

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/335600087>

# DeepHealth: Deep Learning for Health Informatics

Preprint · September 2019

---

CITATIONS

0

READS

4,123

2 authors:



Hyunjung Gloria Kwak

Harvard Medical School

12 PUBLICATIONS 169 CITATIONS

[SEE PROFILE](#)



Pan Hui

The Hong Kong University of Science and Technology

663 PUBLICATIONS 20,712 CITATIONS

[SEE PROFILE](#)

# DeepHealth: Deep Learning for Health Informatics reviews, challenges, and opportunities on medical imaging, electronic health records, genomics, sensing, and online communication health

GLORIA HYUN-JUNG KWAK, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology

PAN HUI, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology and Department of Computer Science, University of Helsinki

---

Machine learning and deep learning have provided us with an exploration of a whole new research era. As more data and better computational power become available, they have been implemented in various fields. The demand for artificial intelligence in the field of health informatics is also increasing and we can expect to see the potential benefits of artificial intelligence applications in healthcare. Deep learning can help clinicians diagnose disease, identify cancer sites, identify drug effects for each patient, understand the relationship between genotypes and phenotypes, explore new phenotypes, and predict infectious disease outbreaks with high accuracy. In contrast to traditional models, its approach does not require domain-specific data pre-process, and it is expected that it will ultimately change human life a lot in the future. Despite its notable advantages, there are some challenges on data (high dimensionality, heterogeneity, time dependency, sparsity, irregularity, lack of label) and model (reliability, interpretability, feasibility, security, scalability) for practical use.

This article presents a comprehensive review of research applying deep learning in health informatics with a focus on the last five years in the fields of medical imaging, electronic health records, genomics, sensing, and online communication health, as well as challenges and promising directions for future research. We highlight ongoing popular approaches' research and identify several challenges in building deep learning models.

CCS Concepts: •**Computing methodologies** → **Machine learning approaches; Machine learning;** •**Social and professional topics** → **Computing/technology policy; Medical information policy;** •**Applied computing** → **Life and medical sciences; Health care information systems; Health informatics; Computational biology; Genomics; Bioinformatics;**

Additional Key Words and Phrases: deep learning, machine learning, medical informatics, health informatics, medical image, electronic health records, genomics, biosensor, online communication, social media, personalized medicine, personalized clinical care

---

## 1 INTRODUCTION

Machine learning and deep learning have been newly become a trend and opened a whole new research era. In fact, they have been implemented in various fields. Among the various types of academia and industry, the demand for artificial intelligence in the field of health informatics has increased, and the potential benefits of artificial intelligence applications in healthcare have also been proven. Previous studies attempted to have the right treatment, delivered to the right patient at the right time by taking into account several aspects of patient's data, including variability in molecular traits, medical images, environmental factors, electronic health records (EHRs) and lifestyle [1, 10, 58, 62, 76, 81, 88, 98, 142, 161, 164, 166, 169, 180, 193, 195, 216, 238, 286, 293].

Health informatics and how deep learning can be used in health informatics can be explained with clinical informatics and decision support. Using data aggregation and analysis from multiple

data sources, researchers train models to learn what clinicians do when they see the patients and let them produce supportive clinical information. It includes how to read clinical images, predict outcomes, discover the relationship between genotype and phenotype or phenotype and disease, analyze treatment response, track a lesion or structural change (ex. decreased hippocampal volume). Predicting outcomes (ex. disease) or readmissions can be expanded to an early warning system with risk scoring. Identifying correlations and patterns can be extended to global pattern research and population healthcare, such as providing predictive treatment for the entire population.

Deep learning in health informatics has many advantages that it can be trained without *a priori*, which combats the lack of labelled data and burden on clinicians. For example, medical imaging dealt with data complexity, overlapped detection target points and 3- or 4-dimensional medical images. Researchers provided more sophisticated and elaborative outcomes with data augmentation, un-/semi-supervised learning, transfer learning, and multi-modality architectures [37, 186, 225, 285, 287]. Second of all, it was also studied to discover nonlinear relationships between variables and help clinicians and patients with an objective and personalized definition of disease and solutions, since decisions are basically made up of data itself rather than human intervention and models divide the cohort into subgroups according to their clinical information. In bioinformatics, DNA or RNA sequences were studied to identify gene alleles and environmental factors that contribute to diseases, investigate protein interactions, understand higher-level processes (phenotype), find similarities between two phenotypes, design targeted personalized therapies and more [149]. In particular, deep learning algorithms were implemented to predict the splicing activity of exons, the specificities of DNA-/RNA- binding proteins and DNA methylation [5, 9, 284]. Third, it demonstrates its usefulness, especially when predicting rapidly developing diseases such as acute renal failure. It was expected to be sufficient for the use of new phenotypes and subtypes discovery, personalized patient-level risk predictions and real-time level prediction rather than regularly scheduled health screenings and helpful to guide treatment [17, 64, 91, 112, 134, 210, 231, 255, 268]. Fourth, it is expected to be widely used for first-time inpatients, transferred patients, weak healthcare infrastructure patients, and outpatients without chart information [18, 279]. For example, portable neurophysiological signals such as Electroencephalogram (EEG), Local Field Potentials (LFP), Photoplethysmography (PPG) [123, 189], accelerometer data from above ankle and mobile apps were used to monitor individual health status, to predict freezing from Parkinson's disease, rheumatoid arthritis, chronic diseases such as obesity, diabetes and cardiovascular disease to provide health information before hospital admission and to prepare the emergency intervention. In addition, mobile health technologies for resource-poor and marginalized communities were also studied with reading X-ray images taken by a mobile phone [32]. Clinical notes including discharge notes summarization studies were aimed to study how summarization notes express reliable, effective and accurate information timely, comparing the information with medical records. Finally, disease outbreaks, social behavior, drug/treatment review analysis and research on remote surveillance systems have also been studied to prevent disease, prolong life, and monitor epidemics [4, 27, 36, 70, 87, 203, 258].

Among various observation methods, range from comparatively simple statistical projections to machine learning (ML) and deep learning (DL) algorithms, several architectures stood out in popularity. Researchers started with a mass univariate analysis model such as t-test, F-test and chi-square test to prove the contrast of a null hypothesis and continued to the methods such as feature extraction, classification, prediction, and de-identification. For example, Support Vector Machine (SVM) trained a classifier to make a maximal value on the margin of separation between two groups [263]. Even though SVM has given researchers many choices of experiments, it has the disadvantage that it requires experts' feature selection from insights. And DL, a remarkable record-making ML family member throughout many fields, solved this problem. DL algorithm

is a deep neural network with neuro-inspired techniques to generate optimal weights, abstract high-level features on its own, and extract information factors non-manually, resulting in more objective and unbiased classification results [146, 227, 244].

In the sense of trust and expectation, the number of papers grew rapidly, and this is illustrated in Fig. 1. The number of hospitals that have adopted at least a basic EHR system drastically increased. Indeed, according to the latest report from the Office of the National Coordinator for Health Information Technology (ONC), nearly over 75% of office-based clinicians and 96% of hospitals in the United States using an EHR system, nearly all practices have an immediate, practical interest in improving the efficiency and use of their EHRs [24, 104]. With the rapid development of imaging technologies (MRI, PET, CT), wearable sensors, genomic technologies (microarray, next-generation sequencing), information about patients can now be more readily acquired. Thus far, deep learning architectures have developed with computation power support in Graphics Processing Units (GPUs) which have been a significant impact on practical uptake and acceleration of deep learning. Therefore, plenty of experimental works have implemented deep learning models for health informatics, reaching alternative techniques that have been used by most clinicians. Nevertheless, the application of deep learning to health informatics raises a number of challenges that need to be resolved, including data informativeness (high dimensionality, heterogeneity, multi-modality), lack of data (missing values, class imbalance, expensive labelling), data credibility and integrity, model interpretability and reliability (tracking and convergence issues as well as overfitting), feasibility, security and scalability.

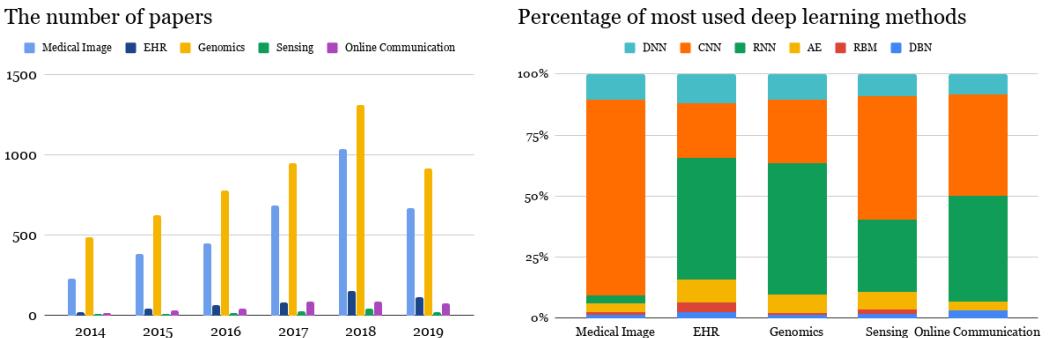


Fig. 1. Left: Distribution of published papers that use deep learning in subareas of health informatics from PubMed, Right: Percentage of most used deep learning methods in health informatics. (DNN: Deep Neural Network, CNN: Convolutional Neural Network, RNN: Recurrent Neural Network, AE: Autoencoder, RBM: Restricted Boltzmann Machine, DBN: Deep Belief Network)

In the following sections of this review, we examine a rapid surge of interest in recent health informatics studies including bioinformatics, medical imaging, electronic health records, sensing, and online communication health, with practical implementations, opportunities, and challenges.

## 2 MODELS OVERVIEW

This section reviews the most common models used in the studies reviewed in this paper. There are a variety of architectures available today, and they were developed quickly, so only a brief introduction to the main base models applied to health informatics. We begin by introducing some common non-deep learning models used in many studies to compare or combine with deep learning

models. Subsequently, deep learning architectures are reviewed, including CNN, RNN, AE, RBM, DBN, and their variants with transfer learning, attention learning, and reinforcement learning.

## 2.1 Support Vector Machine

SVM aims to define an optimal hyperplane which can distinguish groups from each other. In a training phase, when data itself is linearly separable [263], SVM finds a hyperplane with the longest distance between support vectors of each group (ex. disease case and healthy control group). If training data is not linearly separable, SVM can be extended to a soft-margin SVM and kernel-trick methods [42, 236].

For an original SVM, with training data  $n$  points of the form  $\{\mathbf{x}_i, y_i\} i = 1, \dots, l$ ,  $\mathbf{x}_i \in \mathbb{R}^n$ , and  $y_i \in \{-1, 1\}$  where  $y_i$  are either 1 or -1 and  $x_i$  is a  $n$ -dimensional real vector, minimizing  $\|\mathbf{w}\|$  is aimed subject to

$$y_i(\mathbf{x}_i \mathbf{w} + b) - 1 \geq 0, \forall i \quad (1)$$

for all  $i = 1, \dots, l$ . Unlike a hard-margin algorithm, an extended SVM with a soft-margin introduces a different minimizing problem (hinge loss function) with a trade-off parameter (Equation 2). Having a regularization term  $\|\mathbf{w}\|^2$  and a small value parameter makes data can be finally linearly classifiable.

$$\left[ \frac{1}{n} \sum_{i=1}^l \max(0, 1 - y_i(\mathbf{x}_i \mathbf{w} - b)) \right] + \lambda \|\mathbf{w}\|^2 \quad (2)$$

For another option for non-linearly classifiable data to become linearly separable, kernel methods helps with a feature map  $\varphi : \mathcal{X} \rightarrow \mathcal{V}$  which satisfies  $k(x, x') = \langle \varphi(x), \varphi(x') \rangle_v$  and two of the popularly used methods are polynomial kernel  $k(x_i, x_j) = (x_i \cdot x_j)^2$  and Gaussian radial basis kernel  $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ . Kernel-trick methods seek a certain dimension which helps data can be linearly separable.

## 2.2 Matrix/Tensor Decomposition

A tensor is a multidimensional array. More formally, an  $n$ -way or  $n^{th}$ -order tensor is an element of a tensor product of  $n$  vector spaces, each of which has its own coordinate system. A first-order tensor is a vector and a second-order tensor is a matrix. So, normally second-order tensor decomposition is called as matrix decomposition, and three or higher-order tensor decomposition is called tensor decomposition. One of the tensor decompositions is CANDECOMP/PARAFAC (CP) decomposition, and a third-order tensor is factorized into a sum of component rank-one tensors, as shown in Fig. 2. For a third-order tensor  $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ , it can be also stated in Equation 3 and 4 for a positive integer  $R$  and  $a_r \in \mathbb{R}^I$ ,  $b_r \in \mathbb{R}^J$ ,  $c_r \in \mathbb{R}^K$  [136].

$$\mathbf{X} \approx [\mathbb{A}, \mathbb{B}, \mathbb{C}] \equiv \sum_{r=1}^R a_r \circ b_r \circ c_r \quad (3)$$

$$\mathbf{x}_{ijk} \approx \sum_{r=1}^R a_{ir} b_{jr} c_{kr} \text{ for } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K \quad (4)$$

## 2.3 Word Embedding

Word embedding is a technique to map words to vectors with real numbers, and word2vec is a group of models to produce word embedding [276]. It is considered because it allows a model to have more informative and condensed features. Conceptually, with similarity and co-occurrence, words are mapped to a binary space with many dimensions first and then to a continuous vector

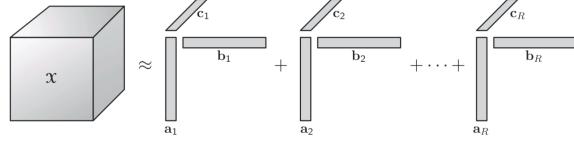


Fig. 2. CP decomposition of a three-way array [136].

space with a much lower dimension. Word2vec is introduced with two distributed representations of words such as continuous bag-of-words (CBOW) and skip-gram. CBOW predicts a current word with surrounding context words, and skip-gram uses a current word to predict a surrounding window of context words as Fig. 3 [176].

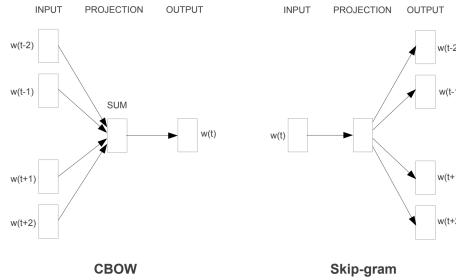


Fig. 3. CBOW and Skip-gram [176].

## 2.4 Multilayer Perceptron

Perceptron is an ML algorithm that researchers refer to as the first online learning algorithm. Multilayer perceptron (MLP) is a feedforward neural network that has perceptrons (neurons) for each layer [146]. When a model has three layers which are the minimum amount of layers, the network is called either a vanilla or shallow neural network, and when it is deeper than three layers, the network is called a deep neural network (DNN). In the case of  $n$ -layers, the first layer is an input layer (when 1-d data is trained, a list of voxels' intensity corresponds to an input data), the last layer is an output layer and  $n - 2$  layers are hidden layers. In contrast to SVM, MLP does not require prior feature selection, since it combines some features and finds optimal ones by itself.

As an online learning based algorithm which trains data line by line (sample by sample), for every sample, the model compares the expected value and the labelled value. The difference between the expected value and the given labelled value reflects the cost or error, and the amount and direction of weights are changed with backpropagation, toward minimizing the error and preventing overfitting with dropout [146, 221, 244].

## 2.5 Convolutional Neural Networks

Convolutional neural network (CNN) is an algorithm inspired by biological processing of the animal visual cortex [139, 146, 147]. Unlike the original fully connected neural network, the algorithm eventually implements how the animal visual cortex works, with convolutional layers which have shared sets of 2-dimensional weights for 2D CNN case that recognize the spatial information and pooling layers to filter comparatively more important knowledge and only transmit

concentrated features (Fig. 4. Left) [116, 147]. As other deep learning algorithms have a way of preventing overfitting, CNN classifies whether images have specific labels which they look for or not with convolutional and pooling layers. For 3D CNN, three-dimensional weights are used (Fig. 4. Right) [122], and for 2.5D CNN, two-dimensional weights with multi-angle learning architectures are used.

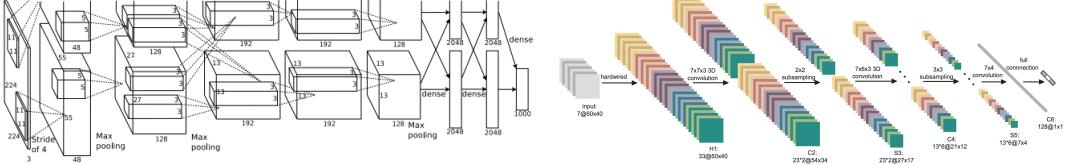


Fig. 4. Left: The architecture of AlexNet (2D CNN), Right: The architecture of 3D CNN [122, 139].

## 2.6 Recurrent Neural Networks

Recurrent neural network (RNN) is a class of ANN specialized for streams of data such as time-series data and natural language [49, 59, 94, 277]. RNN operates by sequentially updating a hidden state based on the activation of the current input  $x$  at the time and the previous hidden state  $h_{t-1}$ . Likewise,  $h_{t-1}$  is updated from  $x_{t-1}$  and  $h_{t-2}$ , and each output values are dependent on the previous computations. Even though RNN showed significant performance on temporal data, RNN had limitations in terms of vanishing gradient and exploding gradient [23]. For that, RNN variants have been developed, and some well-known examples are Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) networks, which addressed these problems by capturing long-term dependencies (Fig. 5) [110].

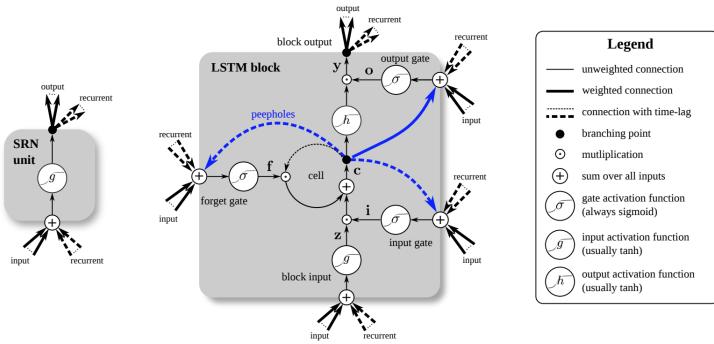


Fig. 5. Left: Detailed schematic of the Simple Recurrent Network (SRN) unit, Right: The architecture of a Long Short-Term Memory block as used in the hidden layers of a recurrent neural network [94].

## 2.7 Autoencoders

When the layers in neural networks are very deep, the amount of weight update is obtained by the multiplication of small gradient descents and may reach to 0. Calling the phenomenon of ‘vanishing gradient’, a greedy layerwise training was proposed for this problem which is a foundation of stacked autoencoders and deep belief networks [107]. Autoencoder (AE) is one of the

unsupervised learning methods and it consists of an encoder  $\phi : \mathcal{X} \rightarrow \mathcal{F}$  and a decoder  $\psi : \mathcal{F} \rightarrow \mathcal{X}$  which performs as generating high-level or latent value and reconstructing input data (Fig. 6). It aims to find  $\phi$  and  $\psi$ , which can make the minimal difference between the given input data and the reconstructed input data  $\phi, \psi = \arg \min_{\phi, \psi} \|X - (\psi \circ \phi)X\|^2$  [93].

Several variants exist to the basic model, with the aim of forcing learned representations of input to assume useful features, which are regularized autoencoders (sparse, denoising, stacked denoising, and contractive autoencoders) and variational autoencoders [209, 217, 265, 266, 275]. In particular, sparse autoencoder (SAE) learns representations by allowing only a small number of hidden units to be active and others inactive for sparsity, so that a sparsity penalty encourages the model to learn with some specific areas of the network. Denoising autoencoder (DAE) is trained to reconstruct corrupted input after first denoising input, minimizing the same reconstruction loss between a clean input and its reconstruction from hidden representation features. Finally, stacked denoising autoencoder is introduced to make a deep network, in a way like stacking RBMs in deep belief networks [107, 145, 265, 266], only corrupting input and using the highest level output representation of each autoencoder as another input for the next one to study, which can be found in Fig. 6. Unlike classical autoencoders, variational autoencoders (VAEs) are generative models, like Generative Adversarial Networks (GANs) with encoders which form latent vectors with the mean and standard deviation from sampled inputs and decoders to reconstruct/generate the training data.

