# A Survey on Applications of Machine Learning in Bioinformatics and Neuroscience

Shahla Mousavi[1], Narges Habibi[2]*
1- Faculty of Engineering, Khorasgan Branch, Islamic Azad University, Isfahan, Iran.
Email: shahla.mousavi@khuisf.ac.ir
2- Department of Computer Engineering, Faculty of Engineering, Khorasgan Branch, Islamic Azad University, Isfahan, Iran.
Email: N.habibi@khuisf.ac.ir (Corresponding author)

Received: January 2022 Revised: March 2022 Accepted: May 2022

**ABSTRACT:**
Machine learning is one of the most practical branches of artificial intelligence that tries to provide algorithms by which the system can analyze a set of data in different formats. Machine learning algorithms are widely used in biomedicine, bioinformatics and neuroscience. The main goal of this paper is to propose the latest applications of machine learning in bioinformatics and neural imaging and to introduce new branches of research. In this article, the application of four indicators of machine learning techniques in the field of bioinformatics is examined. The four categories of techniques studied include clustering, classification, dimensionality, and deep learning. In this paper, we also show that machine learning techniques can be successfully used to address common bioinformatics challenges such as gene expression, DNA methylation identification, mRNA expression, patient classification, brain network analysis, protein chain identification, clustering, and biomarker identification. In each section, some efficient articles with technical details are discussed separately. The results of some papers are also reported in terms of accuracy, database and techniques used.

**KEYWORDS:** Machine Learning, Bioinformatics, Biomedical, Neuroimaging, Classification, Clustering.

## 1. INTRODUCION

In the last two decades, the amount of biomedical data has increased dramatically. More than 90% of such data is collected by bioinformatics and neuroimaging laboratories. This data is usually generated in various forms such as genome, gene expression, protein structure, DNA data, and many functional imaging techniques in neuroscience. This large amount of data has made the need for effective computing tools to analyze them more than ever. Bioinformatics is an interdisciplinary field in which new computational methods are developed for the analysis of biological data and the discovery of new concepts [1]. In other words, in genetics and genomics, bioinformatics' tools are used to find sequencing and annotation processes of genomes [2]. Bioinformatics also plays a key role in understanding and regulating gene and protein expression [3, 4]. Bioinformatics at the level of systems biology helps to list biological pathways and analyze the underlying networks of specific biological mechanisms [5, 6]. In neuroscience, different neuroimaging techniques, such as computed tomography (CT) scans, positron emission tomography (PET) scans, functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI), are used to understand brain function in vivo. One of the fascinating areas of research in this area is the analysis of the human brain network, often referred to as the human connection. The main purpose of this research sub-field is to understand the anatomical and functional organization of the brain. Understanding the brain's functional process is crucial for early detection of neurological disorders and increasing the effectiveness of treatment methods for this disease.

Machine learning (ML) refers to techniques and methods that are able to learn from data for the purpose of predictive analysis. In other words, in many applications, machine learning is considered as the develop models to explain optimally data by tuning their specific settings. In other words, the main purpose of learning a machine is to extract knowledge from input data to predict new patterns that have not been seen before. In many references, machine learning algorithms are usually divided into the following three groups:

•Supervised [7]



Paper type: Research paper
DOI: https://doi.org/ 10.30486/mjtd.2022.695920
How to cite this paper: Sh. Mousavi, N. Habibi, "**A Survey on Applications of Machine Learning in Bioinformatics and Neuroscience**", Majlesi Journal of Telecommunication Devices, Vol. 11, No. 2, pp. 95-111, 2022.

•Unsupervised [8]
•Semi-supervised [9]

In supervised learning, each observation has a label (interest result). In these models, the main purpose is to design a model that can assign the labels to the related observations. If the data has no labels and the purpose is only to analyze how the data is structured and statistically, that model is defined as unsupervised learning. In the field of semi-supervised learning, it is possible to have labeled and unlabeled data in a same database.

Machine learning algorithms and techniques have been widely used to solve problems in the field of bioinformatics and neural imaging. Most of researches in this field is devoted to bioinformatics and neural data mining, which analyzes various problems such as classification [10], clustering [11], network analysis [6] and dimensionality reduction [12]. The main purpose of this survey paper is to examine advanced approaches based on machine learning and their application in the fields of bioinformatics and neural imaging. Some effective methods are also discussed in more detail to suggest ideas for future research.

### 1.1. Paper Organization

This reminder of this paper is organized as follows: Section 2 discusses about the dimensionality reduction and feature selection techniques in the field of bioinformatics. In this subsection, univariate and multivariate methods are compared. Also, for each of the above two strategies, practical examples of feature selection methods are examined. Various applications such as biomarker prioritization, mononucleotide analysis, mutation analysis, detection of functional brain networks, and dimensional embedding of fMRI data are presented in this section. Section 3 discusses clustering techniques and their applications. Some examples of clustering are examined in more detail to identify genes co-expression. In addition, clustering applications for brain segmentation and feature extraction in fMRI data are also discussed. Section 4 describes some of the most efficient supervised classification methods for drug repositioning, proteomic classification, patient classification of neuroimaging data, and multifaceted segmentation of the human cerebral cortex. Section 5 discusses the differences between deep learning and classical low-deep neural networks. Examples of CNN applications in situ hybridization of RNA, DNA binding proteins, and predicting neural growth outcomes from the structural network of the brain are reviewed. The discussion and conclusion are in Section 6.

