

Review

The Application of Deep Learning in Cancer Prognosis Prediction

Wan Zhu ^{1,2,*}, Longxiang Xie ^{1,†} , Jianye Han ³ and Xiangqian Guo ^{1,*} 

¹ Department of Preventive Medicine, Institute of Biomedical Informatics, Cell Signal Transduction Laboratory, Bioinformatics center, School of Basic Medical Sciences, Henan University, Kaifeng 475004, China; xielongxiang123@126.com

² Department of Anesthesia, Stanford University, 300 Pasteur Drive, Stanford, CA 94305, USA

³ Department of Computer Science, University of Illinois, Urbana Champions, IL 61820, USA; mr.neohan@gmail.com

* Correspondence: wanzhu@stanford.edu (W.Z.); xqguo@henu.edu.cn (X.G.)

† These authors contributed equally to this work.

Received: 4 February 2020; Accepted: 2 March 2020; Published: 5 March 2020



Abstract: Deep learning has been applied to many areas in health care, including imaging diagnosis, digital pathology, prediction of hospital admission, drug design, classification of cancer and stromal cells, doctor assistance, etc. Cancer prognosis is to estimate the fate of cancer, probabilities of cancer recurrence and progression, and to provide survival estimation to the patients. The accuracy of cancer prognosis prediction will greatly benefit clinical management of cancer patients. The improvement of biomedical translational research and the application of advanced statistical analysis and machine learning methods are the driving forces to improve cancer prognosis prediction. Recent years, there is a significant increase of computational power and rapid advancement in the technology of artificial intelligence, particularly in deep learning. In addition, the cost reduction in large scale next-generation sequencing, and the availability of such data through open source databases (e.g., TCGA and GEO databases) offer us opportunities to possibly build more powerful and accurate models to predict cancer prognosis more accurately. In this review, we reviewed the most recent published works that used deep learning to build models for cancer prognosis prediction. Deep learning has been suggested to be a more generic model, requires less data engineering, and achieves more accurate prediction when working with large amounts of data. The application of deep learning in cancer prognosis has been shown to be equivalent or better than current approaches, such as Cox-PH. With the burst of multi-omics data, including genomics data, transcriptomics data and clinical information in cancer studies, we believe that deep learning would potentially improve cancer prognosis.

Keywords: cancer prognosis; deep learning; machine learning; multi-omics; prognosis prediction

1. Current Development in Cancer Prognosis Prediction

In the United States, approximately 1 in 10 adults have been diagnosed with cancer [1]. Cancer causes 1 in 6 deaths around the world [1]. While new therapies can improve cancer treatment and increase survival rate, cancer prognosis is to estimate cancer development, to provide survival estimation and to improve clinical management. One major task in cancer prognosis is to provide better survival estimation based on patients' clinical features and molecular profile.

Current state-of-the-art analytic methods in cancer prognosis for survival analysis are statistical approaches, including Cox proportional hazard regression [2,3], Kaplan Meier estimator [4] and log-ranks test [5–7]. The main data sources for these approaches in cancer prognosis for survival prediction are mainly clinical data, including cancer diagnosis, cancer types, tumor grades, molecular

profile, etc. In recent years, more data types are available to better understand disease status. These data are high throughput and high dimensional multi-omics data of patient samples [8]. Multi-omics data include genomic data (i.e., whole genome data, single nucleotide polymorphism (SNP) data, copy number alternation (CNA) data, etc.), expression data (i.e., mRNA and miRNA data), proteomic data and epigenetic data (i.e., methylation and other chromosomal modifications). The volume of multi-omic data has poses challenges to use purely statistical methods to perform prediction. Other methods, including machine learning approaches, have been applied or established to solve these problems. So far, some machine learning methods, including principal component analysis (PCA), clustering and autoencoder, have been tested to classify cancer types [9,10]. Also, machine learning methods, including support vector machine (SVM), Bayesian network, semi-supervised learning and decision tree, have been applied to cancer prognosis prediction and shown some good success [11–16].

The establishment of public accessible large-scale cancer databases indeed provides an open-source platform for researchers and clinicians to share and analyze patients' multi-omics data. The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and Genotype-Tissue Expression (GTEx) databases are the major ones. The Cancer Genome Atlas (TCGA) database has clinical and molecular data of over 11,000 tumor patients across 33 different tumor types [17,18], including genomic (whole genome and/or exome sequencing, WGS/WES), transcriptomic (RNAseq, small RNAseq), epigenomic (HumanMethylation450 BeadChip) and proteomic profiling (reverse-phase protein arrays, RPPAs) data. There are several public portals of TCGA, such as TCGA data portal [19], cBioPortal [20], the University of California, Santa Cruz cancer genome browser (UCSC Xena) [21] and FIREHOSE [22]. The Gene Expression Omnibus (GEO) database is a public data repository storing microarray and next-generation sequencing (NGS) data, as well as other high-throughput functional genomic datasets, such as genome methylation, chromatin structure, genomic mutation/copy number variation, protein profiling, and genome–protein interactions [23,24]. The Genotype-Tissue Expression (GTEx) database contains whole-genome sequencing and RNA-sequencing profiles from ~960 postmortem adult donors of many tissue samples that have tissue images stored in an image library for public access [25,26]. These public data not only provide unprecedented opportunities to better illustrate the molecular mechanism of cancers and normal tissues, but also become the major resources to apply novel methods, build models and perform predictions in cancer prognosis.

2. Overview of Deep Learning

Deep learning, also known as deep neural network (DNN), is a branch of machine learning that has made some major breakthrough in recent years due to the increase of computation power, the improvement in model architecture [27] and the exponential growth of data captured by cellular and other devices. There are three basic machine learning paradigms, supervised learning, unsupervised learning and reinforcement learning. Supervised learning algorithm are those should be fed in a set of training data containing features (inputs) and labels (outputs). Some popular supervised learning algorithms include linear and logistic regression [28], SVM [29], naive bayes [30], gradient boosting [31,32], classification trees, and random forest [33,34]. These methods are commonly used in classification and regression studies. Unsupervised learning, on the other hand, does not require pre-existing output/labels and aim to find patterns based on the input data distributions. Clustering (e.g., hierarchical clustering [35,36], K-means [37,38]) is the most common unsupervised learning method. Latent Dirichlet Allocation (LDA) [39], PCA [40] and word2vec [41], are among the most recent popular unsupervised learning approaches. Neural network (NN) can either be supervised, unsupervised or semi-supervised learning, suggesting its flexibility. Reinforcement learning [42] can be summarized as a reward system for the computer program to maximize the rewards in order to search for the best solution [27].

Deep learning (or DNN) consists of multiple layers of artificial neurons that mimic neurons in human brain. Similar to linear regression, each neuron has a weight value that is updated by gradient descent algorithm during backpropagation to minimize global loss function [43]. By applying

nonlinearity using activation function, such as sigmoid, tanh, or relu, to the multiple layers of each neuron, more abstract mathematical relationship was extracted from the input data to map to the output [44]. A well trained model can therefore be used to predict new unlabeled data. Deep learning is a branch of machine learning, and therefore inherits some common knowledge foundation in machine learning, including basic probability and statistics, loss/cost function and etc., but in the meantime has more flexibility and can be built towards more complex layers and multiple neurons in each layer to have better predictive power [45–50]. The most commonly used NN in medical research includes fully connected NN (or simplified as NN) for structured data, convolutional NN (CNN) for image data, and recurrent NN (RNN) for text and sequence data.

In recent years, deep learning has been applied to biomedical research to annotate pathogenicity of genetic variants [51,52], show state-of-the-art performance in the task of genomic variant calling [53] and improve protein folding prediction [54,55]. Compared to other methods, deep learning is more flexible and generic to be applied on discrete or continuous data [56], requires less feature engineering with expertise knowledge compared to machine learning in general [27] and works better than many state-of-the-art methods [53].

3. Current Application of Deep Learning in Cancer Prognosis

To review the application of deep learning in the field of cancer prognosis, we used key words, including “deep learning”, “neural networks” and “cancer prognosis”, and searched literature on PubMed. To better understand the development of the field and for better comparison, we have included studies that built simple NN models which consist of 3–4 layers and studies that built DNNs which consist of more than 4 layers. We reviewed and summarized these studies and models. Based on the types of NN and whether feature extraction has been used, the publications that we reviewed could be grouped into three classes: (1) NN models with no feature extraction, (2) Feature extraction from multi-omics data to build fully connected NNs, and (3) CNN based models. Here, we reviewed and summarized these studies and models.

