deep NN architectures for adverse drug effects (ADE) classification. In particular, they proposed two new neural network models, convolutional recurrent neural network (CRNN) by concatenating convolutional neural networks with recur-

rent neural networks, and convolutional neural network with attention (CNNA) by adding attention weights into convolutional neural networks. Various deep NN architectures were evaluated on a Twitter dataset containing informal language and an adverse drug effects (ADE) dataset constructed by sampling from MEDLINE case reports. All the deep NN models outperformed traditional ones, where CNN came out to be the best model. The AUC of CNN on Twitter dataset is 0.88, compared to the best traditional method with AUC of 0.85. On MEDLINE dataset, it is 0.97 versus 0.95.

**Discussion.** The works on drug & compound analysis are listed in Table 3. Chemical structure and ADE related online texts were adopted to predict the chemical properties of the drug & compound. First of all, the data should be preprocessed for the DL model. DNN is the most popular model for chemical structure analysis. CNN is the best model for ADE detection task, even it can be compared to some carefully designed DL models, which implies that CNN is so powerful that it is difficult to improve it.

## 3.7 Genomics analysis

Genomics analysis is the identification, measurement or comparison of genomic features such as deoxyribonucleic acid (DNA) sequence, structural variation, gene expression, or regulatory and functional element annotation at a genomic scale. Despite model architecture complexity and mathematical connotation reconditeness, deep learning methods have shown promising practical values in genomic researches, especially in cancer detection and survival prediction.

Chaudhary et al.[58] studied survival expectations among different subgroups of hepatocellular carcinoma (HCC) by integrating multi-omics data of various patient cohorts. They built a deep learning model (autoencoder) by training 360 HCC patients′ data gathered from the Cancer Genomic Atlas (TCGA) database. They found that mutations of TP53, KRT19, EPCAM and BIRC5 genes, and activated Wnt and Akt signaling pathways are strongly associated with patient′s survival expectation, which was validated by five external datasets containing various cohorts and omics types. They had expected their workflow to be practical for HCC prognosis prediction, as it is the first study employing deep learning to identify multi-omics features in order to predict differential survival of HCC patients.

In another study which was focused on detecting and identifying breast cancer biomarkers, deep learning approaches showed their practical values as well. Danaee et al.[13] used SDAE to deeply extract functional features from gene expression profiles, and then evaluated the performance of the extracted representation through supervised classification models to verify the usefulness of the new features in cancer detection. As a result, they identi-

Table 4    Insights of DL in healthcare

| Architecture | Applications | Scope | Comments |
| --- | --- | --- | --- |
| DNN | EHR, EEG, community, drug, genomics | General | Conventional, various tasks |
| AE | EHR, ECG, community, wearable, drug, genomics | General | Feature extraction, usually combined with other models |
| DBN&DBM | EHR, ECG, EEG, wearable | General | Various tasks, sometimes as feature extractor |
| CNN | EHR, ECG, EEG, wearable, drug | Specific | Uniformly sampled data, such as image & time series |
| RNN | EHR, ECG, wearable | Specific | Sequences |

fied a set of highly interactive genes to be useful cancer biomarkers for the detection of breast cancer.

Deep learning methods have to digest high-dimensional data generated by genomic platforms in most cases, which conveys either low accuracies of predicting outcomes or expensive costs in selecting, labeling and verifying such numerous data. Yousefi et al.[59] had a strong confidence in that deep learning methods could be remarkably successful in cancer prediction tasks by using general high-dimensional data. Hence, they compared the result of survival analysis between Bayesian optimized deep survival models and other state-of-the-art machine learning methods. They eventually improved prognostic accuracy by transferring information across diseases.

Nevertheless, deep learning methods can be employed to pursue precision and personalized medical care by finding comprehensive and reasonable biomarkers of certain diseases. Deep learning architectures have great potential to integrate and analyze various data from different sources, e.g., DNA sequence data, gene expression data, protein structure data, etc.

**Discussion.** The works on genomics analysis are listed in Table 4. Genomics analysis is employed for cancer detection and survival prediction. AE and DNN are used to model the genomic data for these tasks. Novel models such as CNN and RNN should be attempted for these tasks. Because genomic data has strong relationship with its neighborhood, CNN and RNN may work better for genomics analysis.