## 2. FETURE SELECTION AND DIMENSIONALITY REDUCTION

Bioinformatics and biomedical data usually extracted in high dimensions/features. Therefore, the use of a preprocessing step with the aim of reducing the size of the data, before performing any analysis, is strongly required. There are two main strategies for achieving this goal. First one is dimensionality reduction approaches. Second is feature selection technique.

Dimension reduction methods involve converting a large data set into a low-dimensional representation so that most of the data information is still preserved. One of the simplest and most commonly used methods is principal component analysis (PCA), which involves a linear transformation that displays the original data in a new space where the variable with the highest variance in the first axis is displayed. Be, shows. The second variable with the highest variance is mapped to the second axis, and so on.

In PCA technique, dimensional reduction is achieved only by considering the principal components, that is, it keeps a subset of the features that constitute the most variation in the data. One of the limitations of PCA is that it assumes that the data follow the Gaussian distribution, so it cannot be used for data that follow more complex distribution functions. While PCA is based on orthogonal conversion to obtain non-correlated linear properties, the Independent Component Analysis (ICA) technique [13] is described based on identifying statistically independent components in the data. Some other approaches are based on factor analysis, projection pursuit, regression, and continuous topological mapping [14].

The main limitation of dimensional reduction techniques is that some information is inevitably lost in the process. They may also prevent interpretability in the case of irreversible (one-sided) maps that do not allow a return to the original representation. Feature selection methods may be preferred when the main problem is the need to remain key features. In feature selection methods, the main purpose is to select a low-dimensional subset of high-dimensional data to reveal basic information. This group of techniques is also mainly used as a preprocessing step for other computational methods .Three different strategies are most used in this area and are commonly used:

- Univariate or multivariate filtering methods
- Embedded methods
- Multivariate wrapper [12]

In filter-based methods, features (attributes) are sorted ranked based on a predefined measure, then a percentage of the top-rated features are selected and the rest are discarded. These approaches are independent of the classifier. In univariate methods, each dimension (feature) is evaluated and ranked separately. Examples of univariate filters are the t-test, the $\chi 2$ test, or the Wilcoxon total rank [15, 16]. These approaches are

relatively fast because the computational complexity is linear based on the number of features. The main limitation of these methods is the lack of consideration for dependencies between features. To deal with this limitation, multivariate methods have been proposed that have the ability to separate groups of features and evaluate them together. These techniques are slower but more scalable than univariate methods (in this family, too, the number of possible subsets that can be extracted increases exponentially as the number of features increases). This group of methods does not depend on classification process at all.

Examples of multivariate feature ranking approaches are Markov coating filters [17, 18] and fast-correlation and correlation-based methods [19]. While filtering techniques identify the best features independently of the model selection stage, wrapping methods combine the model selection stage with a feature subset search. In fact, the goodness of each feature set is assessed by training and testing a specific classification model. In this method, the feature selection process is strongly related to the selected classifier, and compared to filtering methods, these methods are computationally more expensive and have a higher risk of over-fitting, but can generally achieve higher accuracy because Practically, from the beginning, the features are selected based on the same final criterion (classification accuracy).

Greedy forward selection and backward elimination strategies are two popular examples of wrapping selection approaches [20, 21]. Embedded techniques for feature selection methods look for the optimal subset of features based on the classification accuracy. These methods search in the hybrid space of feature subsets and hypotheses. Like wrapping approaches, embedded techniques are strongly related to the classifier and the learning algorithm. Compared to the wrapper methods, they are less computationally compact. Examples of these applications are the use of random forest (RF) internal criteria such as mean reduction accuracy and Gini index or feature selection based on support vector machine (SVM) weights. Both univariate and multivariate approaches have the common goal of finding the smallest set of useful features to classify objects correctly. Accuracy and stability are the two main requirements for selecting a feature selection method. In last two decade, the main challenge in this area has been to find high-precision methods to improve discriminative potential of selected features.

### 2.1. Feature Selection to Find the Biomarker of Diseases

Identification of biomarkers for a disease is an important research area in bioinformatics. In most methods, they have to find a sub-set of features in different phases. Fortino et al., [22] proposed an envelope feature selection method that combines stochastic forest and fuzzy logic and offers a stable performance compared to other methods in this field. The proposed algorithm in [22] consists of three steps:

- Gene expression data is converted into fuzzy patterns (FPs). This phase is called discretization.
- Second phase, prior knowledge of FPs is used to train random forest and classify data.
- In the third phase, the selected features are ranked by a permutation variable significance criterion.