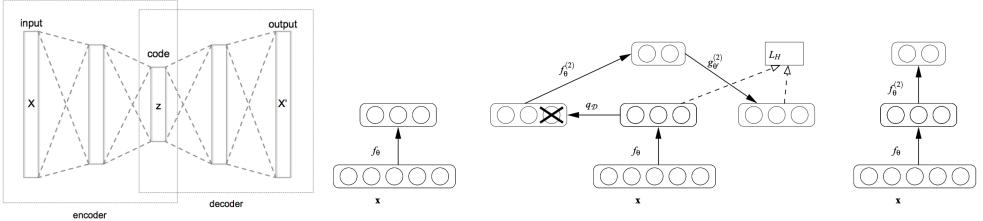


Fig. 6. Left: Autoencoder, Right: Stacking Denoising Autoencoder.

## 2.8 Deep Belief Networks

Deep belief network (DBN) is composed of a stacked Restricted Boltzmann Machine (RBM) and a belief network [106, 108, 197]. RBM has a similar concept to Autoencoder, but AE has three layers (input, hidden and output layer) and is deterministic, and RBM has two layers (visible and hidden layer) and is stochastic. As pre-training, the first RBM is trained with a sample  $v$ , a hidden activation vector  $h$ , a reconstruction  $v'$ , a resampled hidden activations  $h'$  (Gibbs sampling) and weight is updated with  $\Delta W = \epsilon(vh^T - v'h'^T)$  (single-step version of contrastive divergence) to get a maximum probability of  $v$ . The hidden layer represents a new input layer for the second RBM followed by the first RBM and the network can start from learning high-level features. When stacked RBMs are all trained, a belief network is added onto the last hidden layer from RBMs and trained to provide a label corresponding to the input label (Fig. 7) [106, 108, 224].

## 2.9 Attention Learning

Attention mechanism can be described by mapping a query and a set of key-value pairs to an output. The output is calculated as the weighted sum of the values, and the weight assigned to each value is calculated by the compatibility function of the query with that key [264]. The mechanism differs in the way of the process, including scaled dot-product attention and multi-head attention.

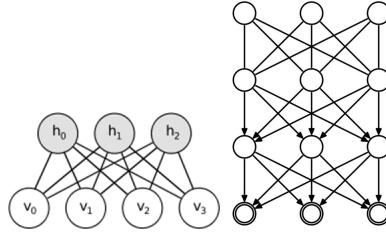


Fig. 7. Left: Restricted Boltzmann Machine with four visible nodes and three hidden nodes, Right: Three-layer Deep Belief Network [93, 106, 108, 224].

Bahdanau et al. [15] introduced Neural Machine Translation (NMT) with attention mechanisms to help memorize long source sentences. The authors proposed a neural machine translation, which consists of an RNN or a bidirectional RNN as an encoder with hidden states and a decoder with a sum of hidden states weighted by alignment scores to emulate searching through a source sentence during decoding a translation (Fig. 8).

A RNN with attention has hidden states  $h_i$ , and a BiRNN with attention has forward  $\tilde{h}_i$  and backward  $\tilde{h}_i$  hidden states (ex.  $h_i = [\tilde{h}_i^T; \tilde{h}_i^T]^T$ ). Each annotation  $h_i$  contains information about the whole input sequence with a strong focus on the parts surrounding the  $i^{th}$  word of the input sequence. The probability  $\alpha_{ij}$  reflects the importance of the annotation  $h_j$  with respect to the previous hidden state  $s_{i-1}$  and the context vector  $c_i$ . The context vector  $c_i$  is computed as a weighted sum of the hidden annotations  $h_j$ . And the weight  $\alpha_{ij}$  of each annotation  $h_j$  is computed by how well the inputs around position  $j$  and the output at position  $i$  match [15]. The motivation is for the decoder to decide words to pay attention to. With the context vector which has access to the entire input sequence, rather than forgetting, the alignment between input and output is trained and controlled by the context vector.

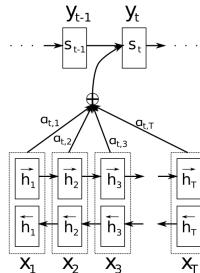


Fig. 8. The encoder-decoder model with additive attention mechanism [15].

## 2.10 Transfer Learning

In transfer learning, a base network on a base dataset was trained first and then using the learned features, a target network is also trained on a target dataset [20, 21, 34, 294]. This process is generally meaningful and making a significant improvement when the target dataset is small to train and researchers intend to avoid overfitting. Usually, after training a base network, the first  $n$  layers are copied and used for the target network and the remaining layers of the target network are randomly initialized. The transferred layers can be left as frozen or fine-tuned, which

means either locking the layers so that there is no change during training the target network or backpropagating the errors for both copied and newly initialized layers of the target network [294].

## 2.11 Reinforcement Learning

Reinforcement learning was introduced as an agent learning policy  $\pi$  to take action in the environment to maximize cumulative rewards. At each time stamp  $t$ , an agent observes a state  $s_t$  from its environment and takes an action  $a_t$  in state  $s_t$ . The environment and the agent then transition to a new state  $s_{t+1}$  based on the current state  $s_t$  and the chosen action, and it provides a scalar reward  $r_{t+1}$  to the agent as feedback [250] (Fig. 9).

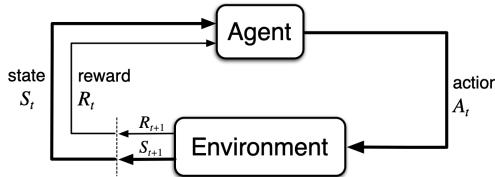


Fig. 9. The agent-environment interaction in reinforcement learning [250].

Markov decision process (MDP) is the mathematical formulation of the RL problem. The MDP formulation consists of:

- a set of states  $S$
- a set of actions  $A$
- a transition function  $P_a(s, s') = Pr(s_{t+1} = s' | s_t = s, a_t = a)$   
: from state  $s$  to state  $s'$  under action  $a$
- a reward after transition  $R_a(s, s')$   
: from state  $s$  to state  $s'$  with action  $a$
- a discount factor  $\gamma \in [0, 1]$   
: lower values place more emphasis on immediate rewards (ex.  $\sum_{t=0}^{T-1} \gamma^t r_{t+1}$ )

$$\pi^* = \arg \max_{\pi} \mathbb{E}[R|\pi] \quad (5)$$

The goal of RL is to find the best policy with the maximum expected return, and the RL algorithm class includes value function based, policy search based and actor-critic based methods using both of the preceding [74, 95, 126]. Deep reinforcement learning (DRL) is based on extending the previous work of RL to higher-dimensional problems. The low-dimensional feature representation and powerful function approximation of the neural network allow the DRL to handle the curse of dimensionality and optimize the expected return by a stochastic function [12, 22, 101, 228, 250].

## 3 APPLICATION OF DEEP LEARNING METHODS

The use of deep learning for medicine is recent and not thoroughly explored. In order to estimate performances of deep learning algorithms on health care, a search was conducted across several databases with the combination of search terms: ('deep learning' OR 'neural network' OR 'machine learning') and (i) medical imaging (ii) EHR (iii) genomics (iv) sensing and online communication health. Among the articles found, significantly relevant papers regarding each part with applying DL algorithms were briefly reviewed.

### 3.1 Medical Imaging

The first applications of deep learning on medical datasets were medical images including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), X-ray, Microscopy, Ultrasound (US), Mammography (MG), Hematoxylin & Eosin Histology Images (H&E), Optical Images and etc. PET scans show regional metabolic information through positron emission, unlike CT and MRI scans, which reveal the structural information of organs or lesions within the body in perspective with radio waves with X-rays and magnets. Medical imaging technology has been chosen for purposes, and in terms of potential health risks to the human body due to X-rays, low-dose CT scans have also been considered, but have disadvantages such as image quality and diagnostic performance. Applications included pathology, psychiatry, brain, lungs, abdomen, heart, breasts, etc., and have been studied in image classification (classify disease present/absent), object detection (detect disease with location), image segmentation (detect disease and label pixels), image registration (transform one image set into another set of coordinate systems) and other tasks.

Image classification is still the preferred approach for medical image research by classifying one or several classes for each image. Its limitations are in particular the lack of labelled training samples, which have been addressed by transfer learning and multi-stream learning. To track disease progression and make full use of 3D data, a combination of RNN and CNN was also studied. Deep learning has also been extremely quickly implemented in all other aspects of medical image analysis, such as pixel, edge and region-based image segmentation, class imbalance studies, image registration (ex. registration of brain CT/MRI images or whole-body PET/CT images for tumor localization), image generation, image reconstruction, and etc.



Fig. 10. Slices of an MRI scan of an AD patient, from Left to Right: in axial view, coronal and sagittal view [196].

#### 3.1.1 *Image Classification and Object Detection.*

Since the relatively shallow LeNet and AlexNet [139, 147], there has been an exploration for novel architectures such as [99, 158, 242, 251, 282] that are still popular in medical data. Researchers trained the model with or without a pre-trained network. Nevertheless, some of the problems with computer-aided diagnostics (CAD) using medical imaging still remain. The challenge is how to use all the features of different shapes and intensities of the detection points, even within the same imaging modality, overlapping detection points, and 3D or 4D medical images.

To deal with this data complexity, traditional machine learning or deep learning approaches using hand-designed feature extraction were used [11, 80, 186, 219, 223, 261, 285]. In a deep learning approach, CNN essentially learns the hierarchical structure of more and more complex features, so it can work directly on image patches centered on abnormalities. Disease classification has evolved into 2D as well as 3D CNN, transfer learning through feature extraction with DBN and AE,

multi-scale/multi-modality learning, RCNN, and f-CNN. In recent years, a clear transition to deep learning approaches, in particular, transfer learning and multi-stream learning with 3D images and visual attention mechanisms, can be observed, and the application of these methods is very extensive, from brain MRI to retinal imaging and digital pathology to lung computed tomography (CT).

### (1) Transfer Learning

Transfer learning is a popular method in which a model developed for a task is reused as a starting point for a model in other tasks so that researchers do not start the learning process from scratch. Pre-processing with images of similar distribution is still a crucial step influencing classification performance, but performance is still limited because of a lack of ground-truth/annotated data. The cost and time to collect and manually annotate medical images by experts are enormous and manual annotation is also subjective. To alleviate the limitations of the study, strategies can be identified in two categories: (i) using pre-trained networks as feature extractors with unsupervised learning based methods, and (ii) fine-tuning pre-trained networks with either natural images or other medical domain data with supervised learning methods.

For the first category, RBM, DBN, AE, VAE, SAE, and CSAE [105–107, 132, 145, 209, 217, 265, 266] are unsupervised architectures which constitute a hidden layer with input or visible layers and latent feature representation vectors. The medical imaging community also focused on unsupervised learning. After training the layers of unsupervised learning first, a linear classifier is added to the top layer of the algorithm. With a combination of unsupervised learning and classifier (ex. AE with regression, AE with CNN), the methods were applied to the automatic biomarkers extraction and outperformed traditional CAD approaches [28, 46, 207, 234, 247, 248, 262].

In addition, in relation to avoiding lack of training samples and overfitting, transfer learning via fine-tuning has been proposed in medical imaging applications, using a database of labelled natural images or other labelled medical field images [37, 43, 188, 202, 225, 241, 252, 285]. Pre-training supervised learning’s layers and copying the first few layers into the new algorithm with the target dataset firstly be done, and fine-tuning is performed by optimizing the whole algorithm. There was concern about using natural or other field medical image datasets for fine-tuning since there is a profound difference between those. Nevertheless, previous studies have shown that CNN, fine-tuned based on natural image/other medical field data, improves the performance of algorithms, such as the shape, edges, and etc. Even if the base and target datasets are dissimilar, unless the target dataset is significantly smaller than the base dataset, transfer learning is likely to give us a powerful model without overfitting, in general [188, 294]. For example, in [188], CNN with and without transfer learning was compared with natural image datasets for classification between benign nodule, primary lung cancer, and metastatic lung cancer, and pre-trained model outperformed others with around 13% difference of accuracy. Although mostly transfer learning is to combat lack of data with natural images and medical images, recently, Shan et al. [233] proposed a 3D convolutional encoder-decoder network for Low-Dose CT (LDCT) via transfer learning from a 2D trained network. LDCT newly has been used in the medical imaging field because of health risk, however, it makes low diagnostic performance. The authors introduced a 3D conveying path based convolutional encoder-decoder to denoise LDCT to normal dose CT. Putting a trained 2D convolutional layers in the middle of the 3D convolutional layers, they incorporated the 3D spatial information from the adjacent slices of images, since a radiologist also needs to scan adjacent slices to extract pathological information more accurately.

## (2) Multi-stream Architectures

Whereas CNN is fundamentally designed for fixed-size and one type of 2D images, medical images are inherently 3D or 4D images, image sizes are varied, imaging techniques produce different images, and coordinates of detection points are different and comparably small.

Dimensional problems can be solved using the 3D image itself. In fact, 3D Volume of Interest (VOI) was initially used for classification problems [111, 196]. In general, researchers have developed 3D kernel or convolutional layers and several new layers that formed the basis of their network, and those have shown to outperform existing methods. However, there is a computational burden in processing 3D medical scans, and they are not efficient and effective dense training schemes. The voxel size difference can be solved by data interpolation, but it can result in severely blurred images. Therefore, a dilated convolution and multi-stream learning (multi-scale, multi-angle, multi-modality) are suggested as another solution. For multi-stream learning, the default CNN architecture is trained and the channels can be merged at any point in the network, but in most cases, the final feature layers are concatenated or summarized to make the final decision on the classifier. Although there have been some studies including 3D Faster-RCNN [311] for nodule detection with 3D dual-path blocks and U-Net shaped AE structures, the two most widely used main approaches were multi-scale analysis and 2.5D classification [47, 127, 129, 182, 186, 237, 243, 288, 290]. It has become widespread in multi-stream image analysis, especially localization which often requires parsing of 3D volumes in medical imaging and better approaches for classification and segmentation problems, following clinicians' workflow that they rotate, zoom in/out 3D images and check adjacent images during diagnosis.

First of all, multi-scale image analysis reliably detected detection points for irregularly shaped diseases with various intensity distributions and densities [127, 129, 182, 220, 237, 243]. For example, Kamnitsas et al. [127] used multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. In [129], multi-resolution two-stream CNN was proposed with hybrid pre-trained and skin-lesion trained layers. The authors first trained the original images and each stream with the highest resolution images and low-resolution images created by average pooling, then concatenated the last layers for final prediction. In [57, 237], both used multi-scale CNN architecture in multiple streams for pulmonary nodule classification. Shen et al. [237] investigated the problem of diagnostic pulmonary nodule classification, which primarily relies on nodule segmentation for regional analysis, and proposed a hierarchical learning framework that uses multi-scale CNNs to learn class-specific features without nodule segmentation. In recent studies, researchers used reinforcement learning to improve detection efficiency and performance [3, 89, 200]. Among them, Ghesu et al. [89] proposed a combination of multi-scale and reinforcement learning by reformulating the detection problem as a behavior learning task for an agent in reinforcement learning. That is, the artificial agent was trained to distinguish the target anatomical object in 3D with the optimal navigation path and scale.

Furthermore, the 2.5D classification was to address the trade-off between 2D and 3D image classifiers [47, 57, 152, 288, 308]. The method used 3D Volume of Interest (VOI), but 2D slices were trained as input images, so it was able to use important 3D features without compromising computational complexity. It sliced the 3D spatial information in the middle for three or more than three orthogonal views of an input image, or it transformed grayscale images to color images. When sliced into three parts based on the intersection of three axial, sagittal and coronal planes, the three 2D slices were generally selected, as shown in Fig. 11. Otherwise, image slices were

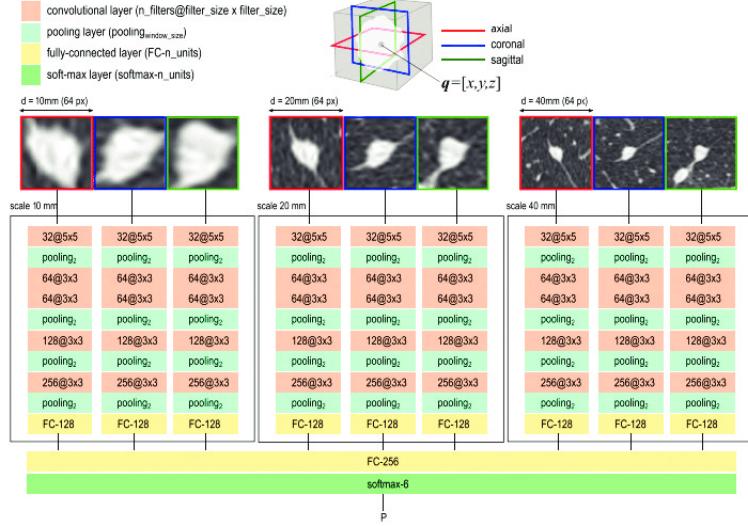


Fig. 11. The multi-angle and multi-scale CNN architecture for pulmonary nodule classification [57].

collected variously through scale, random translation, and rotation. For instance, in [57], different angle and scale images were used for each stream for 9 different kinds of perifissural lung nodules classification problems. Without feeding information about nodule size, each stream was trained, followed by a concatenation of the last layers for classifiers including SVM and KNN.

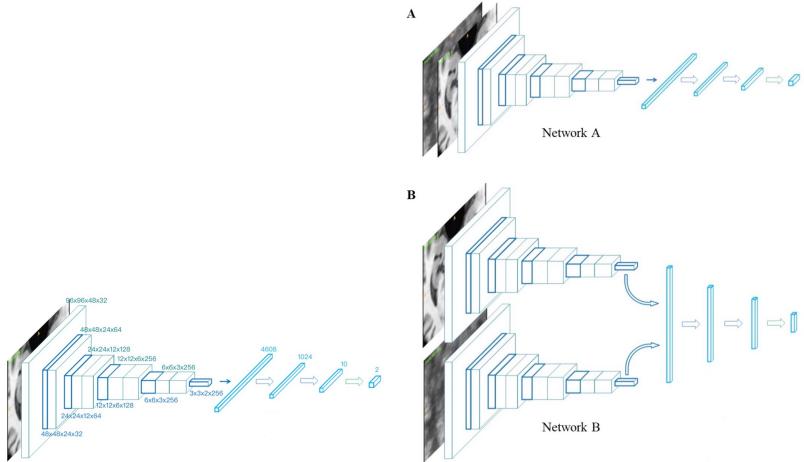


Fig. 12. The architecture of the single- and multi- modality network for Alzheimer's disease [115].

Third of all, multi-modality was also considered to solve the issues, since each medical imaging techniques have different advantages. In general, PET captures the metabolic information and CT/MRI does the structural information of organs. As metabolic changes occur before any functional and structural changes in tissues, organs, and bodies, PET facilitates early disease detection [115, 186, 254]. In [115], the authors looked into paired FDG-PET and T1-MRI to catch different biomarkers

for Alzheimer's disease. PET indicates the regional cerebral metabolic rate of glucose to evaluate the metabolic activity of tissues, and MRI provides high-resolution structural information of the brain to measure the structural metrics such as thickness, volume, and shape. In particular, the study measured the shrinkage of cerebral cortices (brain atrophy) and hippocampus (memory-related), enlargement of ventricles, and change of regional glucose uptake. To use different modality images together, image registration needed to be held, and then the authors compared single-modality, multi-modality (sharing weights like 3D CNN) and multi-modality (multiple streams without sharing weights) (Fig. 12).

### 3.1.2 *Image Segmentation.*

Image segmentation is a process of partitioning an image into multiple meaningful segments (sets of pixels) in a bottom-up approach. And CNN-based models are still the most commonly used to classify each pixel in an image, and it was welcomed by researchers in terms of shared weights compared to a fully connected network. Nevertheless, a drawback of this approach is huge overlaps from neighboring pixels and repeated computations of the same convolutions. In order to have a more efficient convolutional layer, the concepts of fully connected layers and convolutional neural networks were combined, and a fully convolutional network (fCNN) was proposed to an entire input image in an efficient way, by rewriting the fully connected layers as convolutions. While 'Shift-and-stitch' [165] was proposed to boost up the performance of fCNN, U-Net, an image segmentation architecture, was proposed for biomedical images [218]. Inspired by fCNN, Ronneberger et al. [218] proposed a U-Net architecture with upsampling (upconvolutional layers) and skip-connection, and made a better output.

