Artificial Intelligence In Medicine xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

A survey of neural network-based cancer prediction models from microarray data

Maisa Daoud[a,*], Michael Mayo[b]

[a] University of Waikato, P.O. Box 1212, Hamilton, New Zealand
[b] University of Waikato, Hamilton, New Zealand

ARTICLE INFO

Keywords:
Cancer prediction models
Neural networks
Classification
Clustering
Filtering

ABSTRACT

Neural networks are powerful tools used widely for building cancer prediction models from microarray data. We review the most recently proposed models to highlight the roles of neural networks in predicting cancer from gene expression data. We identified articles published between 2013–2018 in scientific databases using keywords such as cancer classification, cancer analysis, cancer prediction, cancer clustering and microarray data. Analyzing the studies reveals that neural network methods have been either used for filtering (data engineering) the gene expressions in a prior step to prediction; predicting the existence of cancer, cancer type or the survivability risk; or for clustering unlabeled samples. This paper also discusses some practical issues that can be considered when building a neural network-based cancer prediction model. Results indicate that the functionality of the neural network determines its general architecture. However, the decision on the number of hidden layers, neurons, hypermeters and learning algorithm is made using trail-and-error techniques.

1. Introduction

Microarray technology is one of the most widely used tools for analyzing genetic diseases. Standardized microarray dataset consists of thousands of gene expressions and a few hundred of samples. Each expression measures the level of activity of genes within a given tissue so comparing the genes expressed in abnormal cancerous tissues with those in normal tissues gives a good insight into the disease pathology and allows for better diagnosis and predictions for future samples.

The high dimensionality of the gene expression profiles is a crucial issue when building a cancer predictive model. This problem affects the accuracy of the model and increases the computational time [11,27]. Two general approaches have been suggested to reduce the dimensionality of the gene expressions and to overcome its consequent problems. These are: (i) feature selection methods, which select the most relevant discriminating features and eliminate the non-relevant dependent features; and (ii) feature creation methods, which generate new low dimensional features (codes) that best represent the original high dimensional features (as indicated in [77]).

Neural networks are powerful machine learning methods that are used widely to learn data representations (features) at multiple levels of abstractions. These representations are useful for many applications such as reconstruction, classification, clustering and recognition. Predictive models such as cancer prediction models use the generated features for classifying, clustering or applying statistical analysis on the samples.

Based on our analysis of the most recent studies on cancer prediction models, we categorize the current neural network methods according to their functionality into: (i) filtering (preprocessing) methods, (ii) predicting (classification) methods, and (iii) clustering methods. Neural network filtering methods are used for extracting representations that best describe the gene expressions without any direct consideration to the prediction goal, such as the networks used in [67,28]. Alternatively, predicting and clustering methods extract the representations that, respectively, maximize the prediction accuracy [12,20,50] or best divide the genes or samples according to their mutual similarities into groups respectively [81,83,9].

In this study, we review the most recent neural network-based cancer prediction models by presenting the data preprocessing tools and the adopted architectures. We also provide a brief discussion to highlight some important issues that can be considered when building new cancer prediction models. This work is distinguished from others by presenting neural networks models that were specifically designed for predicting cancer using gene expression data. Previous works such as [48] focused on applications of deep learning in different bioinformatic related fields. Deep learning applications in regulatory genomics and biological image analysis was introduced in [3] which also presents some practical points to start with deep architectures. Ravi et al. [60]

* Corresponding author.
E-mail addresses: mtdd1@waikato.ac.nz (M. Daoud), michael.mayo@waikato.ac.nz (M. Mayo).

https://doi.org/10.1016/j.artmed.2019.01.006
Received 1 October 2017; Received in revised form 22 October 2018; Accepted 27 January 2019

highlighted computational biology problems in a way that is accessible to machine learning researchers.

The rest of the paper is organized as follows. In Section 2 we present a basic background to neural networks and cancer prediction models. Section 3 presents our methodology including a review of the most commonly used neural network models for preprocessing, filtering, prediction and clustering gene expressions. Section 4 summarizes the reviewed studies. The discussion is presented in Section 6 and the conclusion in Section 7.

2. Background

This section presents a basic introduction to neural networks and cancer prediction models.

2.1. Neural Networks

Neural networks are powerful tools capable of solving non-linear complex problems and discovering universal input–output mappings [46]. To get a better understanding of the concept, consider a fully connected feedforward Multi-Layer Perceptron (MLP) network with L layers ordered as: input layer, a sequence of hidden layers, and an output layer. The layers are indexed as $l = \{0, \ldots, L-1\}$ and each layer has a number of neurons equal to n_l. We will denote each input training example x as $I(x) = [I_1, I_2, ...,I_{n_0}]$ and its output $O(x) = [O_1, O_2, ...,O_{n_{L-1}}]$.

The network is trained by feeding the inputs forward to calculate the activation value for every neuron. At the output layer, neuron activation values are calculated and aggregated to get $O(x)$.

The difference between $O(x)$ and a desired output, i.e. the error, is calculated using a predefined objective function. Using a backpropagation algorithm, the objective function is optimized by propagating the error derivatives back through the network to fine-tune the weights for optimal error value. The discussed feed forward layered networks and backpropagation mechanism are one of the most commonly used architectures and training algorithms. We focus on them as they are widely used with gene expression data. Prieto et al. [58] presented a comprehensive overview of modelling, simulation and implementation of neural networks with some examples to models used for solving real-world problems.

In the next sections, we review previous works proposing neural network-based cancer prediction models. These models adopt the MLP, convolutional neural network or generative adversarial network architectures for learning the gene expression features. All methods use a similar training algorithm to the one described above with some differences in the number of neurons and networks architectures. Convolutional neural networks [47], for example, perform feature extraction by scanning a set of weight matrices across the input; these weight matrices learn to recognize the relevant patterns. Another example is generative adversarial networks [59] which consist of a generative network and a discriminator network. The generative network learns to generate output samples, given random noise as input, while the discriminator network learns to discriminate the true data samples from the generated fake data samples.

2.2. Cancer prediction models

Cancer is a serious worldwide health problem usually associated with genetic abnormalities. These abnormalities can be detected using microarray techniques which measure the expression and the activity of thousands of genes. Generating a microarray involves hybridizing two DNA strands collected from two samples, e.g. diseased and healthy tissues. Each of these samples is originally a reverse transcript of mRNA and labeled with a dye. The two samples are mixed into a single microarray and scanned with an appropriate source of light to provide an image with an array of features. The intensity of each spot or the average difference between matches and mismatches can be related to the amount (expression) of the mRNA presents in the tissue, i.e the amount of protein produced by the gene corresponding to the given feature. The ultimate goal of this genomic analysis technology is to get better insight into the disease and to improve cancer diagnosis [5].

Table 1
classification metrics, TP is true positive, TN is true negative, FP is false positive and FN is false negative.

Metric name	Definition
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
Sensitivity, Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$
Negative Prediction Rate	$\frac{TN}{TN + FN}$
Matthews Correlation Coefficient (MCC)	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
F1	$\frac{precision \times recall}{precision + recall}$

Cancer prediction models consist of one or more methods working collaboratively to achieve high prediction accuracy. Statistical and machine learning methods have been widely used for building cancer prediction models that help physicians to provide more accurate prognosis, individualized treatments, and reduce the cost per patient.

The accuracy of cancer prediction models is affected by the characteristics of the input data. Gene expressions are high dimensional and include irrelevant noisy features which degrades classification accuracy. They also exhibit spatial structure and, hence, incorporating information about this may increase the model's discriminating ability [68].

The predictive performance of cancer prediction models can be measured by different metrics such as the accuracy, recall, specificity, precision, negative prediction rate, Matthew correlation coefficient and F1 (shown in Table 1). Receiver Operating Characteristic (ROC) curve [29] is another measure represented as a plot of the true positive (sensitivity) rate against the false positive rate (specificity) [38,18]. It reflects the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. A good prediction model is the one achieving good balance between sensitivity and specificity [18]. Statistically, this corresponds to ROC ≥ 0.7 [18]. However, the ROC value should be paired with the confidence interval value to test its validity.

Survival analysis is a different field concerned with predicting the time until a medical condition occurs. From a machine learning perspective, survival analysis is a ranking problem in which data points are ranked on their survival times rather than predicting the actual survival times [64]. The Concordance Index (CI) is one of the standard performance measuring tools for assessing the quality of ranking in survival analysis studies. CI can be interpreted as the probability of concordance between the predicted and the observed survival where a value close to 0.7 indicates a good model and a value close to 0.5 means random concordance. Brier score is another metric measuring the mean of the difference between observed and estimated survival over a certain time. Brier score ranges between 0 and 1 and a larger score indicates higher inaccuracy [12].