The method presented in this paper has been evaluated on several multi-class sets of gene expression and compared with several other RF-based methods [23]. Two criteria F and G (in 30 epochs) have been considered to evaluate the efficiency of the proposed method, which are especially suitable for unbalanced multi-class classification problems [24]. Performing iterations, subsets of efficient features have been selected, and the ratio between the number of compatible features and the total number of selected features has been considered as the final evaluation metric. Results show that the proposed system in [24] has similar or better results than other compared methods.

### 2.2. Single Nucleotide Polymorphisms Analysis Methods

SNPs are single-position nucleotide mutations that are produced by evolution and are inherited. The most genetic differences between individuals are SNPs. They play an important role in many studies related to disease based on genes. Because of the large number of SNPs in the genome [25], it is important to identify relevant subtypes of SNPs that are sufficiently discriminant to differentiate between patients. So far, several different methods have been proposed for selecting SNPs based on different criteria. For example, Charlon et al. [26] proposed a PCA-based algorithm to identify SNPs that are effective in diagnosing autoimmune diseases. Using several transform functions, they were able to maintain SNPs with the largest absolute prediction values in the 100 major components. A Gaussian mixture distribution function is set up for each one of the major component. The probability that each SNP belongs to each of the Gaussian functions is calculated, and based on the probability range obtained; the classification uncertainty can be calculated. Only strong participants have zero uncertainty, so SNPs with empty category uncertainty are selected. Researchers in [26] were able to classify SNPs with good accuracy. They were also able to reduce the effect of ancestral information, which is known as one of the main sources of genetic variation among individuals.

### 2.3. Independent Component Analysis for Brain Networks

As mentioned above, in functional neuroimaging, due to the high dimensional of the data, the use of dimensionality reduction methods is essential. ICA methods are often applied to fMRI data with the aim of extracting spatially independent signal sources and detecting noise components [27-29].

Given the time series associated with brain voxels, which are often represented by matrix structure, the ICA identifies a number of independent sources whose combined contributions produce the fMRI signal (Figure 1). Depending on the need for independent spatial patterns or time periods, we can distinguish between spatial and temporal ICAs. But spatial ICA is used a lot because there are more voxels in it. One of the most relevant applications of ICA is the identification of resting-state networks, that is, areas of the brain that are functionally connected when the brain is not engaged in a specific task. Six main networks can be distinguished: the default mode network, the visual network, the fronto-parietal network, the sensori-motor network, the auditory network, and the self-referential network [30]

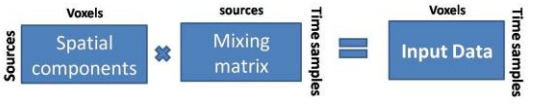


**Fig. 1.** Component analysis process on FMRI data.

### 2.4. Embedding fMRI Data in Low-Dimensional Space

In 2020, Shen et al., [31] proposed a machine learning-based approach to the analysis of resting-state functional connection patterns in schizophrenia. The method presented in [31] uses several hybrid steps including a feature selection algorithm, dimensional embedding algorithm and self-organizing clustering.

In the proposed method, resting functional networks are considered as points distributed in a high-dimensional feature space, and the aim is to identify the spatiotemporal patterns associated with schizophrenia symptoms that are presumed to be located on a dimensional manifold embedded in the feature.

- In the first step, functional connection networks are extracted from each subject and the correlation coefficient method is used to extract the most distinctive features.
- In the second step, a dimensional step based on local linear embedding is applied [32]
- In the third stage, K-means clustering is applied in a small space to identify two groups (patients versus healthy individuals) and the two obtained clusters are optimized to maximize the classification rate.

The method proposed in [32], provided both high classification accuracy and good generalization ability together. This article is an example of the ability of machine learning to diagnose and evaluate the treatment of schizophrenia.

### 2.5. Feature Selection in Sequence Analysis

As mentioned in the introduction, sequence analysis of genomes and proteins is an important issue in bioinformatics. Usually, two types of challenges can be identified in terms of feature selection: content analysis and signal analysis.

Content analysis can be performed on the broad features of a sequence, such as the tendency to encode in a protein sequence or to perform a specific biological function. But signal analysis focuses on identifying important motifs in the sequence, such as gene structural elements or regulatory elements. Apart from the basic properties that show only nucleotides or amino acids in each position in a sequence, many other properties can be derived, such as the higher order compositions of these building blocks, the number of which increases exponentially. Since many of them are unrelated or redundant, attribute selection techniques are applied to focus on the effective subset of variables

.