A similar approach has been studied by some researchers, and there have been a variety of variant algorithms [14, 56, 75, 177]. More specifically, Çiçek et al. [56] expanded U-Net from 2D to 3D architecture with introducing use cases of (i) semi-annotation and (ii) full-annotation of training sets. Full annotation of the 3D volume is not only difficult to obtain but also leads to rich training. Therefore, the authors focused on how to generate 3D models to learn image segmentation with only a few annotated 2D slices for training. In [177], the authors proposed a 3D-variant of U-Net architecture, called V-net, performing 3D image segmentation using 3D convolutional layers and Dice coefficient optimization. Since it is not uncommon to have a strong imbalance between the number of foreground and background voxels, previous researchers did re-weighting, but the authors proposed an objective function based on dice coefficients. Drozdzal et al. [75] investigated the use of short and long ResNet-like skip connections, and Badrinarayanan et al. [14] proposed a SegNet, which reused pooling indices of the decoders to perform up-sampling of the low-resolution feature maps. That was one of the most important key elements of the SegNet, gaining high-frequency details and reducing the number of parameters to train in decoders. Still, the architecture upon the U-Net architecture was also built with the nearest neighbor interpolation for up-sampling, one down-sampling and squeeze-and-excitation (SE) residual building blocks, multi-scale and 3D convolutional kernels for adjacent images network [97, 153, 156, 310].

Although these specific segmentation architectures offered compelling advantages, many authors have also achieved excellent segmentation results by combining RNN, MRF, CRF, RF, dilated convolutions and others with segmentation algorithms [6, 8, 44, 138, 246, 283]. P. K. Poudel et al. [192] combined 2D U-Net architecture with GRU to perform 3D segmentation, and Chen et al. [44] applied it several times in multiple directions to incorporate bidirectional information from neighbors. And 3D fCNN with sum instead of concatenation operation and 4D fully convolutional structured LSTM was studied [86, 296], and those outperformed to the 2D U-Net method using RNN. Several fCNN methods have also been tried, using graphical models such as MRF and CRF, applied on

top of the likelihood map produced by CNN or fCNN [2, 31, 55, 84, 85, 232, 312]. Finally, researchers have shown dilated convolutional layers and attention mechanisms [45, 181, 271, 295]. Yu and Koltun [295] and Chen et al. [45] employed dilated convolution to handle the problem of segmenting objects at multiple scales and systematically aggregate multi-scale contextual information. Wang et al. [271] proposed their automatic prostate segmentation in transrectal ultrasound images, using 3D deep neural network equipped with attention modules. The attention module was utilized to selectively leverage the multilevel features integrated from different layers and refine the features at each individual layer. In addition, Mishra et al. [181] used fCNN with attention module for the automatic and accurate segmentation of the ultrasound images which has broken boundaries.

### 3.1.3 Others (*Class Imbalance, Image Registration, Generation and etc.*)

One of the challenges in image classification/detection/segmentation is class imbalance since most voxels/pixels in the image are from non-disease class and often have a different number of images for each disease. Researchers attempted to solve this problem by adapting a loss function [29] and performing data augmentation on positive samples [127, 162, 198], and etc. The loss function was defined as a larger weight for the specificity to make it less sensitive to data imbalance. In addition, often scientists tried to use images from multiple experiments and multiple tomography techniques, but the resolution, orientation, even dimensionality of the dataset was not the same. The researchers made use of algorithms that attempt to find the best image alignment transformation (registration), generation, reconstruction, and combination of image and text reports [41, 128, 133, 157, 233, 239, 240, 270]. Liao et al. [157] presented a 3D medical image registration method along with an agent, trained end-to-end to perform the registration task coupled with attention-driven hierarchical strategy, and Huang et al. [115] paired FDG-PET and T1-MRI for two different biomarkers for Alzheimer's disease with image registration and multi-modality classifiers. In [41, 233], the authors tried to overcome the low diagnostic performance from low-dose CT and low-dose Cone-beam CT (CBCT) images. For low-dose CBCT image, they developed a statistical iterative reconstruction (SIR) algorithm using pre-trained CNN to overcome the data deficiency problem, the noise level and resolution of images. For low-dose CT, they proposed segmentation technology for denoising LDCT and generating normal dose CT.

## 3.2 Electronic Health Records

The terms ‘electronic medical record’ and ‘electronic health record’ has been often used interchangeably. EMRs are a digital version of the paper charts in the clinician’s office, focusing on the medical and treatment history and EHRs are designed for sharing the total health information of patients with other health care providers, such as laboratories and specialists. Since EHR was primarily designed for the internal purpose in the hospital, medical ontologies schema already exists such as the International Statistical Classification of Diseases (ICD) (Fig. 13), Current Procedural Terminology (CPT), Logical Observation Identifiers Names and Codes (LOINC), Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), Unified Medical Language Systems (UMLS) and RxNorm medication code. These codes can vary from institution to institution, and even in the same institution, the same clinical phenotype is represented in different ways across the data [238]. For example, in EHR, patients diagnosed with acute kidney injury can be identified with a laboratory value of serum creatinine (sCr) level 1.5 times or 0.3 higher than the baseline sCr, presence of 584.9 ICD-9 code, ‘acute kidney injury’ mentioned in the free text clinical notes and so on.

Diseases Of The Genitourinary System 580-629
Nephritis, Nephrotic Syndrome, And Nephrosis 580-589
renal (kidney) failure 586 -
- acute 584.9
- necrosis cortical 584.6
- necrosis medullary 584.7
- necrosis tubular 584.5
- specified pathology NEC 584.8
- chronic 585.9

Fig. 13. ICD-9-CM Diagnosis Codes of Acute Kidney Injury.

subject_id	date	prescriptions					input events			labevents									
		Heparin Sodium	Nitro glycerin	Docusate Sodium	Insulin	...	vanco mycin	vanco mycin trough	...	Creatinine high	Creatinine normal	Creatinine low	Creatinine	Lymphocytes	CD8	WBC, CSF	WBC, Joint Fluid	...	
1	2029.01.23 07:23:00	1	0	0	0		1	0		1	1	1	1	1	0	1	0		
2	2029.01.23 07:23:00	1	1	0	0		1	1		0	0	0	0	1	0	0	1		
1		0	0	1	0		1	1		1	1	1	1	0	0	1	1		

Fig. 14. Sample of pre-processed labevents data. Each row represents a visit to the clinic.

EHR systems include structured data (demographics, diagnostics, physical exams, sensor measurements, vital signs, laboratory tests, prescribed or administered medications, laboratory measurements, observations, fluid balance, procedure codes, diagnostic codes, hospital length of stay) and unstructured data (notes charted by care providers, imaging reports, observations, survival data, and more) (Fig. 14). Challenges in EHR research contain high-dimensionality, heterogeneity, temporal dependency, sparsity and irregularity [81, 113, 121, 164, 166, 169, 180, 216, 238]. EHRs are composed of numerical variables such as 1mg/dl, 5% and 5kg, datetime such as admission time, date of birth and date of death, and categorical values such as gender, ethnicity, insurance, ICD-9/10 codes (approx. 68,000 codes) and procedure codes (approx. 9,700 codes), and free-text from clinical notes. In fact, the data is not only heterogeneous but also very different in distribution. Previous studies have applied deep learning on electronic health records for diseases/admission prediction, information extraction, representation, phenotyping, and de-identification.

### 3.2.1 Outcome Prediction.

In studies using deep learning to predict disease, mortality, and admission from patients' medical records, several studies have shown that one of the main contributions was the characterization of features. Avati et al. [13] proposed to improve the palliative care system with a deep learning approach using observation window and slices. The authors used DNN on EHR data of patients from previous years to predict mortality of patients within the next 3-12 month period. On the other hand, researchers are increasingly using word embeddings in vectorized representations to predict the outcomes. Choi et al. [52] proposed a new method using skip-gram to represent heterogeneous medical concepts (ex. diagnoses, medications and procedures) based on co-occurrence and predict

heart failure with 4 classifiers (LR, NN, SVM, K-nearest neighbors). Because higher-order clinical features may be intuitively meaningful and reduce the dimension of data, but fail to capture some inherent information. And raw data may contain all important information, but be represented by a heterogeneous and unstructured mix of elements. Based on the thoughts that related events would occur in a short time difference, the authors used skip-gram for medical concept vectors and use the patient vector with adding the occurred medical vectors to use for heart failure classifier. Using this proposed representation, the area under the ROC curve (AUC) increased by 23% improvement, compared to the one-hot coding vector representation. Nguyen et al. [185], Stojanovic et al. [245] and Liu et al. [163] also treated medical data as language model inputs. Liu et al. [163] found that the bag-of-word embedding representing better for their chronic disease prediction case. Nguyen et al. [185] also used CBOW with CNN that captures and exploits the spatial local correlation of the inputs (for input images, pixels are more relevant to closer pixels than faraway pixels). Their system, the DeepR, used word2vec (CBOW) and CNN to predict unplanned readmission and motif detection. Not only did they predict discrete clinical event codes as other methods which outperformed than the bag of words and logistic regression models, but they also showed the clinical motif of the convolution filter. Stojanovic et al. [245] generated inpatient representation with both CBOW and skip-gram for each diagnosis and procedure to predict important indicators of healthcare quality (ex. length of stay, total incurred charges and mortality rates) with regression and classification models.

There were studies on RNNs, LSTMs, and GRUs for continuous time signals including structured data (physical exams, vital signs, laboratory tests, medications) and unstructured data (clinical notes, discharge summary), toward the automatic prediction of diseases and readmission. For example, Pham et al. [201] and Xiao et al. [281] predicted future risks via deep contextual embedding of clinical concepts. In [201], the DeepCare framework used clinical concept word embedding with diagnoses and interventions (medications, procedures) and demonstrated the efficacy of the LSTM based method for disease progression modeling, intervention recommendation, and future risk prediction with time parameterizations to handle irregular timing. In terms of different rates of disease progress for each patient, the model was trained with recency attention (weight) via multi-scale pooling (12 months, 24 months, all available history). The attention scheme weighted recent events more than old ones. Readmission prediction study was followed by Xiao et al. [281] via contextual embedding of clinical concepts and a hybrid Topic Recurrent Neural Network (TopicRNN) model. Emergency room visit prediction was also studied by Qiao et al. [211], with two non-linear models (XGBoost, RNN) using yearly EHRs. Esteban et al. [78, 79] studied kidney transplantation related complications' prediction with RNN based approaches (RNN/LSTM/GRU). They converted static and dynamic features (time-dependent), but in binned formats as low, normal, or high, into latent embedded variables respectively, and then combined together. The RNN based models, logistic regression, temporal latent embeddings model and random prediction models were used to predict the transplantation three main endpoints: (i) kidney rejection, (ii) kidney loss and (iii) patient death. Additionally, they pointed out that encoding laboratory measurements were decided to use in a binary way by representing each of them such as high/normal/low, compared to mean or median imputation and normalization/standardization. In addition, a combination of GRU and the residual network was used [222] to develop a hybrid NN for joint prediction of present and period assertions of medical events in clinical notes. They used the clinical notes (ex. discharge summaries and progress notes), and the prediction outcomes were presence assertions with six categories (ex. present, absent, possible, conditional, hypothetical, and not associated) and the period assertions including four categories (ex. current, history, future, and unknown).

In order to settle the missing value problem, three types of studies were conducted: (i) missing value imputation, (ii) using the percentage of missing values as an input, and (iii) using clustering/similarity-based algorithms. Weng et al. [273] analyzed the percentage of missing values such as demographic details, health status, prescriptions, acute medical outcomes, hospital records, did missing value imputation and assessed whether machine learning could improve cardiovascular risk prediction with LR, RF, and NN. The cohort of patients is from 30 to 84 years of age at baseline, with complete data on eight core baseline variables (gender, age, smoking status, systolic blood pressure, blood pressure treatment, total cholesterol, HDL cholesterol, diabetes) used in the established ACC/AHA 10-year risk prediction model. Similarly, Che et al. [40] held experiments on pediatric ICU datasets for Acute Lung Injury (ALI) and proposed a combinatorial architecture of DNN and GRU models in an interpretable mimic learning framework with missing value imputation. The DNN was to take static input features, and the GRU model was to take temporal input features. After training a set of 27 static features such as demographic information and admission diagnosis, and another set of 21 temporal features such as monitoring features and discretized scores made by experts with simple missing value imputation, the authors showed its performance with baseline machine learning methods such as Linear SVM, Logistic Regression (LR), Decision Trees (DT) and Gradient Boosting Trees (GBT). In recent studies, a hierarchical fuzzy classifier DFRBS was proposed using a deep rule-based fuzzy classifier and Gaussian imputation to predict mortality in intensive care units (ICUs) [65], and Golas et al. [90] used a clinical text note to show a model for predicting re-hospitalization within 30 days of heart failure patients with interpolation techniques. The risk prediction model was based on the proposed model deep unified network (DUN) with attention units, a new mesh-like network structure of deep learning designed to avoid over-fitting. Moreover, Che et al. [39] developed GRU-D network which was a variation of the recurrent GRU cell for ICD-9 classification and mortality prediction. For missing values, they measured the percentage of them and showed the correlation between the percentage and mortality. With the demonstration of informative missingness, to fundamentally address the missing value problem, the authors introduced the missing value imputation and decay rates in GRU-D to utilize the missingness directly with the input feature values and implicitly in the RNN states. Input decay was to use last observation with time passed information, and hidden state decay was to capture richer knowledge from missingness.

### 3.2.2 Computational Phenotyping.

With the development of electronic health records, including more structured data, we are able to retain a large number of patient datasets and well-organized data. This allows us to reassess existing traditional disease definitions/explanations and to more closely investigate new definitions and subtypes of diseases. Whereas existing diseases have been defined by clinical experts along with manuals, but the newly developing computational phenotyping aims to find phenotypes and etiology with a data-driven bottom-up approach. And by obtaining new clusters that can represent new phenotypes, it is expected to understand the structure and relationships between diseases and provide better prescriptions and medications with fewer side effects and accompanying diseases.

To discover and stratify new phenotypes or subtypes, unsupervised learning approaches including AE and its variants have been broadly used. For example, Beaulieu-Jones et al. [19] suggested denoising autoencoders (DAEs) phenotype stratification and random forest (RF) classification. They simulated scenarios of missing and unlabelled data which is common in EHR, as well as four case/control labelling methods (all case, one case, percentage case, rule-based case). They randomly corrupted the data and then entered it into the DAE algorithm to extract meaningful features and trained classifiers including RF with DAE hidden nodes. Through different classifiers and scenarios,

the best generalised algorithm was chosen. In addition, in terms of using unlabelled and missing data, they generated those data and conducted trials to see the usefulness of the algorithm in current EHR-based studies. Furthermore, in DeepPatient [179], a deep neural network consisting of a stack of denoising autoencoders (SDA) was used to capture the stable structure and regular pattern in EHR data representing patients. Some general demographic details (age, gender, ethnicity), ICD-9 codes, medications, procedures, and lab tests, as well as free-text clinical notes, were collected and pre-processed, differed by data type, using the Open Biomedical Annotator and SNOMED-CT, UMLS, RxNorm, NegEx and etc. like other researchers [206]. Then, topic modeling and SDA were applied to generalize clinical notes and improve automatic processing. In particular, clinical notes latent variables were produced with latent dirichlet allocation (LDA) [26] as topic modeling, and the frequency of presence of diagnostics, drugs, procedures, and laboratory tests to extract summarized biomedical concepts and normalized data versions was calculated. Finally, SDA was used to derive a general purpose patient representation for clinical predictive modeling, and the performance of DeepPatient was evaluated by disease and patient level. Pivovarov et al. [206] also presented the LDA-based Unsupervised Phenome Model (UPhenome), a probabilistic graphical model for large-scale discovery of computational models of disease or phenotypes with notes, laboratory tests, medications, and diagnosis codes.

	rs185393250			rs143225517			rs7518545			rs116013160		
Allele	AA	Aa	aa	AA	Aa	aa	AA	Aa	aa	AA	Aa	aa
Control 1	1	0	0	0.087	0.912	0.001	0	1	0	0	0.891	0.109
Control 2	1	0	0	0	0.033	0.067	0.998	0	0.002	0.963	0.036	0.001
...												
Case 1	0	1	0	0.999	0	0.001	0	1	0	0	1	0
Case 1298	0	1	0	0.998	0.002	0	0	1	0	0	1	0

Fig. 15. Sample of the frequency of allele combination. ex. 0.087 for AA (TT) and 0.912 for Aa (TC) or aA (CT).  $p + q + r = 1$  for three alleles.

On the other hand, there were computational phenotype studies with different approaches from AE based, but other machine learning models. The association research between genetic variants and phenotypes has been studied. Zhao et al. [307] looked into the single nucleotide polymorphism (SNP) rs10455872 which is associated with increased risk of hyperlipidemia and cardiovascular diseases (CVD) and the minor allele frequency (MAF) of the rs10455872 G allele was measured for the SNP (Fig. 15). Meanwhile, ICD-9 codes from EHRs were mapped into disease phecodes [272, 307] and the phecodes were used as their input for topic modeling. Topic modeling via non-negative matrix factorization (NMF) was used to extract a set of topics from individuals' phenotype data. The association between topic and LPA SNP was shown with pearson correlation coefficient (PCC) and LR to find out the most relevant topic for the SNP and the disease. There have been researches on computational phenotyping to produce clinically interesting phenotypes with matrix/tensor factorization, and Henderson J [102] incorporated auxiliary patient information into the phenotype derivation process and introduced their phenotyping through semi-supervised tensor factorization (PSST). In particular, tensors were described with three dimensions (patients, diagnoses, medication), and semi-supervised clustering was proposed with using pairs of data points that must be clustered together and pairs that must not be clustered together in the same cluster. Deliu et al. [71] and Seymour et al. [230] addressed asthma and sepsis, a kind of heterogeneous disease comprising a number of subtypes and new phenotypes, caused by different pathophysiologic mechanisms. They stressed that the precise identification of their subtypes and their pathophysiological mechanisms

with phenotypes may lead to enable more precise therapeutic and prevention approaches. In particular, both considered non-hierarchical clustering (k-means), but Deliu et al. [71] additionally considered hierarchical clustering, latent class analysis, and mixture modeling. Both evaluated the outcomes with phenotype size, clear separation (distance between clusters, soft or hard decision) and characteristics analysis with the distribution. Similar to those, van den Berge et al. [260] suggested log-likelihood and Kyeong et al. [143] proposed topological data analysis for subtypes of Tinnitus and Attention Deficit Hyperactivity Disorder (ADHD).

Phenotyping algorithms were implemented to identify patients with specific disease phenotypes with EHRs, and the unsupervised based feature selection methods were broadly suggested. However, due to the lack of labelled data, some researchers suggested a fully automated and robust unsupervised feature selection from medical knowledge sources, instead of EHR data. Yu et al. [297] suggested surrogate-assisted feature extraction (SAFE) for high-throughput phenotyping of coronary artery disease, rheumatoid arthritis, Crohn's disease, and ulcerative colitis, which was typically defined by phenotyping procedure and domain experts. The SAFE contained concept collection, NLP data generation, feature selection and algorithm training with Elastic-Net. For UMLS concept collection, they used 5 publicly available knowledge sources including Wikipedia, Medscape, Merck Manuals Professional Edition, Mayo Clinic Diseases and Conditions, and Medline-Plus Medical Encyclopedia, followed by searching for mentions of candidate concepts. For feature selection, they used majority voting, frequency control, and surrogate selection. Surrogate selection was based on the fact that when S relates to a set of features F only through Y, it is statistically plausible to infer the predictiveness of F for Y based on the predictiveness of F for S. Using low and high threshold for the main NLP and ICD-9 counts, the features were selected and then trained by fitting an adaptive Elastic-Net penalized logistic regression. Also, SEMantics-Driven Feature Extraction (SEDFE) [187] showed the performance, compared with other algorithms based on EHR for five phenotypes including coronary artery disease, rheumatoid arthritis, Crohn's disease, ulcerative colitis, and pediatric pulmonary arterial hypertension, and algorithms yielded by SEDFE.

Moreover, there were studies to find new phenotypes and sub-phenotypes and improve current phenotypes by using the supervised learning approach. For example, Cheng et al. [48] used a four-layer CNN model with temporal slow fusion (slowly fuses temporal information throughout the network such that higher layers get access to progressively more global information in temporal dimensions) to solve an issue that still remained after performing matrix/tensor-based algorithms, extracted phenotypes and predicted Congestive Heart Failure (CHF) and Chronic Obstructive Pulmonary Disease (COPD). Lipton et al. [160] and Che et al. [38] framed phenotyping problem as a multilabel classification problem with LSTM and MLP. Che et al. [38]'s pre-trained architecture with DAE also showed the usefulness with structured medical ontologies, especially for rare diseases with few training cases. They also developed a novel training procedure to identify key patterns for circulatory disease and septic shock.