3.1. NN Models with no Feature Extraction

As mentioned, Cox proportional hazards model (Cox-PH) is a multivariate semi parametric regression model that has been used widely in cancer studies to compare survival characteristics between two or more treatment groups [2,57]. Some early attempts in cancer prognosis have either used clinical tumor and patient data [58], cellular features from tissue slides [14] or some genes expression data [13] to build the models. To show the performance, these studies compared the performance of NN to Cox-PH and/or Kaplan Meier methods, and showed that simple NN models have achieved similar performance compared to these methods (Table 1). Also, in these studies, because the number of features was relatively small without omics data, feature selection was not necessary.

Table 1. Summary of neural network models with no feature extraction.

Publication ^a	Type of Cancer	Type of Data	Sample Size	Methods	Architecture	Outputs	Hyperparameters	Validation	NN Model Performance
Joshi et al., 2006 [58]	Melanoma	Clinical data of tumors	1946 (1160 females and 786 males)	3 layers NN	Normalized input	Survival time	Sigmoid activation	Not reported	Achieved similar performance as Cox and Kaplan Meier statistical methods
Chi et al., 2007 [14]	Breast cancer	Cell images to measure 30 nuclear features	Dataset 1: 198 cases; Dataset 2: 462 cases	3 layers NN	30 input nodes, 20 hidden nodes	Survival time	Epoch = 1000, Sigmoid activation	10 fold cross validation	As good as conventional methods
Petalidis et al., 2008 [13]	Astrocytic brain tumor	A list of genes expression from microarray data	65	A single layer perceptron and an output (multiple binary models)	Number of inputs equals to the number of classifier genes in different models	Tumor grades	Lr ¹ = 0.05, Epoch = 100	Leave-one-out cross validation	44, 9 and 7 probe sets have achieved 93.3%, 84.6%, and 95.6% validation success rates, respectively.
Ching et al., 2018 [59]	10 types of cancer	TCGA gene expression data, clinical data and survival data	5031	NN	Input normalization and log-transformed, 0–2 hidden layers (143 nodes)	Survival time	L1, L2, or MCP ² regularization, tanh activation for hidden layer(s), dropout, Cox regression as output layer	5-fold cross validation	Similar or in some cases better performance than Cox-PH, Cox-boosting or RF
Katzman et al. 2018 [60]	Breast cancer	METABRIC ³ , 4 genes data and clinical information, GBSG ⁴ : clinical data	METABRIC: 1980, GBSG: 1546 training, 686 testing	NN	METABRIC: 1 dense layer, 41 nodes GBSG: patients clinical information, 1 dense layer, 8 nodes	Survival	SELU ⁵ activation, Adam optimizer, dropout, LR decay, momentum	20% of METABRIC patients used as test set GBSG has split test dataset	C-index: 0.654 for METABRIC and 0.676 for GBSG, both are better than CoxPH
Jing et al. 2019 [61]	Breast cancer, nasopharyngeal carcinoma	METABRIC, GBSG, NPC ⁷ : 8–9 clinical features	METABRIC: 1980 NPC: 4630	DNN	METABRIC: 4 hidden layers, 45 nodes of each; GBSG: 3 hidden layers, 84, 84 and 70 nodes, respectively NPC, 3 layers, 120 nodes of each layer in model 1, 108, 108 and 90 nodes respectively in model 2.	Survival	ELU ⁶ , dropout, L1 and L2, momentum, LR decay, batch size. Loss function equals to mean square error and a pairwise ranking loss	After removed patients with missing data, 20% used as test set	C-index: 0.661 for METABRIC and 0.688 for GBSG, both are better than CoxPH and DeepSurv. c-index ranges 0.681–0.704 depends on input data for NPC, better than CoxPH.

Abbreviation:¹ Lr; learning rate; ² MCP: minimax concave penalty. ³ METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; ⁴ GBSG: the German Breast Cancer Study Group; ⁵ SELU: scaled exponential linear unit; ⁶ SELU: exponential linear unit; ⁷ NPC: nasopharyngeal carcinoma. ^aLinks to source codes if available from publications: Petalidis et al. [13]: <http://www.imbb.forth.gr/people/poirazi/software.html>. Ching et al. [59]: <https://github.com/lanagarmire/cox-nnet>. Katzman et al. [60]: <https://github.com/jaredleekatzman/DeepSurv>. Jing et al. 2019 [61]: <http://github.com/sysucc-ailab/RankDeepSurv>.

Since the wide acceptance of Cox regression model in survival prediction, Cox regression was used as the output layer to build NNs to predict cancer survival. Cox-nnet [59] is a NN network which used genomic data from TCGA as input and Cox regression as the output layer. To avoid overfitting, they tested ridge regularization, dropout, reduction of NN complexity by using 0 to 2 hidden layers and a combination of ridge and dropout in training the NN (Table 1). They reported that dropout and reduction of NN complexity by using 1 hidden layer worked the best to avoid overfitting in their experimental setting. To measure the performance, they showed that Cox-net performed better than Cox-PH, Cox-boost (based on gradient boosting) or random forest in the TCGA datasets that they have tested (Table 1).

Katzman et al. has built a neural network model, named DeepSurv, to perform survival analysis. DeepSurv is a feed forward NN that uses patient's clinical data as input and applied dropout, learning rate decay, regularization, and other commonly used hyperparameters to optimize for different datasets [60]. Their results showed that this model performed better than CoxPH models (Table 1). Another neural network model, named RankDeepSurvival, has adapted the basic architecture of DeepSurv and increased the depth of the network to build 3–4 hidden layers' DNN to perform survival analysis in multiple datasets, including cancer datasets [61]. More importantly, they have updated the loss function by using the sum of mean squared error loss and a pairwise ranking loss based on ranking information on survival data [61]. They reported that RankDeepSurvival model outperformed CoxPH models and DeepSurv model in breast cancer datasets from Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and the German Breast Cancer Study Group (GBSG) (Table 1). Both of these studies have further validated their models performed better than CoxPH models in other disease datasets, such as heart disease and diabetes, which suggested that deep learning models can be generalized for different tasks.

3.2. Feature Extraction from Gene Expression Data to Build Fully Connected NNs

Health data has the characteristics of high-dimension, small sample size and complex non-linear effects between biological components [62,63]. Dimension reduction assists the integrative analysis of multi-omics data [64]. These following studies have tested different algorithms to reduce dimension of sequencing data, extract a smaller number of features and train a fully connected NN.

In a study to predict breast cancer prognosis, Sun et al. used a method named minimum redundancy maximum relevance (mRMR) [65] to reduce the dimensionality of gene expression data and copy number alternation (CNA) data by extracting 400 and 200 genes, respectively, from these datasets [66]. Next, 3 NN models were built using features selected from gene expression data, CNA data or clinical data, respectively. Finally, prediction outputs of these three NN models were added up based on a weighted linear aggregation to calculate a final prediction score. They named this model as Multimodal Deep Neural Network by integrating Multi-dimensional Data (MDNNMD). When they selected threshold of 0.443–0.591, a high specificity (0.95–0.99), yet low sensitivity (0.2–0.45) were reported (Table 2). To show model performance, they reported that ROC (0.845), accuracy, and precision, and Matthew's correlation coefficient (MCC) of MDNNMD outperformed other methods, including SVM, random forest, and logistic regression (Table 2). One of the reasons that the model has a big performance difference between specificity and sensitivity is likely due to the imbalanced data in training the NN (491 short term survival versus 1489 long term survival cases).

Table 2. Summary of neural network models that used feature extraction.