#### 2.5.1. Content analysis

Predicting protein sequences has long been an important issue in bioinformatics. Because many properties can be derived from a sequence, and most dependencies occur between adjacent situations, Markov models are widely used. To deal with the large number of features and the number of samples, the Markov model of interpolation (IMM) was introduced by Delcher et al. [33], which used interpolation between different levels of the Markov model to deal with small samples. In another study, Delcher et al. [33] extended the IMM framework to deal with non-adjacent feature dependencies, leading to the Interpolation Tissue Model (ICM), which uses the Bayesian decision tree by filtering crosses.

Another class of methods focuses on predicting protein function by sequencing. One of the influential articles in this group is [34] which combines a genetic algorithm in combination with a gamma test to score feature subsets to classify large rRNA subunits. An interesting technique [35] using selective core scaling for support vector machines (SVMs) is presented as a way to evaluate feature weights and eliminate low weight features. In last years, new applications are defined in the sequence analysis field which uses variation feature selection (FS) techniques such as promoter regions [35] and microRNA recognition [36].

### 2.5.2. Signal analysis

Many sequence analysis methods involve identifying short or protected signals in the sequence. These type of signals mainly indicate the binding sites of different proteins or protein complexes. A common approach to finding regulatory motifs is to relate the motifs to gene expression levels using regression techniques. You can then use the feature selection to search for motifs that best fit the regression model [37-38]. In [39], a classification-based approach has been proposed to find differentiation patterns. This method is inspired by the paper [40] which uses the wrong classification threshold number (TNoM, in the section on microarray analysis) to score genes for correlation with tissue classification. From the TNoM score, a value of P is calculated, which indicates the importance of each motif. The motifs are then sorted by P-value

Another part of the research is done on gene prediction settings. In these cases, structural elements such as translation starting point (TIS) and link locations can be considered as different classification problems. The issue of feature selection for the recognition of structural elements was raised in [41]. For the junction prediction problem, combining a sequential back method with an embedded SVM evaluation criterion was used to evaluate the feature association.

## 3. CLUSTERING

The term clustering in machine learning means cluster analysis of unsupervised data that is able to identify structures of data without prior knowledge of their distribution. Clustering is a strategy used to group a set of objects in such a way that the patterns of a group (called a cluster) are more similar to each other than the patterns in other clusters. The results of clustering strongly depend on the similarity criterion adopted. Choosing the best similarity criterion is just one of the considerations in this area. So far, many different clustering algorithms have been proposed. Here we review some of the most popular and practical algorithms in the field of biomedicine and review examples of applications.

### 3.1. Portioning Clustering

Some clustering algorithms use partitioning strategy. In all of them, the main goal is to obtain a partition of the data where each point belongs to a unique cluster. Hartigan and Wong [42] introduced one of the most well-known clustering algorithms, known as K-means. If we assume that X = {x1,…., Xn} is a set of N points in a multidimensional space and K is an integer value, then the K-means method try to find a set of K vectors $\mu_k$, which minimize the within cluster sum of squares (WCSS).

$$WCSS = \sum_{h=1} k \sum_{x_i \in c_h} d(x_i, \mu_h) \quad (1)$$

In the above equation, $C_h$ represents the cluster h and consequently $\mu_h$ is the center of the cluster h. d(x, y) also shows the distance between the two vectors x and y. The K-means algorithm works well on some practical problems. The K-means method provides better performance in cases where the clusters are produced in a hyper-spherical shape and their variance is less. Its main limitations are the possibility of getting stuck in local minima during the optimization process, sensitivity to initial starting points and sensitivity to noise. Another problem with the K-means algorithm is that the number of clusters is constant during the process. While this parameter is unknown in most databases and should be estimated using clustering analysis. This is a problem for some other partitioning methods as well. Many efforts have been made to overcome the limitations of K-means clustering. Some of these versions are described in detail in [43].

### 3.2. Hierarchical clustering

Hierarchical clustering strategy is one of the most common clustering approaches that are widely used to identify data structures in bioinformatics. Hierarchical methods usually produce a hierarchical tree (called a dendrogram) which includes a hierarchical, nested and related set of partitions. By cutting the dendrogram at a certain level, the partition of that level converts to K number of discontinuous clusters. Depending on the method used to create the cluster distance measurement, the clustering results may be different. For example, single linkage technique merges clusters that have the closest distance between two pairs of samples in each cluster. The complete linkage technique merges the clusters based on the farthest distance between a sample pair. This method is effective for compact clusters with unconventional structures. The center linkage technique calculates the Euclidean distance squared between the cluster centers. However, this method assumes that the data can be represented in Euclidean space. Hierarchical clustering has been widely used in bioinformatics and neural imaging applications. For example, [44] used this method to identify gene patterns that differentiate breast cancer subclasses. Examples of applications of hierarchical clustering in gene expression analysis are discussed in [44]. In the field of neuroimaging, hierarchical clustering has been used to measure binding in fMRI resting state data [45] and to construct a brain atlas [46]. A hierarchical method for sequencing microbiomes including RNA and DNA [47] is also proposed.