### 3.2.3 Knowledge Extraction.

Clinical notes contain dense information about patient status, and information extraction from clinical notes can be a key step towards semantic understanding of EHRs. It can be started with the sequence labelling or annotation, and Conditional Random Field (CRF) based models have been widely proposed in previous studies. However, DNN was newly suggested by researchers, and Jagannatha and Yu [119] was the first group that explored RNN frameworks. EHR of cancer patients diagnosed with hematological malignancy was used, and the annotated events for notes were broadly divided into two: (i) medication (drug name, dosage, frequency, duration, and route) and (ii) disease (adverse drug events, indication, other sign, symptom or disease), and their RNN

based architecture was found to significantly surpass the CRF model. Wu et al. [280] also showed that DNN outperformed CRFs at the minimal feature setting, achieving the highest F1-score (0.93) to recognize clinical entities in Chinese clinical documents. They developed a deep neural network (DNN) to generate word embeddings from a large unlabelled corpus through unsupervised learning and another DNN for the Named Entity Recognition (NER) task. Unlike word-based maximum likelihood estimation of conditional probability having CRFs, NER used the sentence level log-likelihood approach, which consisted of a convolutional layer, a non-linear layer, and linear layers. On the other hand, Qiu et al. [213] implemented CNN to extract ICD-O-3 topographic codes from a corpus of breast and lung cancer pathology reports, using TF-IDF as a baseline model. Consistently, CNN outperformed the TF-IDF based classifier, however not for well-populated classes but for low prevalence classes, pre-training with word embeddings features on differing corpora achieved better performance. In addition, Luo et al. [167] applied subgraph augmented non-negative tensor factorization (SANTF). That is, the authors converted sentences from clinical notes into a graph representation and then identified important subgraphs. Then the patients were clustered, and simultaneously latent groups of higher-order features of patient clusters were identified, as in clinical guidelines, compared to the widely used non-negative matrix factorization (NMF) and k-means clustering methods. Although several methods of information extraction have already been introduced, Scheurwegs et al. [226] focused on minimal annotation dependent method with unsupervised and semi-supervised techniques for the extraction of multi-word expressions that conveyed a generalizable medical meaning. In particular, they used annotated and unannotated corpus of dutch clinical free text and used a linguistic pattern extraction method based on point-wise linguistic mutual information (LMI), and a bootstrapped pattern mining method (BPM), as introduced by Gupta and Manning [96], comparing with a dictionary-based approach (DICT), a majority voting and a bag of words approach. The performance was assessed with a positive impact on diagnostic code prediction.

Unlike above, in [83], the authors extracted time-related medical information (events and corresponding times) from a document collection of clinic and pathology notes from Mayo Clinic with a joint inference-based approach which outperformed RNN, and then found a combination of date canonicalization and distant supervision rules to find time relations with events, using Stanford’s DeepDive application [300]. DeepDive based system made the best labelling entities to encode domain knowledge and sequence structure into a probabilistic graphical model. Also, the temporal relationship between an event mention and corresponding document creation time was represented as a classification problem, assigning event attributes from the label set (before, overlap, before/overlap, after).

As much as it is important to study how medical concepts and temporal events can be explained, relation extraction on medical data including clinical notes, medical papers, Wikipedia and any other medical related documents is also a key step of building medical knowledge graph. Lv et al. [168] proposed a CRF model for a relation classification model and three deep learning models for optimizing extracted contextual features of concepts. Among the three models, deepSAE was chosen, which was developed for contextual feature optimization with both autoencoder and sparsity limitation remedy solution. They divided the clinic narratives such as discharge summaries or progress notes into complete noun phrases (NPs) and adjective phrases (APs), and relation extraction aimed to determine the type of relationship such as ‘treatment improves medical problem’, ‘test reveals medical problem’, and etc. Ling et al. [159] extracted clinical concepts from free clinical narratives with relevant external resources (Wikipedia, Mayo Clinic), and trained Deep Q-Network (DQN) with two states (current clinical concepts, candidate concepts from external

articles) to optimize the reward function to extract clinical concepts that best describe a correct diagnosis.

In [151], 9 entity types such as medications, indications, and adverse drug events (ADEs) and 7 types of relations between these entities are extracted from electronic health record (EHR) notes via natural language processing (NLP). They used a bidirectional long short-term memory (BiLSTM) conditional random field network to recognize entities and a BiLSTM-Attention network to extract relations, and then proposed with multi-task learning to improve performance (HardMTL, RegMTL, and LearnMTL for hard parameter sharing, parameter regularization, and task relation learning in multi-task learning, respectively). HardMTL further improved the base model, whereas RegMTL and LearnMTL failed to boost the performance. Munkhdalai et al. [184] and Zhang et al. [306] also showed models for clinical relation identification, especially for long-distance intersentential relations. Munkhdalai et al. [184] exploited SVM, RNN and attention models for 9 named entities (ex. medication, indication, severity, ADE) and 7 different types of relations (ex. medication-dosage, medication-ADE, severity-ADE). They showed that the SVM model achieved the best average F1-score outperforming all the RNN variations, however the bidirectional LSTM model with attention achieved the best performance among different RNN models. In [306], they aimed to recognize relations between medical concepts described in Chinese EMRs to enable the automatic processing of clinical texts, with an Attention-Based Deep Residual Network (ResNet) model. Although they used EMRs as input data instead of notes for information extraction, the residual network-based model reduced the negative impact of corpus noise to parameter learning, and the combination of character position attention mechanism enhanced the identification features from different types of entities. More specifically, the model consisted of a vector representation layer (character embedding pre-trained by word2vec, position embedding), a convolution layer, and a residual network layer. Of all other methods (SVM, CNN based, LSTM based, Bi-LSTM based, ResNet based models), the model achieved the best performance on F1-score and efficiency when matched with annotations from clinical notes.

### 3.2.4 *Representation Learning.*

Modern EHR systems contain patient-specific information including vital signs, medications, laboratory measurements, observations, clinical notes, fluid balance, procedure codes, diagnostic codes, and etc. The codes and their hierarchy were initially implemented for internal administrative and billing tasks with their relevant ontologies by clinicians. However, recent deep learning approaches have attempted to project discrete codes into vector space, get inherent similarities between medical concepts, represent patients' status with more details, and do more precise predictive tasks. In general, medical concepts and patient representations have been studied through word embedding and unsupervised learning (temporal characteristics, dimension reduction and dense latent variables).

For medical concepts, Choi et al. [54] showed embeddings of a wide range of concepts in medicine, including diseases, medications, procedures, and laboratory tests. The three types of medical concept embedding with skip-gram were respectively learned from medical journals, medical claims, and clinical narratives. The one from medical journals was used as the baseline for their two new medical concept embeddings. They identified medical relatedness and medical conceptual similarity for embeddings and performed comparisons between the embeddings. Choi et al. [51] addressed the challenges such as (i) combination of sequential and non-sequential information (visits, medical codes and demographic information), (ii) interpretable representations of RNN, (iii) frequent visits, and proposed Med2Vec based on skip-gram, compared with popular baselines such as original skip-gram, GloVe, and stacked autoencoder. For each visit, they generated the corresponding visit

representation with a multi-layer perceptron (MLP), concatenating the demographic information to the visit representation information. Similarly, once such vectors were obtained, clustered diagnoses, procedures, and medications were shown with qualitatively analysis. Choi et al. [52, 53] also vectorized representations, but used skip-grams for clinical concepts relying on the sequential ordering of medical codes. In particular, they represented heterogeneous medical concepts including diagnoses, medications, and procedures based on co-occurrence, evaluated whether they were generally well grouped by their corresponding categories, and captured the relations between medications and procedures as well as diagnoses with similarity. With mapping medical concepts to similar concept vectors, they predicted heart failure with 4 classifiers (ex. LR, neural network (NN), SVM, K-nearest neighbors) [52], and an RNN model with a 12- to 18- month observation window [53]. Henriksson et al. [103] proposed the approaches with ensembles of randomized trees using skip-gram for representations of clinical events. Meanwhile, Tran et al. [256] analyzed patients who had at least one encounter with the hospital services and one risk assessment with their EMR-driven nonnegative restricted Boltzmann machines (eNRBM) for suicide risk stratification, using two constraints into model parameters: (i) nonnegative coefficients, and (ii) structural smoothness. Their framework has led to medical conceptual representations that facilitate intuitive visualizations, automated phenotypes, and risk stratification.

Likewise, for patient representations, researchers tried to consider word embedding [50, 176, 185, 201, 281]. The Deepr system used word embedding and pre-trained CNN with word2vec (CBOW) to predict unplanned readmission [185]. They mainly focused on diagnoses and treatments (which involve clinical procedures and medications). Before applying CNN on a sentence, discrete words were represented as continuous vectors with irregular-time information. For that, one-hot coding and word embedding were considered, and a convolutional layer became on top of the word embedding layers. The Deepr system predicted discrete clinical event codes and showed the clinical motif of the convolution filter. Pham et al. [201] developed their DeepCare framework to predict the next disease stages and unplanned readmission. After exclusion criteria, they had 243 diagnosis, 773 procedure, and 353 medication codes in total, and to be embedded into a vector space. They extended a vanilla LSTM by (i) parameterizing time to enable irregular timing, (ii) incorporating interventions to reflect their targeted influence in the course of illness and disease progression, (iii) using multi-scale pooling over time (12 months, 24 months, and all available history), and finally (iv) augmenting a neural network to infer about future outcomes. Doctor AI system [50] utilized sequences of (event, time) pairs occurring in each patient’s timeline across multiple admissions as input to a GRU network. Patients’ observed clinical events for each timestamp were represented with skip-gram embeddings. And the vectorized patients information was fed into a pre-trained RNN based model, from which future patient statuses could be modeled and predicted. Xiao et al. [281] predicted readmission via contextual embedding of clinical concepts and a hybrid TopicRNN model.

Aside from simple vector aggregation with word embedding, it was also possible to directly model the patient information using unsupervised learning approaches. Some unsupervised learning methods were used to either get dimensionality reduction or latent representation for the patient especially with words such as ICD-9, CPT, LOINC, NDC, procedure codes and diagnostic codes. Zhou et al. [309] analyzed patients’ health data, using unsupervised deep learning-based feature learning (DFL) framework to automatically learn compact representations from patient health data for efficient clinical decision making. Mehrabi et al. [173] and Miotto et al. [179] (Deep Patient) used stacked RBM and stacked denoising autoencoder (SDA) trained on each patient’s temporal diagnosis codes to produce patient latent representations over time, respectively. Mehrabi et al.

[173] paid special attention to temporal aspects of EHR data, constructing a diagnosis matrix for each patient with distinct diagnosis codes per a given time interval.

Finally, in [249], similarity learning was proposed for patients' representation and personalized healthcare. With CNN, they captured local important information, rank the similarity, and then do disease prediction and patient clustering for diabetes, obesity, and chronic obstructive pulmonary disease (COPD).

### 3.2.5 *De-identification.*

EHRs including clinical notes contain critical information for medical investigations, however, most researchers can only access de-identified records, in order to protect the confidentiality of patients. For example, the Health Insurance Portability and Accountability Act (HIPAA) defines 18 types of protected health information (PHI) that need to be removed in clinical notes. A covered entity should be not individually identifiable for the individual or of relatives, employers, or household members of the individual and all the information such as name, geographic subdivisions, all elements of dates, contact information, social security numbers, IP addresses, medical record numbers, biometric identifiers, health plan beneficiary numbers, full-face photographs and any comparable images, account numbers, any other unique identifying number, characteristic, or code, except for some which is required for re-identification [238, 267]. De-identification leads to information loss which may limit the usefulness of the resulting health information in certain circumstances. So, it has been desired to cover entities by de-identification strategies that minimize such loss [238], with manual, cryptographical and machine learning methods. In addition to human error, the larger EHR, the more practical, efficient, and reliable algorithms to de-identify patients' records are needed.

Li et al. [154] applied the hierarchical clustering method based on varying document types (ex. discharge summaries, history and physical reports, and radiology reports) from Vanderbilt University Medical Center (VUMC) and i2b2 2014 de-identification challenge dataset discharge summaries. Instead, Dernoncourt et al. [72] introduced the de-identification system based on artificial neural networks (ANNs), comparing the performance of the system with others including CRF based models on two datasets: the i2b2 and the MIMIC-III ('Medical Information Mart for Intensive Care' de-identification dataset. Their framework consisted of a bidirectional LSTM network (Bi-LSTM) as well as the label sequence optimization, utilizing both the token and character embeddings. Recently, three ensemble methods, combining multiple de-identification models trained from deep learning, shallow learning, and rule-based approaches represented the stacked learning ensemble were more effective than other methods for de-identification processing through the i2b2 dataset, and GAN was also considered to show the possibility of de-identifying EHR with natural language generation [131, 148].

## 3.3 Genomics

Human genomic data contains vast amounts of data. In general, identifying genes themselves with exploring the function and information structure, investigating how environmental factors affect phenotype, protein formation, interaction without DNA sequence modification, the association between genotype and phenotype, and personalized medicine with different drug responses have been aimed to study [1, 10, 76, 88, 193]. More specifically, DNA sequences are collected via microarray or next-generation sequencing (NGS) for specific SNPs only based on a candidate or total sequence as desired. In order to understand the gene itself after the extraction of the genetic data, what kinds of mutation can be done in replication and splicing can be done in transcription have been studied. This is because some mutations and alternative splicings can

cause humans to have different sequences and are associated with diseases. Indeed, the absence of the SMN1 gene for infants has been shown to be associated with spinal muscle atrophy and mortality in North America [33]. In addition, environment factor does not change genotype but phenotype such as DNA methylation or histone modification, and both genotype and phenotype data can be used to understand human biological processes and disclose environmental effects. Furthermore, it is expected to use analysis to enable disease diagnosis and design of targeted therapies [58, 76, 88, 149, 169, 174](Fig. 16).

The genetic datasets are extremely high dimensional, heterogeneous, and unbalanced. Therefore, pre-processing and feature extraction were often needed by domain experts, and recently, machine learning and deep learning approaches were tried to solve the issues [130, 305]. In terms of feature selection and identifying genes, deep learning helped researchers to capture nonlinear features.

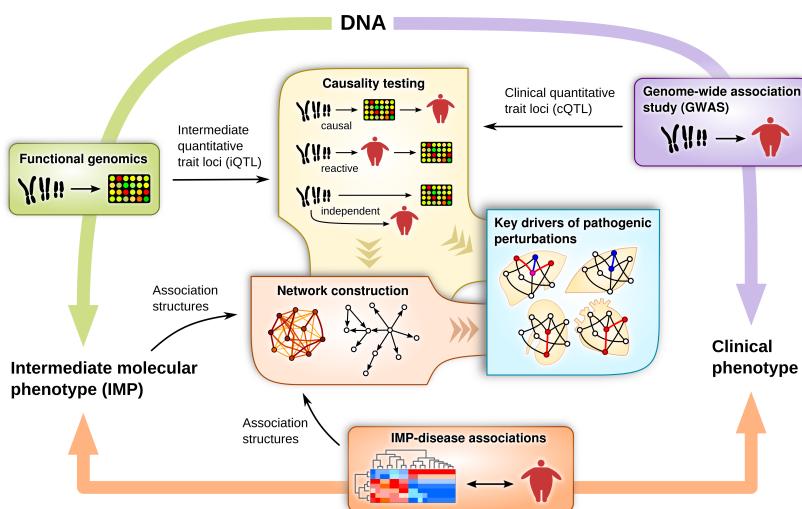


Fig. 16. Systems biology strategies that integrate large-scale genetic, intermediate molecular phenotypes and disease phenotypes [174].

### 3.3.1 Gene Identification.

Genomics involves DNA sequencing exploration of the function and information structure of genes, and it leads researchers to understand the creation of protein sequences and the association between genotype and phenotype. Analysis of genes or alleles identification could help in the diagnosis of disease and in the design of targeted therapies [149]. After genetic data extraction, to understand the gene itself, mutations in replication and splicings in transcription are studied. DNA mutation is an alteration in the nucleotide sequence of the genome which may and may not produce phenotypic changes in an organism. It can be caused by risk factors such as errors or radiation during DNA replication. Gene splicing is a form of post-transcriptional modification processing, of which alternative splicing is the splicing of a single gene into multiple proteins. During RNA splicing, introns (non-coding) and exons (coding) are split and exons are joined together to be transformed into an mRNA. In the meanwhile, unusual splicing can happen, including skipping exons, joining introns, duplicating exons, back-splicing and etc, as shown in Fig. 17. Predicting

mutations and splicing code patterns and identifying genetic variations are critical for shaping the basis of clinical judgment and classifying diseases [35, 114, 117, 144, 155, 178, 214, 253, 271, 299].

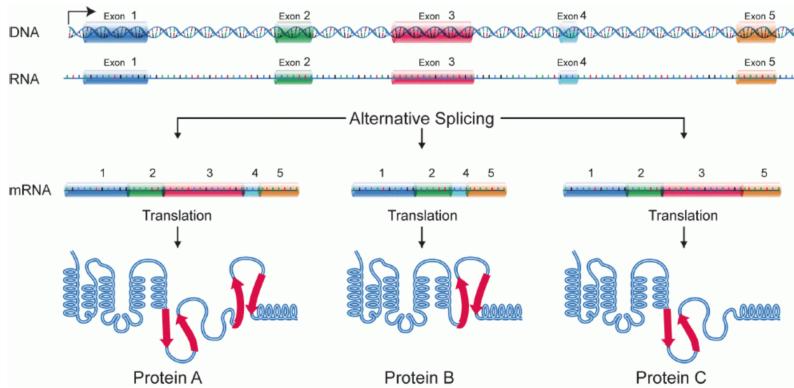


Fig. 17. Alternative splicing produces three protein isoforms [274].

Quang et al. [214] proposed a DNN-based model and compared the performance of the model. The traditional combined annotation-dependent depletion (CANN) annotated both coding and non-coding variants, and trained SVM to separate observed genetic variants from simulated genetic variants. With DNN based DANN, they focused on capturing non-linear relationships among the features and reduced the error rate. In [35, 253], the authors focused on circular RNA, produced through back-splicing and one of the focus of scientific studies due to its association with various diseases including cancer. Chaabane et al. [35] distinguished non-coding RNAs from protein-coding gene transcripts, and separated short and long non-coding RNAs to predict circular RNAs from other long non-coding RNAs (lncRNAs). They proposed ACNN-BLSTM, which used an asymmetric convolutional neural network that described the sequence using k-mer and a sliding window approach and then Bi-LSTM to describe sequence, compared to other CNN and LSTM based architectures.

To understand the cause and phenotype of the disease, unsupervised learning, active learning, reinforcement learning, attention mechanisms, etc. were used. For example, to diagnose the genetic cause of rare Mendelian diseases, Li et al. [155] proposed a highly tolerant phenotype similarity scoring method of noise and imprecision in clinical phenotypes, using the amount of information content concepts from phenotype terms. Zhao et al. [307] identified relationships between phenotypes which result from topic modeling with EHRs and the minor allele frequency (MAF) of the single nucleotide polymorphism (SNP) rs104455872 in lipoprotein which is associated with increased risk of hyperlipidemia and cardiovascular disease (CVD). Baker et al. [16] focused on the development and expression of the midbrain dopamine system since dopamine-related genes are partially responsible for vulnerability to addiction. They adopted a reinforcement learning based method to bridge the gap between genes the behaviour in drug addiction and found a relationship between the DRD4-521T dopamine receptor genotype and substance misuse.

Meanwhile, AE based methods were applied to generalize meaningful and important properties of the input distribution across all input samples. In [63], the authors adopted SDAE to detect functional features and capture key biological principles and highly interactive genes in breast cancer data. Also, in [235], to predict prostate cancer, two DAEs with transfer learning for labelled and unlabelled dataset's feature extraction were introduced. In order to capture information for both labelled and labelled data, they trained two DAEs separately and apply transfer learning to

bridge the gap between them. In addition, Ibrahim et al. [117] proposed a DBN with an active learning approach to find the most discriminative genes/miRNAs to enhance disease classifiers and to mitigate the dimensionality curse problem. Considering group features instead of individual ones, they showed the data representation in multiple levels of abstraction, allowing for better discrimination between different classes. Their method outperformed classical feature selection methods in hepatocellular carcinoma, lung cancer, and breast cancer. Moreover, Hu et al. [114] developed an attention-based CNN framework for human immunodeficiency virus type 1 (HIV-1) genome integration with DNA sequences with and without epigenetic information. Their framework accurately predicted known HIV integration sites in the HEK293T cell line, and attention-based learning allowed them to make which 8-bp sequences are important to predict sites. And they also calculated the enrichment of binding motifs of known mammalian DNA binding proteins to further exploit important sequences. In addition to transcription factors prediction, motif extraction strategies were also studied [144].

