Lyallpur Khalsa College
Jalandhar-144001, Punjab

Research & Reviews in Biotechnology & Biosciences
Homepage: http://www.biotechjournal.in
Volume: 6, Issue: 1, Year: 2019 PP: 19-30

Review

A review on current and future prospective of Cancer Classification through Deep Learning

Prakash B[a], Srihariakash K[b], Poornima R M[b], Gayathri.,V.R[a] , Yashicka J [a], Shen-Ming Chen[c]

[a]Department of Biotechnology, Vivekanandha College of Arts & Science for Women (Autonomous) Namakkal, India

[b]Department of Computer Applications (MCA), Kongu Engineering College, Perundurai, Erode, India

[c]Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taipei 106, Taiwan

https://doi.org/10.6084/m9.figshare.9411167.v1

Article History

Received: 02/01/2019
Revised: 25/01/2019
Accepted: 15/02/2019

Abstract

Cancer is the second foremost origin of death in the world, next to heart disease. The name cancer refers to more than a thousand sicknesses illustrate by out of direct development & replication of multiple cells. Due to this reason of cancer analysis, utilization of microarray datasets along with machine learning methods escalating in the current research scenario. Classification is one of the very broadly used datamining techniques to build a model that describes & distinguishes data classes in a manner to be used to predict the class of unseen instances. In machine learning, features are chosen manually for a classifier. With Deep learning features, extraction and modelling steps are automatic. Deep learning is one of the most significant among machine learning that requires computing system to iteratively perform calculations to identified patterns by itself. Deep learning use training data to discover underlying patterns, build models & make predictions based on the best fit model. In the last decades, there has been a growing interest of addressing cancer classification using deep learning due to their positive revival of neural networks and connectionism from the genuine integration of the latest advances in parallel processing enabled by coprocessors. Here the review of deep learning for classification in bioinformatics presenting examples of current research. Additionally, we discuss Deep learning and convolutional neural network working principles to provide a useful and comprehensive perspective, this paper presents three works DeepGen, SDAE, Enhance Feature learning in a brief description of each study. We believe that this review will provide valuable insights and serve as a starting point for the researcher to apply deep learning approaches for classification in Gene expression dataset.

**Corresponding Author:*
E-Mail: prakazbt@gmail.com

Keywords : Deep learning, DNA Microarray, Conventional Neural Networks, cancer cells, DeepGene, SDAE

I. INTRODUCTION

All cancers derive from single cells that have acquired the characteristics of continually dividing in an unrestrained manner & invading surrounding cells. Cancer cells behave in this abnormal behavior because of change in the DNA(Deoxyribonucleic Acid) sequence of key genes, which are known as cancer genes. cancer cells have modified genomes that lead arising from the uncontrolled growth of cells. In a cancer cell, several chromosomes are present in more than two copies mutation plays a vital role in cancer cells. Two types of mutation occur in cancer cells. First, the germline mutation changes in the DNA Sequence that can be inherited from either parent. The second the somatic mutation changes in the DNA sequence in cells other than sperm or eggs. The mutation is present in the cancer cell and its offspring, but not in the patients. Mutation in cancer genes can occur somatically or can be inherited. Cancer genes are classified into two types tumor suppressors genes and oncogene. Genes which normally function to promote cell growth/division in a controlled manner.oncogene acquire mutations that result in continually active proteins which leads to uncontrolled cell growth and division. Tumor suppressor gene normally functions to prevent cell growth /division. It codes for a protein that slows down cell growth. Due to the enhancing nature of the database, it leads to issues such as efficient and effective retrieval of data, meaning full data from a huge dataset. Parallel clustering reduces the time for processing the huge set of data for efficient data retrieval [3]. In cancer medical diagnosis classification of the several cancer types in the most significant. A precise prediction of several cancer types provides better treatment and toxicity minimization of patients.