3. Neural network-based cancer prediction models

An extensive search relevant to neural network-based cancer prediction was conducted using Google scholar and two other electronic databases namely PubMed and Scopus. Search was performed using keywords such as “Neural Networks” AND “Cancer classification”, “Neural Networks” AND “Gene expressions”, “Neural Networks” AND

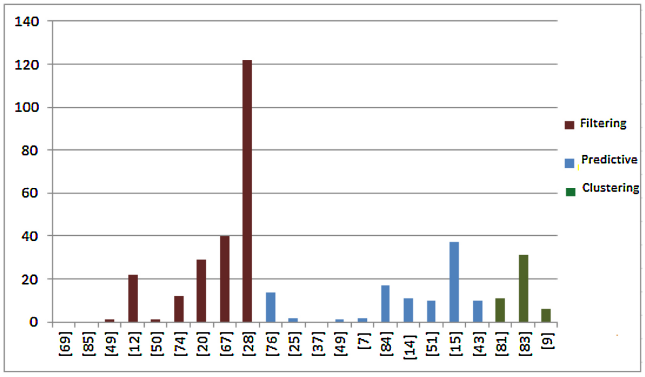

Fig. 1. Number of citation for each of the considered datasets.

"Microarray", "Neural Networks AND Cancer Prediction" and "Neural Networks AND gene expression clustering". Only articles published between 2013–2018, publicly available for free, and including one or more neural network models in their approach were considered. The chosen papers covered cancer classification, discovery, survivability prediction, and statistical analysis models. Papers using imaging or text (record) inputs were excluded.

Fig. 1 shows a graph representing the number of citations for each of the considered papers. We grouped the papers according to their functionality and ordered them chronologically by the year of publication. Considered papers have reasonable number of citations ranges between 5–120, however, papers published by 2018 are the least cited ones, some of them have zero citation, but they were considered to present some state-of-the-art approaches in the field.

In the next subsections, we review the chosen papers by presenting the adopted preprocessing techniques and the proposed models configuration.

3.1. Datasets and preprocessing

Most studies investigating automatic cancer prediction and clustering used publicly available datasets such as the TCGA [70], UCI [22], NCBI Gene Expression Omnibus (GEO) [55] and Kentridge biomedical [39] databases. These repositories were used by [20,69,51,37], [85,51,7], [85,14,15,37] and [43] respectively. In rare cases such as [9,50], studies are conducted using a specific dataset that was collected and prepared for a problem under investigation.

Removing the genes that have zero expression value across all samples is one of the simple and most straight forward preprocessing technique and is used by [20,84,12]. Chaudhary et al., [12] followed this step with removing the samples that have 20% of the features removed.

Normalization is also an essential technique in some cases. MAS5.0 affymatrix normalization was used by [69,15,14]. Mapping to Entrez Gene ID and averaging were used by [85]. Fragments per kilobase per million (FPKM) normalization was used by [76]. Funnorm normalization to remove the unwanted technical variation in methylation arrays and filtering with $p\ value > 10^{-05}$ was used in [69]. Datasets should be also normalized for training purposes using methods such as the zero mean one unit variance normalization which was used in [74,44,12]. Other kinds of transformations such as log-fold change transformation and log_2 transformations can be also used such as in [50] and [12] respectively.

Different techniques were adopted to reduce the dimensionality of the gene expressions by selecting a subset of genes. Statistical methods were applied by [20,74]. Choosing a subset of related genes as indicated by other studies or data repositories such as the Online Mendelian Inheritance in Man (OMIM) [35] is another simple feature selection technique adopted by [15,14,51]. Chen et. al, [15,14] applied, in both of their studies, the same preprocessing methods including Chi-square feature selection, selecting the top-10 ranked lung cancer-related gene signatures and combining them with T-stage and N-stage clinical data. Instead of selecting a subset of genes, Tan et al. [67] set the expression of a randomly chosen number of genes to zeros. Xiao et al. [76] applied the DESeq method to select a set of differentially expressed genes, most informative genes, based on their read count.

Even though neural networks are used for extracting the datasets features by reducing the dimensionality of the data, Principle Component Analysis (PCA) can be used as an initial preprocessing step [28,85]. The PCA method linearly transforms the dataset features into a lower dimensional space without capturing the complex relationships between the features. Therefore, a good technique can be adopted by merging the PCA components with a random number of raw features to allow the networks to capture further useful relationships [28,85].

In cases where researchers are interested in studying datasets from different sources such as [85,67]. Dataset features can be forced to have the same dimensionality for training and testing purposes, Zhang et al [85] padded the features with zero values, but this technique might increase the sparsity of the data. However, [67] overcame the problem by simply removing the genes that were not measured by the other datasets.

The class imbalance problem has only been considered once by [20] who used Synthetic Minority Class Over Sampling (SMOTE) method to generate synthetic minority class samples. Liu et al. [49] had considered the problem of small sized dataset and proposed an over-sampling technique that simply duplicates 20% of the data and used it for training. The duplication rate was proportional to the dimensionality of the data and, to force variability in the generated samples, they set a number of randomly chosen features into zeros. However, this technique can lead to sparse matrix problem.

Clustering was also applied in some studies for labeling the data by grouping the samples into high-risk, low-risk groups, or more [14]. In [84] a clustered gene filtering technique was used based on the mutation occurrence frequency of the gene data and to reduce the sparsity and the dimensionality of the data they proposed indexed sparsity reduction (ISR) procedure (see [84] for detailed description of both methods).

Table 2 presents the dataset used by each of the considered references, the applied normalization technique, the cancer type and the dimensionality of the datasets.

3.2. Neural network architecture

Our analysis to the most recent studies reveals that neural network methods have been used, in cancer prediction models, for: (i) filtering the gene expressions by removing their noise or reducing their dimensionality. The resulted features are used with statistical methods or machine learning classification and clustering tools such as decision trees, K Nearest Neighbor (KNN) and Self Organizing Maps (SOM) (see Fig. 2). (ii) predicting methods which extract the features that increase the prediction (classification) accuracy. These methods combine both the dimensionality reduction with the prediction goals in the same learning algorithm, and (iii) Clustering methods which divide gene expressions or the samples, based on their similarity.

It is important to note here that all neural network's neurons work as feature detectors (filters) that learn (extract) the inputs' features. However, our categorization into filtering, predicting and clustering methods was based on the overall role that a neural network performs in the cancer prediction model. Filtering methods are trained to remove the input's noise and to extract the most representative features that best describe the unlabeled gene expressions, so the used learning objective function has no direct relation with prediction. Nevertheless, predicting methods are trained, using labeled inputs, to extract the

Table 2
Datasets and Preprocessing. Dimension column shows the number of *features* × *samples* used for prediction, "-" indicates a missing value.

Dataset	Reference	Normalization	Cancer type	Dimensions
[33]	[25]	AML	-	7129 × 46
TCGA	[76]	FPKM	lung	1385 × 162
			stomach	801 × 271
			breast	934 × 878
TCGA, GEO and others	[37]	-	Prostate	15 × 466
TCGA	[69]	Funnorm	Breast	300, 000 × 862
GEO	[85]	MAS 5.0	Breast	881 × 13698 + 69/75
TCGA	[12]	Unit-variance	Liver	− ×365
[75]	[49]	Unit-variance	Breast	30006 × 20
[2]			lukemia	12600 × 60
[16]			Brain	2000 × 62
TCGA	[74]	Unit variance	33 different cancer	5000 × 10, 459
[36]	[81]	-	breast	1213 × 49
[36]			DLBCLA	661 × 141
[54]			leukemia	985 × 248
[54]			Multi-tissue	1000 × 103
[21]			brain	1379 × 142
[21]			multi-tissue	1363 × 190
[21]			endometrium	1771 × 42
[21]			multi-tissue	1571 × 174
home made	[50]	-	Breast	21 × 222
TCGA	[84]	-	12 different cancer	3122 × 22, 834
TCGA	[7]	Unit variance	Breast	40992 × 20
			prostate	12600 × 136
GEO	[20]	-	breast cancer	1210× −
[19]	[67]	-	METABRIC	2136 × 2520
TCGA			breast cancer	547 × 2520
[53]	[28]	-	AML	2341 × 54613
[30]			adenocarcinoma	193 × 34749
[75]			breast cancer	1047 × 30006
[40]			leukemia	2284 × 54675
[16]			leukemia	658 × 12600
[78]			AML	625 × 12625
[73]			breast cancer	2301 × 22277
[31]			seminoma	618 × 12625
[56]			ovarian cancer	153 × 15154
[2]			colon Cancer	32 × 2000
[57]			medulloblastoma	30 × 7129
[62]			prostate Cancer	102 × 12600
[71]			leukemia	2389 × 54613
[4]	[83]	-	leukemia	2194 × 72
[26]			bladder	1203 × 40
[61]			endometrial	1771 × 42
[79]			leukemia	248 × 2526
[8]			lung	1543 × 203
[17]			breast	182 × 104
[34]			lung	1626 × 181
[45]			colorectal	2202 × 37
TCGA	[14]	MAS5	lung	14 × 280
GEO	[51]	-	lung	15 × 107
TCGA	[15]	MAS5	lung	14 × 280
Kentridge biomedical	[43]	-	ALL/AML [39]	7129× 133
			CNS [39]	12533× 274
			lung cancer [39]	500× 12600
			ovarian cancer [39]	93 × 7129 93
			prostate cancer [39]	642× 15154
home made	[9]	-	bladder	10 × 104