## 4  Insights

After reviewing all the works, we give some insights of DL models in healthcare in Table 4. The DL models are divided into general models and specific models. General models refer to those models that are capable to be used in various tasks, and compatible with input data in various formats. Specific models refer to those models that are designed to accomplish certain task, and the input data is strictly restricted.

DNN is conventionally a general model, which has been used in nearly all the tasks. As mentioned in Section 2, AE is generally used for feature extraction, which is an unsupervised learning model. However, it is usually combined with other supervised learning models to accomplish certain tasks. AE is the most general deep learning model, it could be used in almost all kinds of tasks. DBN\DBM is general model, too. DBN\DBM can be used in various tasks, and is compatible with input data in various formats.

CNN and RNN are specific models, which are designed for specific tasks. CNN is first designed for image classification, and it is easily extendible to time series, thus we conclude that CNN is good at handling uniformly sampled data. Data in other formats has to be transformed to similar structure if CNN is adopted to handle it. As for RNN, it is designed to handle sequences, such as speech, language and text. RNN is easy to be extended to any event sequences like EHR. If the input data is not a sequence, it has to be transformed to a sequence first. It is worth noting that, since RNN is more robust to local (short time) change, the input of RNN is not necessary to be uniformly sampled.

## 5  Challenges and limitations

Although deep learning models have shown their power in so many healthcare applications, there are still a few major challenges. We summarize them as follows:

**Data.** The deep NN models are data driven. The number of model parameters is much higher than that of conventional models. Huge volume of data is needed to train the models. However, in healthcare applications, data collection is not easy, a dataset with 10 000 samples is often considered large and is hard to get large. This scale is small if compared to Imagenet dataset, which has 14 197 122 images. Furthermore, data in healthcare is often unbalanced. For disease screening tasks, patients with target outcome are typically scarce compared to healthy cohorts. Building big, representative dataset is an important and time-consuming task.

**Interpretability.** In the tasks of image recognition and speech recognition, we care more about whether the model works accurately, and care less about why it works. Although some visualization of the feature maps may help us understand the intermediate results, most of the deep NNs are end-to-end black-boxes and not interpretable. In healthcare task, interpretability is far more important than other applications. We need to analyze risk factors, and find out what is the best treatment for certain disease. Che et al.[22] attempted to distill knowledge from deep NN by Interpretable Mimic Learning.

This idea is helpful for model interpretation. Furthermore, data visualization is still the most powerful tool for model interpretation.

**Data representation.** In many traditional learning tasks, such as image or speech, the data is homogeneous and neat. In healthcare, some data (such as EHR) are irregular and poorly structured. Moreover, many data often have missing values. For example, vital signs manually collected at hospitals often have missing fields. At present, EHR data is represented as a temporal-events matrix and missing values are often handled by simple interpolation[27, 30]. A better representation may help to improve the performance.

**Generalization ability.** In other applications, the difference between training dataset and testing dataset is not significant, because the training and testing datasets are often generated by the same or similar underlying distribution. Therefore, the learned models generalize well. In healthcare, the difference between training and testing data is prone to be significant. For example, model training on dataset from Americans would not suit for Chinese[39]. Transfer learning would be helpful. For deep NN, if the model is trained based on Chinese population, corresponding adjustment is needed when the model is applied to other populations. Transfer learning is a powerful method for such settings and its uses in healthcare can be a future direction of research.

**Computational complexity.** The deep NNs are one of the most complex machine learning models. They are time, space, and memory consuming. Lots of parameters need to be stored in memory. Huge number of operations takes time for model running. This problem is especially severe for wearable devices with limited memory, computing power and battery. Simplified model may be helpful, Che et al.[22] have given a good attempt. Chen et al.[60, 61] have tried to reduce complexity in deep NN, which shrinks the storage requirements of neural networks substantially while mostly preserving generalization performance. Another solution is to design energy-efficient hardware accelerators for deep NNs. For example, on a number of representative neural network layers, it is possible to achieve a speedup of 450.65x over a graphics processing unit (GPU), and reduce the energy by 150.31x on average for a 64-chip DaDianNao system[54].