Monitoring the characteristics of DNA microarray offered a deep insight into cancer classification. It introduced a lavish amount of data equipped to investigate. Besides, it helps in understanding the function of gene and the interactions between genes in disease and normal conditions. This is done by the analysis of DNA microarray dataset for gene-gene communication under various conditions [1]. DNA Microarray is one such technology which enables the researchers to investigate and address issues which were once thought to be nontraceable. It has empowered the scientific community to understand the fundamental aspects underlying the growth and development of life as well as explore the genetic cause of anomalies occurring in the functioning of the human body. Three types of microarrays can be categorized as microarray expression analysis, microarray for mutation analysis and comparative genome hybridization. Among this microarray Expression analysis is used for comparison of the expression pattern of the gene responsible for a disease. Applications of microarrays are gene discovery, disease diagnosis, drug discovery and Toxicological Research. Thus, DNA Microarray procedures have been useful to extract patterns and build classification model& it supports to predicted Cancer disease [2]

Classification is a task which assigns objects to classes or group on the basis of measurements made on the objects. it must have labels for some points and its need a rule that accurately assign labels to new points and it is supervised learning. The role of classification a sample at first as disease or free from disease and subsequently if diseased the find the particular type of disease. Initially, research starts cancer classification based on morphological and clinical based. The conventional cancer classification procedures are confirmed to have a lot of limitations. Recent research demonstrated that DNA microarray could provide useful information for cancer classification at the gene expression level. Several soft computing algorithms have already been applied to classifying cancer using microarray data, but there are some issues that make it a nontrivial task. Nature-inspired computing are admired for solving real-world problems are enhancing huge, difficult of the problems, it essential to locate as optimization method [4].Many classification algorithms have been proposed in the past. Building classifier based gene expression data is a promising approach, yet the selection of non-redundant but relevant gene is difficult [5].

II. DEEP LEARNING

Deep learning (DL) is termed as universal approximated because of its mapping from input to output as y = f(x) to find out correlation among attributes x & y present in the dataset [6]. Deep learning concept evolves from a neural network of the machine learning algorithm. Deep learning algorithm overcomes the issues faced by the neural network. DL diverges from the traditional neural network in terms of depth having of more than one hidden layer apart from the input & output layer. Thus, a deep learning algorithm is known as "stacked neural network". DL must have a minimum of three hidden layers. DL has feature hierarchy so that end-to-end learning was achieved, where a network is given raw data and a task to perform, such as classification and it learns how to do this automatically. Since they combine and aggregate the features from one layer to next [7]. thus, it gains complexity and abstraction level to makes the best choice for handling very large and high dimensional complex dataset. DL has shown great representation learning can discover effective features as well as their mapping from data for given tasks. In other words, with artificial neural networks of multiple nonlinear layers, referred to as deep learning architectures, hierarchical representations of data can be discovered with increasing levels of abstraction [8].The deep learning architectures classify into four groups. They are Deep Neural network [9,10,11,12]. Convolutional neural networks [13,14,15,16], Recurrent Neural network [17] and emergent architectures [18,19,20,21]. The deep learning architectures is shown in Figure 1.