features that are significant to prediction, so its objective function measures how accurately the network is able to predict the class of an input. Clustering methods are trained using unlabeled samples to divide them into groups based on their similarities.

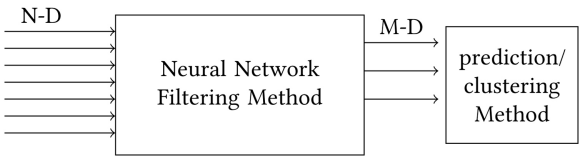

Fig. 2. Neural networks for filtering the gene expressions in cancer prediction models.

3.3. Building neural networks-based approaches for gene expression prediction

This section surveys a group of recently published studies to highlight the basic role that neural networks perform in cancer prediction. Our classification of the studies was based on investigating the learning algorithm, the objective function and its relationship with the prediction goal.

Based on our investigation of the related studies and categorization to the networks into filtering methods, prediction methods and clustering methods. We found that filtering methods learn how to generate representative codes with dimensionality $M \leq N$ (where N is the dimensionality of the input) that can be used with other machine learning

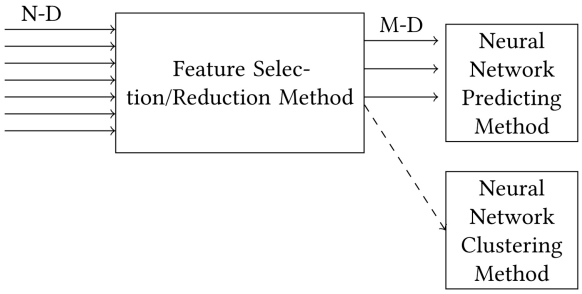

Fig. 3. Neural networks for predicting or clustering the gene expressions in cancer prediction models.

algorithms such as naïve Bayes and k-means for prediction or clustering purposes. These methods can be proceeded with preprocessing methods, but used to apply further processing on the data and learn in a purely unsupervised way with no direct relationship with prediction (Fig. 2). Predictive neural networks are trained using a supervised learning algorithm to maximize classification accuracy. Clustering methods, on the other hand, are trained using an unsupervised algorithm to set similar samples or genes in groups (Fig. 3). prediction and classification approaches are usually proceeded by dimensionality reduction methods but the goal of training them is to increase the network's capability to classify or to find the most similar group to a new testing instance with unknown label respectively.

3.3.1. Neural network filters for cancer prediction

There is a growing interest in using autoencoders to extract generic genomic features in a preprocessing step to classification, clustering and statistical analysis [24,32,23,72]. The autoencoder, in its simplest form, consists of three layers: input layer, hidden layer and output layer, divided into two parts: the encoder part to learn the mapping between high-dimensional unlabeled input data $I(x)$ and low-dimensional representations (in the bottleneck layer), and the decoder part which learns the mapping from the middle layer's representation to the high-dimensional reconstructed output $O(x)$. More complicated deep architectures are built by adding more hidden layers to both halves of the architecture.

Autoencoder-based approaches learn how to reconstruct the input examples by optimizing an objective function such as the Root Mean Squared Error (RMSE) (shown in Eq. 1):

$$\text{RMSE} = \sqrt{\frac{\sum (I(x) - O(x))^2}{n}} \tag{1}$$

or the Logloss function (Eq. 2):

$$\text{Logloss} = \sum (I(x)\log(O(x)) + (1 - I(x))(\log(1 - O(x)))) \tag{2}$$

where the number of neurons at the input layer n_0 equals to the number of neurons at the output layer n_{L-1} and L is the total number of layers.

Different types of autoencoders such as stacked denoising autoencoders used by [67,20], contractive autoencoders [50], sparse autoencoders [28], regularized autoencoders [12] and variational autoencoders [69,74] have been recently used for filtering microarray gene expressions (see Table 3). The architecture of the networks varied in depth (shallow and deep architecture), loss function and other parameters. A single hidden layer architecture and sigmoid activation function was used in [28,67]. Tan et al. [67] used a cross-entropy similarity function and stochastic gradient descent optimization. In 2016, Danee et al. [20] used a deep architecture of four layers (15,000, 10,000, 2000, and 500 neurons respectively) with stochastic gradient descent optimization algorithm. Both Way et al. [74] and Titus et al. [69] used the same variational autoencoder consisting of three hidden layers; two of the layers were in the same hierarchal level such that they receive their inputs from the input layer and send their activations forward to the third hidden layer in the higher level of the network. Titsu et al. [69] used the same implementation used by [74] except that they adapted the model to take 300,000 inputs instead of 5000. Both studies set the batch size to 50, learning rate to 0.0005 and epochs to 50. A simple network of two stacked hidden layers (500, 100 neurons) autoencoder was used in [12] and trained for 10 epochs. Zhang et al. [85] used an interesting yet simple approach for predicting the clinical outcomes of breast cancer patients. They used two hidden layers (64, 32 neurons) and added a number of neurons to the input layer which has 13698 neurons for the PCA components which were extracted in a preprocessing step. The layers were activated with the exponential linear unit activation function and optimized using Adam optimizer for backpropagation. Learning rate was set to 0.001, batch size 64, epochs 10000 and each layer was initialized with uniform distribution. Note that we are counting the number of hidden layers in the encoder side. The decoder part is a reflection to the encoder.

Overfitting is one of the major challenges affecting the efficiency of the extracted features. To overcome this problem, Chaudhary et al. [12] used regularization technique. Dropout was used by [20,12], and [28,85] added sparsity penalty to the similarity functions.

However, the autoencoders features were used by different statistical methods and classifiers to solve both binary and multi-class classification problems. [28] applied softmax regression for binary classification. Macias et al. [50] applied Cox regression model analysis and showed that autoencoders are valuable statistical tool for noise reduction in breast cancer data. They indicated that this result could be generalized for other biomedical data. Danee et al. [20] used Support Vector Machine (SVM) and a shallow neural network classifier with a softmax layer to classify the breast samples into two classes. K-means clustering was used in [12] and results showed that clustering the resulted autoencoder features into 2 clusters was the best to give optimum survival analysis results. They next used the resulting two clusters as labels to build an SVM classifier. Results demonstrated that using cluster labels is robust to predict survival-specific clusters better than the PCA extracted components. The overall performance was measured using CI and Brier error values tools. Zhang et al. [85] used an AdaBoot classifier to classify breast cancer patients into good prognosis and poor prognosis groups according to whether distant metastasis had occurred within 5 years or not. [69] applied t-SNE on the extracted representations for further dimensionality reduction to 3-D, 2-D and 1-D spaces. Logistic regression using a “one v.s rest” multi-class approach was used for subtype cancer classification and results showed that the 3-D features were significantly better than the others. Tan et al. [67] applied sample characteristics classification, transcription factor enrichment, survival analysis, and pathway analysis both on binary and multi-class levels. In [67] the author tested the performance of independent hidden nodes in discriminating tumors from normal samples and in [20] they used another simple neural network classifier consisting of an input layer connected directly to an output layer and compared its performance with an SVM. Most of the studies used 10-fold [28,67,12] and 5-fold [20] cross validation for estimating the classification error.