## 6 Conclusions

In this paper, we have surveyed recent applications of deep learning in healthcare areas, including EHR, ECG, EEG, community healthcare, data from wearable devices, drug analysis and genomics analysis. We have shown that deep learning has achieved remarkable results in these areas. We have shown that, although deep learning first got its success in computer vision and speech, it also has shown great potential in promoting the revolution in the healthcare industry. The unmatched learning ability of

deep learning has made it an attractive and indispensable technology for analyzing clinical and healthcare data. One interesting future direction is using deep learning to learn from multi-dimensional, complex, and non-structural personal data, such as demographics, diets, habits, sleeping, mental health, medical imaging, vital signs, medication, lab tests, etc. Such fusion of information can lead to new breakthrough in data-driven healthcare decision making.

## Acknowledgements

## References

[1] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature 14539.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[3] H. Greenspan, B. van Ginneken, R. M. Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016. DOI: 10.1109/TMI.2016.2553401.

[4] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G. Z. Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017. DOI: 10.1109/JBHI.2016. 2636665.

[5] F. Rosenblatt. The Perceptron: A Perceiving and Recognizing Automaton, Report 85-60-1. Cornell Aeronautical Laboratory, Buffalo, New York, USA, 1957.

[6] P. J. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, Ph. D. dissertation, Harvard University, Harvard, USA, 1974.

[7] D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning representations by back-propagating errors. *Cognitive Modeling*, vol. 5, no. 3, pp. 533–536, 1988.

[8] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[9] G. E. Hinton, S. Osindero, Y. W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006. DOI: 10.1162/neco.2006.18.7. 1527.

[10] A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems*, Lake Tahoe, USA, pp. 1097–1105, 2012.

[11] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury. Deep neural networks for acoustic

modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. DOI: 10.1109/MSP.2012.2205597.

[12] R. Miotto, L. Li, B. A. Kidd, J. T. Dudley. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, vol. 6, Article number 26094, 2016. DOI: 10.1038/srep26094.

[13] P. Danaee, R. Ghaeini, D. A. Hendrix. A deep learning approach for cancer detection and relevant gene identification. In *Proceedings of Pacific Symposium on Biocomputing*, World Scientific, Kohala Coast, USA, 2017.

[14] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R. R. Yager. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, vol. 345, pp. 340–354, 2016. DOI: 10.1016/j.ins.2016.01.082.

[15] Z. Xu, S. Wang, F. Y. Zhu, J. Z. Huang. Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Boston, USA, pp. 285–294, 2017. DOI: 10.1145/3107411.3107424.

[16] G. E. Hinton, T. J. Sejnowski. Learning and relearning in Boltzmann machines. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, Eds., Cambridge, USA: MIT Press, pp. 1, 1986.

[17] R. Salakhutdinov, G. E. Hinton. Deep Boltzmann machines. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, Clearwater Beach, USA, pp. 448–455, 2009.

[18] R. J. Williams, D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989. DOI: 10.1162/neco.1989.1.2.270.

[19] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[20] J. Chung, C. Gulcehre, K. Cho, Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. https://arxiv.org/abs/1412.3555, 2014. (*arXiv: 1412.3555*)

[21] Z. Liang, G. Zhang, J. X. Huang, Q. V. Hu. Deep learning for healthcare decision making with EMRs. In *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine*, Belfast, UK, pp. 556–559, 2014. DOI: 10.1109/BIBM.2014.6999219.

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

[23] A. N. Jagannatha, H. Yu. Bidirectional RNN for medical event detection in electronic health records. In *Proceedings of Conference Association for Computational Linguistics*, North American Chapter, Berlin, Germany, pp. 473–482, 2016.

[24] Z. C. Lipton, D. C. Kale, C. Elkan, R. Wetzel. Learning to diagnose with LSTM recurrent neural networks. https://arxiv.org/abs/1511.03677, 2015.

[25] C. Esteban, O. Staeck, S. Baier, Y. C. Yang, V. Tresp. Predicting clinical events by combining static and dynamic information using recurrent neural networks. In *Proceedings of IEEE International Conference on Healthcare Informatics*, Chicago, USA, pp. 93–101, 2016. DOI: 10.1109/ICHI.2016.16.