### 3.3.2 *Epigenomics.*

Epigenomics aims to investigate the epigenetic modifications on the genetic material such as DNA or histones of a cell, which affect gene expression without altering the DNA sequence itself. Understanding how environmental factors and higher level processes affect phenotypes and protein formation and predicting their interactions such as protein-protein and compound-protein interactions on structural molecular information are important. This is because those are expected to perform virtual screening for drug discovery so that researchers are able to discover possible toxic substances and provide a way for how certain drugs can affect certain cells. DNA methylation and histone modification are one of the best characterized epigenetic processes. DNA methylation is the process by which methyl groups are added to a DNA molecule, altering gene expression without changing the sequence. Also, histones do not affect sequence changes but affect the phenotype. It became even much further possible, according to the development of biotechnology to reduce the cost of collecting genome sequencing and analyze the processes.

In previous studies, DNN was used to predict DNA methylation states from DNA sequence and incomplete methylation profiles in single cells, and they provided insights with the parameters into the effect of sequence composition on methylation variability [9]. Likewise, in [5], CNN was applied to predict specificities of DNA- and RNA-binding proteins, chromatin marks from DNA sequence and DNA methylation states, and Koh et al. [135] applied a convolutional denoising algorithm to learn a mapping from suboptimal to high-quality histone ChIP-sequencing data, which identifies the binding with chromatin immunoprecipitation (ChIP).

While DNN and CNN were the most widely used architectures for extracting features from DNA sequences, other unsupervised approaches have been proposed. In [82], clustering was introduced to identify single gene (Mendelian) disease as well as autism subtypes and discern signatures. They conducted pre-filtration for most promising methylation sites, iterated clustering the features to identify co-varying sites to further refine the signatures to build an effective clustering framework. Meanwhile, in view of the fact that RNA-binding proteins (RBPs) are important in the post-transcriptional modification, Zhang et al. [305] developed the multi-modal DBN framework to identify RNA-binding proteins (RBPs)' preferences and predict the binding sites of them. The multi-modal DBN was modelled with the joint distribution of the RNA base sequence and structural profiles together with its label (1D, 2D, 3D, label), to predict candidate binding sites and discover potential binding motifs.

### 3.3.3 Drug Design.

Identification of genes opens the era for researchers to enable the design of targeted therapies. Individuals may react differently to the same drug, and drugs that are expected to attack the source of the disease may result in limited metabolic and toxic restriction. Therefore, an individual's drug response by differences in genes has been studied to design more personalized treatment drugs whilst reducing side effects and also develop the virtual screening by training supervised classifiers to predict interactions between targets and small molecules [149, 215].

Kearnes et al. [130] showed the structural information, considering a molecular structure and its bonds as a graph and edges. Although their graph convolution model did not outperform all fingerprint-based methods, they represented a new potential research paradigm in drug discovery. Meanwhile, Segler et al. [229] and Yuan et al. [298] reported their studies using RNNs to generate novel chemical structures. In [229], they employed the SMILES format to get sequences of single letters, strings or words. Using SMILES, molecular graphs were described compactly as human-readable strings (ex. c1ccccc), to input strings as input and get strings as output from pre-trained three stacked LSTM layers. The model was first trained on a large dataset for a general set of molecules, then retrained on the smaller dataset of specific molecules. In [298], their library generation method was described, Machine-based Identification of Molecules Inside Characterized Space (MIMICS) to apply the methodology toward drug design applications.

On the other hand, there were studies of other deep learning based methods in chemoinformatics. In [25, 73, 92, 125], VAE based models were applied, and among them, Gómez-Bombarelli et al. [92] generated map chemical structures (SMILES strings) into latent space, and then used the latent vector to represent the molecular structure and transformed again in SMILES format. In [73], their framework was presented to extract latent variables for phenotype extraction using VAEs to predict response to hundreds of cancer drugs based on gene expression data for acute myeloid leukemia. Also, Jaques et al. [120] applied reinforcement learning with RNN to generate high yields of synthetically accessible drug-like molecules with SMILES characters. In [215], the authors examined several aspects of the multi-task framework to achieve effective virtual screening. They demonstrated that more multi-task networks improve performance over single-task models and the total amount of data contributes significantly to the multi-task effect.

## 3.4 Sensing and Online Communication Health

Biosensors are wearable, implantable, and ambient devices that convert biological responses into electro-optical signals and make continuous monitoring of health and wellbeing possible, even with a variety of mobile apps. Since EHRs often lack patients' self-reported experiences, human activities and vital signs outside of clinical settings, tracking those continuously is expected to improve treatment outcomes by closely analyzing patient's condition [7, 124, 259, 292, 293]. Online environments, including social media platforms and online health communities, are expected to help individuals share information, know their health status, and also provide a new era of precision medicine as well as infectious diseases and health policies [60, 170, 199, 205, 302–304].

### 3.4.1 Sensing.

For decades, various types of sensors have been used for signal recording, abnormal signal detection, and more recent predictions. With the development of feasible, effective and accurate wearable devices, electronic health (eHealth) and mobile health (mHealth) applications and telemonitoring concepts have also recently been incorporated into patient care [32, 171, 195, 257, 286](Fig. 18).

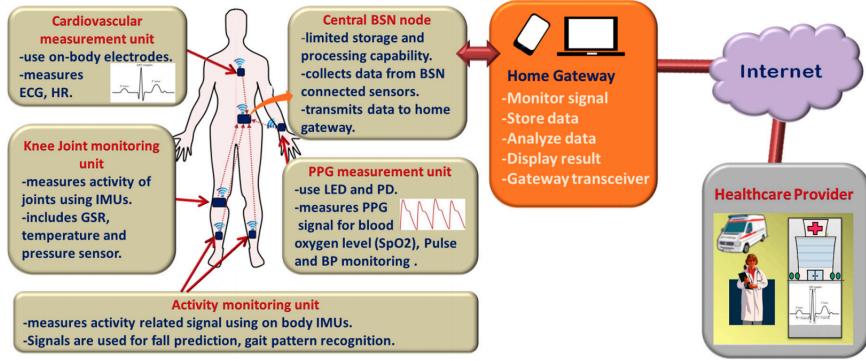


Fig. 18. Schematic overview of the remote health monitoring system with the most used possible sensors worn on different locations (ex. the chest, legs, or fingers) [171].

Especially for elderly patients with chronic diseases and critical care, biosensors can be utilized to track vital signs, such as blood pressure, respiration rate, and body temperature. It can detect abnormalities in vital signs to anticipate extreme health status in advance and provide health information before hospital admission. Even though continuous signals (EEG, ECG, EMG, etc) vary from patient to patient and are difficult to control due to noise and artifacts, deep learning approaches have been proposed to solve the problems. In addition, emergency intervention apps were also being developed to speed the arrival of relevant treatments [195, 286].

For example, Iqbal et al. [118] pointed out that some cardiac diseases, such as myocardial infarction (MI) and atrial fibrillation (Af), require special attention, and classified MI and Af with three steps of the deep deterministic learning (DDL). First, they detected an R peak based on fixed threshold values and extracted time-domain features. The extracted features were used to recognize patterns and divided into three classes with ANN and finally executed to detect MI and Af. Munir et al. [183] presented an anomaly detection technique, Fuse AD with streaming data. The first step was forecasting models for the next time-stamp with Autoregressive integrated moving average (ARIMA) and CNN in a given time-series. The forecasted results were fed into the anomaly detector module to detect whether each time-stamp was normal or abnormal.

For monitoring Parkinson's disease symptoms, accelerometer, gyroscope, cortical and subcortical recordings with the mobile application has been used to detect tremor, freezing of gait, bradykinesia, and dyskinesia. Parkinson's disease (PD) arises with the death of neurons which produce dopamine controlling the movement of the body. Hence, to detect the brain abnormality and early diagnose PD, 14 channels from EEG were used in [190] with CNN. As neurons die, the amount of dopamine produced in the brain is reduced and different patterns are created for each channel to classify PD patients. In particular, Eskofier et al. [77] specifically focused on the detection of bradykinesia with CNN. They made 5 seconds non-overlapping segments from sensor data for each patient and used eight standard features which widely used as a standard set (total signal energy, maximum, minimum, mean, variance, skewness, kurtosis, frequency content of signals) [18, 194]. After normalization and training classification, CNN outperformed any others in terms of classification rate. For cardiac arrhythmia detection, CNN based approach was also used by Yıldırım et al. [291], based on long term ECG signal analysis with long duration raw ECG signals. They used 10 seconds segments and trained the classifier for 13, 15 and 17 cardiac arrhythmia diagnostic classes. Li et al. [150] used pre-trained DBN to classify spatial hyperspectral sensor data, with logistic

regression as a classifier. Amengual-Gual et al. [7] and Ulate-Campos et al. [259] also covered the potentiality of automatic seizure detection with detection modalities such as accelerometer, gyroscope, electrodermal activity, mattress sensors, surface electromyography, video detection systems, and peripheral temperature.

Obesity has been identified as one of the growing epidemic health problems and has been linked to many chronic diseases such as type 2 diabetes and cardiovascular disease. Smartphone-based systems and wearable devices have been proposed to control calorie intake and emissions [109, 140, 175, 208]. For instance, a deep convolutional neural network architecture, called NutriNet, was proposed [175]. They achieved a classification accuracy of 86.72%, along with an accuracy of 94.47% on a detection dataset, and they also performed a real-world test on datasets of self-acquired images, combined with images from Parkinson's disease patients, all taken using a smartphone camera. This model was expected to be used in the form of a mobile app for the Parkinson's disease dietary assessment, so it was important to enable real situations for practical use. In addition, mobile health technologies for resource-poor and marginalized communities were also studied with reading X-ray images taken by a mobile phone [32].

### 3.4.2 *Online Communication Health.*

Based on online data that patients or their parents wrote about symptoms, there were studies that helped individuals, including pain, fatigue, sleep, weight changes, emotions, feelings, drugs, and nutrition [67–69, 172, 191, 204, 278, 289, 302]. For mental issues, writing and linguistic style and posting frequency were important to analyze symptoms and predict outcomes.

Suicide is among the 10 most common causes of death, as assessed by the World Health Organization. Kumar et al. [141] and Coppersmith et al. [61] pointed out that social media can offer new types of data to understand the behavior and pervasiveness and prevent any attempts and serial suicides. Both detected quantifiable signals around suicide attempts and how people are affected by celebrity suicides with natural language processing. Kumar et al. [141] used n-gram with topic modeling. The contents before and after celebrity suicide were analyzed, focusing on negative emotion expressions. Topic modeling with latent dirichlet allocation (LDA) was held on posts shared during two weeks preceding and succeeding the celebrity suicide events to measure topic increases in post-suicide periods. In [61], they initialized the model pre-trained GloVe embeddings, and sequences of word vectors were processed via a bidirectional LSTM layer with using skip connections into a self-attention layer, to capture contextual information between words and apply weights to the most informative subsequences. Meanwhile, in the fact that patients tend to discuss the diagnosis at an early stage [302], and the emotional response to the patient's posts can affect the emotions of others [30, 212, 301], social media data can be used to (i) analyze and identify the characteristics of patients and (ii) help them have good eating habits, stable health condition with proper medication intake, and mental support [66, 100, 269].

Furthermore, investigating infectious diseases such as fever, influenza and systemic inflammatory response syndrome (SIRS) were suggested to uncover key factors and subgroups, improve diagnosis accuracy, warn the public in advance, suggest appropriate prevention, and control strategies [4, 27, 70, 258]. For instance, in [36], the authors addressed that infectious disease reports can be incomplete and delayed and used the search engines' data (both the health/medicine field search engine and the highest usage search engine), weather data from the Korea Meteorological Administration's weather information open portal, Twitter, and infectious disease data from the infectious disease web statistics system. It showed the possibility that deep learning can not only supplement current infectious disease surveillance systems but also predict trends in infectious disease, with immediate responses to minimize costs to society.

## 4 CHALLENGES AND FUTURE DIRECTIONS

Deep learning gives us an exploration of a new era in recent years in machine learning and pattern recognition. And we reviewed how deep learning can be implemented for different types of clinical data and health informatics. Despite the notable advantages, there are some challenges.

### 4.1 Data

Medical data describes patients' health conditions over time, however, it is challenging to identify the true signals from the long-term context due to the complex associations among the clinical events. Data is high-dimensional, heterogeneous, temporal dependent, sparse and irregular. Although the amount of data increases, still lack of labelled data remains a problem. Accordingly data pre-processing and data credibility and integrity can be also thought of.

#### 4.1.1 *Lack of Data and Labelled Data.*

Although there are no hard guidelines about the minimum number of training sets, more data can make stable and accurate models. However, in general, there is still no complete knowledge of the causes and progress of the disease, and one of the reasons is the lack of data. In particular, the number of patients is limited in a practical clinical scenario for rare diseases, certain age-related diseases, or in a case of cutting out patients with missing values.

In addition, health informatics requires domain experts more than any other domain to label complex data and test whether the model performs well and is practically usable. Although labels generally help to have good performance of clinical outcomes or actual disease phenotypes, label acquisition is expensive.

The basis for achieving this goal is the availability of large amounts of data with well-structured data store system guidelines. Also, we need to attempt to label EHR data implicitly with unsupervised, semi-supervised and transfer learning, as previous articles. In general, the first admission patient, disabled or transferred patient may be in worse health status and emergent circumstance, but with no information about medication allergy or any history. If we can use simple tests and calculate patient similarity to see the potential for each risk factor, modifiable complications and crises will be reduced.

Furthermore, to train the target disease using different disease data, especially when the disease is class imbalanced, transfer learning, multi-task learning, reinforcement learning, and generalized algorithms can be considered. In addition, data generation and reconstruction can be other solutions besides incorporating expert knowledge from medical bibles, online medical encyclopedias, and medical journals.

#### 4.1.2 *Data Preprocessing.*

Another important aspect to take into account when deep learning tools are employed is preprocessing. It is pointed out that encoding laboratory measurements in EHRs are decided to use in binary or low/medium/high or minimum/average/maximum ways, missing value interpolation, normalization or standardization is normally considered for pre-processing. Although it is a way to represent the data, especially when the data is high-dimensional, sparse, irregular, biased and multi-scale, none of DNN, CNN and RNN based models with one-hot encoding or AE or matrix/tensor factorization fully settled the problem. Thus, preprocessing, normalization or change of input domain, class balancing and hyperparameters of models are still a blind exploration process.

In particular, considering temporal data, RNN/LSTM/GRU based models with vector-based inputs as well as attention models have already been used in previous studies and are expected to play a

significant role toward better clinical deep architectures. However, what we should point out is that some patients with acute and chronic diseases have different time scales to investigate, and it can take a very long (5 years) time to track down for chronic diseases. Also, depending on the requirements, variables are measured with different timestamps (hourly, monthly, yearly time scale), and we need to understand how to handle those irregular time scale data.

#### *4.1.3 Data Informativeness (high dimensionality, heterogeneity, multi-modality).*

To cope with the lack of information and sparse, heterogeneous data and low dose radiation images, unsupervised learning for high-dimensionality and sparsity and multi-task learning for multi-modality have been proposed. Especially, in the case of multi-modality, these were studies that combined various clinical data types, such as medications and prescriptions in lab events from EHR, CT, and MRI from medical imaging. While deep learning research based on mixed data types is still ongoing, to the best of our knowledge, not so many previous literatures provided attempts with different types of medical data, and the multi-modality related research is needed in the future with more reasons.

First of all, even if we use long term medical records, sometimes it is not enough to represent the patients' status. It can be because of the time stamp to record, hospital discharge, or data itself characteristics (ex. binary, low dose radiation image, short-term information provision data). In addition, even for the same CT or EHR, because hospitals use a variety of technologies, the collected data can be different based on CT equipment and basic or certified EHR systems. Furthermore, the same disease can appear very differently depending on clinicians in one institution when medical images are taken, EHRs are recorded, and clinical notes (abbreviations, ordering, writing style) are written.

With regard to outpatient monitoring and sharing of information on emergency and transferred patients, tracking the health status and summary for next hospital admission, it is necessary to obtain more information about patients to have a holistic representation of patient data. However, indeed, there are not much matched and structured data storing systems yet, as well as models. In addition, we need to investigate whether multi-task learning for different types of data is better than one task learning and if it is better, how deeply dividing types and how to combine the outcomes can be other questions. A primary attempt could be a divide-and-conquer or hierarchical approach or reinforcement learning to dealing with this mixed-type data to reduce dimensionality and multi-modality problems.

#### *4.1.4 Data Credibility and Integrity.*

More than any other area, healthcare data is one of those areas that is heterogeneous, ambiguous, noisy, and incomplete but requires a lot of clean, well-structured data. For example, biosensor data and online data are in the spotlight as they can be a useful data source to track health conditions continuously even outside of clinical settings, extract people's feedback, sentiments and detect abnormal vital signs. Furthermore, investigating infectious diseases can also improve diagnosis accuracy, warn the public in advance and suggest appropriate prevention and management strategies. Accordingly, data credibility and integrity from biosensors, mobile applications, and online sources are required to be controlled.

First of all, if patients collect data and record their symptoms on websites and social media, it may not be worthwhile to use them in forecasting without proper instructions and control policies. In the point of data generation and collection, patients may not be able to collect consistent data and it may be affected by the environment. Patients with chronic illnesses may need to wear sensors

and record their symptoms almost chronically, and it is difficult to expect consistently clean data collection. Not only considering how it can be easy to wear devices, collect clear data and combine clear and unclear data but also it would be helpful to study how we can educate patients in the most efficient way. In addition, online community data can be written in unstructured languages such as misspellings, jokes, metaphors, slang, sarcasm, and etc. Despite these challenges, there is a need for research that bridges the gap between all kinds of clinical information collected from hospitals and patients. And analyzing the data are expected to empower patients and clinicians to provide better health and life for clinicians and individuals.

Second, the fact that patient signals are always detectable can be a privacy concern. People may not want to share data for a long time, which is one of the reasons why most of the research in this paper use either a few de-identified publicly available hospital data or their own institution's privately available dataset. In addition, for mental illness patients, there is a limitation that patients who may and may not want to disclose their data have different starting points. In particular, when using the online community and social media, researchers should take into account the side effects. It can be much easier to try to use information and platform abusively in political and commercial purposes than any other data.

## 4.2 Model

### 4.2.1 *Model Interpretability and Reliability.*

Regardless of the data type, model credibility, interpretability and how we can apply in practice will be another big challenge. The model or framework has to be accurate without overfitting and precisely interpretable to convince clinicians and patients to understand and apply the outcomes in practice. Especially, when training data is small, noisy and rare, a model can be easily fooled. Sometimes it seems that a patient has to have surgery with 90% of certain diseases, but the patient is an unusual case, there may be no disease. However, opening the body can lead to high mortality due to complications, surgical burden, and the immune system. Because of concerns and assurances, there were studies to use multi-modal learning and test normal images trained the model with images taken by PD patients, so that the model could be precise and also generalized. The accuracy is important to convince users because it is related to cost, life-death problem, reliability, and others. At the same time, even if the prediction accuracy is superior to other algorithms, the interpretability is still important and should be taken care of.

Despite recent works on visualization with convolutional layers, clusters using t-SNE, word-cloud, similarity heatmaps, or attention mechanisms, deep learning models are often called as black boxes which are not interpretable. More than any other deterministic domains, in health care, such model interpretability is highly related to whether a model can be used practically for medications, hospital admissions, and operations, with convincing both clinicians and patients. It would be a real hurdle if the model provider does not fully explain to the non-specialist why and how certain patients will have a certain disease with a certain probability on a certain date. Therefore, model credibility, interpretability, and application in practice should be equally important to health care issues.

### 4.2.2 *Model Feasibility and Security.*

Building deep learning models and sharing models with other important research areas without leaking patient sensitive information will be an important issue in the future. If a patient agrees to share data with one clinical institution, but not publicly available to all institutions, our next question might be how to share data on what extent. In particular, deep learning based systems

for cloud computing based biosensors and smartphone applications are growing, and we are emphasizing the importance of model interpretability. It can be a real concern if it is clearer to read the model with parameters and there are attacks that violate the model and privacy. Therefore, we must consider research to protect the privacy of the deep learning models.