Publication ^a	Type of Cancer	Type of Data	Sample Size	Methods Used in Feature Extraction	Architecture	Outputs	Hyperparameters	Validation	NN Model Performance
Sun et al., 2018 [66]	Breast cancer	Gene expression profile, CAN ¹ profile and clinical data	1980 (1489 LTS ² , 491 non-LTS)	mRMR (extracted 400 features from gene expression and 200 features from CNA)	4 hidden layers (1000, 500, 500, and 100 nodes, respectively)	Survival time	Lr ³ = 1e-3, Tanh activation, epoch 10–100, batch size = 64	10-fold cross validation	ROC ⁴ : 0.845 (better than SVM, RF ⁵ , and LR ⁶), Sp ⁷ : 0.794–0.826, Pre ⁸ : 0.749–0.875, Sn ⁹ : 0.2–0.25, Mcc ¹⁰ : 0.356–0.486
Huang et al., 2019 [67]	Breast cancer	mRNA, miRNA, CNB ¹¹ , TMB ¹² , clinical data	583 (80% for training, 20% for testing in each fold of cross validation)	ImQCM ¹³ , Epigenome matrix to extract 57 dimensions from mRNA data and 12 dimensions from miRNA data	Hybrid network, mRNA and miRNA dimension reduction inputs have 1 hidden layer (8 and 4 nodes, respectively), CNB, TMB and clinical data have no hidden layer	Survival time	Adam optimizer, LASSO ¹⁴ regularization, Epoch = 100, sigmoid activation, Cox regression as output, batch size = 64	5-fold cross validation	Multi-omics data network reached a median c-index ¹⁵ of 0.7285
Hao et al., 2018 [62]	Glioblastoma multiforme	Gene expression (TCGA), pathway (MsigDB ¹⁶)	475 (376 non-LTS, 99 LTS)	Pathway based analysis (12,024 genes from mRNA data to 574 pathways and 4359 genes)	4 layers NN: gene layer—pathway layer—hidden layer—output	Survival time	Lr = 1e-4, L2 = 3e-4, dropout, softmax output	5-fold validation	AUC ¹⁷ = 0.66 ± 0.013, F1 score = 0.3978 ± 0.016
Chaudhary et al., 2018 [68]	Liver cancer	mRNA, miRNA, methylation data, and clinical data (TCGA)	360 samples training, (5 additional cohorts, 230, 221, 166, 40 and 27 samples for validation)	Autoencoder unsupervised NN to extract 100 features from mRNA, miRNA and methylation data	3 hidden layers NN (500, 100, 500 nodes, respectively) and a bottleneck layer	Feature reduction	Epoch = 10, Dropout = 0.5, SGD ¹⁸	Not reported	NN outputs were used for K means clustering.
Shimizu and Nakayama, 2019, [69]	Breast cancer	METABRIC ¹⁹	1903 (METABRIC, 952 samples for training)	Select 23 genes by statistical methods	3 layers NN	Survival time	Lr = 0.001, Epoch = 1000, Cross entropy for loss function Relu activation for hidden nodes, softmax function for output layer	951 samples from METABRIC	NN node weights were used to calculate a mPS ²⁰

Abbreviation: ¹ CNA: copy number alteration; ² LTS: long term survivals; ³ Lr: learning rate; ⁴ ROC: receiver operating characteristic; ⁵ RF: random forest; ⁶ LR: logistic regression; ⁷ Sp: specificity; ⁸ Pre: precision; ⁹ Sn: sensitivity; ¹⁰ Mcc: Mathew's correlation coefficient. The equation is $(TP \cdot TN - FP \cdot FN) / \sqrt{[(TP + FN) \cdot (TP + FP) \cdot (TN + FN) \cdot (TN + FP)]}$; ¹¹ CNB: copy number burden; ¹² TNB: tumor mutation burden; ¹³ ImQCM: local maximum Quasi Clique Merger [67]; ¹⁴ LASSO: also known as L1 regularization; ¹⁵ c-index (concordance index): Steck et al. [70] suggested that c-index is equivalent to AUC. Specifically, c-index closes to 0.5 suggested random prediction. The closer c-index gets to 1, the better the model is. ¹⁶ MsigDB: Molecular Signatures Database; ¹⁷ AUC: area under the curve of ROC; ¹⁸ SGD: stochastic gradient descent; ¹⁹ METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; ²⁰ mPS: molecular prognostic score. ^a Links to source codes if available from publications: Sun et al., 2018 [66]: <https://github.com/USTC-Hllab/MDNNMD>. Huang et al., 2019 [67]: <https://github.com/huangzhii/SALMON/>. Hao et al., 2018 [62]: <https://github.com/DataX-JieHao/PASNet>. Shimizu and Nakayama, 2019, [69]: https://hideyukishimizu.github.io/mPS_breast.

There are many ways to reduce data dimensionality. Huang et al. have obtained five omics data, including gene expression (mRNA) data, miRNA data, copy number burden data, tumor mutation burden data and clinical data, performed feature extraction from these data and built a deep learning model to predict breast cancer patient survival [67]. They also applied a Cox proportional hazards model to develop a survival analysis learning with a multi-omics NN (or SALMON) model [67]. In this model, input layers were comprised of features extracted from mRNA and miRNA data using a local maximal Quasi-Clique Merger (lmQCM) algorithm inspired by spectral clustering [70]. A matrix, named eigengene, was generated from lmQCM algorithm and used to represent 57 and 12 dimensions from mRNA and miRNA data, respectively (Table 2). In the hidden layer, mRNA and miRNA data comprises 8 and 4 neurons, respectively. Adam optimizer and lasso regularization were used as hyperparameters in training (Table 2). Sigmoid function was used as activation function after each forward propagation to introduce non-linearity and Cox proportional hazards regression and was used as the output to predict survival time. This model achieved a median concordance index (c-index) [71] of 0.728 which has been suggested to outperform other models that didn't include high dimensional features extracted from mRNA and miRNA data (Table 2), suggesting feature extraction improves model performance.

In addition to reducing data dimension using algorithm, feature extraction by application of domain knowledge as selection criteria has also been tested. Hao et al. used gene expression data from 475 glioblastoma multiforme patients with ~12 k genes that contained survival information to build a prognosis model [62] (Table 2). They grouped the samples into two groups, long term survival (LTS, survival time \geq 24 months) and non-long term survival (non-LTS, survival time $<$ 24 months). Next, they used pathway data from the Molecular Signatures Database (MSigDB) and mapped 4359 genes to 574 pathways. They constructed a NN using the 4,359 genes as input and 574 pathways as the first hidden layer and applied dropout and L2 regularization to avoid overfitting. Since 20% of the samples are LTS, the training data suffered from imbalanced data. It is a common problem in handling patient data. They suggested PASNet achieved AUC of 0.66 that is better than the performance of logistic LASSO, random LASSO or SVM model. The advantage of PASNet is that it took biological pathways into consideration when building a NN model.

NN itself can be used to extract features from multi-omics data. Hepatocellular carcinoma (HCC) is the most common type of liver cancer. High heterogeneity of the disease makes the prognosis prediction challenging. Chaudhary et al. built a NN model using multi-omics data of 360 HCC samples from TCGA database [68]. The multi-omics data includes mRNA expression, miRNA expression, CpG methylation and clinical data. They used unsupervised autoencoder NN to transform features and perform dimension reduction [68] and extract 100 feature nodes from miRNA, mRNA and methylation data (Table 2). Next, they used a Cox-PH model to identify 37 significant features, applied K-means clustering to identify survival risk and used ANOVA to get feature ranking. Finally, prognosis prediction was built using a SVM model. In another study, Shimizu et al. picked 23 genes from 184 prognosis related genes based on the statistical significance of these individual genes on the overall survival of breast cancer patients [69]. They used gene expression levels of these 23 genes to build a NN to get gene weights from NN's nodes and generate a molecular prognostic score (mPS) (Table 2). The mPS was then applied to evaluate prognosis. Although both studies didn't report the performance of the NNs in their study, these studies suggested that NN can also be a useful tool for dimension reduction of multi-omic data for prognosis prediction.

3.3. CNN-Based Models

In recent years, deep learning approach has been made the most significant progress because state-of-the-art networks have been built using convolutional NN (CNN) [45–48] and recurrent NN (RNN) [49,50]. Many success have been shown in the areas of image recognition/classification and computer vision by CNN, and natural language processing (NLP) and sequencing data investigation by RNN. Specifically, great performance has also been witnessed in many medical areas, including

classification of skin cancer types [72,73], identification of pathological histological slides [74], identification of A β plaque region in Alzheimer's patients, classification of cancer cells from normal cells using nuclear morphometric measure [75], and extraction information from electronic health records (EHR) to predict hospital readmission [76,77], mortality [78], and clinical outcome [79]. In cancer prognosis studies, CNN has been applied to the classification of cancerous tissue for survival prediction or extraction of feature for downstream prognosis. Some of these studies also added RNN layers to extract sequential information from the data.