To conclude, neural network filtering methods were used for three different purposes: (1) to learn low dimensional representations; (2) as a transformation function that removes the noise from the input, and (3) to initialize a neural network classifier by replacing the autoencoder's output layer with a new output layer and re-training the classifier using the previously learned weights and biases, this leads to a better generalization performance. Table 3 lists the type of the autoencoder used by each reference, overfitting elimination technique, the number of hidden layers, the predictor type (classification/clustering algorithm), the number of classes and the used evaluation metric.

3.3.2. Neural networks prediction methods for cancer prediction

Neural network-based prediction (classification) involves building a network that maps the input features to an output with one or two

Table 3
Neural network filtering methods used in cancer prediction models.

neural network type	Reference	Overfitting	No.features	No.h layers	Predictor	Class	Metric
Variational Autoencoder	[69]	Kullback-Leibler divergence	100	2	Logistic regression	5	accuracy
Stacked Autoencoder	[85]	noising sparsity	64	4	AdaBoot	2	accuracy specifity sensitivity MCC
Stacked Denoising Autoencoder	[49]	noising	50 30 100	2	neural network	2 4 2	accuracy
Regularized Autoencoder	[12]	regularization dropout	37	2	K-means SVM	2 2	c-index $log-rank\ p-vlaue$ Brier score
Contractive Autoencoder	[50]	Frobenius norm	22	1	statistical analysis	-	cox regression Friedman test Holm step-down
Variational Autoencoder	[74]	Kullback-Leibler divergence	100	2	statistical analysis	-	$p-value$
Stacked Denoising Autoencoder	[20]	noising dropout	500	5	NN classifier SVM classifier	2	accuracy sensitivity specificity precision F1
Stacked Denoising Autoencoder	[67]	noising	100	1	characteristics classification	2 5	accuracy
Sparse Autoencoder	[28]	noising sparsity	183 28 1047 125 60 27 143 20 100 30 30 34 230	1	softmax regression classifier	2	accuracy

neurons (binary classification) or multiple neurons (multi-class classification). A different approach for solving the multi-class classification problem is by using multiple independent binary neural networks. However, a predictive network can be configured using an input layer, one or more hidden layers and an output layer with $n_{L-1} = k$ where k equals to the number of classes that the training inputs belong to. Before training, a binary string C'_k of length k (called the "codeword") is assigned to each training example such that the codeword for the j-th class is assigned by setting the bit in the codeword at the j-th position equals to 1, and the remaining bits to 0. For example, a 4-class problem can be modeled by assigning the (0,0,1,0) codeword to all training examples belonging to the third class of the problem, this technique is called "one-hot encoding".

A neural network can learn by feeding the labeled (or coded) inputs, calculating the neurons activations and passing them forward through the network. At the output layer, the activation function sums together the contributions of all sending units. This sum is then further modified by adjusting the activation sum to a value between 0 and 1 (in binary-class problems) or by setting the activation value at a specific position to zero unless a threshold level is reached. Through this process, the network iteratively learns the mapping between each input example $I(x)$ and its class codeword by minimizing an objective function such as the Mean Squared Error (MSE) (Eq. 3). In the testing phase, a sample is assigned to the j-th class if the network output C'_k (predicted class) at the j-th position has the highest confidence value.

$$\mathrm{MSE} = E\left[C_j(x) - C'_j(x)\right]^2 \tag{3}$$

Hence, the MSE estimates the posterior probability function for the classification problem.

Neural network-based cancer classifiers have been used with both binary-class and multi-class problems to identify cancerous/non-cancerous samples, a specific cancer type, or the survivability risk.

Our analysis revealed that the architecture of the recent predictive neural networks range between deep MLP models [14,51,84,37] and single-layered networks [15,7,43,25]. In 2014, Chen et al. [15] predicted the survival risk of lung cancer patients. The decision on the model configuration was made based on trail-and-error and best results achieved with one hidden layer and eight hidden nodes. The input to the neural network was formed by combining six gene expressions, T-stage and N-stage data to form 14-D inputs. The approach achieved high prediction accuracy in classifying the patients into low-risk and high-risk groups. The same MLP model was adopted by Chen et al. [14] in 2015 to improve the accuracy of lung cancer survival multi-class prediction where patients were classified into five classes (very low, low, normal, high and very high). Labels were assigned to the samples using a clustering technique in a preprocessing step. The network was used to learn these labels and tested in terms of its ability to predict the survivability class of new testing samples. The model achieved superior classification accuracy compared to Bayesian network, SVM, and K-Nearest Neighbor (KNN).

Mandal and Banerjee [51] used another MLP network and experimentally tested the model using two different datasets for breast and lung cancer. Only the results of the latter dataset are considered in this review, as the former consisted of 10 non-genome features. Mandal and Banerjee [51] used 15 genes that are responsible for the lung cancer, as indicated by an earlier study [52]. The best accuracy level (94.0%) and was achieved when the number of hidden layers was 3.

In 2016, Bhat et al., [7] suggested a deep generative model called

Table 4
Neural network predicting methods used in cancer prediction models.

Model name	Reference	Neural network type	Overfitting	No.layers	No. hidden nodes	No.output	Metric
MLP	[76]	MLP	regularization	5	-	1	AUC
MLP	[25]	MLP	-	1	20	1	F1
GA-ANN	[37]	MLP	-	3	-	1	AUC
SE1DCNN	[49]	Convolutional	-	7	21,4,21,4,-	2 4 2	accuracy
DeepCancere (RBM-Logistic) (RBM-SVM)	[7]	convolutional GAN	-	7	-	2	Precesion recall F1
DeepGene	[84]	Convolutional network	-	4	-	12	accuracy
Cancer adjuvant chemotherapystrategic classification	[14]	MLP	-	1	8	2	Acccuracy sensitivity specificity
MLP	[51]	MLP	-	3	-	1	accuracy
MLP	[15]	MLP	-	1	8	5	accuracy
ELM-based classifier	[43]	ELM	-	1	-	1	accuracy

DeepCancer for binary (cancerous/non-cancerous) classification. DeepCancer uses a convolutional GAN with an input dense layer, four hidden layers in the generator side, three hidden layers and an output layer with two hidden nodes in the discriminator side. Interestingly, Bhat et al. [7] also used two other RBM-based regression models, called RBM-SVM and RBM-logistic, as baselines. In both models they used one hidden layer and experimented different number of hidden nodes to get the best classification precession. No feature selection method was applied before classification, and the average performance of DeepCancer was better than the baseline models. This indicates that the generator learned how to accurately represent the features of the datasets.

DeepGene [84] is a convolutional feedforward model for somatic point mutation-based cancer type classification. DeepGene's approach includes converting the gene data into indexes of its non-zero elements using Indexed Sparsity Reduction (ISR), and feeding the output into a deep neural network classifier. The approach was tested on classifying twelve types of cancers and compared with other machine learning approaches such as SVM, KNN and naïve Bayes. The model achieved the best level of accuracy when using four hidden layers, ReLU activation function, softmax output layer and logarithmic loss function.

A new approach was suggested by [43] using Extreme Learning Machines (ELM). ELM utilizes the concept of generalized single hidden layer feedforward neural networks. In this network, neurons are generated using an analytical approach so the hidden layer does not require any tuning. The model was tested using five diseases framed as binary class problems. The correlation coefficient mechanism was used to select 2-4% relevant genes and the model achieved high classification accuracies.

Convolutional networks are most commpnly used with images datasets, however, [49] introduced one dimensional convolutional framework that is applicable with the 1-D gene expression dataset. The model, named SE1DCNN after the sample expansion method used to oversample the training split of the data and the 1-D convolutional model which consists of 7 layers (input, 2 convolutional layers, 2 max pooling layers, one fully connected layer and an output layer)

Dwivedi [25] compared the performance of six machine learning classifiers including: neural networks, SVM, logistic regression, naïve Bayes, classification trees and KNN, in classifying a lukemia dataset into two groups. He used an MLP architecture (one hidden layer with 20 neurons, 30 epochs) and proved that neural network outperformed other considered classifiers. No preprocessing, filtering or feature selection was applied on the 6817 genes which were obtained from a previous study [33].

Hou et al. [37] integrated genetic algorithms with artificial neural network in a model called GA-ANN. The model consisted of 3 layers, 1000 nodes as the input layer, and one node in the output layer. The learning rate was 0.1 and the goal was to classify the clinical phenotype into binary groups. The role of the genetic algorithm was to select the best number of input variables that maximizes the classification accuracy. The population size for the genetic algorithm was set to 100, the maximum evolutionary generations was 50, and the algorithm selected 15 candidate input variables.