For cloud computing-based biosensors and smartphone applications, where and when model trained is another challenge. Training is a difficult and expensive process. Our biosensor and mobile app typically send requests to web services along with newly collected data, and the service stores data to train and replies with the prediction outcomes. However, some diseases progress quickly, and patients need immediate clinical care to avoid intensive care unit (ICU) admission (Fig. 19). There have been studies for deep learning on mobile devices, reinforcement learning and edge computing which focus on bringing computing to the source of data closely, and we should study how to implement this system in health care as well as the development of algorithms for both acute and chronic cases.

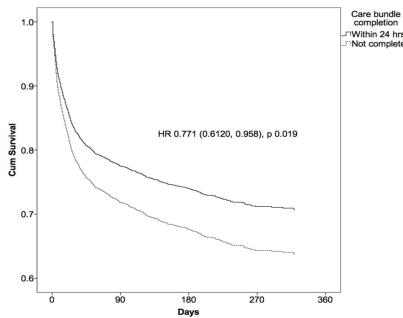


Fig. 19. Adjusted survival curve stratified by timing of completion of AKI Care Bundle [137].

#### 4.2.3 Model Scalability.

Finally, we want to emphasize the opportunity to address the scalability of the model. In most of the previous studies, either a few de-identified publicly available hospital datasets or their own institution's privately available datasets were used. However, patient health conditions and data at public hospitals or general practitioners (GP) clinics or disabled hospitals can be very different due to accessibility to the hospital and other reasons. In general, these hospitals may have less data information stored in the hospitals, but patients may be in an emergency with greater potential. We need to consider how our model with private hospitals or one hospital or one country can be extended for global use.

To conclude, while there are several limitations, we believe that healthcare informatics with deep learning can ultimately change human life. As more data becomes available, system supports, more researches are underway, deep learning can open up a new era of diagnosing disease, locating cancer, predicting the spread of infectious diseases, exploring new phenotypes, predicting strokes in outpatients and etc. This review provides insights into the future of personalized precision medicine and how to implement deep learning methods for clinical data to support better health and life.

## REFERENCES

- [1] James A Diao, Isaac S Kohane, and Arjun K Manrai. 2018. Biomedical Informatics and Machine Learning for Clinical Genomics. *Human Molecular Genetics* 27 (03 2018). <https://doi.org/10.1093/hmg/ddy088>
- [2] Amir Alansary, Konstantinos Kamnitsas, Alice Davidson, Rostislav Khlebnikov, Martin Rajchl, Christina Malamateniou, Mary Rutherford, Joseph V Hajnal, Ben Glocker, Daniel Rueckert, et al. 2016. Fast fully automatic segmentation of the human placenta from motion corrupted MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 589–597.
- [3] Amir Alansary, Ozan Oktay, Yuanwei Li, Loic Le Folgoc, Benjamin Hou, Ghislain Vaillant, Konstantinos Kamnitsas, Athanasios Vlontzos, Ben Glocker, Bernhard Kainz, and Daniel Rueckert. 2019. Evaluating Reinforcement Learning Agents for Anatomical Landmark Detection. *Medical Image Analysis* (2019).
- [4] Ilseyar Alimova, Elena Tutubalina, Julia Alferova, and Guzel Gafiyatullina. 2017. A Machine Learning Approach to Classification of Drug Reviews in Russian. 64–69. <https://doi.org/10.1109/ISPRAS.2017.00018>
- [5] Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. 2015. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology* 33, 8 (2015), 831.
- [6] Md Zahangir Alom, Chris Yakopcic, Mst Shamima Nasrin, Tarek M Taha, and Vijayan K Asari. 2019. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *Journal of digital imaging* (2019), 1–13.
- [7] Marta Amengual-Gual, Adriana Ulate-Campos, and Tobias Loddenkemper. 2018. Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients. *Seizure* (2018).
- [8] Simon Andermatt, Simon Pezold, and Philippe Cattin. 2016. Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data. In *Deep Learning and Data Labeling for Medical Applications*. Springer, 142–151.
- [9] Christof Angermueller, Heather J Lee, Wolf Reik, and Oliver Stegle. 2016. Accurate prediction of single-cell DNA methylation states using deep learning. *BioRxiv* (2016), 055715.
- [10] Christof Angermueller, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. 2016. Deep learning for computational biology. *Molecular systems biology* 12, 7 (2016).
- [11] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on medical imaging* 35, 5 (2016), 1207–1216.
- [12] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866* (2017).
- [13] Anand Avati, Kenneth Jung, Stephanie Harman, Lance Downing, Andrew Ng, and Nigam H Shah. 2018. Improving palliative care with deep learning. *BMC medical informatics and decision making* 18, 4 (2018), 122.
- [14] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.
- [15] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- [16] Travis E Baker, Tim Stockwell, Gordon Barnes, Roderick Haesevoets, and Clay B Holroyd. 2016. Reward sensitivity of ACC as an intermediate phenotype between DRD4-521T and substance misuse. *Journal of cognitive neuroscience* 28, 3 (2016), 460–471.
- [17] Oluwatoyin Bamgbola. 2016. Review of vancomycin-induced renal toxicity: an update. *Therapeutic advances in endocrinology and metabolism* 7, 3 (jun 2016), 136–147. <https://doi.org/10.1177/2042018816638223>
- [18] Jens Barth, Jochen Klucken, Patrick Kugler, Thomas Kammerer, Ralph Steidl, Jürgen Winkler, Joachim Hornegger, and Björn Eskofier. 2011. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 868–871.
- [19] Brett K Beaulieu-Jones, Casey S Greene, et al. 2016. Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of biomedical informatics* 64 (2016), 168–178.
- [20] Yoshua Bengio. 2012. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML workshop on unsupervised and transfer learning*. 17–36.
- [21] Yoshua Bengio, Frédéric Bastien, Arnaud Bergeron, Nicolas Boulanger-Lewandowski, Thomas Breuel, Youssouf Chherawala, Moustapha Cisse, Myriam Côté, Dumitru Erhan, Jeremy Eustache, et al. 2011. Deep learners benefit more from out-of-distribution examples. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 164–172.
- [22] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 8 (2013), 1798–1828.

- [23] Yoshua Bengio, Patrice Simard, Paolo Frasconi, et al. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* 5, 2 (1994), 157–166.
- [24] Guthrie S Birkhead, Michael Klompass, and Nirav R Shah. 2015. Uses of electronic health records for public health surveillance to advance public health. *Annual review of public health* 36 (2015), 345–359.
- [25] Thomas Blaschke, Marcus Olivecrona, Ola Engkvist, Jürgen Bajorath, and Hongming Chen. 2018. Application of generative autoencoder in de novo molecular design. *Molecular informatics* 37, 1-2 (2018), 1700123.
- [26] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [27] Todd Bodnar, Victoria C Barclay, Nilam Ram, Conrad S Tucker, and Marcel Salathé. 2014. On the ground validation of online diagnosis with Twitter and medical records. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 651–656.
- [28] Tom Brosch, Roger Tam, Alzheimer’s Disease Neuroimaging Initiative, et al. 2013. Manifold learning of brain MRIs by deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 633–640.
- [29] Tom Brosch, Lisa YW Tang, Youngjin Yoo, David KB Li, Anthony Traboulsee, and Roger Tam. 2016. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE transactions on medical imaging* 35, 5 (2016), 1229–1239.
- [30] Ngot Bui, John Yen, and Vasant Honavar. 2016. Temporal causality analysis of sentiment change in a cancer survivor network. *IEEE transactions on computational social systems* 3, 2 (2016), 75–87.
- [31] Yunliang Cai, Mark Landis, David T Laidley, Anat Kornecki, Andrea Lum, and Shuo Li. 2016. Multi-modal vertebrae recognition using transformed deep convolution network. *Computerized medical imaging and graphics* 51 (2016), 11–19.
- [32] Yu Cao, Chang Liu, Benyuan Liu, Maria J Brunette, Ning Zhang, Tong Sun, Peifeng Zhang, Jesus Peinado, Epi-fanio Sanchez Garavito, Leonid Lecca Garcia, et al. 2016. Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 274–281.
- [33] Luca Cartegni and Adrian R Krainer. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature genetics* 30, 4 (2002), 377.
- [34] Rich Caruana. 1995. Learning many related tasks at the same time with backpropagation. In *Advances in neural information processing systems*. 657–664.
- [35] Mohamed Chaabane, Robert M Williams, Austin T Stephens, and Juw Won Park. 2019. circDeep: Deep learning approach for circular RNA classification from other long non-coding RNA. *Bioinformatics* (2019).
- [36] Sangwon Chae, Sungjun Kwon, and Donghyun Lee. 2018. Predicting infectious disease using deep learning and big data. *International journal of environmental research and public health* 15, 8 (2018), 1596.
- [37] Hang Chang, Ju Han, Cheng Zhong, Antoine M Snijders, and Jian-Hua Mao. 2017. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. *IEEE transactions on pattern analysis and machine intelligence* 40, 5 (2017), 1182–1194.
- [38] Zhengping Che, David Kale, Wenzhe Li, Mohammad Taha Bahadori, and Yan Liu. 2015. Deep computational phenotyping. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 507–516.
- [39] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 8, 1 (2018), 6085.
- [40] Zhengping Che, Sanjay Purushotham, Robinder Khemani, and Yan Liu. 2016. Interpretable deep models for ICU outcome prediction. In *AMIA Annual Symposium Proceedings*, Vol. 2016. American Medical Informatics Association, 371.
- [41] Binbin Chen, Kai Xiang, Zaiwen Gong, Jing Wang, and Shan Tan. 2018. Statistical iterative CBCT reconstruction based on neural network. *IEEE transactions on medical imaging* 37, 6 (2018), 1511–1521.
- [42] Di-Rong Chen, Qiang Wu, Yiming Ying, and Ding-Xuan Zhou. 2004. Support vector machine soft margin classifiers: error analysis. *Journal of Machine Learning Research* 5, Sep (2004), 1143–1175.
- [43] Hao Chen, Dong Ni, Jing Qin, Shengli Li, Xin Yang, Tianfu Wang, and Pheng Ann Heng. 2015. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE journal of biomedical and health informatics* 19, 5 (2015), 1627–1636.
- [44] Jianxu Chen, Lin Yang, Yizhe Zhang, Mark Alber, and Danny Z Chen. 2016. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In *Advances in neural information processing systems*. 3036–3044.
- [45] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).

- [46] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. 2016. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific reports* 6 (2016), 24454.
- [47] Ruida Cheng, Holger R Roth, Le Lu, Shijun Wang, Baris Turkbey, William Gandler, Evan S McCready, Harsh K Agarwal, Peter Choyke, Ronald M Summers, et al. 2016. Active appearance model and deep learning for more accurate prostate segmentation on MRI. In *Medical Imaging 2016: Image Processing*, Vol. 9784. International Society for Optics and Photonics, 978421.
- [48] Yu Cheng, Fei Wang, Ping Zhang, and Jianying Hu. 2016. Risk prediction with electronic health records: A deep learning approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 432–440.
- [49] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
- [50] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. 2016. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference*. 301–318.
- [51] Edward Choi, Mohammad Taha Bahadori, Elizabeth Seales, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016. Multi-layer representation learning for medical concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1495–1504.
- [52] Edward Choi, Andy Schuetz, Walter F Stewart, and Jimeng Sun. 2016. Medical concept representation learning from electronic health records and its application on heart failure prediction. *arXiv preprint arXiv:1602.03686* (2016).
- [53] Edward Choi, Andy Schuetz, Walter F Stewart, and Jimeng Sun. 2016. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association* 24, 2 (2016), 361–370.
- [54] Youngduck Choi, Chill Yi-I Chiu, and David Sontag. 2016. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings* 2016 (2016), 41.
- [55] Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin DfiAnastasi, et al. 2016. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 415–423.
- [56] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*. Springer, 424–432.
- [57] Francesco Ciompi, Kaman Chung, Sarah J Van Riel, Arnaud Arindra Adiyoso Setio, Paul K Gerke, Colin Jacobs, Ernst Th Scholten, Cornelia Schaefer-Prokop, Mathilde MW Wille, Alfonso Marchiano, et al. 2017. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific reports* 7 (2017), 46479.
- [58] Francis S Collins and Harold Varmus. 2015. A new initiative on precision medicine. *New England journal of medicine* 372, 9 (2015), 793–795.
- [59] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural Language Processing (almost) from Scratch. *CoRR* abs/1103.0398 (2011). arXiv:1103.0398 <http://arxiv.org/abs/1103.0398>
- [60] Lisa M Cunningham and Ginny L Ryan. 2015. The impact of social media on the sexual and social wellness of adolescents. *Journal of pediatric and adolescent gynecology* 28, 1 (2015), 2–5.
- [61] Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights* 10 (2018), 1178222618792860.
- [62] Christopher V Cosgriff, Leo Anthony Celi, and David J Stone. 2019. Critical Care, Critical Data. *Biomedical Engineering and Computational Biology* 10 (2019), 1179597219856564.
- [63] Padideh Danaee, Reza Ghaeini, and David A Hendrix. 2017. A deep learning approach for cancer detection and relevant gene identification. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017*. World Scientific, 219–229.
- [64] Sharon E. Davis, Thomas A. Lasko, Guanhua Chen, Edward D. Siew, and Michael E. Matheny. 2017. Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association* 24, 6 (nov 2017), 1052–1061. <https://doi.org/10.1093/jamia/ocx030>
- [65] Raheleh Davoodi and Mohammad Hassan Moradi. 2018. Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. *Journal of biomedical informatics* 79 (2018), 48–59.
- [66] Munmun De Choudhury. 2015. Anorexia on tumblr: A characterization study. In *Proceedings of the 5th international conference on digital health 2015*. ACM, 43–50.
- [67] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3267–3276.
- [68] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth International AAAI Conference on Weblogs and Social Media*.

- [69] Munmun De Choudhury, Sanket S Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM, 353–369.
- [70] Ed de Quincey, Theocharis Kyriacou, and Thomas Pantin. 2016. # hayfever; A Longitudinal Study into Hay Fever Related Tweets in the UK. In *Proceedings of the 6th international conference on digital health conference*. ACM, 85–89.
- [71] Matea Deliu, Matthew Sperrin, Danielle Belgrave, and Adnan Custovic. 2016. Identification of asthma subtypes using clustering methodologies. *Pulmonary therapy* 2, 1 (2016), 19–41.
- [72] Franck Dernoncourt, Ji Young Lee, Ozlem Uzuner, and Peter Szolovits. 2017. De-identification of patient notes with recurrent neural networks. *Journal of the American Medical Informatics Association* 24, 3 (2017), 596–606.
- [73] Ayse Berceste Dincer, Safiye Celik, Naozumi Hiranuma, and Su-In Lee. 2018. DeepProfile: Deep learning of cancer molecular profiles for precision medicine. *bioRxiv* (2018), 278739.
- [74] Kenji Doya, Kazuyuki Samejima, Ken-ichi Katagiri, and Mitsuo Kawato. 2002. Multiple model-based reinforcement learning. *Neural computation* 14, 6 (2002), 1347–1369.
- [75] Michal Drozdzał, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. 2016. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*. Springer, 179–187.
- [76] Gökçen Eraslan, Žiga Avsec, Julien Gagneur, and Fabian J Theis. 2019. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics* (2019), 1.
- [77] Bjoern M Eskofier, Sunghoon I Lee, Jean-Francois Daneault, Fatemeh N Golabchi, Gabriela Ferreira-Carvalho, Gloria Vergara-Diaz, Stefano Sapienza, Gianluca Costante, Jochen Klucken, Thomas Kautz, et al. 2016. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson’s disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 655–658.
- [78] Cristóbal Esteban, Danilo Schmidt, Denis Krompaß, and Volker Tresp. 2015. Predicting sequences of clinical events by using a personalized temporal latent embedding model. In *2015 International Conference on Healthcare Informatics*. IEEE, 130–139.
- [79] Cristóbal Esteban, Oliver Staack, Stephan Baier, Yinchong Yang, and Volker Tresp. 2016. Predicting clinical events by combining static and dynamic information using recurrent neural networks. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 93–101.
- [80] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017), 115.
- [81] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. 2019. A guide to deep learning in healthcare. *Nature medicine* 25, 1 (2019), 24.
- [82] Mohammad Firouzi, Andrei Turinsky, Sanaa Choufani, Michelle T Siu, Rosanna Weksberg, and Michael Brudno. 2018. An Unsupervised Learning Method for Disease Classification Based on DNA Methylation Signatures. *bioRxiv* (2018), 492926.
- [83] Jason Alan Fries. 2016. Brundlefly at SemEval-2016 Task 12: Recurrent Neural Networks vs. Joint Inference for Clinical Temporal Information Extraction. In *SemEval@NAACL-HLT*.
- [84] Huazhu Fu, Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, and Jiang Liu. 2016. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International conference on medical image computing and computer-assisted intervention*. Springer, 132–139.
- [85] Mingchen Gao, Ziyue Xu, Le Lu, Aaron Wu, Isabella Nogues, Ronald M. Summers, and Daniel J. Mollura. 2016. Segmentation label propagation using deep convolutional neural networks and dense conditional random field. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (2016), 1265–1268.
- [86] Yang Gao, Jeff M Phillips, Yan Zheng, Renqiang Min, P Thomas Fletcher, and Guido Gerig. 2018. Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 1104–1108.
- [87] Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingmar Weber. 2016. Social media image analysis for public health. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5543–5547.
- [88] Erik Gawehn, Jan A Hiss, and Gisbert Schneider. 2016. Deep learning in drug discovery. *Molecular informatics* 35, 1 (2016), 3–14.
- [89] Florin-Cristian Ghesu, Bogdan Georgescu, Yefeng Zheng, Sasa Grbic, Andreas Maier, Joachim Hornegger, and Dorin Comaniciu. 2017. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE transactions on pattern analysis and machine intelligence* 41, 1 (2017), 176–189.
- [90] Sara Bersche Golas, Takuma Shibahara, Stephen Agboola, Hiroko Otaki, Junpei Sato, Tatsuya Nakae, Toru Hisamitsu, Go Kojima, Jennifer Felsted, Sujay Kakarmath, et al. 2018. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical*

- informatics and decision making* 18, 1 (2018), 44.
- [91] Stuart L Goldstein. 2017. Nephrotoxicities. *F1000Research* 6 (jan 2017), 55. <https://doi.org/10.12688/f1000research.10192.1>
  - [92] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. 2018. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science* 4, 2 (2018), 268–276.
  - [93] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
  - [94] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2016. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2016), 2222–2232.
  - [95] Ivo Grondman, Lucian Busoniu, Gabriel AD Lopes, and Robert Babuska. 2012. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1291–1307.
  - [96] Sonal Gupta and Christopher Manning. 2014. Improved pattern learning for bootstrapped entity extraction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*. 98–108.
  - [97] SM Kamrul Hasan and Cristian A Linte. 2018. A Modified U-Net Convolutional Network Featuring a Nearest-neighbor Re-sampling-based Elastic-Transformation for Brain Tissue Characterization and Segmentation. In *2018 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*. IEEE, 1–5.
  - [98] Mohammad Havaei, Nicolas Guizard, Hugo Larochelle, and Pierre-Marc Jodoin. 2016. Deep learning trends for focal brain pathology segmentation in MRI. In *Machine learning for health informatics*. Springer, 125–148.
  - [99] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385 (2015). arXiv:1512.03385 <http://arxiv.org/abs/1512.03385>
  - [100] Ling He and Jiebo Luo. 2016. fiWhat makes a pro eating disorder hashtagfi: Using hashtags to identify pro eating disorder tumblr posts and Twitter users. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 3977–3979.
  - [101] Nicolas Heess, Gregory Wayne, David Silver, Timothy Lillicrap, Tom Erez, and Yuval Tassa. 2015. Learning continuous control policies by stochastic value gradients. In *Advances in Neural Information Processing Systems*. 2944–2952.
  - [102] Malin BA Denny JC Kho AN Ghosh J Ho JC Henderson J, He H. 2018. Phenotyping through Semi-Supervised Tensor Factorization (PSST). *AMIA Annu Symp Proc* (2018), 564–573.
  - [103] Aron Henriksson, Jing Zhao, Hercules Dalianis, and Henrik Boström. 2016. Ensembles of randomized trees using diverse distributed representations of clinical events. *BMC medical informatics and decision making* 16, 2 (2016), 69.
  - [104] J Henry, Yuriy Pylypczuk, Talisha Searcy, and Vaishali Patel. 2016. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2015. *ONC Data Brief* 35 (2016), 1–9.
  - [105] Geoffrey E Hinton. 2012. A practical guide to training restricted Boltzmann machines. In *Neural networks: Tricks of the trade*. Springer, 599–619.
  - [106] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18, 7 (2006), 1527–1554.
  - [107] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *science* 313, 5786 (2006), 504–507.
  - [108] Geoffrey E Hinton, Terrence J Sejnowski, et al. 1986. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition* 1, 282–317 (1986), 2.
  - [109] Irit Hochberg, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Elad Yom-Tov. 2016. Encouraging physical activity in patients with diabetes through automatic personalized feedback via reinforcement learning improves glycemic control. *Diabetes care* 39, 4 (2016), e59–e60.
  - [110] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
  - [111] Ehsan Hosseini-Asl, Georgy Gimel'farb, and Ayman El-Baz. 2016. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. *arXiv preprint arXiv:1607.00556* (2016).
  - [112] Eric A. J. Hoste, Kianoush Kashani, Noel Gibney, F. Perry Wilson, Claudio Ronco, Stuart L. Goldstein, John A. Kellum, Sean M. Bagshaw, and on behalf of the 15 ADQI Consensus Group. 2016. Impact of electronic-alerting of acute kidney injury: workgroup statements from the 15th ADQI Consensus Conference. *Canadian Journal of Kidney Health and Disease* 3, 1 (26 Feb 2016), 10. <https://doi.org/10.1186/s40697-016-0101-1>
  - [113] George Hripcsak and David J Albers. 2012. Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association* 20, 1 (2012), 117–121.
  - [114] Hailin Hu, An Xiao, Sai Zhang, Yangyang Li, Xuanling Shi, Tao Jiang, Linqi Zhang, Lei Zhang, and Jianyang Zeng. 2018. DeepHINT: Understanding HIV-1 integration via deep learning with attention. *Bioinformatics* 35, 10 (2018), 1660–1667.