[26] Z. P. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu. Recurrent neural networks for multivariate time series with missing values. [Online], Available: https://arxiv.org/abs/1606.01865, 2016.

[27] S. Mehrabi, S. Sohn, D. H. Li, J. J. Pankratz, T. Therneau, J. L. S. Sauver, H. F. Liu, M. Palakal. Temporal pattern and association discovery of diagnosis codes using deep learning. In *Proceedings of International Conference on Healthcare Informatics*, Dallas, USA, pp. 408–416, 2015. DOI: 10.1109/ICHI.2015.58.

[28] J. Futoma, J. Morris, J. Lucas. A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, vol. 56, pp. 229–238, 2015. DOI: 10.1016/j.jbi.2015.05.016.

[29] E. Putin, P. Mamoshina, A. Aliper, M. Korzinkin, A. Moskalev, A. Kolosov, A. Ostrovskiy, C. Cantor, J. Vijg, A. Zhavoronkov. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *AGING*, vol. 8, no. 5, pp. 1021–1033, 2016. DOI: 10.18632/aging.100968.

[30] Y. Cheng, F. Wang, P. Zhang, J. Y. Hu. Risk prediction with electronic health records: A deep learning approach. In *Proceedings of SIAM International Conference on Data Mining*, Miami, USA, pp. 432–440, 2016.

[31] E. Choi, A. Schuetz, W. F. Stewart, J. M. Sun. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 361–370, 2017. DOI: 10.1093/jamia/ocw112.

[32] T. Pham, T. Tran, D. Phung, S. Venkatesh. DeepCare: A deep dynamic memory model for predictive medicine. In *Proceedings of the 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, Auckland, New Zealand, pp. 30–41, 2016. DOI: 10.1007/978-3-319-31750-2_3.

[33] A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, N. H. Shah. Improving palliative care with deep learning. [Online], Available: https://arxiv.org/abs/1711.06402, 2017.

[34] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. B. Liu, J. Marcus, M. M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G. E. Duggan, J. Irvine, Q. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. D. Howell, C. Cui, G. S. Corrado, J. Dean. Scalable and accurate deep learning with electronic health records. *Nature Partner Journals Digital Medicine*, vol. 1, pp. 1–10, 2018. DOI: 10.1038/s41746-018-0029-1.

[35] F. Dernoncourt, J. Y. Lee, O. Uzuner, P. Szolovits. De-identification of patient notes with recurrent neural networks. *Journal of the American Medical Informatics Association*, vol. 24, no. 3, pp. 596–606, 2017. DOI: 10.1093/jamia/ocw156.

[36] S. Chauhan, L. Vig. Anomaly detection in ECG time signals via deep long short-term memory networks. In *Proceedings of IEEE International Conference on Data Sci-*

*ence and Advanced Analytics*, Paris, France, 2015. DOI: 10.1109/DSAA.2015.7344872.

[37] Y. Yan, X. B. Qin, Y. G. Wu, N. N. Zhang, J. P. Fan, L. Wang. A restricted Boltzmann machine based two-lead electrocardiography classification. In *Proceedings of the 12th IEEE International Conference on Wearable and Implantable Body Sensor Networks*, Cambridge, USA, pp. 1–9, 2015. DOI: 10.1109/BSN.2015.7299399.

[38] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, vol. 415-416, pp. 190–198, 2017. DOI: 10.1016/j.ins.2017.06.027.

[39] Z. J. Yao, Z. Y. Zhu, Y. X. Chen. Atrial fibrillation detection by multi-scale convolutional neural networks. In *Proceedings of the 20th International Conference on Information Fusion*, Xi′an, China, pp. 1–6, 2017. DOI: 10.23919/ICIF.2017.8009782.

[40] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. https://arxiv.org/abs/1707.01836, 2017.

[41] D. Wulsin, J. Blanco, R. Mani, B. Litt. Semi-supervised anomaly detection for EEG waveforms using deep belief nets. In *Proceedings of the 9th International Conference on Machine Learning and Applications*, Washington DC, USA, pp. 436–441, 2010. DOI: 10.1109/ICMLA.2010.71.