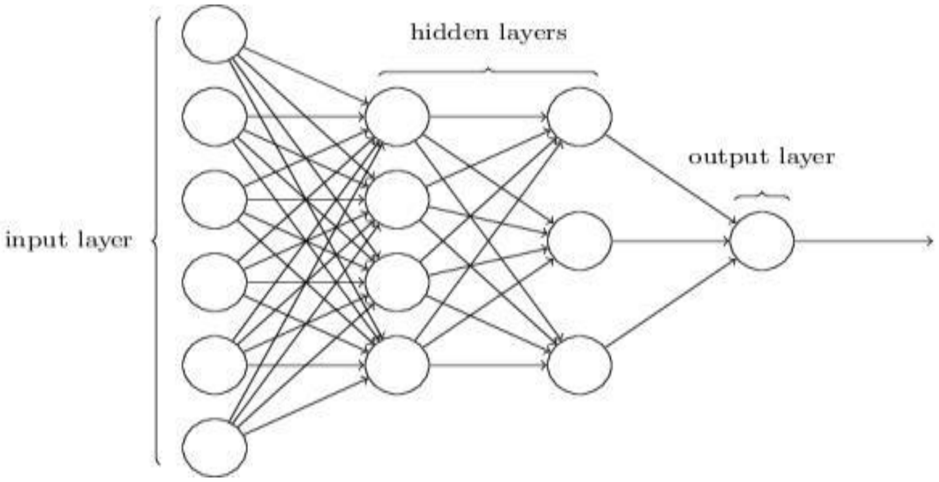

Figure 1 The architecture of Deep Neural network

A. Working principle of Deep learning

The purpose of training DL architecture is the optimization of theweight parameters in each layer, which step by step combines simpler features into intricate features so that mainly appropriate hierarchal representation can be learned from data. A single cycle of the optimization process is organized as follows [22]. First, a training dataset is given, the forward pass sequentially computes the output in each layer and propagates the function signals forward through the network. The finishing output layer, an objective loss function measures the error between the inference outputs and given labels. To reduces the training error, the backward pass uses the chain rule to backpropagate error signals and compute gradients with respect to all weights throughout the neural network [23]. Finally, the weight parameters are updated using an optimization algorithm based on stochastic gradient descent [24]. Another core element in the training of deep learning architecture is regularization, which refers to strategies intended to avoid overfitting and thus achieve the best generalization performance. Presently the most widely used regularization approach is dropout [25]. This method randomly eliminates hidden units from the neural networks during training and can be considered as an ensemble of possible subnetworks [26]. Newly proposed batch normalization [27] provides a new regularization method through normalization of scalar features for each activation within a mini-batch and learning each mean and variance as a parameter. The convolutional neural network (CNN) is a deep learning model with a key idea of using convolutional layers to extract features from input data.

III. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNN) is biologically-inspired variants of MLPs. From Hubel and Wiesel's early work on the cat's visual cortex [28], we know the visual cortex contains a complex arrangement of cells. These cells are sensitive to small sub-regions of the visual field, called a receptive field. The sub-regions are tiled to cover the entire visual field. These cells act as local filters over the input space and are well-suited to exploit the strong spatially local correlation present in natural images. The animal visual cortex being the most powerful visual processing system in existence, it seems natural to emulate its behaviour.

Additionally, two basic cell types have been identified: Simple cells respond maximally to specific edge-like patterns within their receptive field. Complex cells have larger receptive fields and are locally invariant to the exact position of the pattern. It was inspired by the visual mechanism of living organisms. The fundamental structure of CNNs consists of convolutional layers, nonlinear layers, and pooling layers shown in figure 2. This is achieved with local connections and tied weights followed by some form of pooling which results in translation-invariant features.

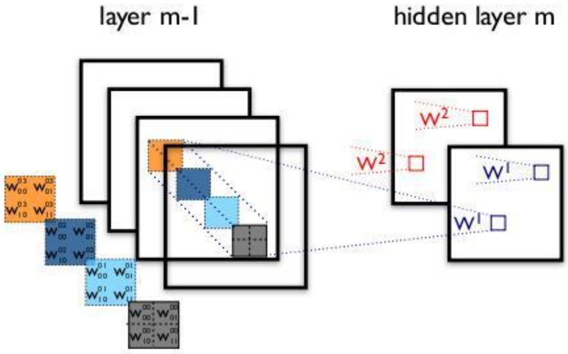
layer m-1
hidden layer m
W^2
W^2
W^1
W^1
W^{03}_{00} W^{03}_{01}
W^{03}_{10} W^{03}_{11}
W^{02}_{00} W^{02}_{01}
W^{02}_{10} W^{02}_{11}
W^{01}_{00} W^{01}_{01}
W^{01}_{10} W^{01}_{11}
W^{00}_{00} W^{00}_{01}
W^{00}_{10} W^{00}_{11}

Figure 2. Structure of Convolutional Neural network.

Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. CNN's exploit spatially-local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. In other words, the inputs of hidden units in layer m are from a subset of units in layer m-1, units that have spatially contiguous receptive fields.

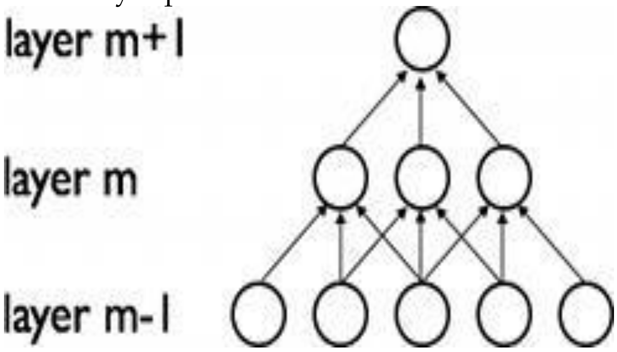

Figure 3 Structure of Convolutional Neural network

Imagine that layer m-1 is the input retina. In the above figure 3, units in layer m have receptive fields of width 3 in the input retina and are thus only connected to 3 adjacent neurons in the retina layer. Units in layer m+1 have similar connectivity with the layer below. We say that their receptive field with respect to the layer below is also 3, but their receptive field with respect to the input is larger (5). Each unit is unresponsive to variations outside of its receptive field with respect to the retina. The architecture thus ensures that the learnt “filters” produce the strongest response to a spatially local input pattern.