- [115] Yechong Huang, Jiahang Xu, Yuncheng Zhou, Tong Tong, and Xiahai Zhuang. [n.d.]. Diagnosis of Alzheimer's Disease via Multi-modality 3D Convolutional Neural Network. *CoRR* abs/1902.09904 ([n. d.]). arXiv:1902.09904 <http://arxiv.org/abs/1902.09904>
- [116] David H Hubel and Torsten N Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* 160, 1 (1962), 106–154.
- [117] Rania Ibrahim, Noha A Yousri, Mohamed A Ismail, and Nagwa M El-Makky. 2014. Multi-level gene/MiRNA feature selection using deep belief nets and active learning. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3957–3960.
- [118] Uzair Iqbal, Teh Wah, Muhammad Habib ur Rehman, Ghulam Mujtaba, Muhammad Imran, and Muhammad Shoaib. 2018. Deep Deterministic Learning for Pattern Recognition of Different Cardiac Diseases through the Internet of Medical Things. *Journal of Medical Systems* 42 (12 2018). <https://doi.org/10.1007/s10916-018-1107-2>
- [119] Abhyuday N Jagannatha and Hong Yu. 2016. Structured prediction models for RNN based sequence labeling in clinical text. In *Proceedings of the conference on empirical methods in natural language processing, conference on empirical methods in natural language processing*, Vol. 2016. NIH Public Access, 856.
- [120] Natasha Jaques, Shixiang Gu, Dzmitry Bahdanau, José Miguel Hernández-Lobato, Richard E Turner, and Douglas Eck. 2017. Sequence tutor: Conservative fine-tuning of sequence generation models with kl-control. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1645–1654.
- [121] Peter B Jensen, Lars J Jensen, and Søren Brunak. 2012. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics* 13, 6 (2012), 395.
- [122] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 2012. 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* 35, 1 (2012), 221–231.
- [123] Vasu Jindal, Javad Birjandtalab, M Baran Pouyan, and Mehrdad Nourani. 2016. An adaptive deep learning approach for PPG-based identification. In *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 6401–6404.
- [124] Alistair EW Johnson, Mohammad M Ghassemi, Shamim Nemati, Katherine E Niehaus, David A Clifton, and Gari D Clifford. 2016. Machine learning and decision support in critical care. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers* 104, 2 (2016), 444.
- [125] Artur Kadurin, Sergey Nikolenko, Kuzma Khrabrov, Alex Aliper, and Alex Zhavoronkov. 2017. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular pharmaceutics* 14, 9 (2017), 3098–3104.
- [126] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4 (1996), 237–285.
- [127] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis* 36 (2017), 61–78.
- [128] Andrej Karpathy and Li Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3128–3137.
- [129] Jeremy Kawahara and Ghassan Hamarneh. 2016. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In *International Workshop on Machine Learning in Medical Imaging*. Springer, 164–171.
- [130] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. 2016. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* 30, 8 (2016), 595–608.
- [131] Younghun Kim, Paul Heider, and Stéphane Meystre. 2018. Ensemble-based Methods to Improve De-identification of Electronic Health Record Narratives. In *AMIA Annual Symposium Proceedings*, Vol. 2018. American Medical Informatics Association, 663.
- [132] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
- [133] Pavel Kisilev, Eli Sason, Ella Barkan, and Sharbell Hashoul. 2016. Medical image description using multi-task-loss CNN. In *Deep Learning and Data Labeling for Medical Applications*. Springer, 121–129.
- [134] William A. Knaus and Richard D. Marks. 2019. New Phenotypes for Sepsis. *JAMA* (may 2019). <https://doi.org/10.1001/jama.2019.5794>
- [135] Pang Wei Koh, Emma Pierson, and Anshul Kundaje. 2017. Denoising genome-wide histone ChIP-seq with convolutional neural networks. *Bioinformatics* 33, 14 (2017), i225–i233.
- [136] Tamara Gibson Kolda. 2006. *Multilinear operators for higher-order decompositions*. Technical Report. Sandia National Laboratories.
- [137] Nitin V Kolhe, David Staples, Timothy Reilly, Daniel Merrison, Christopher W McIntyre, Richard J Fluck, Nicholas M Selby, and Maarten W Taal. 2015. Impact of compliance with a care bundle on acute kidney injury outcomes: a prospective observational study. *PloS one* 10, 7 (2015), e0132279.

- [138] Bin Kong, Yiqiang Zhan, Min Shin, Thomas Denny, and Shaoting Zhang. 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network. In *International conference on medical image computing and computer-assisted intervention*. Springer, 264–272.
- [139] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
- [140] Pallavi Kuhad, Abdulsalam Yassine, and Shervin Shimohammadi. 2015. Using distance estimation and deep learning to simplify calibration in food calorie measurement. In *2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, 1–6.
- [141] Mrinal Kumar, Mark Dredze, Glen Coppersmith, and Munmun De Choudhury. 2015. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM conference on Hypertext & Social Media*. ACM, 85–94.
- [142] Upendra Kumar, Esha Tripathi, Surya Prakash Tripathi, and Kapil Kumar Gupta. 2019. Deep Learning for Healthcare Biometrics. In *Design and Implementation of Healthcare Biometric Systems*. IGI Global, 73–108.
- [143] Sunghyun Kyeong, Jae-Jin Kim, and Eunjoo Kim. 2017. Novel subgroups of attention-deficit/hyperactivity disorder identified by topological data analysis and their functional network modular organizations. *PloS one* 12, 8 (2017), e0182603.
- [144] Jack Lanchantin, Ritambhara Singh, Zeming Lin, and Yanjun Qi. 2016. Deep motif: Visualizing genomic sequence classifications. *arXiv preprint arXiv:1605.01133* (2016).
- [145] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Pascal Lamblin. 2009. Exploring strategies for training deep neural networks. *Journal of machine learning research* 10, Jan (2009), 1–40.
- [146] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.
- [147] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
- [148] Scott Lee. 2018. Natural language generation for electronic health records. In *npj Digital Medicine*.
- [149] Michael KK Leung, Andrew Delong, Babak Alipanahi, and Brendan J Frey. 2015. Machine learning in genomic medicine: a review of computational problems and data sets. *Proc. IEEE* 104, 1 (2015), 176–197.
- [150] Chenming Li, Yongchang Wang, Xiaoke Zhang, Hongmin Gao, Yao Yang, and Jiawei Wang. 2019. Deep Belief Network for Spectralf?Spatial Classification of Hyperspectral Remote Sensor Data. In *Sensors*.
- [151] Fei Li, Weisong Liu, and Hong Yu. 2018. Extraction of information related to adverse drug events from electronic health record notes: design of an end-to-end model based on deep learning. *JMIR medical informatics* 6, 4 (2018), e12159.
- [152] Feng Li, Loc Tran, Kim-Han Thung, Shuiwang Ji, Dinggang Shen, and Jiang Li. 2015. A robust deep model for improved classification of AD/MCI patients. *IEEE journal of biomedical and health informatics* 19, 5 (2015), 1610–1616.
- [153] Jiayun Li, Karthik V Sarma, King Chung Ho, Arkadiusz Gertych, Beatrice S Knudsen, and Corey W Arnold. 2017. A multi-scale u-net for semantic segmentation of histological images from radical prostatectomies. In *AMIA Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 1140.
- [154] Muqun Li, David Carrell, John Aberdeen, Lynette Hirschman, and Bradley A Malin. 2014. De-identification of clinical narratives through writing complexity measures. *International journal of medical informatics* 83, 10 (2014), 750–767.
- [155] Qigang Li, Keyan Zhao, Carlos D. Bustamante, Xin Ma, and Wing H. Wong. 2019. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genetics in Medicine* (01 2019). <https://doi.org/10.1038/s41436-019-0439-8>
- [156] Zhang W Zhou H Xu X Qian TW Wan YJ Li MX, Yu SQ. 2019. Segmentation of retinal fluid based on deep learning: application of three-dimensional fully convolutional neural networks in optical coherence tomography images. *Int J Ophthalmol* (2019), 1012–1020.
- [157] Rui Liao, Shun Miao, Pierre de Tournemire, Sasa Grbic, Ali Kamen, Tommaso Mansi, and Dorin Comaniciu. 2017. An artificial agent for robust image registration. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [158] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125.
- [159] Yuan Ling, Sadid A Hasan, Vivek Datla, Ashequul Qadir, Kathy Lee, Joey Liu, and Oladimeji Farri. 2017. Diagnostic inferencing via improving clinical concept extraction with deep reinforcement learning: A preliminary study. In *Machine Learning for Healthcare Conference*. 271–285.
- [160] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. 2015. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677* (2015).
- [161] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.

- [162] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermens, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports* 6 (2016), 26286.
- [163] Jingshu Liu, Zachariah Zhang, and Narges Razavian. 2018. Deep ehr: Chronic disease prediction using medical notes. *arXiv preprint arXiv:1808.04928* (2018).
- [164] Siqi Liu, Kee Yuan Ngiam, and Mengling Feng. [n.d.]. Deep Reinforcement Learning for Clinical Decision Support: A Brief Survey. *arXiv e-prints*, Article arXiv:1907.09475 (Jul [n. d.]), arXiv:1907.09475 pages. arXiv:cs.LG/1907.09475
- [165] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.
- [166] Jake Luo, Min Wu, Deepika Gopukumar, and Yiqing Zhao. 2016. Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights* 8 (2016), BII-S31559.
- [167] Yuan Luo, Yu Xin, Ephraim Hochberg, Rohit Joshi, Ozlem Uzuner, and Peter Szolovits. 2015. Subgraph augmented non-negative tensor factorization (SANTF) for modeling clinical narrative text. *Journal of the American Medical Informatics Association* 22, 5 (2015), 1009–1019.
- [168] Xinbo Lv, Yi Guan, Jinfeng Yang, and Jiawei Wu. 2016. Clinical relation extraction with deep learning. *International Journal of Hybrid Information Technology* 9, 7 (2016), 237–248.
- [169] Gary H Lyman and Harold L Moses. 2016. Biomarker Tests for Molecularly Targeted Therapies—The Key to Unlocking Precision Medicine. *The New England journal of medicine* 375, 1 (2016), 4.
- [170] Will WK Ma and Albert Chan. 2014. Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. *Computers in Human Behavior* 39 (2014), 51–58.
- [171] Sumit Majumder, Tapas Mondal, and M Deen. 2017. Wearable sensors for remote health monitoring. *Sensors* 17, 1 (2017), 130.
- [172] Sarah A Marshall, Christopher C Yang, Qing Ping, Mengnan Zhao, Nancy E Avis, and Edward H Ip. 2016. Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. *Quality of Life Research* 25, 3 (2016), 547–557.
- [173] Saeed Mehrabi, Sunghwan Sohn, Dingheng Li, Joshua J Pankratz, Terry Therneau, Jennifer L St Sauver, Hongfang Liu, and Mathew Palakal. 2015. Temporal pattern and association discovery of diagnosis codes using deep learning. In *2015 International Conference on Healthcare Informatics*. IEEE, 408–416.
- [174] Mkinen VP, Luk H. et al. Meng, Q. 2013. Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases. In *Curr Cardiovasc Risk Rep*.
- [175] Simon Mezgec and Barbara Korouší Seljak. 2017. NutriNet: a deep learning food and drink image recognition system for dietary assessment. *Nutrients* 9, 7 (2017), 657.
- [176] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
- [177] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 565–571.
- [178] Xu Min, Wanwen Zeng, Ning Chen, Ting Chen, and Rui Jiang. 2017. Chromatin accessibility prediction via convolutional long short-term memory networks with k-mer embedding. *Bioinformatics* 33, 14 (2017), i92–i101.
- [179] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. 2016. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports* 6 (2016), 26094.
- [180] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. 2017. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics* 19, 6 (2017), 1236–1246.
- [181] Deepak Mishra, Santanu Chaudhury, Mukul Sarkar, and Arvinder Singh Soin. 2018. Ultrasound Image Segmentation: A Deeply Supervised Network With Attention to Boundaries. *IEEE Transactions on Biomedical Engineering* PP (10 2018), 1–1. <https://doi.org/10.1109/TBME.2018.2877577>
- [182] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S de Vries, Manon JNL Benders, and Ivana Isgum. 2016. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE transactions on medical imaging* 35, 5 (2016), 1252–1261.
- [183] Mohsin Munir, Shoib Ahmed Siddiqui, Muhammad Ali Chattha, Andreas Dengel, and Sheraz Ahmed. 2019. FuseAD: Unsupervised Anomaly Detection in Streaming Sensors Data by Fusing Statistical and Deep Learning Models. *Sensors* 19, 11 (2019), 2451.
- [184] Tsendsuren Munkhdalai, Feifan Liu, and Hong Yu. 2018. Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning. *JMIR public health and surveillance* 4, 2 (2018), e29.
- [185] Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, and Svetha Venkatesh. 2016. DeepR: a convolutional net for medical records. *IEEE journal of biomedical and health informatics* 21, 1 (2016), 22–30.

- [186] Dong Nie, Han Zhang, Ehsan Adeli, Luyan Liu, and Dinggang Shen. 2016. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 212–220.
- [187] Wenxin Ning, Stephanie Chan, Andrew Beam, Ming Yu, Alon Geva, Katherine Liao, Mary Mullen, Kenneth D. Mandl, Isaac Kohane, Tianxi Cai, and Sheng Yu. 2019. Feature extraction for phenotyping from semantic and knowledge resources. *Journal of Biomedical Informatics* 91 (2019), 103122. <https://doi.org/10.1016/j.jbi.2019.103122>
- [188] Mizuho Nishio, Osamu Sugiyama, Masahiro Yakami, Syoko Ueno, Takeshi Kubo, Tomohiro Kuroda, and Kaori Togashi. 2018. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS one* 13, 7 (2018), e0200721.
- [189] Ewan Nurse, Benjamin S Mashford, Antonio Jimeno Yepes, Isabell Kiral-Kornek, Stefan Harrer, and Dean R Freestone. 2016. Decoding EEG and LFP signals using deep learning: heading TrueNorth. In *Proceedings of the ACM International Conference on Computing Frontiers*. ACM, 259–266.
- [190] Shu Lib Oh, Yuki Hagiwara, U Raghavendra, Rajamanickam Yuvaraj, N Arunkumar, M Murugappan, and U Rajendra Acharya. 2018. A deep learning approach for Parkinson’s disease diagnosis from EEG signals. *Neural Computing and Applications* (2018), 1–7.
- [191] Thomas Opitz, Jérôme Azé, Sandra Bringay, Cyrille Joutard, Christian Lavergne, and Caroline Mollevi. 2014. Breast cancer and quality of life: medical information extraction from health forums. In *MIE: Medical Informatics Europe*. 1070–1074.
- [192] Rudra P. K. Poudel, Pablo Lamata, and Giovanni Montana. 2017. Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation. *Lecture Notes in Computer Science* 10129, 83–94. [https://doi.org/10.1007/978-3-319-52280-7\\_8](https://doi.org/10.1007/978-3-319-52280-7_8)
- [193] Lucas Pastur-Romay, Francisco Cedron, Alejandro Pazos, and Ana Porto-Pazos. 2016. Deep artificial neural networks and neuromorphic chips for big data analysis: pharmaceutical and bioinformatics applications. *International journal of molecular sciences* 17, 8 (2016), 1313.
- [194] Shyamal Patel, Konrad Lorincz, Richard Hughes, Nancy Huggins, John Growdon, David Standaert, Metin Akay, Jennifer Dy, Matt Welsh, and Paolo Bonato. 2009. Monitoring motor fluctuations in patients with Parkinson’s disease using wearable sensors. *IEEE transactions on information technology in biomedicine* 13, 6 (2009), 864–873.
- [195] Shyamal Patel, Hyung-Soon Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. 2012. A Review of Wearable Sensors and Systems with Application in Rehabilitation. *Journal of neuroengineering and rehabilitation* 9 (04 2012), 21. <https://doi.org/10.1186/1743-0003-9-21>
- [196] Adrien Payan and Giovanni Montana. 2015. Predicting Alzheimer’s disease: a neuroimaging study with 3D convolutional neural networks. *arXiv preprint arXiv:1502.02506* (2015).
- [197] Judea Pearl. 2014. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.
- [198] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. 2016. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging* 35, 5 (2016), 1240–1251.
- [199] Andrew Perrin. 2015. Social media usage: 2005–2015. (2015).
- [200] Emanuele Pesce, Petros-Pavlos Ypsilantis, Samuel Withey, Robert Bakewell, Vicky Goh, and Giovanni Montana. 2017. Learning to detect chest radiographs containing lung nodules using visual attention networks. *arXiv preprint arXiv:1712.00996* (2017).
- [201] Trang Pham, Thruyen Tran, Dinh Phung, and Svetha Venkatesh. 2016. Deepcare: A deep dynamic memory model for predictive medicine. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 30–41.
- [202] Ha Tran Hong Phan, Ashnil Kumar, Jinman Kim, and Dagan Feng. 2016. Transfer learning of a convolutional neural network for HEp-2 cell image classification. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 1208–1211.
- [203] NhatHai Phan, Dejing Dou, Brigitte Piniewski, and David Kil. 2015. Social restricted boltzmann machine: Human behavior prediction in health social networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 424–431.
- [204] Qing Ping, Christopher C Yang, Sarah A Marshall, Nancy E Avis, and Edward H Ip. 2016. Breast cancer symptom clusters derived from social media and research study data using improved  $k$ -medoid clustering. *IEEE transactions on computational social systems* 3, 2 (2016), 63–74.
- [205] Matthew Pittman and Brandon Reich. 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior* 62 (2016), 155–167.
- [206] Rimma Pivarov, Adler J Perotte, Edouard Grave, John Angiolillo, Chris H Wiggins, and Noémie Elhadad. 2015. Learning probabilistic phenotypes from heterogeneous EHR data. *Journal of biomedical informatics* 58 (2015), 156–165.
- [207] Sergey M Plis, Devon R Hjelm, Ruslan Salakhutdinov, Elena A Allen, Henry J Bockholt, Jeffrey D Long, Hans J Johnson, Jane S Paulsen, Jessica A Turner, and Vince D Calhoun. 2014. Deep learning for neuroimaging: a validation