### 3.3. Combinational Methods

Combined methods usually provide an algorithm that has a hierarchical structure but in some cases partitioning is done and the whole data is seen in a unique space. This group of algorithms is also used in bioinformatics due to their combined nature. For example, it has been used to identify molecular properties [48], clustering gene partners, discovering molecular pathways in PPI networks [49]. It has been used in neuroinformatics for clustering and quantification of fiber device data [50] and for time series clustering [51].

### 3.4. Clustering based on Density Function

Density-based clustering strategy assumes that clusters are represented by dense regions of points in the data space that are separated by less density regions. The most famous density-based clustering algorithm is the DBSCAN method proposed by Ester et al. [52]. The optimal performance of this algorithm depends on two parameters: the distance threshold (ε) and the minimum number of objects to form a cluster (minPts). In the first step, neighbors are first found anywhere less than ε. In the following, points with more neighbors than minPts are called principal points. . Then, by navigating the adjacent diagram, the components connected to the main points are identified. Finally, if a neighbor cluster is ε, each non-principal point is assigned to the nearest cluster. If the cluster is not a neighbor ε, that point is considered as a noise sample. The DBSCAN method has several advantages over the partitioning strategy: it does not need to specify the number of retrieval clusters, it can find clusters of various shapes (not just Gaussian), and it is resistant to outliers and data samples. In particular, this clustering algorithm is widely used in the fields of bioinformatics and neuroscience [53-55].

### 3.5. Clustering based on Spectral Analysis

Most spectral clustering algorithms in the first step use the spectrum (eigenvalues and eigenvectors) of similarity matrices to reduce the dimensions of the data. Objects are then clustered in a space with lower dimensions. The similarity matrix in this strategy is a matrix in which each element, such as $A_{i,j}$, represents the similarity between samples i and j. The strategy is to compute the associated Laplace matrix and then apply the clustering method only to the corresponding special vectors. One of the most common algorithms in this field has been proposed by Meila and Shi [56] in which clustering is performed on special vectors associated with the highest eigenvalues of the normalized Laplace matrix. Spectral clustering has many applications in bioinformatics. For example, building protein fragments libraries [57], multi-view clustering of patients subtyping [58], DNA methylation study [59] and so on. In the field of neural image analysis, resting state imaging has recently been used to identify biomarkers in autism spectrum disorder. In addition, it has been used to identify time-varying networks for magnetic resonance imaging data [60].

### 3.6. Fuzzy clustering

Unlike most clustering strategies, in fuzzy clustering methods, samples can belong to more than one single cluster. In fact, a membership point is assigned to each data point for each cluster. Therefore, the edge points of a cluster (with a lower degree of membership) may belong to the same cluster to a lesser extent than the center points of the cluster. The most common method of fuzzy clustering is fuzzy c-means algorithm, which acts exactly like K-means, except that it adds membership values to the objective function. Fuzzy clustering has been widely used in gene expression and clustering of gene expression networks [61, 62]. It has also been used in neuroimaging for different issues such as tumor segmentation [63], lesion diagnosis [64], and autoimmune brain segmentation [65].

### 3.7. Sub-space clustering

As mentioned earlier, biomedical data usually are high-dimensional, so analyzing it often has problems such as visualizing data and reducing the accuracy of similarity criteria. Also in such a space, the complete count of all subspaces becomes insoluble. Also, due to the large number of features, it is possible that some features are related and this relationship is not statistically detectable. As a result, clusters may exist in arbitrary dependent subspaces. Subspace-based clustering is an extended version of traditional clustering that locates the search for the most informative dimensions. It also allows clusters to exist in multiple subspaces (overlapping). One of the main limitations of this strategy is that as the number of dimensions increases, the sub-spaces may increase exponentially by a factor of 2. Hence, some innovative algorithms have been developed that try to solve this problem by using the downward closure technique to create higher dimensional subspaces. Examples of these algorithms are CLIQUE [66] and SUBCLU.

### 3.8. Application of Clustering in Clustering

#### 3.8.1. Co-Expressed genes identification based on clustering techniques

The most widely used field of clustering in bioinformatics is its use in grouping genes in expression data. Gene expression is the process by which information encoded in genes is converted into active functional structures in the cell. Some evidence of gene activation has been presented in [1]. This activation is measured, for example, in next-generation microarray or sequencing (NGS) experiments. In

microarray articles, the expression value of thousands of genes is obtained in a set of samples [67], while in the NGS method, the whole genome is scanned, which allows new cases to be identified, but with a higher computational load than the previous type [68]. From this type of data, information of coexpressed genes can be extracted using clustering techniques. This is an example of a clustering application in which genes with the same expression level in all samples are grouped in a cluster.

For the first time since the presentation of the paper by Alizadeh et al. [69], hierarchical clustering has been widely used in gene expression clustering. In a hierarchical strategy, there is no need for a predefined number of clusters to select. It is useful for visualization purposes because it computes a complete hierarchy of data displayed as a dendrogram [69]. It is then possible to determine the division of the beds in the clusters by cutting the dendrogram at a certain depth. The choice of height can be arbitrary and the efficiency of clustering accuracy varies according to its value.