Glioblastoma multiforme (GBM) is a type of brain tumor. Methylation of O6-methylguanine methyltransferase (MGMT) gene promoter has been found to associate with longer survival and better response to a drug, temozolomide. Therefore, methylation of MGMT gene has been considered as a biomarker. However, verification of MGMT gene promoter in the brain is difficult and invasive. Using high quality MRI images from patients that have labeled information of methylation status of MGMT promoter, a 50-layer pre-trained CNN model, ResNet50 [80] was used for transfer learning and achieved the highest accuracy of ~95% compared to ResNet18 and ResNet34 [81] (Table 3). Similarly, another research group used brain MRI images from a different cohort of GBM patients to build a bidirectional convolutional recurrent NN (CRNN) model to predict methylation status of MGMT gene promoter and suggested patient's sensitivity to temozolomide based on the prediction of methylation status [82]. RNN layers were added into this model to capture sequential information of MRI images [82], but the effect was not well studied since the model performance was not compared with or without RNN layer. In this study, the authors applied many techniques to reduce overfitting, such as L2 regularization, dropouts, and data augmentation (Table 3). Although the training accuracy is high (0.97), but validation and test accuracies were only 0.67 and 0.62, respectively, suggesting the model was still overfit to the training data. Instead of predicting methylation status of MGMT gene promoter in glioblastoma cancer, Mobadersany et al. trained a survival convolutional NN (SCNN) using histology images, clinical data with or without genomic markers in glioma and glioblastoma and showed the prediction power of this NN has surpassed the prognostic accuracy of the WHO genomic classification and histologic grading in 2018 [83]. Using H&E-stained tissue sections of 1,061 samples from 769 patients, regions of interest (ROIs) that contain viable tumor cells by a web-based platform were identified in tissue images to train a CNN with Cox proportional hazard regression as the output layer to predict patient outcomes (Table 3). They also compared the performance of the NN with or without inclusion of some genomic data (i.e., IDH gene mutation and 1p/19q codeletion). They showed that with the addition of genomic data, the performance has improved the median of c-index from 0.754 to 0.801 (Table 3).

Table 3. Summary of CNN based models.

Publication ^a	Type of Cancer	Type of Data	Sample Size	Architecture	Outputs	Hyperparameters	Validation	NN Model Performance
Korfiatis et al., 2017 [81]	Glioblastoma multiforme	MRI images	155 (66 methylated and 89 unmethylated tumors) Training: 7856 images (934 methylated, 1621 unmethylated, 5301 no tumor) Testing: 2612 images (250 methylated, 335 unmethylated, 2027 no tumor)	Base model: ResNet18 ResNet34 ResNet50	3 classes, methylated, unmethylated, or no tumor	Lr ¹ = 0.01, mini Batch = 32, momentum = 0.5, weight decay = 0.1, Relu activation, Epoch = 50, SGD ² as optimizer, batch normalization	Stratified cross-validation	ResNet50 based model validation dataset performance: Accuracy = 94.9%, Precision = 96%, Recall = 95% ResNet34 Accuracy = 80.72%, Precision = 93%, Recall = 81%, ResNet18 Accuracy = 76.75%, Precision:80%, Recall = 77%
Han et al., 2018 [81]	Glioblastoma multiforme	MRI images	458,951 image frames from 5235 MRI scans of 262 patients (TCIA ³)	3 convolutional layers, 2 fully connected layers, 1 bi-directional GRU ⁴ layer (RNN), 1 fully connected layer, softmax output	2 classes (positive and negative methylation status)	Data augmentation, (rotation and flipping, 90-fold increase of the dataset), Lr = 5e−6 – 5e−1, dropout (0–0.5), Adam optimizer, Epoch = 10, L2 regularization, batch norm, relu activation	Validation set reached a precision of 0.67, an AUC of 0.56.	Training data set obtained 0.97 accuracy. 0.67 and 0.62 accuracies on the validation and test set, respectively
Mobadersany et al., 2018 [83]	Low grade glioma and glioblastoma	H&E images, genomics data, clinical data	769 gliomas from TCGA, containing genomics data (IDH mutation and 1 p/19 q codeletion), clinical data and 1061 slides.	VGG19 is the base model and cox regression used as output, Built 2 models with or without genomics data	Survival	Data augmentation, Lr = 0.001, epoch = 100, exponential learning decay	Monte Carlo cross-validation	SCNN median c-index is 0.754, GSCNN (adding IDH mutation and 1 p/19 q codeletion as features) improved the median c-index to 0.801
Kather et al., 2019 [74]	Colorectal cancer	H&E tissue slides	Training set (tissue): 86 H&E slides to create 100,000 image patches Testing set (tissue): 25 H&E slides of 7180 image patches Training set (OS ⁵): 862 H&E slides from 500 TCGA patients Validation set (OS): 409 H&E slides from 409 DACHS patients	Base models: VGG19, AlexNet, GoogLeNet, SqueezeNet and Resnet50, add an output softmax layer	9 tissue type classification	Lr = 3e−4, Iteration = 8, Batch size = 360, softmax function	An independent cohort of 409 samples	VGG19 gets the best results, 94–99% accuracy in tissue class prediction

Table 3. Cont.

Publication ^a	Type of Cancer	Type of Data	Sample Size	Architecture	Outputs	Hyperparameters	Validation	NN Model Performance
Bychkov et al., 2018 [84]	Colorectal cancer	H&E images of tumor tissue microarray	420 patients (equal number of survived or died within five years after diagnosis)	VGG16 to generate a 16 × 16 feature from input data, followed with 3 layers LSTM ⁶ (264, 128 and 64 LSTM cells, respectively)	Survival	Default hyperparameters in VGG16, LSTM used hyperbolic tangent as activation, binary cross entropy loss function, Adadelta optimizer	60 samples for validation, 140 samples for testing	CNN + LSTM model reached an AUC ⁷ of 0.69, better than CNN + SVM, CNN + LR ⁸ , or CNN + NB ⁹
Courtiol et al., 2019 [85]	Mesothelioma	H&E slides	2981 patient slides (MESOPATH/MESOBANK, 2300 training, 681 testing) Validation: 56 patients (TCGA)	Divided each slide to up to 10,000 tiles as input data 3 classes of each tile: epithelioid, sarcomatoid or biphasic. ResNet50 for feature extraction	Survival	Multi-layer perceptron with sigmoid activation, Autoencoder	56 patient slides	MesoNet outperformed histology-based classification but no better than a linear regression based model (Meanpool)
Wang et al., 2019 [86]	High grade serous ovarian cancer	CT scanning venous phase images	Feature learning cohort: 8917 CT images from 102 patients	Five convolutional layers (24, 16, 16, 16, 16 filters, respectively)	16 dimensional feature vector	Batch normalization, average pooling between adjacent convolutional layers	Not reported	CNN outputs were used to build Cox-PH model

Abbreviation: ¹ Lr: learning rate; ² SGD; stochastic gradient descent; ³ TCIA: The Cancer Image Archive; ⁴ GRU: gated recurrent unit, which is similar to LSTM and is used in building RNN models; ⁵ OS: overall survival; ⁶ LSTM: long short term memory cell; ⁷ AUC: area under the curve of ROC; ⁸ LR: logistic regression; ⁹ NB: naive bayes; ¹⁰ c-index: also known as Harell's concordance index. ^a Links to source codes if available from publications: Han et al., 2018 [81]: <http://onto-apps.stanford.edu/m3crnn/>. Kather et al., 2019 [74]: <http://dx.doi.org/10.5281/zenodo.1214456>, <http://dx.doi.org/10.5281/zenodo.1420524>, <http://dx.doi.org/10.5281/zenodo.1471616>, Wang et al., 2019 [86]: <http://www.radiomics.net.cn/post/111>.