However, as shown above, stacking many such layers lead to (non-linear) “filters” that become increasingly “global” (i.e. responsive to a larger region of pixel space). For example, the unit in hidden layer m+1 can encode a non-linear feature of width 5 (in terms of pixel space). In addition, inCNN's, each filter is replicated across the entire visual field. These replicated units share the same parameterization (weight **vector and bias) and form a feature map.**

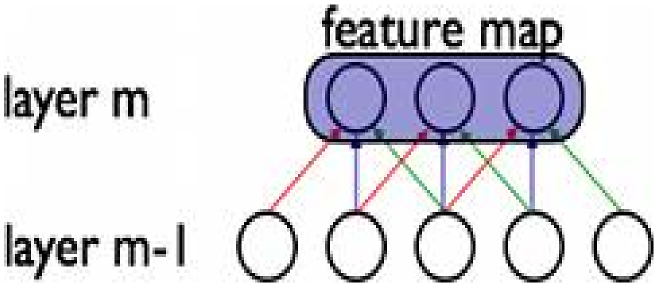

Figure 3 Shared weights in Convolutional Neural network

In the above figure, we show 3 hidden units belonging to the same feature map. Weights of the same colour are shared—constrained to be identical. Gradient descent can still be used to learn such shared parameters, with only a small change to the original algorithm[29]. The gradient of a shared weight is simply the sum of the gradients of the parameters being shared. Replicating units in this way allows for features to be detected regardless of their position in the visual field. Additionally, weight sharing increases learning efficiency by greatly reducing the number of free parameters being learnt. The constraints on the model enable CNNs to achieve better generalization on vision problems.

A. Working principles of CNN

CNN is applied to imitate three key ideas: local connectivity, invariance to location and invariance to local transition [30]. To use this highly correlated with sum regions of data, groups of locally weighted sums a called feature maps are obtained at each convolutional layer by computing convolutional between local paths and weight vectors called filters. Identical patterns can appear regardless of the location in the data. Filters are applied repeatedly across the entire datasets which also improves training efficiency by reducing the number of parameters to learn. Then nonlinear layers increase the non-linear properties of features maps. At each pooling layer maximum or average subsampling enables CNNs to handle somewhat different but semantically similar features and thus aggregate local features to identify more complex features

IV. RELATED WORK

A. Somatic point mutation-based cancer classification (SMCC)

The SMCC method, named Deepgene developed to simultaneously address the three identified problems. DeepGene is a DNN based classification model contains three major steps. Initially, it conducts two pre-processing methods. The first steps are clustered gene filtering & second on is indexed Sparsity reduction (ISR). The final outcomes of two pre-processing are combined for the final DNN classifier.

The clustering Gene Filtering

The clustering Gene Filtering (CGF) method finds the discriminatory gene subset based on mutation concurrency frequency through this method, the relationship among the genes can be properly summarize the discriminatory subset of the gene dataset. It is based on the mutation occurrence frequency of the gene data.CFG finds effective discriminatory subset from entire datasets.using mean and standard deviation.CFG doesn't require any prior knowledge from the actual dataset.

Algorithm for CGF

Step 1: sum the row and add the result with row index.

Step 2: list out descending order by higher count frequency genes.

Step 3: initialize each as ungrouped.

Step 4: use Jacrad coefficient to group the sum of row for finding inter-gene similarity.

Step 5: Set the thershold value

Step 6: check its similarity with the threshold.

Step 7: select the top most gene having highest mutation occurrence count to get the filtered genes to create a discriminatory subset.

Indexed Sparsity Reduction (ISR)

The CGF method removed irrelevant genes but still discriminatory gene group suffers mostly sparse. It contains most of the gene value is zero. To overcome the data sparsity problem, ISR method was proposed. It reduces the sparsity by changing the gene value into non-zero genes. Monitor the count of non-zero elements from each sample in the data. The ISR threshold value is set to create an index which contains non-zero elements. Count the top non zero elements having most occurrence frequency then fills zero padding to the tail of the output data that is has length ISR performs elimination of data sparsity from the taken dataset. Thus, it helps to given a better classifier.