[42] A. Page, J. Turner, T. Mohsenin, T. Oates. Comparing raw data and feature extraction for seizure detection with deep learning methods. In *Proceedings of the 27th International Flairs Conference*, AAAI, Pensacola Beach, USA, pp. 284–287, 2014.

[43] X. W. Jia, K. Li, X. Y. Li, A. D. Zhang. A novel semi-supervised deep learning framework for affective state recognition on EEG signals. In *Proceedings of IEEE International Conference on Bioinformatics and Bioengineering*, Boca Raton, US, pp. 30–37, 2014. DOI: 10.1109/BIBE.2014.26.

[44] I. Sturm, S. Lapuschkin, W. Samek, K. R. Muller. Interpretable deep neural networks for single-trial EEG classification. *Journal of Neuroscience Methods*, vol. 274, pp. 141–145, 2016. DOI: 10.1016/j.jneumeth.2016.10.008.

[45] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017. DOI: 10.1002/hbm.23730.

[46] L. Q. Nie, M. Wang, L. M. Zhang, S. C. Yan, B. Zhang, T. S. Chua. Disease inference from health-related questions via sparse deep learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2107–2119, 2015. DOI: 10.1109/TKDE.2015.2399298.

[47] L. Zhao, J. Z. Chen, F. Chen, W. Wang, C. T. Lu, N. Ramakrishnan. Simnest: Social media nested epidemic simulation via online semi-supervised deep learning. In *Proceedings of IEEE International Conference on Data Mining*, IEEE, Atlantic City, USA, pp. 639–648, 2015. DOI: 10.1109/ICDM.2015.39.

[48] B. Zou, V. Lampos, R. Gorton, I. J. Cox. On infectious intestinal disease surveillance using social media content. In *Proceedings of the 6th International Conference on Digital Health Conference*, ACM, Montreal, Canada, pp. 157–161, 2016. DOI: 10.1145/2896338.2896372.

[49] A. Benton, M. Mitchell, D. Hovy. Multi-task learning for mental health using social media text. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, Spain, pp. 1–11 2017.

[50] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, T. Ploetz. PD disease state assessment in naturalistic environments using deep learning. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, Austin, USA, pp. 1742–1748, 2015.

[51] D. Ravi, C. Wong, B. Lo, G. Z. Yang. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *Proceedings of the 13th IEEE International Conference on Wearable and Implantable Body Sensor Networks*, San Francisco, USA, pp. 71–76, 2016. DOI: 10.1109/BSN.2016.7516235.

[52] A. Aliamiri, Y. C. Shen. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In *Proceedings of IEEE EMBS International Conference on Biomedical & Health Informatics*, Las Vegas, USA, pp. 442–445, 2018. DOI: 10.1109/BHI.2018.8333463.

[53] Q. Zhang, X. X. Chen, Q. Y. Zhan, T. Yang, S. H. Xia. Respiration-based emotion recognition with deep learning. *Computers in Industry*, vol. 92–93, pp. 84–90, 2017. DOI: 10.1016/j.compind.2017.04.005.

[54] Y. J. Chen, T. S. Chen, Z. W. Xu, N. H. Sun, O. Temam. DianNao family: Energy-efficient hardware accelerators for machine learning. *Communications of the ACM*, vol. 59, no. 11, pp. 105–112, 2016. DOI: 10.1145/2996864.

[55] T. Unterthiner, A. Mayr, G. Klambauer, S. Hochreiter. Toxicity prediction using deep learning. [Online], Available: https://arxiv.org/abs/1503.01445, 2015.

[56] J. S. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik. Deep neural nets as a method for quantitative structure-activity relationships. *Journal of Chemical Information and Modeling*, vol. 55, no. 2, pp. 263–274, 2015. DOI: 10.1021/ci500747n.

[57] T. Huynh, Y. L. He, A. Willis, S. Ruger. Adverse drug reaction classification with deep neural networks. In *Proceedings of the 26th International Conference on Computational Linguistics*, Osaka, Japan, pp. 877–887, 2016.