Colorectal cancer (CRC) is a type of solid tumors. H&E images are the major tool to diagnose CRC and determine the stage of CRC. In H&E slide of CRC patients, it is important to differentiate normal tissues from cancer regions. Kather et al. [74] hand labeled 100,000 image patches using 86 CRC H&E slides into 9 tissue classes, including adipose, background, debris, lymphocytes, mucus, smooth muscle, normal mucosa, stroma and cancer epithelium [74]. They used these images as the training data with an additional of 7,180 images from 25 patients as the testing data to build a CNN model using state-of-the-art CNN networks, such as VGG19 and Resnet50, to perform transfer learning and have reached 94–99% accuracy in classifying tissue types (Table 3). By calculating the hazard ratios (HRs) for shorter overall survival (OS) and selecting optimal cutoffs based on the ROC curve, the authors defined a deep stromal score and suggested although not significant correlated, deep stromal score shows a trend of correlation to shorter OS. In another CRC study, Bychkov et al. [84] used CNN models as a tool for feature extraction and built an RNN (LSTM) model to predict CRC patient survival. They used VGG16 as the base model to perform transfer learning and extracted a 256-tile feature vector from each input H&E image of tumor tissue microarray (Table 3). They then input these feature vectors of 220 patients (equal number of patients in short and long term survival group) to train a LSTM-cell RNN model. They also trained SVM, naive bayes and logistic regression models to compare the performance. They showed that LSTM model reached an AUC of 0.69, while SVM, naive bayes, and logistic regression reached AUCs of 0.64, 0.61, 0.65, respectively. They also reported that human experts can only reach an AUC of 0.57–0.58, suggesting that the performance of this model is better than human.

Malignant mesothelioma is a type of rare and highly lethal cancer of the pleural lining. According to the WHO classification, patients tissue biopsy can be classified into epithelioid, sarcomatoid and biphasic types. Prognosis of mesothelioma is closely associated with tissue types as epithelioid type has the longest overall survival, sarcomatoid type has the shortest overall survival and biphasic type is in-between [85]. Based on the clinical knowledge, Courtiol et al. built a MesoNet model using 100 to 10,000 tiles of histological tissue from 2,300 H&E slide from the MESOPATH/MESOBANK database. By transfer learning of ResNet50 and performing feature extraction, a matrix of features (2,048) was extracted from each tile to train MesoNet. C-index showed that MesoNet performed better than histological based classification methods, but not as good as a linear regression based model, named Meanpool (Table 3) [85].

Similarly, CNN models can be used to extract features from images for building other machine learning model to predict cancer prognosis. High-grade serous ovarian cancer (HGSOC) is the most common and most lethal histological type of ovarian cancer. Wang et al. [86] used CT-based images and trained a CNN model to extract image features for building a Cox-PH survival prediction model. In this study, 102 HGSOC patients, who underwent debulking surgery and have remained in 2-year follow-up study, were used as a feature extraction cohort (Table 3). A total of 8,917 tumor images were used to train an unsupervised CNN model for feature extraction of a 16-dimensional feature vector. Next the feature vector was fed to a multivariate Cox-PH regression model to identify the association of feature vector and recurrence of HGSOC. This study provides an example of using NN, particularly CNN, to extract image features for downstream studies.

4. Challenges in the Application of Deep Learning in Cancer Prognosis

By reviewing literature, we have noticed that many state-of-the-art deep learning techniques have been applied to cancer prognosis prediction, indicating the great potential and the urgent need of utilizing multi-omics data from cancer patients to test new algorithm and improve model performance (Figure 1). Meanwhile, we found that there are seven main challenges in applying deep learning approach in cancer prognosis prediction to achieve high performance. We also suggested some potential solutions for these challenges.

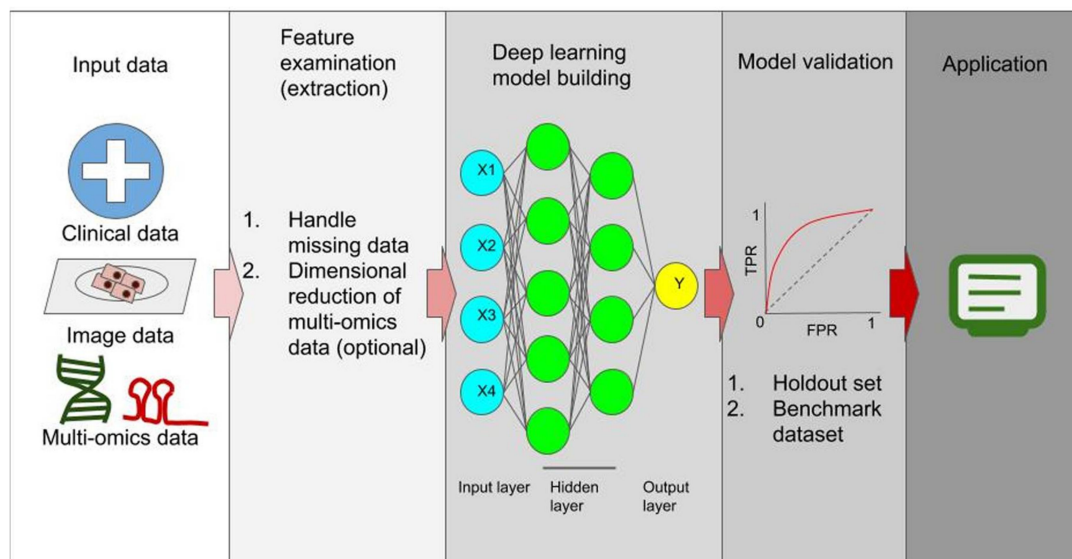


Figure 1. Workflow of building deep learning models for cancer prognosis prediction. The sources of input data include clinical data which could be text data and/or structured data (numeric and/or categorical data), clinical images which could be tissue slides in H&E staining or immune-histological staining. MRI, CT, etc, and genomic data which could be expression data (i.e., mRNA expression data, miRNA expression data), genomic sequence data (i.e., whole genome sequence, SNP data, CNA data, etc), epigenetic data (i.e., methylation data), etc. In the next step, researchers will examine the data to handle missing data and imbalanced data. Reduction of high dimensional genomic data is an optional step here. Features are then used to build a deep learning (neural network) model. The type of models to use depends on the input data. For example, fully connected NN is commonly used for structured datasets. Image data is used to build CNN models. Sequence data is often used to build RNN models. If multiple types of data exist, hybrid models can be built to accept different data types. After the model is built, the model will be tested in the holdout (or validation) datasets. It will also be important to test and compare the models using benchmark datasets. Finally, the model can be used in applications. Abbreviations: FPR: false positive rate; TPR: true positive rate.

First, the amount of patient data is still relatively small. Majority models were built on hundreds of patient samples (Tables 1–3). It is common to see sub-optimal performance and overfitting problems in these studies. The performance of deep learning models is leveraged by the amounts of data [27]. To combat overfitting, researchers applied regularization methods (ridge and lasso or L1 and L2), dropout, data augmentation, reduction of NN complexity to improve model performance, but the effect is still limited by the amount of data. To improve model performance with small datasets, transfer learning with pretrain models on large amounts of datasets have shown success solving some of the problems [87–89]. In addition, newer methods and algorithms have also been proposed and tested to combat small sample size problem, such as few-shot or one-shot learning in CNN [90,91]. Another direction is to perform data simulation. It will be interesting to test these methods in the field of cancer prognosis.

Second, imbalanced patient data is commonly found. For some high mortality cancers, it is very common to find less survivors in the study groups. Imbalanced data in training will reduce the model performance. While under-sampling in the majority group is suboptimal, generation of synthetic data could be one of the solutions. In image classification problem, data augmentation is also one way to increase sample size to adjust the groups that have fewer sample sizes. Also, reporting model performance should use additional algorithms, such as precision, recall, F1 score and confusion matrix, rather than just reporting accuracy to better reflect the model performance.

Third, handling sparse or missing data from noisy patient clinical profiles is also a challenge. Missing data in building a model reduce the power of the model in prediction. Common ways to

handle missing data include exclusion of missing data observation, but this is very costly when patient samples are already very limited. A better way to overcome this problem is to do data imputation based on known data. Rendleman et al. proposed to perform imputation using Multivariate Imputation by Chained Equations (MICE) [92] to overcome the problem of missing or sparse data in cancer patient outcome [93]. MICE is a multiple imputation technique [94] that works under the assumption that the missing data are missing at random. In this study, they showed that prediction using naive bayes or random forest both works slightly better after imputation, suggesting imputation could be a useful way to improve prediction.

Fourth, health care data, particularly sequencing data, is high dimensional, feature extraction could be the solution to improve model performance. As we showed in Table 2, studies have performed feature extraction by using algorithm or applying domain knowledge to improve model performance. NN can also be used for feature extraction and dimension reductions [68,86]. It will be interesting to test and apply new way of data embedding for high dimensional data.