DNN (Deep Neural Network) based Classifier

Combine both CGF & ISR as the pre-processing for our DNN based classifier lead to perform a effective classifier. The output of the CGF is given to ISR to perform taking away zero value genes from the dataset. Thus dense sparsity is minimize The output of two CGF & ISR is concatenated as the input of DNN classifer.DNN based classifier performed feed forward artificial neural network which contains fixed size of input and output. It also has multiple hidden layers for pre-processing the data. Compute activation function for calculating activation function for hidden layers and the total the weighted sum of the input using softmax layer and logarithm loss function. Train feed-forwarding and compute the loss l is transfer from the previous layer to the next layer through back propagation.The parameters of W and b are updated for each layer. Then training enter into next fixed point and back propagation and feed forwarding are carryout and get softmax probability. Take up the cancer type correspond to minimum sofmax probability. Finally, it exhibits best performance improvements and generates the best accuracy result for classification problem.

B. Stacked Denoising Autoencoder (SDAE)

Deep learning method uses SDAE for extracting the significant gene expression relationship. Complexity the training with SDAE selected a layer which contains both low validation errors and low dimension. Compared both with stacked to help of validation dataset independent of both test set and training set[30].In SDAE [31] have four layers of dimensions .the weights used for each layer to extract genes having information content. Stochastic gradient descent (SGD) algorithm is used for getting weight for each model in SDAE. When the training set is large, one could reach the desired accuracy ρ measured on the whole training set without even visiting all the training examples. This is in fact a kind of generalization bound.[32]. SDAE contains both encoder & decoder. The encoder of the SDAE decrease the dimensionality of the gene expression data stack by stack which leads to reduced loss of information compared to reducing the dimension in one step.[33].In other side of decoder increases the dimensionality to finally to finally achieve the full reconstruction of the actual input as close as possible. In this method the output of the one layer is the input of the next layer. For overfitting in SDAE, dropout regularization factor is used. This method outcome the low rate of noise in the gene expression dataset.

Deep feature Extraction & Deeply Connected Genes

The importance of genes by considering combined effect of each stack of the deep architecture. To find deduce these genes computing the product of the weight matrices for each layer of our SDAE. Dimensional matrix is found weight of each layer of SDAE is computed with nonlinear model. Thus, genes with largest weights are strongly connected to the extracted and highly predictive features thus these genes names as deeply connected genes.

C. Enhance Feature Learning Method for Deep Learning

Dimensionality problem in gene expression dataset is nowadays a complex for researchers. To find a sparse representation, multi-layer sparse encode is used. To improve the classification accuracy and computational time, we need a efficient feature learning process is need. In this paper [34] it has two phases. In the first phase it reduces the dimensionality of feature space and in the second phase sparse encoding of the data samples to find out high level and complex features. through this duplicate and noisy data can be reduced. thus, it improves the performance of classification by providing the significant features for classification. In order to detect cancer genes & cancer type classification, features are learned. Softmax regression as learning approach for classifier is used. The sparse autoencoder is constructed by three layers in the neural network (i.e input layer, hidden layer, and output layer) in which the hidden layer contains K nodes. The units in the hidden layer force the network to learn a representation of the input with only K hidden unit activations, representing K features. To train the network it uses the back-propagation method to minimize the squared reconstruction error with an additional sparsity penalty [35]. A sparse autoencoder with one layer and one with two layers (aka. stacked autoencoder) has been used as the unsupervised feature learning method to learn a sparse representation from unlabeled data which then served as the input representation for classifier learning using the softmax regression classifier. In addition, we performed an additional experiment in which we used the fine-tuning method in order to tune the weights of the features of the stacked autoencoder to better match the requirements of the classification task. In this method, the weights of the features learned by the unsupervised feature learner are tuned through the classifier using labeled data. While this makes the features less generic by tuning them towards the specific classification task, it also promises the possibility of higher classification accuracy in some situations. It is unsupervised non-linear sparse feature learning for the fetching of effective features for general classification tasks. This methodology allows for the effective use of unlabeled data, and thus of microarray data unrelated to the specific classification task, to assist in and improve the classification accuracy. As mentioned earlier, since the number of gene expression data samples for a specific cancer type is generally low, other cancer data from the same platform (i.e. with the same genes in the microarray)are a good candidate to be used in this method as unlabeled data for feature learning. One of the significant advantages of this approach as compared to most previous work is that it generalizes the feature sets across different types of cancers. For instance, data from prostate, lung, and other cancers can be used as unlabeled data for feature learning in a breast cancer detection or classification problem. The potential of this is further demonstrated by the results of Lu et.al in [36] who showed via comprehensive gene analyses that it is possible to find common cancer genes across different cancer data. This finding intensifies our argument for having generalized feature sets across data from various cancer types.