Another approach called PVClust [70] is proposed to solve this problem.In this paper, a classical hierarchical clustering approach is proposed that can evaluate the uncertainty in the analysis. For each cluster, it evaluates the value of p, which indicates how much the cluster is supported by the data. Pvclust is a freely available R package and it has been widely applied in many bioinformatics applications [71-73].

Classical clustering methods, such as K-means, has also been widely used [74-76]. The main advantage of K-means is that it is simple and fast, but, unlike hierarchical clustering, K-means requires an initial determination of the number of clusters. Because the number of gene clusters is usually unknown in advance, this is considered a limitation. To identify the optimal number of clusters, K-means are usually performed for different values of k and then the clustering results are compared [77].

Each clustering algorithm that is applied to different datasets can offer different performance, so there is no definite choice between the algorithms described in this section. For example, if the data set contains numeric values and the number of clusters is known, K-means or SOM may work better than other methods. For high-noise gene expression datasets without any prior knowledge of the number of clusters, CAST or CLICK algorithms may work best.

In all clustering methods, the clusters must be validated when they are obtained independently of the applied algorithm. In the validation process, the quality and reliability of the clusters are evaluated. Clustering validation can be done in terms of homogeneity, when objects in the same cluster are closer to each other than those in different clusters [78]. In addition, the cloning of gene expression can be examined from a biological perspective. For example, Tavazoie, et al., [79] created a mapping of genes in each cluster resulting in 199 known functional categories. For each cluster, p values were calculated to measure functional group enrichment.

### 3.8.2. Patient subtypes identification based on clustering

In many diseases - for example, cancer, neuropsychiatric disorders and autoimmunity - it is difficult to provide definitive treatment due to the wide variety of symptoms of patients [80]. The exact medical solution tries to solve this problem by personalizing the medical operation. In order to increase the accuracy of predicting disease progression and determining the most appropriate medical treatments, these methods take into account the diversity and differences of people in different terms such as genes, DNA, lifestyle and environment [81]. In precision medicine, patient sub-typing plays an important role. In this field, the main goal is to identify sub-populations of similar patients that can lead to more accurate diagnostic and treatment strategies. Identifying subtypes of diseases helps both medical science and the efficiency of surgery. In fact, from a clinical point of view, correcting the prognosis for similar individuals can reduce the uncertainty about the expected outcome of treatment for each individual. Recently, methods based on data integration approaches for detecting patient subtypes using supervised classification and unsupervised clustering have been proposed [82-84]. To improve the accuracy of the model for patient classification, bioinformatics data can be used, such as miRNA expression, methylation, or changes in RNA copy number and gene expression. For example, somatic copy number changes provide good biomarkers for cancer subclassification [85]. Data integration approaches to effectively identify subgroups among existing examples have recently been considered. The basic idea is based on the assumption that identifying groups of samples that share relevant molecular properties is the most influential factor.

For example, SNF is an intermediate integration network fusion method, which is able to collect multiple genomic properties such as DNA methylation, mRNA expression, and miRNA expression data, to identify relevant patient subtypes.

A late integration methodology for classify patient subtypes in cancer data-sets called MVDA is proposed in [86]. The approach consists of four main phases:

- Step 1: The prototype extracted is a sample in which features are clustered to reduce the size of the data.
- Step 2: prototypes are ranked, based on their class resolution scores.

- Step 3: single-view clustering in each view
- Step 4: The last case is the integration of single-view clustering results with the matrix factorization approach.
-

### 3.8.3. Application of clustering in fMRI data analysis

Functional neuroimaging data include volumetric images of the brain that are extracted over time. This type of data can describe brain activity. Clustering methods on raw time series data are used to identify areas of the brain that have similar functional patterns [87]. However, as the spatial and temporal resolution of the existing data set increases, the efficiency of this approach decreases, which can be measured by the signal-to-noise ratio criterion (this criterion describes raw time series data). In this regard, another method is to use clustering methods to identify structures in the data, after pre-processing the raw time series [88]. An example of this is the use of clustering techniques on spatial maps derived from the analysis of independent components on resting fMRI data [74]. Most methods that follow this approach cluster voxels using the Pearson correlation coefficient. The quality of the generated clusters is low in some issues, so in some methods different partitions are combined in a final clustering to increase the clustering efficiency of individual data.

## 4. CLASSIFICATION

In machine learning science, classification is one of the key issues of supervised-learning methods. In other words, classification means mapping one of the default possible variables to an input instance. Depending on the type of outcome variable (output), supervised learning is divided into two subsets:

- If the output variables are categorized, or it can only assume a finite set of discrete values, the problem is called classification.
- If the result variable can assume values in a continuous range (for example, the amount of water behind a barrier or the level of glucose in the blood), the problem is known as regression.