Xiao et al. [76] applied deep learning to an ensemble approach that incorporates five different machine learning models. Informative gene selection was applied on differentially expressed genes. Then, a deep learning method was employed to ensemble the outputs of the five classifiers KNN, SVM, decision trees, random forests, and gradient boosting decision trees. The model of five hidden layers was used for binary (tumor/normal) classification hence one output unit was used in the output layer, ReLU activation functions and SGD optimizer was used to minimize the MSE similarity function; regularization, to constrain the magnitudes of the weights, was added to the function to overcome the overfitting problem. This paper used neural networks for discovering the relationship between the 5 different classifiers which were used to classify the training sample into tumor/normal class. The approach built a new dataset consisting to $m \times 5$ items where m is the number of training samples and 5 denotes the binary label predicted by a specific classifier. The new dataset was used as input to the neural network model to avoid using the weighted averaging and majority voting algorithms which is widely used in general ensemble strategies.

Table 4 lists the models' name, reference, the type of the used neural network, overfitting technique, the number of hidden layers, and the number of classes i.e. the number of output layer's nodes and the used evaluation metric. The researchers experimentally tested different neural network configurations and parameters but we only listed the number of hidden layers that achieved best results. We also listed the number of nodes in the output layer to indicate the category of the problem (binary-class/multi-class). We want to note that some of the listed, parenthesized, approaches have been used as baselines for comparison purposes and we considered them here to show their configuration.

3.3.3. *Neural network clustering methods in cancer prediction*

Clustering is an unsupervised learning technique which divides the input data examples based on their feature similarity into groups.

Neural networks, particularly SOM, are traditional model-based clustering techniques that are widely used with gene expression data. SOM [41] is a single-layered neural network projecting high dimensional data inputs onto a grid space. SOM's output layer neurons are organized in a two or three-dimensional map where each neuron represents a cluster and similar clusters are placed near each other using a simple neighborhood function. SOM associates each of its output neurons with a reference vector, learned during training, and each data

Table 5
Neural network clustering methods used in cancer prediction models.

Model Name	reference	neural network type	No. Clusters	metric
PCE	[81]	SOM	4	RI
(SOM)			3	Purity
(RDCFCE(SOM))			6	
			4	
			5	
			14	
			4	
			10	
RDCCE	[83]	SOM / NG	3	AR
RDCFCE			3	ARI
(SOM^2CE)			4	NMI
(NG^2CE)			6	
			5	
			2	
			2	
			2	
SOM	[9]	SOM	16	-

point is mapped to the neuron with the closest reference vector.

SOM learns using a pure unsupervised algorithm with unlabeled data and without backpropagation mechanism. Its accuracy can be measured using various evaluation matrices such as the Rand Index (RI, defined by Eq. 5) which is one of the most commonly used techniques to asses clustering in gene expression datasets. [81,83].

$$RI = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Adjusted Random Index (ARI) and Normalized Mutation Information (NMI) are improved variations of RI metric, equations can be found in [83].

In terms of cancer related applications, clustering is an analysis tool used to divide the samples or the genes into groups. Generally, gene expression clustering approaches separately group the genes or the samples by either considering the relevance, probability, of a sample belonging to a cluster (samples-to-cluster assignment) or the relevance of a gene belonging to a cluster (gene-to-cluster assignment). However, recent approaches tried to improve the quality of clustering by taking both kinds of assignments into account.

The high dimensionality of gene expression samples is a problem facing clustering algorithms. Traditional clustering techniques such as k-means separate the samples (or genes) based on a distance function but this approach is not suitable for high dimensional datasets as the distance between the samples is isometric [65]. Neural network-based approaches are able to discover sample-cluster mapping and, consequently, they improve clustering accuracy. However, these methods may suffer from noisy genes and improper setting of parameters. Two solutions have been suggested to overcome the high dimensionality problem: (i) clustering ensembles by repeatedly running a single clustering algorithm such as SOM or Neural Gas (NG) with different initializations or numbers of parameters, and (ii) projective clustering by only considering a subset of the high dimensional features.

In 2014, Borkowska et al. [9] evaluated the molecular events that characterize high- and low-grade bladder cancer pathways in bladder cancer patients. Low-grade and high-grade tumors are two distinct pathways of urothelial carcinogenes resulting from the deletion and mutation of some markers. However, many tumors have mutual aspects of low- and high-grade biology. The aim of the study was to discriminate future tumor behavior using molecular alterations. The authors collected samples from 104 random patients and measured the expression of ten genes which were proved to be related to bladder cancer as indicated by previous studies. The number of the network's inputs was ten and the output was set to two-dimensional grid of 16 (4×4) neurons. SOM was compared to other statistical methods and proved its ability to group the patients into 4 groups, each of which consisted of 4 clusters namely $X1$, $X2$, $Y1$, $Y2$. Patients with the worst prognosis were set in $X1$ group. The highest abnormal TP53 expression and heterozygocity loss for 9, 13 and 17 chromosome loci were grouped in $X2$. Samples with negative UroVysion test and high FGFR3 mutation ratio were grouped in $Y1$ and the ones with positive UroVysion test and had FGFR3 gene mutation were grouped in $Y2$. These results were hard to be obtained using classic statistical models which require explicit assumption of certain relationships within the data that are often unproven.

In 2015, Yu et al. [83] introduced two ensemble clustering frameworks, respectively, Random Double Clustering-based Cluster Ensembles (RDCCE) and Random Double Clustering-based Fuzzy Cluster Ensembles (RDCFCE). Both frameworks select a basic clustering algorithm such as SOM or NG to project high dimensional genes into a low grid dimension. As a result, a set of representative features, corresponding to the centers of the clusters, is generated. A new dataset is generated after that based on a subset of representative features and the process repeats for a specific number of times to get a number of clustering solutions B. A consensus matrix is then constructed using the set of clustering solutions $C'1$, $C'2$, . . . , $C'B$. Finally, a normalized cut algorithm [52] is used to partition the consensus matrix, and obtains the final result. Both RDCCE and RDCFCE work in the same way except that RDCFCE extends the model by incorporating the fuzzy algorithm to improve the performance of the framework. The models achieved high accuracy clustering results but they only considered sample-to-cluster assignment and ignored the gene-to-cluster assignment. Clustering performance was measured by RI and Purity measure (more details are provided in [83]).

As shown by Table 5, Yu et al. [83] also used another neural network-based clustering approach as a baseline which is: (i) double SOM-based Clustering Ensemble Approach (SOM^2CE) [82] and (ii) double NG-based Clustering Ensemble Approach (NG^2CE) [13]. Both obtain satisfactory results on most testing datasets and proved that they are robust to noisy genes.

In 2017, Yu et al. [81] suggested Projective Clustering Ensemble (PCE), which combines the advantages of both projective clustering and ensemble clustering, and they compared its performance with a set of clustering techniques including SOM and RDCFCE. Experiments on synthetic datasets showed that the accuracy of SOM decreases with the increase of injected noise. This indicates that SOM can not distinguish clusters in the presence of noisy genes. Results also showed that the PCE outperformed the RDCFCE as it assigns the irrelevant genes weight that explicitly reduces the interference.

4. Summary

Cancer is a world wide genetic-related disease which imposes significant mortality and cost. Analyzing gene expression data is essential for discovering genes abnormalities and increasing survivability as a consequence.

Neural network methods are the backbone of most recent cancer prediction models. Their ability to discover complex input-output relationships assists in obtaining more accurate sample-class (or sample-cluster) assignments than using the traditional machine learning tools which relays on distance functions or statistical assumptions. Our analysis to the most recent research in gene expression analysis tools and cancer prediction models reveals that neural networks are essentially used for filtering the gene expressions, predicting their class, or clustering them.

Neural network filtering methods, more specifically the autoencoders, were used as data engineering methods in a prior step to prediction. Examples of autoencoder-based approaches are contractive autoencoders, regularized autoencoders, sparse autoencoders and stacked denoising autoencoders which was the most widely used one. Most of the suggested filtering approaches have been experimentally tested using shallow architectures. However, deep architectures are

more recommended for best practice as they combine many non-linearities. On the other hand, it has been proven that shallow networks are inefficient in terms required training examples and the number of hidden nodes [6].