PERFORMANCE OF THE ALGORITHMS

DeepGene method conducts cancer type classification for somatic point mutation. From the gene data mutation occurrence frequency carried out by CGF. The gene data sparsity diminishes by performing ISR steps to create index non-zero elements and in the final step. Taking result of CGF and ISR output values was given to DNN based classifier to processes data and produce the classification output with high level data feature learning. CGF, ISR and DNN all methods are Contribute in improving the overall classification performance. DeepGene with three widely adopted classifiers and demonstrate that DeepGene has a better performance improvement in terms of testing accuracy. SDAE method procedure avoids nodes from co-adapting too much and as a result avoids overfitting.[37] For the same purpose, the method provided moderately degraded input values to the SDAE (denoising). The SDAE is robust, and its accuracy does not change upon introducing noise at a low rate. In fact, SDAE with denoising and dropout can find a better representation from the noisy data. Using SDAE method, cancer classification achieving 94.78% accuracy. SDAE extract the meaningful features from the gene dataset this render the classification of cancer genes. Enhance Feature Learning utilizes Principle Component Analysis to address the very high dimensionality of the initial raw feature space followed by sparse feature learning techniques to construct discriminative and sparse features for the final classification step, provides the potential to overcome the problems found traditional approaches with feature dimensionality as well as very limited size data sets. It does this by allowing data from different cancers and other tissue samples to be used during feature learning independently of their applicability to the final classification task. Applying this method to cancer data and comparing it to baseline algorithms, our method not only shows that it can be used to improve the accuracy in cancer classification problems, but also demonstrates that it provides a more general and scalable approach to deal with gene expression data across different cancer types.

VI. CONCLUSION

This study presents the cancer classification using deep learning for gene expression dataset. The DeepGene method able to classify various cancers subtypes with high accuracy. Using this method we find how many cancer subtypes are there, how we can evaluate the robustness of a new classification system, how are classification systems affected by intra tumor heterogeneity and tumor evolution, how we should interpret cancer subtypes, can multiple classification systems co-exist.

While related issues have existed for a long time, we will focus on those aspects that have been magnified by the recent influx of complex multi-omics data. Exploration of these problems is essential for data-driven refinement of cancer classification and the successful application of these concepts in precision medicine. The SDAE methods eliminates overfitting while performing classification. This algorithm encourages for Mutli-objective optimization that can select a more optimal and representative set of genes for discriminating between various cancer subtypes. The Enhance Feature learning gives a promise feature set in large dataset for classification. the redundancy between genes is reduced to a large extent. This helps to build a classifier with smaller subset of genes that distinguish between various cancer types. Deep Learning method has group of models which have non-linear transforming layers used for representing data at successively high abstraction. with these many layers, combined models are expected to be able to solve complicated problems in Cancer classification In future Deep learning algorithm can be enhance of combining the three methods of DeepGen, SDAE and feature learning to increase the performance of the classifier for to spot the differences between the sequences and Identify the type of mutation in the cancer marker gene.

ACKNOWDLEGEMNTS

Authosr want to thank management committee of insthtite for finalcial support

CONFLICT OF INTEREST

Authors declares no conflict of interest

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. This article does not contain any studies involving animals or human participants performed by any of the authors

REFERENCES

1. Author J K. Classification of human cancer diseases by gene expression profiles. App Soft Comput. 2017;124:134.

2. Bard E, Hu W. Identification of a 12 gene signature for lung cancer prognosis through machine learning. J of cancer. 2011; 148-156.

3. Thenmozhi K, Shanthi S. Optimized Data Retrieval in Big Data Environment using PPFC Approach . AJRSSH. 2017; 683-690.

4. Pyingkodi M, Thagarajan R. Meta–Analysis in Autism ene Expression Dataset with biclustering methods using Random Cuckoo search Algorithm", AJRSSH. 2017; 186-194.