- study. *Frontiers in neuroscience* 8 (2014), 229.
- [208] Parisa Pouladzadeh, Pallavi Kuhad, Sri Vijay Bharat Peddi, Abdulsalam Yassine, and Shervin Shirmohammadi. 2016. Food calorie measurement using deep learning neural network. In *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE, 1–6.
- [209] Christopher Poultney, Sumit Chopra, Yann L Cun, et al. 2007. Efficient learning of sparse representations with an energy-based model. In *Advances in neural information processing systems*. 1137–1144.
- [210] M Prendecki, E Blacker, O Sadeghi-Alavijeh, R Edwards, H Montgomery, S Gillis, and M Harber. 2016. Improving outcomes in patients with Acute Kidney Injury: the impact of hospital based automated AKI alerts. *Postgraduate Medical Journal* 92, 1083 (2016), 9–13. <https://doi.org/10.1136/postgradmedj-2015-133496> arXiv:<https://pmj.bmjjournals.com/content/92/1083/9.full.pdf>
- [211] Zhi Qiao, Ning Sun, Xiang Li, Eryu Xia, Shiwan Zhao, and Yong Qin. 2018. Using Machine Learning Approaches for Emergency Room Visit Prediction Based on Electronic Health Record Data. *Studies in health technology and informatics* 247 (01 2018), 111–115.
- [212] Baojun Qiu, Kang Zhao, Prasenjit Mitra, Dinghao Wu, Cornelia Caragea, John Yen, Greta E Greer, and Kenneth Portier. 2011. Get online support, feel better—sentiment analysis and dynamics in an online cancer survivor community. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 274–281.
- [213] John X Qiu, Hong-Jun Yoon, Paul A Fearn, and Georgia D Tourassi. 2017. Deep learning for automated extraction of primary sites from cancer pathology reports. *IEEE journal of biomedical and health informatics* 22, 1 (2017), 244–251.
- [214] Daniel Quang, Yifei Chen, and Xiaohui Xie. 2014. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 31, 5 (2014), 761–763.
- [215] Bharath Ramsundar, Steven Kearns, Patrick Riley, Dale Webster, David Konnerding, and Vijay Pande. 2015. Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072* (2015).
- [216] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. 2016. Deep learning for health informatics. *IEEE journal of biomedical and health informatics* 21, 1 (2016), 4–21.
- [217] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. 2011. Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*. Omnipress, 833–840.
- [218] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
- [219] Holger R Roth, Christopher T Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, and Ronald M Summers. 2015. Anatomy-specific classification of medical images using deep convolutional nets. *arXiv preprint arXiv:1504.04003* (2015).
- [220] Holger R Roth, Le Lu, Jiamin Liu, Jianhua Yao, Ari Seff, Kevin Cherry, Lauren Kim, and Ronald M Summers. 2015. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE transactions on medical imaging* 35, 5 (2015), 1170–1181.
- [221] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. 1988. Learning representations by back-propagating errors. *Cognitive modeling* 5, 3 (1988), 1.
- [222] Li Rumeng, N Jagannatha Abhyuday, and Yu Hong. 2017. A hybrid Neural Network Model for Joint Prediction of Presence and Period Assertions of Medical Events in Clinical Notes. In *AMIA Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 1149.
- [223] Tanzila Saba. [n.d.]. Automated lung nodule detection and classification based on multiple classifiers voting. *Microscopy Research and Technique* 0, 0 ([n.d.]). <https://doi.org/10.1002/jemt.23326> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/jemt.23326>
- [224] Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Deep boltzmann machines. In *Artificial intelligence and statistics*. 448–455.
- [225] Ravi K Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark A Helvie, Jun Wei, and Kenny Cha. 2016. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical physics* 43, 12 (2016), 6654–6666.
- [226] Elyne Scheurwegs, Kim Luyckx, Léon Luyten, Bart Goethals, and Walter Daelemans. 2017. Assigning clinical codes with data-driven concept representation on Dutch clinical free text. *Journal of biomedical informatics* 69 (2017), 118–127.
- [227] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.
- [228] John Schulman, Nicolas Heess, Theophane Weber, and Pieter Abbeel. 2015. Gradient estimation using stochastic computation graphs. In *Advances in Neural Information Processing Systems*. 3528–3536.

- [229] Marwin HS Segler, Thierry Kogej, Christian Tyrchan, and Mark P Waller. 2017. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science* 4, 1 (2017), 120–131.
- [230] Christopher W Seymour, Jason N Kennedy, Shu Wang, Chung-Chou H Chang, Corrine F Elliott, Zhongying Xu, Scott Berry, Gilles Clermont, Gregory Cooper, Hernando Gomez, et al. 2019. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *Jama* 321, 20 (2019), 2003–2017.
- [231] Christopher W. Seymour, Jason N. Kennedy, Shu Wang, Chung-Chou H. Chang, Corrine F. Elliott, Zhongying Xu, Scott Berry, Gilles Clermont, Gregory Cooper, Hernando Gomez, David T. Huang, John A. Kellum, Qi Mi, Steven M. Opal, Victor Talisa, Tom van der Poll, Shyam Visweswaran, Yoram Vodovotz, Jeremy C. Weiss, Donald M. Yealy, Sachin Yende, and Derek C. Angus. 2019. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *Jama* (2019). <https://doi.org/10.1001/jama.2019.5791>
- [232] Mahsa Shakeri, Stavros Tsogkas, Enzo Ferrante, Sarah Lippe, Samuel Kadoury, Nikos Paragios, and Iasonas Kokkinos. 2016. Sub-cortical brain structure segmentation using F-CNN's. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 269–272.
- [233] Hongming Shan, Yi Zhang, Qingsong Yang, Uwe Kruger, Mannudeep K Kalra, Ling Sun, Wenxiang Cong, and Ge Wang. 2018. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. *IEEE transactions on medical imaging* 37, 6 (2018), 1522–1534.
- [234] Juan Shan and Lin Li. 2016. A deep learning method for microaneurysm detection in fundus images. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 357–358.
- [235] Hossein Sharifi-Noghabi, Yang Liu, Nicholas Erho, Raunak Shrestha, Mohammed Alshalalfa, Elai Davicioni, Colin C Collins, and Martin Ester. 2019. Deep Genomic Signature for early metastasis prediction in prostate cancer. *BioRxiv* (2019), 276055.
- [236] John Shawe-Taylor and Nello Cristianini. 2000. Support vector machines. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (2000), 93–112.
- [237] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian. 2015. Multi-scale convolutional neural networks for lung nodule classification. In *International Conference on Information Processing in Medical Imaging*. Springer, 588–599.
- [238] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi. 2018. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics* 22, 5 (Sep. 2018), 1589–1604. <https://doi.org/10.1109/JBHI.2017.2767063>
- [239] Hoo-Chang Shin, Le Lu, Lauren Kim, Ari Seff, Jianhua Yao, and Ronald M Summers. 2015. Interleaved text/image deep mining on a very large-scale radiology database. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1090–1099.
- [240] Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, and Ronald M Summers. 2016. Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2497–2506.
- [241] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging* 35, 5 (2016), 1285–1298.
- [242] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [243] Youyi Song, Ling Zhang, Siping Chen, Dong Ni, Baiying Lei, and Tianfu Wang. 2015. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. *IEEE Transactions on Biomedical Engineering* 62, 10 (2015), 2421–2433.
- [244] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
- [245] Jelena Stojanovic, Djordje Gligorijevic, Vladan Radosavljevic, Nemanja Djuric, Mihajlo Grbovic, and Zoran Obradovic. 2017. Modeling healthcare quality via compact representations of electronic health records. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 14, 3 (2017), 545–554.
- [246] Marijn F Stollenga, Wonmin Byeon, Marcus Liwicki, and Juergen Schmidhuber. 2015. Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In *Advances in neural information processing systems*. 2998–3006.
- [247] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. 2014. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* 101 (2014), 569–582.
- [248] Heung-Il Suk and Dinggang Shen. 2013. Deep learning-based feature representation for AD/MCI classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 583–590.

- [249] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang. 2018. Deep patient similarity learning for personalized healthcare. *IEEE transactions on nanobioscience* 17, 3 (2018), 219–227.
- [250] Richard S Sutton, Andrew G Barto, et al. 1998. *Introduction to reinforcement learning*. Vol. 2. MIT press Cambridge.
- [251] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.
- [252] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging* 35, 5 (2016), 1299–1312.
- [253] Deirdre C Tatomer and Jeremy E Wilusz. 2017. An unchartered journey for ribosomes: circumnavigating circular RNAs to produce proteins. *Molecular cell* 66, 1 (2017), 1–2.
- [254] Atsushi Teramoto, Hiroshi Fujita, Osamu Yamamuro, and Tsuneo Tamaki. 2016. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Medical physics* 43, 6Part1 (2016), 2821–2827.
- [255] Nenad Tomašev, Xavier Glorot, Jack W Rae, Michal Zieliński, Harry Askham, Andre Saraiva, Anne Mottram, Clemens Meyer, Suman Ravuri, Ivan Protsyuk, et al. 2019. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572, 7767 (2019), 116.
- [256] Truyen Tran, Tu Dinh Nguyen, Dinh Phung, and Svetha Venkatesh. 2015. Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBMs). *Journal of biomedical informatics* 54 (2015), 96–105.
- [257] Tran Quang Trung and Nae-Eung Lee. 2016. Flexible and stretchable physical sensor integrated platforms for wearable human-activity monitoring and personal healthcare. *Advanced materials* 28, 22 (2016), 4338–4372.
- [258] Supawong Tuarob, Conrad S Tucker, Marcel Salathe, and Nilam Ram. 2014. An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. *Journal of biomedical informatics* 49 (2014), 255–268.
- [259] A Ulate-Campos, F Coughlin, M Gainza-Lein, I Sánchez Fernández, PL Pearl, and T Lodenkemper. 2016. Automated seizure detection systems and their effectiveness for each type of seizure. *Seizure* 40 (2016), 88–101.
- [260] Minke JC van den Berge, Rolien H Free, Rosemarie Arnold, Emile de Kleine, Rutger Hofman, J Marc C van Dijk, and Pim van Dijk. 2017. Cluster analysis to identify possible subgroups in tinnitus patients. *Frontiers in neurology* 8 (2017), 115.
- [261] Mark JJP Van Grinsven, Bram van Ginneken, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez. 2016. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE transactions on medical imaging* 35, 5 (2016), 1273–1284.
- [262] Gijs van Tulder and Marleen de Brujne. 2016. Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines. *IEEE transactions on medical imaging* 35, 5 (2016), 1262–1272.
- [263] V Vapnik. 1995. Support vector machine. *Mach. Learn* 20 (1995), 273fi?297.
- [264] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.
- [265] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. ACM, 1096–1103.
- [266] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research* 11, Dec (2010), 3371–3408.
- [267] Veronika Vincze and Richard Farkas. 2014. De-identification in natural language processing. *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2014), 1300–1303.
- [268] Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. 2018. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2447–2456.
- [269] Tao Wang, Markus Brede, Antonella Ianni, and Emmanouil Mentzakis. 2017. Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 91–100.
- [270] Xiaosong Wang, Le Lu, Hoo-chang Shin, Lauren Kim, Isabella Nogues, Jianhua Yao, and Ronald Summers. 2016. Unsupervised category discovery via looped deep pseudo-task optimization using a large scale radiology image database. *arXiv preprint arXiv:1603.07965* (2016).

- [271] Yi Wang, Haoran Dou, Xiaowei Hu, Lei Zhu, Xin Yang, Ming Xu, Jing Qin, Pheng-Ann Heng, Tianfu Wang, and Dong Ni. 2019. Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound. *IEEE transactions on medical imaging* (2019).
- [272] Wei-Qi Wei, Lisa A. Bastarache, Robert J. Carroll, Joy E. Marlo, Travis J. Osterman, Eric R. Gamazon, Nancy J. Cox, Dan M. Roden, and Joshua C. Denny. 2017. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenotype-wide association studies in the electronic health record. *PLOS ONE* 12, 7 (07 2017), 1–16. <https://doi.org/10.1371/journal.pone.0175508>
- [273] Stephen F Weng, Jenna Reps, Joe Kai, Jonathan M Garibaldi, and Nadeem Qureshi. 2017. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *Plos one* 12, 4 (2017), e0174944.
- [274] Wikipedia contributors. [n.d.]. Alternative splicing – Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/w/index.php?title=Alternative\\_splicing&oldid=900663821](https://en.wikipedia.org/w/index.php?title=Alternative_splicing&oldid=900663821) [Online; accessed 12-August-].
- [275] Wikipedia contributors. [n.d.]. Autoencoder – Wikipedia, The Free Encyclopedia. <https://en.wikipedia.org/w/index.php?title=Autoencoder&oldid=907982276> [Online; accessed 11-August-].
- [276] Wikipedia contributors. [n.d.]. Word2vec – Wikipedia, The Free Encyclopedia. <https://en.wikipedia.org/w/index.php?title=Word2vec&oldid=909500488> [Online; accessed 11-August-].
- [277] Ronald J Williams and David Zipser. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural computation* 1, 2 (1989), 270–280.
- [278] Max L Wilson, Susan Ali, and Michel F Valstar. 2014. Finding information about mental health in microblogging platforms: a case study of depression. In *Proceedings of the 5th Information Interaction in Context Symposium*. ACM, 8–17.
- [279] Stephen Wilson, Warwick Ruscoe, Margaret Chapman, and Rhona Miller. 2002. General practitioner-hospital communications: A review of discharge summaries. *Journal of quality in clinical practice* 21 (01 2002), 104–8. <https://doi.org/10.1046/j.1440-1762.2001.00430.x>
- [280] Yonghui Wu, Min Jiang, Jianbo Lei, and Hua Xu. 2015. Named entity recognition in Chinese clinical text using deep neural network. *Studies in health technology and informatics* 216 (2015), 624.
- [281] Cao Xiao, Tengfei Ma, Adji B Dieng, David M Blei, and Fei Wang. 2018. Readmission prediction via deep contextual embedding of clinical concepts. *PloS one* 13, 4 (2018), e0195024.
- [282] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1492–1500.
- [283] Yuanpu Xie, Zizhao Zhang, Manish Sapkota, and Lin Yang. 2016. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 185–193.
- [284] Hui Y Xiong, Babak Alipanahi, Leo J Lee, Hannes Bretschneider, Daniele Merico, Ryan KC Yuen, Yimin Hua, Serge Gueroussov, Hamed S Najafabadi, Timothy R Hughes, et al. 2015. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 6218 (2015), 1254806.
- [285] Tao Xu, Han Zhang, Xiaolei Huang, Shaoting Zhang, and Dimitris N Metaxas. 2016. Multimodal deep learning for cervical dysplasia diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 115–123.
- [286] Michal Yablowitz and David Schwartz. 2018. A Review and Assessment Framework for Mobile-Based Emergency Intervention Apps. *Comput. Surveys* 51 (01 2018), 1–32. <https://doi.org/10.1145/3145846>
- [287] Zhennan Yan, Yiqiang Zhan, Zhigang Peng, Shu Liao, Yoshihisa Shinagawa, Shaoting Zhang, Dimitris N Metaxas, and Xiang Sean Zhou. 2016. Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition. *IEEE transactions on medical imaging* 35, 5 (2016), 1332–1343.
- [288] Dong Yang, Shaoting Zhang, Zhennan Yan, Chaowei Tan, Kang Li, and Dimitris Metaxas. 2015. Automated anatomical landmark detection on distal femur surface using convolutional neural network. In *2015 IEEE 12th international symposium on biomedical imaging (ISBI)*. IEEE, 17–21.
- [289] Fu-Chen Yang, Anthony JT Lee, and Sz-Chen Kuo. 2016. Mining health social media with sentiment analysis. *Journal of medical systems* 40, 11 (2016), 236.
- [290] Wei Yang, Yingyin Chen, Yunbi Liu, Liming Zhong, Genggeng Qin, Zhentai Lu, Qianjin Feng, and Wufan Chen. 2017. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Medical image analysis* 35 (2017), 421–433.
- [291] Özal Yıldırım, Paweł P. lawiak, Ru-San Tan, and U Rajendra Acharya. 2018. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in biology and medicine* 102 (2018), 411–420.
- [292] Zhijun Yin, Bradley Malin, Jeremy Warner, Pei-Yun Hsueh, and Ching-Hua Chen. 2017. The power of the patient voice: learning indicators of treatment adherence from an online breast cancer forum. In *Eleventh International AAAI Conference on Web and Social Media*.

- [293] Zhijun Yin, Lina M Sulieman, and Bradley A Malin. 2019. A systematic literature review of machine learning in online personal health data. *Journal of the American Medical Informatics Association* 26, 6 (2019), 561–576.
- [294] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.
- [295] Fisher Yu and Vladlen Koltun. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122* (2015).
- [296] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng. 2017. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *Thirty-first AAAI conference on artificial intelligence*.
- [297] Sheng Yu, Abhishek Chakrabortty, Katherine P Liao, Tianrun Cai, Ashwin N Ananthakrishnan, Vivian S Gainer, Susanne E Churchill, Peter Szolovits, Shawn N Murphy, Isaac S Kohane, et al. 2016. Surrogate-assisted feature extraction for high-throughput phenotyping. *Journal of the American Medical Informatics Association* 24, e1 (2016), e143–e149.
- [298] William Yuan, Dadi Jiang, Dhanya K Nambiar, Lydia P Liew, Michael P Hay, Joshua Bloomstein, Peter Lu, Brandon Turner, Quynh-Thu Le, Robert Tibshirani, et al. 2017. Chemical space mimicry for drug discovery. *Journal of chemical information and modeling* 57, 4 (2017), 875–882.
- [299] Haoyang Zeng, Matthew D Edwards, Ge Liu, and David K Gifford. 2016. Convolutional neural network architectures for predicting DNA–protein binding. *Bioinformatics* 32, 12 (2016), i121–i127.
- [300] Ce Zhang. 2015. DeepDive: a data management system for automatic knowledge base construction. *University of Wisconsin-Madison, Madison, Wisconsin* (2015).
- [301] Shaodian Zhang, Erin Bantum, Jason Owen, and Noémie Elhadad. 2014. Does sustained participation in an online health community affect sentiment?. In *AMIA Annual Symposium Proceedings*, Vol. 2014. American Medical Informatics Association, 1970.
- [302] Shaodian Zhang, Edouard Grave, Elizabeth Sklar, and Noémie Elhadad. 2017. Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. *Journal of biomedical informatics* 69 (2017), 1–9.
- [303] Shaodian Zhang, Tian Kang, Lin Qiu, Weinan Zhang, Yong Yu, and Noémie Elhadad. 2017. Cataloguing treatments discussed and used in online autism communities. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 123–131.
- [304] Shaodian Zhang, Erin OfiCarroll Bantum, Jason Owen, Suzanne Bakken, and Noémie Elhadad. 2017. Online cancer communities as informatics intervention for social support: conceptualization, characterization, and impact. *Journal of the American Medical Informatics Association* 24, 2 (2017), 451–459.
- [305] Sai Zhang, Jingtian Zhou, Hailin Hu, Haipeng Gong, Ligong Chen, Chao Cheng, and Jianyang Zeng. 2015. A deep learning framework for modeling structural features of RNA-binding protein targets. *Nucleic acids research* 44, 4 (2015), e32–e32.
- [306] Zhichang Zhang, Tong Zhou, Yu Zhang, and Yali Pang. 2019. Attention-based deep residual learning network for entity relation extraction in Chinese EMRs. *BMC medical informatics and decision making* 19, 2 (2019), 55.
- [307] Juan Zhao, QiPing Feng, Patrick Wu, Jeremy L. Warner, Joshua C. Denny, and Wei-Qi Wei. [n.d.]. Using topic modeling via non-negative matrix factorization to identify relationships between genetic variants and disease phenotypes: A case study of Lipoprotein(a) (LPA). *PLOS ONE* 14, 2 (02 [n. d.]), 1–15. <https://doi.org/10.1371/journal.pone.0212112>
- [308] Yefeng Zheng, David Liu, Bogdan Georgescu, Hien Nguyen, and Dorin Comaniciu. 2015. 3D deep learning for efficient and robust landmark detection in volumetric data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 565–572.
- [309] Chongyu Zhou, Jia Yao, and Meihul Motani. 2018. Optimizing Autoencoders for Learning Deep Representations From Health Data. *IEEE Journal of Biomedical and Health Informatics* PP (07 2018), 1–1. <https://doi.org/10.1109/JBHI.2018.2856820>
- [310] Wentao Zhu, Yufang Huang, Liang Zeng, Xuming Chen, Yong Liu, Zhen Qian, Nan Du, Wei Fan, and Xiaohui Xie. 2019. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics* 46, 2 (2019), 576–589.
- [311] Wentao Zhu, Chaochun Liu, Wei Fan, and Xiaohui Xie. 2018. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 673–681.
- [312] Wentao Zhu, Xiang Xiang, Trac D Tran, Gregory D Hager, and Xiaohui Xie. 2018. Adversarial deep structured nets for mass segmentation from mammograms. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 847–850.