The extracted features have been used to train different machine learning tools such as SVM [20], neural network classifiers [80], k-means clustering [12] or statistical analysis [50]. Generally, filtering methods were used to reduce the dimensionality of the gene expressions, to transform the gene expressions into a different form, with the same length, or to learn gene expressions' representations and use the learned weights and biases to initialize a predicting network.

Neural network prediction methods have been used for both binary and multi-class problems. In binary problems, the classifier learns how to diagnose a given sample as cancerous, or non-cancerous or to discriminate one type of cancer from another. In this case, the network architecture can only have one [43,42] or two [7,14] output neurons in its output layer. In multi-class problems the network learns how to discriminate between multiple types of cancer or to predict the survivability risk; so the network's output layer has a number of neurons equal to the number of classes that the training data belong to [84], or alternatively multiple binary classifiers can be used. However, the decision on the number of hidden layers and nodes is usually made based on trail-and-error technique. Statistical preprocessing have been also applied in some cases [14]. Deep architecture with convolution layers was the most recently used model, especially with multiclass problems, and proved efficient capability in predicting cancer subtypes as it captures the spatial correlations between gene expressions. However, full Details on the networks configuration, overfitting elimination technique and learning parameters were not provided in most studies. Results generally show that neural network-based approaches outperformed other machine learning tools in classifying gene expression samples in most of the studies.

Clustering is another analysis tool that is used to divide the gene expressions into groups. Neural network-based clustering approaches have been applied as an alternative to traditional clustering techniques, which use a distance function to measure the similarity between gene expressions. Nevertheless, neural network approaches such as SOM and NG are not able to distinguish the noisy genes, so samples are clustered based on both the relevant and irrelevant genes. Ensembling clustering and projective clustering are two general approaches suggested to overcome the high dimensionality related problems. A hybrid approach combining both the ensembling clustering and projective clustering, such as [81], has proved higher accuracy than using single-point clustering algorithm such as SOM.

Considered studies used different preprocessing techniques, datasets, analysis and classification tools and they were used to solve different problems (binary and multi-class classifications). These differences make the decision on the best network configuration and performance hard to be made and unfair. This was also noted by [66] who indicated that there is no machine learning algorithm consistently outperforms any of the other as the nature of the dataset seem to be of a major influence on the performance of the algorithm. There is an opportunity here for a future work to develop a "benchmark" toolkit for gene expression data mining so that modern NN algorithms can be compared in a uniform way. However, we are discussing in the next section some practical issues that can be considered for future models.

5. Discussion

In this paper, we reviewed recent works on neural network based cancer prediction models. Here we discuss some technical problems that can be considered for building new models.

Overfitting: gene expression datasets are high dimensional and have a relatively small number of samples. While most of the studies on cancer prediction models focused on learning good representations that increase the models predictability, some ignored the problem of overfitting which is caused by using small number of training examples as indicated by [63]. Overfitting problem occurs when the network is over trained on the training examples so it properly fits the training examples but not the validation and the testing examples due to the lack of generalization capability. Overfitting can be avoided by: (i) adding weight penalties using regularization; (ii) models combination by averaging the predictions from many models trained on different datasets; (iii) dropout [63].

Model configuration and training: choosing neural network configuring and setting its parameters is a crucial issue for extracting good information that achieve high prediction accuracy. While there is no rule of thumb for setting the networks parameters, there are some recommended issues that can be considered to reduce both the computational and memory expenses for better performance such as: (i) proper initialization: poor initialization cause slow convergence, local minima and model uncertainties problems [1]. (ii) pruning the unimportant connections by removing the neurons that have zero values. Pruning is a good approach for reducing the memory and computational complexities. (iii) using ensemble learning framework by training different models using different parameter settings or using different parts of the dataset for each base model. (iv) class imbalance is a data-related problem that had little to get attention to date when building and analyzing cancer prediction models. This problem makes the classifiers biased toward the classes that has the majority of the data, and consequently poorly classifie the samples belonging to minor classes. Danee et al. [20] used SMOTE for dealing with class imbalance on the high dimensional level. however, oversampling methods can be applied to generate synthetic representations to increase classifiers accuracy.

Model evaluation: Using 10-fold and 5-fold cross validation to estimate the error of classifying small data size leads to severely inaccurate conclusions as proved by [10]. Braga-Neto and Dougherty [10] investigated, in their simulation study the performance of cross-validation, substitution and bootstrap methods and revealed that cross-validation displayed excessive variance and, hence, it is unreliable for small size data. The bootstrap method proved more precise and reliable predictability.

Study reproducibility: Reproducibility of the studies is another important issue that has to be highlighted here. A study is reproducible when others are able to replicate the results using the same algorithms data and methodology. Reproducibility enhances research reliability and requires the authors to publish the used data and to clearly document their methodology. Researchers using genomic databases such as TCGA [70] should at least state the query used for downloading their experimental dataset in supplementary material.

6. Conclusion

This paper reviewed the most recent neural network-based cancer prediction models and gene expression analysis tools. The considered papers were published between 2013-2018 and used gene expression datasets for cancer classification and clustering. This review presented some commonly used architectures, datasets, and the accuracies of each suggested model.

Analysing the considered papers indicated that neural network methods are able to serve as filters, predictors and clustering methods. Neural network filtering methods are used to reduce the dimensionality of the gene expressions and remove their noise. MLP and convolutional neural network classifications methods have been used with binary-class and multi-class classification problems while the number of the networks' hidden layers and hidden nodes have been decided by trial-and-error. A hybrid approach combining both ensembling clustering and projective clustering is the best to achieve high clustering accuracy.

Deciding the network architecture is one of the challenges facing the cancer prediction model designers, as there is no specific rule to guarantee high prediction accuracy. Most of the studies determined the number of hidden layers and neurons based on trail and error.

However, this study indicated that the role of the neural network determines its general architecture.

This study has summarized most recent approaches and their related neural network architectures. We also highlighted some critical points that have to be considered when building a neural network-based prediction model such as overfitting and class imbalance. More powerful neural network-based approaches can be suggested in future by choosing different network's parameters or combining two or more of the presented approaches.

Conflict of interest statement

The authors declare that they have no conflict or competing interests.