5. Xiaoxing L, Ksrishnan A. An Entropy-based gene selection method for cancer classification using microarray data. BMC Bioinformatics. 2005; 6: 76.

6. Seonwoo Min, Byunghan Lee, SungrohYoon,"Deep Learning in Bioinformatics" Briefings in Bioinformatics, Briefings in Bioinformatics, 2017; 18(5), , 851–869

7. Mrutyunjaya Panda, "Elephant Search with Deep Learning for Microarray Data Analysis", Allen institute for artificial intelligence., 2017 ; ArXiv

8. LeCunY, RanzatoM. "Deep learning tutorial", tutorials in international conference on machine learning (Conference Proceedings : ICML'13), 2013.

9.Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neuralnetworks. Chemometr Intell Lab .1997;39(1):43-62.

10. Vincent P, Larochelle H, Bengio Y et al. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning p.1096-103.ACM

11. Vincent P, Larochelle H, Lajoie I et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. JMLR. 2010;11: 3371-408.

12. Hinton G, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504-7.

13. LeCun Y, Boser B, Denker JS et al. Handwritten digit recognition with a back-propagation network. Adv Neural Inf Process Syst. 1990.

14. Lawrence S, Giles CL, Tsoi AC et al. Face recognition: A convolutional neural-network approach. Neural Networks, IEEE Transactions on 1997;8(1):98-113

15. Krizhevsky A, Sutskever I, Hinton G. Image net classification with deep convolutional neural networks. In: Adv Neural Inf Process Syst. 2012;1097-105.

16. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. Neural Comput. 1989;1(2):270-80

17. Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Comput .2000;12(10):2451-71.

18. Lena PD, Nagata K, Baldi PF. Deep spatio-temporal architectures and learning for protein structure prediction. Adv Neural Inf Process Syst. 2012;512-20.

19. Graves A, Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks. Adv Neural Inf Process Syst .2009; 545-52.

20. Hadsell R, Sermanet P, Ben J et al. Learning long-range vision for autonomous off-road driving. J Field Robot 2009;26(2).

21. Masci J, Meier U, Cireşan D et al. Stacked convolutionalal auto-encoders for hierarchical feature extraction. Artificial Neural Networksand Machine Learning–ICANN 2011. Springer, 2011, 52-9.

22. Lena PD, Nagata K, Baldi PF. Deep spatio-temporal architectures and learning for protein structure prediction. Adv Neural Inf Process Syst. 2012; 512-20

23. Lena PD, Nagata K, Baldi PF. Deep spatio-temporal architectures and learning for protein structure prediction. Adv Neural Inf Process Syst. 2012; 512-20

24. Baldi P, Sadowski PJ. Understanding drop out. Adv Neural Inf Process Syst. 2013; 2814-22.

25. Moon T, Choi H, Lee H et al. RnnDrop: A Novel Dropout for RNNs in ASR. Automatic Speech Recognition and Understanding (ASRU) 2015.

26. Hubel, D. and Wiesel, T. Receptive fields and functional architecture of monkey striate cortex. J Physiol (London). 1968;195: 215–243.

27. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. Neural Networks, IEEE Transactions on 1994;5(2):157-66.

28. G.E.Hinton and R.R.Salakhutdinov,"Reducing the Dimensionality of Data with Neural Networks", Science vol 313.[information missing]

29. Padidehdanaee, rezaghaeini "A Deep learning approach for cancer detection and relevant gene identification" Pacific Symposium on Biocomputing,2017.

30. L´eonBottou, Olivier Bousquet, "The Tradeoffs of Large Scale Learning",Advances in neural information processing systems,2008.

31. Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy Layer-Wise Training of Deep Networks Adv Neural Inf Process Syst. 2007; 19:153.

32. Fakoor R, Ladhak F, Nazi A, "Using deep learning to enhance cancer diagnosis and classification", 2013, Conference: The 30th International Conference on Machine Learning (ICML 2013), WHEALTH workshop

33. Coates A, Lee H, and Ng A. Y. An analysis of single-layer networks in unsupervised feature learning. In AISTATS, 2011

34. Lu Y, Yi, Y Liu P, Wen W, James, Wang D, and You M. Common human cancer genes discovered by integrated gene-expression analysis. PLoS ONE. 2007; 11:1149

35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R. The Journal of Machine Learning Research. 2014; 15:1929.