Fifth, more generic deep learning models are needed and model validation in benchmark datasets is critical to validate model performance. The accuracy in model performance is difficult to compare among different studies and different models [95]. Deep learning models with improved algorithms should be built and tested for more generic tasks. For example, a deep recurrent survival analysis which used LSTM cells as the building blocks has been proposed for survival analysis [96]. It will be interesting to test this model in cancer prognosis. Also, building benchmark datasets for model comparison will allow researchers to compare and analyze model performance easier and more efficient. For example, in recent years, ImageNet, a database that contains millions of images from daily life, has been frequently used to evaluate CNN models [97–99], which is a critical contributing factor for the development in the field. The models that were built using daily objects from ImageNet have been widely used for other tasks and reach great success. Also, these models are commonly used in many fields and tasks to perform transfer learning. In the medical field, it has also been shown that a single deep learning model is effective at diagnosis across medical modalities [100]. Therefore, building benchmark databases for model validation is urgently needed. One solution is to start building cancer patients' databases for prognosis analysis [101–110].

Sixth, in addition to technical challenges, building the infrastructure for data storage and establishing the pipeline to build machine learning model may be greatly useful to facilitate the development [8]. Because health care data are sensitive, data safety becomes a concern. How to build a system to safely store and use patients' health care data to build models and also protect the patients' privacy requires the effort of administration, research community and personal awareness. Secure cloud services and relevant infrastructure can be established to support the storage of large amount of health care data. Federated learning that only train and predict user data on their own devices is one innovative way to solve privacy issues [111].

Lastly, there is the urgent need of researchers who have expertise in biomedical research and machine learning. Compared crowdsourced data annotations, such as annotations for ImageNet objects [112], medical data requires annotators who have expertise to label the data. Domain knowledge facilitate the construction of machine learning models. Therefore, research engineers who have domain knowledge are greatly needed to improve research in this area. To solve this need, universities can provide more relevant courses and degrees.

5. Conclusions and Summary

Deep learning has made significant improvement in research and started to make changes in our daily lives. In the medical field, many studies have applied deep learning and shown many great successes [78,113–121]. One advantage of using deep learning to train a model is its capability to continue training when more data is available [27]. In addition, since health care data have different formats, e.g., genomic data, expression data, clinical (structured) data, text and image (unstructured) data, using different NN architectures to solve different types of data problems become more and more

popular and useful [27]. In this review, we summarized recent studies that applied deep learning in studying cancer prognosis (Tables 1–3). Among these studies, many have shown deep learning models performed equally or better than other machine learning models [14,58,59]. Future work should continue focusing on testing and improving the algorithm and building state-of-the-art models to improve cancer prognosis prediction.

6. Key Points

- Deep learning (DNN) models accept lots of data in different formats. It is a great tool to be used in cancer prognosis prediction since patient's health data contain multi-source data.
- Using feature extraction could be one way to efficiently extract data from multi-omics data to train neural networks and possibly improve cancer prognosis prediction.
- Fully connected NN and CNN models have been tested in a number of studies to predict cancer prognosis and showed good performance.
- Current deep learning models in cancer prognosis studies still require further testing and validation in larger datasets.

Funding: This study is supported by the following funding: Kaifeng Science and Technology Major Project (18ZD008), National Natural Science Foundation of China (No.81602362), Program for Science and Technology Development in Henan Province (No.162102310391), Program for Young Key Teacher of Henan Province (2016GGJS-214), Supporting grants of Henan University (No.2015YBZR048, No.B2015151), Yellow River Scholar Program (No.H2016012).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 7–34. [[CrossRef](#)] [[PubMed](#)]
2. Ahmed, F.E.; Vos, P.W.; Holbert, D. Modeling survival in colon cancer: A methodological review. *Mol. Cancer* **2007**, *6*, 15. [[CrossRef](#)] [[PubMed](#)]
3. Michael, K.Y.; Ma, J.; Fisher, J.; Kreisberg, J.F.; Raphael, B.J.; Ideker, T. Visible Machine Learning for Biomedicine. *Cell* **2018**, *173*, 1562–1565.
4. Kaplan, E.L.; Meier, P. Nonparametric Estimation From Incomplete Observations. *Publ. Am. Stat. Assoc.* **1958**, *53*, 457–481. [[CrossRef](#)]
5. NW, M. Evaluation of Survival Data and Two New Rank Order Statistics Arising In Its Consideration. *Cancer Chemother. Rep.* **1966**, *50*, 163–170.
6. Peto, R.; Peto, J. Asymptotically efficient rank invariant test procedures. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 185–198. [[CrossRef](#)]
7. Harrington, D. Linear Rank Tests in Survival Analysis. In *Encyclopedia of Biostatistics*, 2nd ed.; Armitage, P., Colton, T., Eds.; Wiley: New York, NY, USA, 2005.
8. Goossens, N.; Nakagawa, S.; Sun, X.; Hoshida, Y. Cancer biomarker discovery and validation. *Transl. Cancer Res.* **2015**, *4*, 256–269.
9. Tan, M.; Peng, C.; Anderson, K.A.; Chhoy, P.; Xie, Z.; Dai, L.; Park, J.; Chen, Y.; Huang, H.; Zhang, Y.; et al. Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **2014**, *19*, 605–617. [[CrossRef](#)]
10. Alexe, G.; Dalgin, G.; Ganesan, S.; Delisi, C.; Bhanot, G. Analysis of breast cancer progression using principal component analysis and clustering. *J. Biosci.* **2007**, *32*, 1027–1039. [[CrossRef](#)]
11. Hemsley, P.A. An outlook on protein S-acylation in plants: What are the next steps? *J. Exp. Bot.* **2017**. [[CrossRef](#)]
12. Kretowska, M. Computational Intelligence in Survival Analysis. In *Encyclopedia of Business Analytics and Optimization*; IGI Global: Warsaw, Poland, 2014; pp. 491–501.

13. Petalidis, L.P.; Oulas, A.; Backlund, M.; Wayland, M.T.; Liu, L.; Plant, K.; Happerfield, L.; Freeman, T.C.; Poirazi, P.; Collins, V.P. Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol. Cancer Ther.* **2008**, *7*, 1013–1024. [[CrossRef](#)] [[PubMed](#)]
14. Chi, C.L.; Street, W.N.; Wolberg, W.H. Application of Artificial Neural Network-Based Survival Analysis on Two Breast Cancer Datasets. *Amia. Annu. Symp. Proc.* **2007**, *11*, 130–134.
15. Van IJzendoorn, D.G.; Szuhai, K.; Briaire-de Bruijn, I.H.; Kostine, M.; Kuijjer, M.L.; Bovée, J.V. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Comput. Biol.* **2019**, *15*. [[CrossRef](#)] [[PubMed](#)]
16. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [[CrossRef](#)] [[PubMed](#)]
17. Chang, K.; Creighton, C.J.; Davis, C.; Donehower, L.; Drummond, J.; Wheeler, D.; Ally, A.; Balasundaram, M.; Birol, I.; Butterfield, Y.S.N. The Cancer Genome Atlas Pan-Cancer analysis project. *Chin. J. Lung Cancer* **2013**, *45*, 1113–1120.
18. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [[CrossRef](#)]
19. Chandran, U.R.; Medvedeva, O.P.; Michael, B.M.; Blood, P.D.; Anish, C.; Soumya, L.; Antonio, F.; Wong, K.F.; Lee, A.V.; Zhihui, Z. TCGA Expedition: A Data Acquisition and Management System for TCGA Data. *PLoS ONE* **2016**, *11*. [[CrossRef](#)]
20. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **2013**, *6*. [[CrossRef](#)]
21. Haeussler, M.; Zweig, A.S.; Tyner, C.; Speir, M.L.; Rosenbloom, K.R.; Raney, B.J.; Lee, C.M.; Lee, B.T.; Hinrichs, A.S.; Gonzalez, J.N. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **2018**, *47*, D853–D858. [[CrossRef](#)]
22. Deng, M.; Brägelmann, J.; Kryukov, I.; Saraiva-Agostinho, N.; Perner, S. FirebrowseR: An R client to the Broad Institute’s Firehose Pipeline. *Database J. Biol. Databases Curation* **2017**. [[CrossRef](#)]
23. Clough, E.; Barrett, T. The Gene Expression Omnibus Database. *Methods Mol. Biol.* **2016**, *1418*, 93–110. [[PubMed](#)]
24. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)] [[PubMed](#)]
25. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *13*, 307–308. [[CrossRef](#)] [[PubMed](#)]
26. Consortium, T.G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **2015**, *348*, 648–660. [[CrossRef](#)] [[PubMed](#)]
27. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)] [[PubMed](#)]
28. Cramer, J.S. The Origins of Logistic Regression. *Soc. Sci. Electron. Publ.* **2003**. [[CrossRef](#)]
29. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. *Proc. Fifth Annu. Workshop Comput. Learn. Theory* **2008**, *5*, 144–152.
30. Maron, M.E. Automatic Indexing: An Experimental Inquiry. *J. ACM* **1961**, *8*, 404–417. [[CrossRef](#)]
31. Breiman, L.; Friedman, J.H.; Olshen, R.A. *Classification and Regression Trees*; Routledge: New York, NY, USA, 2017.
32. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
33. Breiman, L. *Arcing the Edge*; Technical Report; Statistics Department, University of California: Berkeley, CA, USA, 1997.
34. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
35. Sibson, R. SLINK: An optimally efficient algorithm for the single-link cluster method. *Comput. J.* **1973**, *16*, 30–34. [[CrossRef](#)]
36. Defays, D. An efficient algorithm for a complete link method. *Comput. J.* **1977**, *20*, 364–366. [[CrossRef](#)]
37. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]

38. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Los Angeles, CA, USA, 21 June–18 July 1967; pp. 281–297.
39. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
40. Pearson, K. Principal components analysis. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *6*, 559. [[CrossRef](#)]
41. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3111–3119.
42. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
43. Hinton, G.E. Learning distributed representations of concepts. In ; , 1991. In Proceedings of the Eighth Annual Conference of the Cognitive Science Society, Hillsdale, NJ, USA, 7–10 August 1991; p. 12.
44. Bengio, Y. Learning deep architectures for AI. *Found. Trends[®] Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
45. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
46. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, 15 June 2000; pp. 770–778.
48. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
49. Jordan, M. *Serial Order: A Parallel Distributed Processing Approach*; Technical Report; California University: San Diego, CA, USA, 1986.
50. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
51. Quang, D.; Chen, Y.; Xie, X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **2015**, *31*, 761–763. [[CrossRef](#)]
52. Farahbakhsh-Farsi, P.; Djalali, M.; Koohdani, F.; Saboor-Yaraghi, A.A.; Eshraghian, M.R.; Javanbakht, M.H.; Chamari, M.; Djazayeri, A. Effect of omega-3 supplementation versus placebo on acylation stimulating protein receptor gene expression in type 2 diabetics. *J. Diabetes. Metab. Disord.* **2014**, *13*, 1. [[CrossRef](#)] [[PubMed](#)]
53. Poplin, R.; Varadarajan, A.V.; Blumer, K.; Liu, Y.; McConnell, M.V.; Corrado, G.S.; Peng, L.; Webster, D.R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158. [[CrossRef](#)] [[PubMed](#)]
54. Wang, Y.; Yao, H.; Zhao, S. Auto-encoder based dimensionality reduction. *Neurocomputing* **2016**, *184*, 232–242. [[CrossRef](#)]
55. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865. [[CrossRef](#)]
56. Biganzoli, E.; Boracchi, P.; Mariani, L.; Marubini, E. Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach. *Stat. Med.* **1998**, *17*, 1169–1186. [[CrossRef](#)]
57. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–202. [[CrossRef](#)]
58. Joshi, R.; Reeves, C. Beyond the Cox model: Artificial neural networks for survival analysis part II. In Proceedings of the Eighteenth International Conference on Systems Engineering, Coventry, UK, 3–10 May 2003; pp. 179–184.
59. Ching, T.; Zhu, X.; Garmire, L.X. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput. Biol.* **2018**, *14*. [[CrossRef](#)]
60. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [[CrossRef](#)]
61. Jing, B.; Zhang, T.; Wang, Z.; Jin, Y.; Liu, K.; Qiu, W.; Ke, L.; Sun, Y.; He, C.; Hou, D. A deep survival analysis method based on ranking. *Artif. Intell. Med.* **2019**, *98*, 1–9. [[CrossRef](#)]
62. Hao, J.; Kim, Y.; Kim, T.-K.; Kang, M. PASNet: Pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC Bioinform.* **2018**, *19*, 510. [[CrossRef](#)]

63. Ma, T.; Zhang, A. Multi-view Factorization AutoEncoder with Network Constraints for Multi-omic Integrative Analysis. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018.
64. Meng, C.; Zeleznik, O.A.; Thallinger, G.G.; Kuster, B.; Gholami, A.M.; Culhane, A.C. Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.* **2016**, *17*, 628–641. [[CrossRef](#)]
65. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)]
66. Sun, D.; Wang, M.; Li, A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 841–850. [[CrossRef](#)] [[PubMed](#)]
67. Huang, Z.; Zhan, X.; Xiang, S.; Johnson, T.S.; Helm, B.; Yu, C.Y.; Zhang, J.; Salama, P.; Rizkalla, M.; Han, Z. SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer. *Front. Genet.* **2019**, *10*, 166. [[CrossRef](#)] [[PubMed](#)]
68. Chaudhary, K.; Poirion, O.B.; Lu, L.; Garmire, L.X. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res.* **2017**, *24*. [[CrossRef](#)] [[PubMed](#)]
69. Shimizu, H.; Nakayama, K.I. A 23 gene-based molecular prognostic score precisely predicts overall survival of breast cancer patients. *EBioMedicine* **2019**, *46*, 150–159. [[CrossRef](#)]
70. Zhang, J.; Huang, K. Normalized imqcm: An algorithm for detecting weak quasi-cliques in weighted graph with applications in gene co-expression module discovery in cancers. *Cancer Inform.* **2014**, *13*, CIN-S14021. [[CrossRef](#)]
71. Steck, H.; Krishnapuram, B.; Dehing-oberije, C.; Lambin, P.; Raykar, V.C. On ranking in survival analysis: Bounds on the concordance index. In Proceedings of the Advances in Neural Information Processing Systems, Malvern, PA, USA, 8–10 December 2008; pp. 1209–1216.
72. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
73. Levine, A.B.; Schlosser, C.; Grewal, J.; Coope, R.; Jones, S.J.M.; Yip, S. Rise of the Machines: Advances in Deep Learning for Cancer Diagnosis. *Trends Cancer* **2019**, *5*, 157–169. [[CrossRef](#)]
74. Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **2019**, *16*. [[CrossRef](#)]
75. Radhakrishnan, A.; Damodaran, K.; Soylemezoglu, A.C.; Uhler, C.; Shivashankar, G.V. Machine Learning for Nuclear Mechano-Morphometric Biomarkers in Cancer Diagnosis. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)]
76. Rajkomar, A.; Oren, E.; Chen, K.; Dai, A.M.; Hajaj, N.; Hardt, M.; Liu, P.J.; Liu, X.; Marcus, J.; Sun, M. Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* **2018**, *1*, 18. [[CrossRef](#)]
77. Shameer, K.; Johnson, K.W.; Yahya, A.; Miotto, R.; Li, L.; Ricks, D.; Jebakaran, J.; Kovatch, P.; Sengupta, P.P.; GELIJNS, S. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using Mount Sinai heart failure cohort. *Pac. Symp. Biocomput.* **2017**, *22*, 276–287.
78. Elfiky, A.A.; Pany, M.J.; Parikh, R.B.; Obermeyer, Z. Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw. Open* **2018**, *1*. [[CrossRef](#)]
79. Mathotaarachchi, S.; Pascoal, T.A.; Shin, M.; Benedet, A.L.; Rosa-Neto, P. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiol. Aging* **2017**, *59*, 80. [[CrossRef](#)]
80. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
81. Korfiatis, P.; Kline, T.L.; Lachance, D.H.; Parney, I.F.; Buckner, J.C.; Erickson, B.J. Residual Deep Convolutional Neural Network Predicts MGMT Methylation Status. *J. Digit. Imaging* **2017**, *30*, 622–628. [[CrossRef](#)]
82. Han, L.; Kamdar, M. MRI to MGMT: Predicting Drug Efficacy for Glioblastoma Patients. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2018**, *23*, 331–338.
83. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Velázquez Vega, J.E.; Brat, D.J.; Cooper, L.A.D. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**. [[CrossRef](#)]

84. Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P.E.; Verrill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* **2018**, *8*. [[CrossRef](#)]
85. Courtiol, P.; Maussion, C.; Moarii, M.; Pronier, E.; Pilcer, S.; Sefta, M.; Manceron, P.; Toldo, S.; Zaslavskiy, M.; Le Stang, N. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* **2019**, *25*, 1519–1525. [[CrossRef](#)]
86. Wang, S.; Liu, Z.; Rong, Y.; Zhou, B.; Bai, Y.; Wei, W.; Wang, M.; Guo, Y.; Tian, J. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiother. Oncol.* **2019**, *132*, 171–177. [[CrossRef](#)]
87. Christopher, M.; Belghith, A.; Bowd, C.; Proudfoot, J.A.; Goldbaum, M.H.; Weinreb, R.N.; Girkin, C.A.; Liebmann, J.M.; Zangwill, L.M. Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. *Sci. Rep.* **2018**, *8*. [[CrossRef](#)]
88. Ding, Y.; Sohn, J.H.; Kawczynski, M.G.; Trivedi, H.; Harnish, R.; Jenkins, N.W.; Lituiev, D.; Copeland, T.P.; Aboian, M.S.; Mari Aparici, C. A Deep learning model to predict a diagnosis of alzheimer disease by using 18F-FDG PET of the brain. *Radiology* **2018**, *290*, 456–464. [[CrossRef](#)]
89. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding transfer learning for medical imaging. *Adv. Neural Inf. Process. Syst.* **2019**, 3342–3352.
90. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst.* **2016**, 3630–3638.
91. Triantafyllou, E.; Zemel, R.; Urtasun, R. Few-shot learning through an information retrieval lens. *Adv. Neural Inf. Process. Syst.* **2017**, 2255–2265.
92. Buuren, S.V.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **2010**, 1–68. [[CrossRef](#)]
93. Rendleman, M.C.; Buatti, J.M.; Braun, T.A.; Smith, B.J.; Nwakama, C.; Beichel, R.R.; Brown, B.; Casavant, T.L. Machine learning with the TCGA-HNSC dataset: Improving usability by addressing inconsistency, sparsity, and high-dimensionality. *BMC Bioinform.* **2019**, *20*, 339. [[CrossRef](#)]
94. Raghunathan, T.E.; Lepkowski, J.M.; Van Hoewyk, J.; Solenberger, P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **2001**, *27*, 85–96.
95. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [[CrossRef](#)]
96. Ren, K.; Qin, J.; Zheng, L.; Yang, Z.; Zhang, W.; Qiu, L.; Yu, Y. Deep recurrent survival analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 4798–4805.
97. Deng, J.; Russakovsky, O.; Krause, J.; Bernstein, M.S.; Berg, A.; Fei-Fei, L. Scalable multi-label annotation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26–27 April 2014; pp. 3099–3102.
98. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
99. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
100. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
101. Goswami, C.P.; Nakshatri, H. PROGene: Gene expression based survival analysis web application for multiple cancers. *J. Clin. Bioinform.* **2013**, *3*, 22. [[CrossRef](#)]
102. Anaya, J. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Comput. Sci.* **2016**, *2*. [[CrossRef](#)]
103. Elfilali, A.; Lair, S.; Verbeke, C.; La, R.P.; Radvanyi, F.; Barillot, E. ITTACA: A new database for integrated tumor transcriptome array and clinical data analysis. *Nucleic Acids Res.* **2006**, *34*, D613–D616. [[CrossRef](#)]
104. Wang, Q.; Xie, L.; Dang, Y.; Sun, X.; Xie, T.; Guo, J.; Han, Y.; Yan, Z.; Zhu, W.; Wang, Y. OSIm: A Web Server to Evaluate the Prognostic Value of Genes in Leiomyosarcoma. *Front. Oncol.* **2019**, *9*. [[CrossRef](#)]

105. Wang, Q.; Wang, F.; Lv, J.; Xin, J.; Xie, L.; Zhu, W.; Tang, Y.; Li, Y.; Zhao, X.; Wang, Y. Interactive online consensus survival tool for esophageal squamous cell carcinoma prognosis analysis. *Oncol. Lett.* **2019**, *18*, 1199–1206. [[CrossRef](#)]
106. Zhang, G.; Wang, Q.; Yang, M.; Yuan, Q.; Dang, Y.; Sun, X.; An, Y.; Dong, H.; Xie, L.; Zhu, W. OSblca: A Web Server for Investigating Prognostic Biomarkers of Bladder Cancer Patients. *Front. Oncol.* **2019**, *9*, 466. [[CrossRef](#)]
107. Yan, Z.; Wang, Q.; Sun, X.; Ban, B.; Lu, Z.; Dang, Y.; Xie, L.; Zhang, L.; Li, Y.; Guo, X. OSbrca: A Web Server for Breast Cancer Prognostic Biomarker Investigation with Massive Data from tens of Cohorts. *Front. Oncol.* **2019**, *9*, 1349. [[CrossRef](#)]
108. Xie, L.; Wang, Q.; Dang, Y.; Ge, L.; Sun, X.; Li, N.; Han, Y.; Yan, Z.; Zhang, L.; Li, Y. OSkirc: A web tool for identifying prognostic biomarkers in kidney renal clear cell carcinoma. *Future Oncol.* **2019**, *15*, 3103–3110. [[CrossRef](#)]
109. Xie, L.; Wang, Q.; Nan, F.; Ge, L.; Dang, Y.; Sun, X.; Li, N.; Dong, H.; Han, Y.; Zhang, G. OSacc: Gene Expression-Based Survival Analysis Web Tool For Adrenocortical Carcinoma. *Cancer Manag. Res.* **2019**, *11*, 9145–9152. [[CrossRef](#)] [[PubMed](#)]
110. Wang, F.; Wang, Q.; Li, N.; Ge, L.; Yang, M.; An, Y.; Zhang, G.; Dong, H.; Ji, S.; Zhu, W. OSuvm: An interactive online consensus survival tool for uveal melanoma prognosis analysis. *Mol. Carcinog.* **2020**, *59*, 56–61. [[CrossRef](#)] [[PubMed](#)]
111. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S. Communication-efficient learning of deep networks from decentralized data. *arXiv* **2016**, arXiv:1602.05629.
112. Su, H.; Deng, J.; Fei-Fei, L. Crowdsourcing annotations for visual object detection. In Proceedings of the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
113. Avati, A.; Jung, K.; Harman, S.; Downing, L.; Ng, A.; Shah, N.H. Improving palliative care with deep learning. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 122. [[CrossRef](#)] [[PubMed](#)]
114. Elfiky, A.A.; Elshemey, W.M. Molecular dynamics simulation revealed binding of nucleotide inhibitors to ZIKV polymerase over 444 nanoseconds. *J. Med. Virol.* **2018**, *90*, 13–18. [[CrossRef](#)]
115. Horng, S.; Sontag, D.A.; Halpern, Y.; Jernite, Y.; Shapiro, N.I.; Nathanson, L.A. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE* **2017**, *12*. [[CrossRef](#)]
116. Henry, K.E.; Hager, D.N.; Pronovost, P.J.; Saria, S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci. Transl. Med.* **2015**, *7*, ra122–ra299. [[CrossRef](#)]
117. Culliton, P.; Levinson, M.; Ehresman, A.; Wherry, J.; Steingrub, J.S.; Gallant, S.I. Predicting severe sepsis using text from the electronic health record. *arXiv* **2017**, arXiv:1711.11536.
118. Oh, J.; Makar, M.; Fusco, C.; McCaffrey, R.; Rao, K.; Ryan, E.E.; Washer, L.; West, L.R.; Young, V.B.; Guttag, J. A generalizable, data-driven approach to predict daily risk of *Clostridium difficile* infection at two large academic health centers. *Infect. Control Hosp. Epidemiol.* **2018**, *39*, 425–433. [[CrossRef](#)]
119. Miotto, R.; Li, L.; Kidd, B.A.; Dudley, J.T. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* **2016**, *6*. [[CrossRef](#)]
120. Yang, Z.; Huang, Y.; Jiang, Y.; Sun, Y.; Zhang, Y.-J.; Luo, P. Clinical assistant diagnosis for electronic medical record based on convolutional neural network. *Sci. Rep.* **2018**, *8*. [[CrossRef](#)]
121. De Langavant, L.C.; Bayen, E.; Yaffe, K. Unsupervised machine learning to identify high likelihood of dementia in population-based surveys: Development and validation study. *J. Med. Internet Res.* **2018**, *20*. [[CrossRef](#)]