References

[1] Alhamdoosh Monther, Wang Dianhui. Fast decorrelated neural network ensembles with random weights. Information Sciences 2014;264:104–17.
[2] Alon Uri, Barkai Naama, Notterman Daniel A, Gish Kurt, Ybarra Suzanne, Mack Daniel, Levine Arnold J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci 1999;96(12):6745–50.
[3] Angermueller Christof, Pärnamaa Tanel, Parts Leopold, Stegle Oliver. Deep learning for computational biology. Mol Syst Biol 2016;12(7):878.
[4] Armstrong Scott A, Staunton Jane E, Silverman Lewis B, Pieters Rob, den Boer Monique L, Minden Mark D, Sallan Stephen E, Lander Eric S, Golub Todd R, Korsmeyer Stanley J. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 2002;30(1):41.
[5] Bashiri Azadeh, Ghazisaeedi Marjan, Safdari Reza, Shahmoradi Leila, Ehtesham Hamide. Improving the Prediction of Survival in Cancer Patients by Using Machine Learning Techniques: Experience of Gene Expression Data: A Narrative Review. Iran J Public Health 2017;46(2):165.
[6] Bengio Yoshua, Lamblin Pascal, Popovici Dan, Larochelle Hugo. Greedy layer-wise training of deep networks. In Advances in Nural Information Processing Systems 2007:153–60.
[7] Rana Bhat Rajendra, Viswanath Vivek, Li Xiaolin. DeepCancer: Detecting Cancer through Gene Expressions via Deep Generative Learning. IEEE 2017 International Conference on Big Data Intelligence and Computing (2017) 2017.
[8] Bhattacharjee Arindam, Richards William G, Jane Staunton, Li Cheng, Monti Stefano, Vasa Priya, Ladd Christine, Beheshti Javad, Bueno Raphael, Gillette Michael, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci 2001;98(24):13790–5.
[9] Borkowska Edyta M, Kruk Andrzej, Jedrzejczyk Adam, Rozniecki Marek, Jablonowski Zbigniew, Traczyk Magdalena, Constantinou Maria, Banaszkiewicz Monika, Pietrusinski Michal, Sosnowski Marek, et al. Molecular subtyping of bladder cancer using K ohonen self-organizing maps. Cancer Med 2014;3(5):1225–34.
[10] Braga-Neto Ulisses M, Dougherty Edward R. Is cross-validation valid for small-sample microarray classification? Bioinformatics 2004;20(3):374–80.
[11] Chandrashekar Girish, Sahin Ferat. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28.
[12] Chaudhary Kumardeep, Poirion Olivier B, Lu Liangqun, Garmire Lana. Deep learning based multi-omics integration robustly predicts survival in liver cancer. bioRxiv 2017(2017). 114892.
[13] Chen Hantao, Yu Zhiwen, Han Guoqiang, You Jane, Li Le. NG 2 CE: Double neural gas based cluster ensemble framework. In Computer Science & Education (ICCSE). 2012. p. 26–31.
[14] Chen Yen-Chen, Chang Yo-Cheng, Ke Wan-Chi, Chiu Hung-Wen. Cancer adjuvant chemotherapy strategic classification by artificial neural network with gene expression data: An example for non-small cell lung cancer. J Biomed Inform 2015;56:1–7.
[15] Chen Yen-Chen, Ke Wan-Chi, Chiu Hung-Wen. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. Comput Biol Med 2014;48:1–7.
[16] Cheok Meyling H, Yang Wenjian, Pui Ching-Hon, Downing James R, Cheng Cheng, Naeve Clayton W, Relling Mary V, Evans William E. Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. Nat Genet 2003;34(1):85.
[17] Chowdary Dondapati, Lathrop Jessica, Skelton Joanne, Curtin Kathleen, Briggs Thomas, Zhang Yi, Yu Jack, Wang Yixin, Mazumder Abhijit. Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative. J. Mol Diagnostics 2006;8(1):31–9.
[18] Ciatti Stefano, Gennaro G, Mungai V, Nannini G. Should volumetric breast density be included in breast cancer prediction models? Proposal of an integrated quantitative and reproducible approach. European Congress of Radiology 2015 2015.
[19] Curtis Christina, Shah Sohrab P, Chin Suet-Feung, Turashvili Gulisa, Rueda Oscar M, Dunning Mark J, Speed Doug, Lynch Andy G, Samarajiwa Shamith, Yuan Yinyin, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486(7403):346–52.
[20] Danaee Padideh, Ghaeini Reza, Hendrix David A. A deep learning approach for cancer detection and relevant gene identification. In Pacific Symposium on Biocomputing. 2016. p. 219.
[21] de Souto Marcilio CP, Costa Ivan G, de Araujo Daniel SA, Ludermir Teresa B, Schliep Alexander. Clustering cancer gene expression data: a comparative study. BMC Bioinformatics 2008;9(1):497.
[22] Dua Dheeru and Efi Karra Taniskidou 2017. UCI Machine Learning Repository. (2017). http://archive.ics.uci.edu/ml.
[23] Doersch Carl, Gupta Abhinav, Efros Alexei A. Unsupervised visual representation learning by context prediction. In Proceedings of the IEEE International Conference on Computer Vision 2015:1422–30.
[24] Dosovitskiy Alexey, Springenberg Jost Tobias, Riedmiller Martin, Brox Thomas. Discriminative unsupervised feature learning with convolutional neural networks. In Advances in Neural Information Processing Systems 2014:766–74.
[25] Dwivedi Ashok Kumar. Artificial neural network model for effective cancer classification using microarray gene expression data. Neural Comput Appl 2018;29(12):1545–54.
[26] Dyrskjøt Lars, Thykjaer Thomas, Kruhøffer Mogens, Jensen Jens Ledet, Marcussen Niels, Hamilton-Dutoit Stephen, Wolf Hans, Ørntoft Torben F. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet 2003;33(1):90.
[27] El-Bakry Hazem M. An efficient algorithm for pattern detection using combined classifiers and data fusion. Information Fusion 2010;11(2):133–48.
[28] Fakoor Rasool, Ladhak Faisal, Nazi Azade, Huber Manfred. Using deep learning to enhance cancer diagnosis and classification. In Proceedings of the International Conference on Machine Learning 2013.
[29] Fawcett Tom. An introduction to ROC analysis. Pattern Recognit Lett 2006;27(8):861–74.
[30] Fujiwara Takeshi, Hiramatsu Miyako, Isagawa Takayuki, Ninomiya Hironori, Inamura Kentaro, Ishikawa Shumpei, Ushijima Masaru, Matsuura Masaaki, Jones Michael H, Shimane Miyuki, et al. ASCL1-coexpression profiling but not single gene expression profiling defines lung adenocarcinomas of neuroendocrine nature with poor prognosis. Lung Cancer 2012;75(1):119–25.
[31] Gashaw I, Grümmer R, Klein-Hitpass L, Dushaj O, Bergmann M, Brehm R, Grobholz R, Kliesch S, Neuvians TP, Schmid KW, et al. Gene signatures of testicular seminoma with emphasis on expression of ets variant gene 4. Cell Mol Life Sci CMLS 2005;62(19-20):2359–68.
[32] Girshick Ross, Donahue Jeff, Darrell Trevor, Malik Jitendra. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition 2014:580–7.
[33] Golub Todd R, Slonim Donna K, Tamayo Pablo, Huard Christine, Gaasenbeek Michelle, Mesirov Jill P, Coller Hilary, Loh Mignon L, Downing James R, Caligiuri Mark A, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286(5439):531–7.
[34] Gordon Gavin J, Jensen Roderick V, Hsiao Li-Li, Gullans Steven R, Blumenstock Joshua E, Ramaswamy Sridhar, Richards William G, Sugarbaker David J, Bueno Raphael. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res 2002;62(17):4963–7.
[35] Hamosh Ada, Scott Alan F, Amberger Joanna S, Bocchini Carol A, McKusick Victor A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucl Acids Res 2005;33(suppl_1):D514–7.
[36] Hoshida Yujin, Brunet Jean-Philippe, Tamayo Pablo, Golub Todd R, Mesirov Jill P. Subclass mapping: identifying common subtypes in independent disease data sets. PloS One 2007;2(11):e1195.
[37] Qi Hou, Zhi-Tong Bing, Cheng Hu, Mao-Yin Li, Ke-Hu Yang, Zu Mo, Xiang-Wei Xie, Ji-Lin Liao, Yan Lu, Shigeo Horie, et al. 2018. RankProd Combined with Genetic Algorithm Optimized Artificial Neural Network Establishes a Diagnostic and Prognostic Prediction Model that Revealed C1QTNF3 as a Biomarker for Prostate Cancer. EBioMedicine (2018).
[38] Huang Min-Wei, Chen Chih-Wen, Lin Wei-Chao, Ke Shih-Wen, Tsai Chih-Fong. SVM and SVM Ensembles in Breast Cancer Prediction. PloS One 2017;12(1):e0161501.
[39] Kentridge. 2017. Kentridge biomedical repository. (2017). http://datam.i2r.a-star.edu.sg/datasets/krbd/.
[40] Klein Hans-Ulrich, Ruckert Christian, Kohlmann Alexander, Bullinger Lars, Thiede Christian, Haferlach Torsten, Dugas Martin. Quantitative comparison of microarray experiments with published leukemia related gene expression signatures. BMC Bioinformatics 2009;10(1):422.
[41] Kohonen Teuvo. The self-organizing map. Neurocomputing 1998;21(1):1–6.
[42] Krishnaiah V, Narsimha Dr G, Subhash Chandra Dr N. Diagnosis of lung cancer prediction system using data mining classification techniques. Int J Comput Sci Inform Technol 2013;4(1):39–45.
[43] Arun Kumar C, Ramakrishnan S. Binary Classification of cancer microarray gene expression data using extreme learning machines. In Computational Intelligence and Computing Research (ICCIC). 2014. p. 1–4.
[44] Kumar Vijay, Kumar Dinesh. Gene Expression Data Clustering Using Variance-based Harmony Search Algorithm. IETE J Res 2018:1–12.
[45] Laiho P, Kokko A, Vanharanta S, Salovaara R, Sammalkorpi H, Järvinen H, Mecklin JP, Karttunen TJ, Tuppurainen K, Davalos V, et al. Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis. Oncogene 2007;26(2):312.
[46] Lam Hak-Keung, Ekong Udeme, Liu Hongbin, Xiao Bo, Araujo Hugo, Ling Sai Ho, Chan Kit Yan. A study of neural-network-based classifiers for material classification. Neurocomputing 2014;144:367–77.
[47] Lawrence Steve, Lee Giles C, Tsoi Ah Chung, Back Andrew D. Face recognition: A convolutional neural-network approach. IEEE Trans Neural Netw