[58] K. Chaudhary, O. B. Poirion, L. Q. Lu, L. X. Garmire. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research*, vol. 24, no. 6, pp. 1248–1259, 2017. DOI: 10.1158/1078-0432.CCR-17-0853..

[59] S. Yousefi, F. Amrollahi, M. Amgad, C. L. Dong, J. E. Lewis, C. Z. Song, D. A. Gutman, S. H. Halani, J. E. V. Vega, D. J. Brat, L. A. D. Cooper. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific Reports*, vol. 7, Article number 11707, 2017. DOI: 10.1038/s41598-017-11817-6.

[60] W. L. Chen, J. Wilson, S. Tyree, K. Weinberger, Y. X. Chen. Compressing neural networks with the hashing trick. In *Proceedings of the 23nd International Conference on Machine Learning*, Lille, France, pp. 2285–2294, 2015.

[61] W. L. Chen, J. Wilson, S. Tyree, K. Q. Weinberger, Y. X.

Chen. Compressing convolutional neural networks in the frequency domain. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, San Francisco, USA, pp. 1475–1484, 2016. DOI: 10.1145/2939672.2939839.

**Zhen-Jie Yao** received the B. Sc. degree in instrument science from the Zhejiang University, China in 2007, and the Ph. D. degree in communication engineering from University of Chinese Academy of Sciences, China in 2012. In 2012, he was a researcher at China Mobile Research Institute, China. Currently, he is a researcher in Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, China. He has published about 15 refereed journal and conference papers, and applied about 15 patents.

His research interests include machine learning, signal processing and their application in healthcare.

E-mail: yaozhenjie@gmail.com (Corresponding author)

ORCID iD: 0000-0003-1027-637X
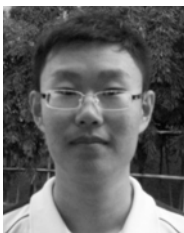
**Jie Bi** received the B. Sc. degree in biotechnology from Shanghai Jiao Tong University, China in 2013, and the M. Sc. degree in bioinformatics from the University of Chinese Academy of Sciences, China in 2016. Currently, he is a front-end engineer in Rhinotech Limited Liability Company, China. He won first prize of National Undergraduate Innovation Program (2012, China).

His research interests include machine learning methods for data visualization, Javascript, CSS3, HTML5.

E-mail: jie.bi@rhinotech.cn

**Yi-Xin Chen** received the Ph. D. degree in computer science from the University of Illinois at Urbana-Champaign (UIUC), USA in 2005. He is a professor of computer science and engineering at the Washington University in St. Louis, USA, which he joined in 2005. He received the Best Paper Award at the *Idea, Development, Exploration, Assessment, Long-term Follow-up, Improving the Quality of Research in Surgery Conference* (2016), Distinguished Paper Award at the *American Medical Informatics Association Conference* (2015), Best Student Paper Runner-up Award at the Association for Computing Machinery (ACM) Special Interest Group on *Knowledge Discovery and Data Mining (SIGKDD) Conference* (2014), Best Paper Award at the *American Association for Artificial Intelligence Conference on Artificial Intelligence* (2010), and the *IEEE International Conference on Tools for AI* (2005). He also received Best Paper Award nominations at *IEEE International Conference on Data Mining* (2013), *IEEE Real-time and Embedded Technology and Applications Symposium* (2012), and *ACM SIGKDD Conference* (2009). His work on planning has won First Prizes in the International Planning Competitions (2004 and 2006). He received an Early Career Principal Investigator Award from the Department of Energy (2006) and a Microsoft Research New Faculty Fellowship (2007). His research has been funded by National Science Foundation, National Institutes of Health, Department of Energy, Microsoft, Fujitsu, and Memorial Sloan-Kettering Cancer Center. He is an associate editor for *ACM Transactions of Intelligent Systems and Technology*, *Annals of Mathematics and Artificial Intelligence*, and *Journal of Artificial Intelligence Research*. He was an associate editor for *IEEE Transactions on Knowledge and Data Engineering* from 2008 to 2012.

His research interests include data mining, machine learning, artificial intelligence, and optimization.

E-mail: chen@cse.wustl.edu