In this section, the focus will be solely on examining classification models. When only two values are assigned to the output variable (label), it is called a binary classification. Predictive models can belong to two families, namely parametric and non-parametric models. Parametric models assume that the function to be estimated belongs to a method that is described by a finite set of parameters. In this case, the learning is in accordance with the estimation of the parameters that optimally describe the estimated function of the data.

Nonparametric models do not limit the relationship between data set input properties and the output variable to a specific function family. The complexity of nonparametric models is automatically adjusted in the training phase. The following are some of the most popular classification models in bioinformatics issues.

### 4.1. Support Vector Machine as Classifier

One of the most widely used nonparametric learning models used for classification is the support vector machine (SVM). This classifier assumes that the output variable (label) is linearly related to the corresponding input features. Assuming linearity, it means that geometrically in a three-dimensional space or more, a hyper-plane can separate observations of one class from instances of another class. In SVM, a hyper-plane is uniquely defined by an orthogonal vector corresponding to the parameters w (parameters to be estimated). Among the infinite separator line, the SVM model finds a separator hyper-plane that maximizes the potential error margin (Figure 2), which is defined as the smallest distance between each sample and the separator hyper-plane. SVM can also be used as a nonlinear non-parametric classifier. In fact, instead of calculating the linear function f, data can be converted by applying a nonlinear kernel function. This will be possible with the help of the kernel. Initially, SVMs were proposed to solve binary classification problems. In many practical applications, especially in the field of biomedicine, the number of classes is generally more than two classes. These multi-class problems are usually solved by breaking them down into binary sub-problems and building SVM classifiers. At this stage, two main strategies are used: classify one class from the other (one-against-all approach) or classify each pair of classes (one-against-one approach) [89, 90]. Because SVMs can cope well with high-dimensional data, it has been widely used in neuroscience and bioinformatics. For example, in [91], a fuzzy version of SVM with multiple hybrid kernels is proposed to detect DNA-binded proteins. More diverse applications can be studied in [92-93].

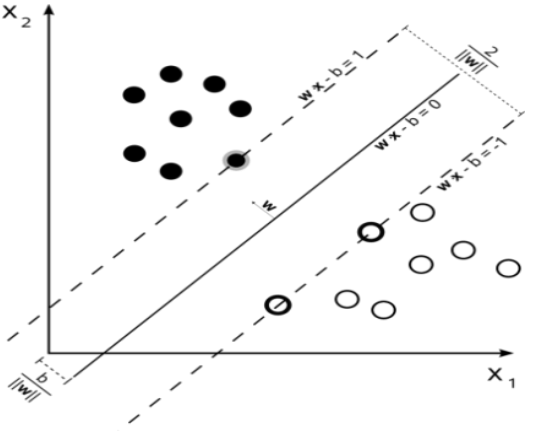


**Fig. 2.** Hyper-plane to maximizing two-class error margin (SVM process).

### 4.2. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a multivariate and parametric statistical learning model. The LDA assumes that the prediction variables must be evenly distributed along the covariance matrices. For this reason, this method is not used in cases where there is no presumption of normalcy or there is outlier data. The LDA is commonly used as a linear classifier. Also it is used to analyze the degree to which variables are more discriminative to distinguish between classes. For example, in neuroscience, the LDA has been used for electroencephalographic (EEG) data analysis to differentiate between healthy elderly. It has also been used to detect Alzheimer's disease early [94]. As another example, an LDA-compliant version has been proposed to classify gene expression data [95-97].

### 4.3. Random forest

Random forest (RF) is an aggregate classifier based on bagging theory. In this classifier, a large set of independent classifiers (decision tree types) are collected to produce a stronger and more accurate classification. The more independent trees used for random forest training, the lower the standard deviation. One of the main advantages of using RF is that the features can be classified based on the average improvement criterion in the purity criterion, which can be considered as an indicator of the relationship between the features for classification. One of the disadvantages of random forest is its sensitivity to input parameters. Random forests have been widely used in both bioinformatics [98] and neural imaging.

### 4.4. Nearest neighbor's Classifier

The nearest neighbor is a non-parametric classifier that performs the following very simple concept in classification process: The result of a prediction for an experimental sample depends on the K labels of its nearest neighbor in the feature space. The Euclidean distance criterion is usually used to identify the nearest neighbor. Class membership is also assigned based on the selected majority tag. The voting scheme does not require any prior knowledge of data distribution so it is considered a lazy category. The choice of K value is up to the user, but setting it can increase the final performance of the system. As a general rule, if the K is too small, the problem will be noise sensitive. But if the K is too large, the probability of selecting examples from other classes (wrong classes) as the nearest neighbors increases. In databases with unbalanced classes, the risk of incorrectly assigning membership to the class is over-sampled. Since now, KNN has been applied in many bioinformatics problems such as: Study protein localization in proteomics [8, 99-101], protein types prediction [99]. For its ability to model the local structure of data, it has also been used to segment brain's texture [102-104].