1997;8(1):98–113.
[48] Leung Michael KK, Delong Andrew, Alipanahi Babak, Frey Brendan J. Machine learning in genomic medicine: a review of computational problems and data sets. Proc. IEEE 2016;104(1):176–97.
[49] Liu Jian, Wang Xuesong, Cheng Yuhu, Zhang Lin. Tumor gene expression data classification via sample expansion-based deep learning. Oncotarget 2017;8(65):109646.
[50] Laura Macías-García, José María Luna-Romera, Jorge García-Gutiérrez, María del Mar Martínez-Ballesteros, José C Riquelme-Santos, and Ricardo González-Cámpora. 2017. A Study of the Suitability of Autoencoders for Preprocessing Data in Breast Cancer Experimentation, Journal of Biomedical Informatics (2017).
[51] Mandal Sudip, Banerjee Indrojit. Cancer classification using neural network. Int J 2015;172.
[52] Mandal S, Saha G, Pal RK. Reconstruction of dominant gene regulatory network from microarray data using rough set and bayesian approach. J Comput Sci Syst Biol 2013;6(5):262–70.
[53] Mills Ken I, Kohlmann Alexander, Williams P Mickey, Wieczorek Lothar, Liu Weimin, Li Rachel, Wei Wen, Bowen David T, Loeffler Helmut, Hernandez Jesus M, et al. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. Blood 2009;114(5):1063–72.
[54] Monti Stefano, Tamayo Pablo, Mesirov Jill, Golub Todd. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach Learn 2003;52(1):91–118.
[55] NCBI repository. ([n. d.]).
[56] Petricoin Emanuel F, Ardekani Ali M, Hitt Ben A, Levine Peter J, Fusaro Vincent A, Steinberg Seth M, Mills Gordon B, Simone Charles, Fishman David A, Kohn Elise C, et al. Use of proteomic patterns in serum to identify ovarian cancer. The Lancet 2002;359(9306):572–7.
[57] Pomeroy Scott L, Tamayo Pablo, Gaasenbeek Michelle, Sturla Lisa M, Angelo Michael, McLaughlin Margaret E, Kim John YH, Goumnerova Liliana C, Black Peter M, Lau Ching, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 2002;415(6870):436–42.
[58] Prieto Alberto, Prieto Beatriz, Ortigosa Eva Martinez, Ros Eduardo, Pelayo Francisco, Ortega Julio, Rojas Ignacio. Neural networks: An overview of early research, current frameworks and new challenges. Neurocomputing 2016;214:242–68.
[59] Alec Radford, Luke Metz, and Soumith Chintala 2015. Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 (2015).
[60] Ravì Daniele, Wong Charence, Deligianni Fani, Berthelot Melissa, Andreu-Perez Javier, Lo Benny, Yang Guang-Zhong. Deep learning for health informatics. IEEE J Biomed Health Inform 2017;21(1):4–21.
[61] Risinger John I, Larry Maxwell G, Chandramouli GVR, Jazaeri Amir, Aprelikova Olga, Patterson Tricia, Berchuck Andrew, Carl Barrett J. Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer. Cancer Res 2003;63(1):6–11.
[62] Singh Dinesh, Febbo Phillip G, Ross Kenneth, Jackson Donald G, Manola Judith, Ladd Christine, Tamayo Pablo, Renshaw Andrew A, D'Amico Anthony V, Richie Jerome P, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002;1(2):203–9.
[63] Srivastava Nitish, Hinton Geoffrey E, Krizhevsky Alex, Sutskever Ilya, Salakhutdinov Ruslan. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–58.
[64] Steck Harald, Krishnapuram Balaji, Dehing-oberije Cary, Lambin Philippe, Raykar Vikas C. On ranking in survival analysis: Bounds on the concordance index. In Advances in Neural Information Processing Systems 2008:1209–16.
[65] Steinbach Michael, Ertöz Levent, Kumar Vipin. The challenges of clustering high dimensional data. In New Directions in Statistical Physics 2004:273–309. Springer.
[66] Patrik Svensson 2016. Machine learning techniques for binary classification of microarray data with correlation-based gene selection. (2016).
[67] Tan Jie, Ung Matthew, Cheng Chao, Greene Casey S. Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders. In Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing. 2015. p. 132.
[68] Jiliang Tang, Salem Alelyani, and Huan Liu 2014. Feature selection for classification: A review. Data Classification: Algorithms and Applications (2014), 37.
[69] Alexander J Titus, Carly A Bobak, and Brock C Christensen 2018. A New Dimension of Breast Cancer Epigenetics. (2018).
[70] Wiznerowicz M Tomczak K, Czerwinska P. 2017. The Cancer Genom Atlas (TCGA). (2017). https://portal.gdc.cancer.gov/.
[71] Verhaak Roel GW, Wouters Bas J, Erpelinck Claudia AJ, Saman Abbas H, Beverloo Berna, Lugthart Sanne, Löwenberg Bob, Delwel Ruud, Valk Peter JM. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009;94(1):131–4.
[72] Wang Xiaolong, Gupta Abhinav. Unsupervised learning of visual representations using videos. In Proceedings of the IEEE International Conference on Computer Vision 2015:2794–802.
[73] Wang Yixin, Klijn Jan GM, Zhang Yi, Sieuwerts Anieta M, Look Maxime P, Yang Fei, Talantov Dmitri, Timmermans Mieke, Meijer-van Gelder Marion E, Yu Jack, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. The Lancet 2005;365(9460):671–9.
[74] Gregory P Way and Casey S Greene. 2017. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. bioRxiv (2017), 174474.
[75] Woodward Wendy A, Krishnamurthy Savitri, Yamauchi Hideko, El-Zein Randa, Ogura Dai, Kitadai Eri, Niwa Shin-ichiro, Cristofanilli Massimo, Vermeulen Peter, Dirix Luc, et al. Genomic and expression analysis of microdissected inflammatory breast cancer. Breast cancer research and treatment 2013;138(3):761–72.
[76] Xiao Yawen, Wu Jun, Lin Zongli, Zhao Xiaodong. A deep learning-based multi-model ensemble method for cancer prediction. Comput Methods Programs Biomed 2018;153:1–9.
[77] Xue Bing, Zhang Mengjie, Browne Will N, Yao Xin. A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Comput 2016;20(4):606–26.
[78] Yagi Tomohito, Morimoto Akira, Eguchi Mariko, Hibi Shigeyoshi, Sako Masahiro, Ishii Eiichi, Mizutani Shuki, Imashuku Shinsaku, Ohki Misao, Ichikawa Hitoshi. Identification of a gene expression signature associated with pediatric AML prognosis. Blood 2003;102(5):1849–56.
[79] Yeoh Eng-Juh, Ross Mary E, Sheila A, Shurtleff W, Williams Kent, Patel Divyen, Mahfouz Rami, Behm Fred G, Raimondi Susana C, Relling Mary V, Patel Anami, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1(2):133–43.
[80] Safoora Yousefi, Congzheng Song, Nelson Nauata, and Lee Cooper. 2016. Learning Genomic Representations to Predict Clinical Outcomes in Cancer. arXiv preprint arXiv:1609.08663 (2016).
[81] Yu Xianxue, Yu Guoxian, Wang Jun. Clustering cancer gene expression data by projective clustering ensemble. PloS One 2017;12(2):e0171429.
[82] Yu Zhiwen, Chen Hantao, You Jane, Li Le, Han Guoqiang. SOM 2 CE: Double Self-Organizing Map Based Cluster Ensemble Framework and its Application in Cancer Gene Expression Profiles. In IEA/AIE 2012:351–60. Springer.
[83] Yu Zhiwen, Chen Hantao, You Jane, Liu Jiming, Wong Hau-San, Han Guoqiang, Li Le. Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data. IEEE/ACM Trans Comput Biol Bioinform 2015;12(4):887–901.
[84] Yuan Yuchen, Shi Yi, Li Changyang, Kim Jinman, Cai Weidong, Han Zeguang, Feng David Dagan. DeepGene: an advanced cancer type classifier based on deep learning and somatic point mutations. BMC Bioinformatics 2016;17(17):476.
[85] Dejun Zhang, Lu Zou, Xionghui Zhou, and Fazhi He. 2018. Integrating Feature Selection and Feature Extraction Methods with Deep Learning to Predict Clinical Outcome of Breast Cancer. IEEE Access (2018).