### 4.5. Rule-based Classifiers

Rule-based classifiers are different type of learning models that, starting from a set of observations, derive rules that identify subgroups of objects. Identification of subgroups is determined based on the features extracted from the samples. Most structural induction rules have the format, IF condition THEN class. In this conditional format, it is a set of attribute-value pairs derived from the attributes that describe the instructional examples. Because of this, they are different from traditional methods. There are basically three categories of law enforcement strategies:

I. Separate-and-conquer strategies: These methods look for a rule that describes the patterns of training data that make up a subset then retrospectively learn more rules until they assign a class to each object.

II. Divide-and-conquer strategies: These are those used by decision trees that create a rule for each path from root to leaf.

III. Comprehensive search strategies: Examine all rules that predict class tags, which can then be filtered using the minimum quality criterion.

Applications of these models include the identification of miRNA regulatory modules [105]. Consider a set of miRNAs and a set of their genetic targets, then the goal is to find the corresponding miRNA subsets and target genes. These modules are commonly known as control modules. In these issues, the main features of miRNA expression profiles and mRNA structure are shown. For each specific gene, a variety of similarity criteria (such as Pearson's correlation coefficient) can be used to calculate similarities between pairs of genes. A threshold is also used to divide a set of genes into two groups of similar and different genes. Next, a segregation and segregation law induction strategy is applied to generate a set of miRNA-mRNA regulatory rules. Finally, only rules containing miRNAs with higher expression rates are preserved. Other applications of rule induction in bioinformatics include disease subtyping [106, 107] and description of gene assemblies [108]. Some efficient approaches for bioinformatics data classification remain briefly in Table 1.

**Table 1.** A summary of ML classifiers applied to gene expression data.

| Ref. | Technique | Data | Accuracy |
|---|---|---|---|
| [109] | Extreme learning machine (ELM) | Central nervous system tumor | 79% |
| [110] | SVM | Breast Cancer (somatic mutation | 69 % |

| | | | |
|---|---|---|---|
| | | profile) | |
| [110] | Tree C4.5 | Breast Cancer (somatic mutation profile) | 60 % |
| [110] | KNN | Breast Cancer (somatic mutation profile) | 49 % |
| [111] | SVM – ANN | Breast Cancer (TCGA samples) | 91.74 % |
| [111] | SVM+RBF | Breast Cancer (TCGA samples) | 94.78% |
| [112] | KNN + SVD | Leukemia cancer (Gene expression) | 92% |
| [112] | KNN + SVD | Colon cancer (Gene expression) | 80% |
| [112] | KNN + EVD) | Breast cancer (Gene expression) | 91% |
| [113] | Recursive feature elimination (RFE) +SVM | Breast Cancer (microarray) | 88.8% |
| [113] | Recursive logistic regression (RLR) +SVM | Breast Cancer (microarray) | 87% |
| [114] | SVM + ICA | Prostate tumor (microarray) | 93 % |
| [115] | Sequential minimal optimization + SVM | Leukemia (gene expression) | 94.11% |
| [116] | Artificial neural network (ANN) | Renal cell cancer (genome) | 89.22% |
| [117] | SVM + feature selection | Breast Cancer (genomic data, RNA sequence) | 82 % |
| [117] | (Naïve Bayes) NGB | Breast Cancer (gene expression ) | 85 % |
| [117] | KNN | Breast Cancer (gene expression ) | 87 % |

## 5. APPLICATION of DEEP LEARNING IN BIOINFORMATICS

Today, deep learning is a large and effective part of a variety of learning methods in the field of artificial intelligence. In a nutshell, deep learning models are composed of several layers in row that are capable of displaying data with a high level of abstraction. The main difference between classical low-deep learning models (such as the perceptron neural network with hidden layers) and deep learning networks is that classical neural networks cannot work with raw data and at least one feature extraction step must be performed before the network learning process is done. But deep learning networks actually receive raw data and act as feature extraction units themselves. As we move forward in layers, each layer extracts more abstract and complex features from its input (which can be raw data such as an image matrix) than the previous layer (Figures 3 and 4).

Since the term deep learning refers to a wide range of techniques, one of the main challenges in deep learning applications is to select the most appropriate model/structure for the intended application. Different models of deep networks can be classified into three categories:

- Supervised learning networks: designed to provide the power of differentiation in classification problems.
- Unsupervised learning networks: designed to identify high-level data correlations.
- Hybrid or semi-regulatory networks: which aim to classify data using the outputs of an unsupervised model, in order to speed up the learning process

One of the main limitations of deep networks is the large number of their parameters, which sometimes complicate the models. These parameters depend on various variables, such as:

- Architectural aspects (such as number of layers or transfer functions)
- Type of optimization (Ex. learning rates and momentum values)
- Regularization type.

In recent years, deep learning networks have been successfully used in a broad range of different applications in bioinformatics and biomedicine [118-121].

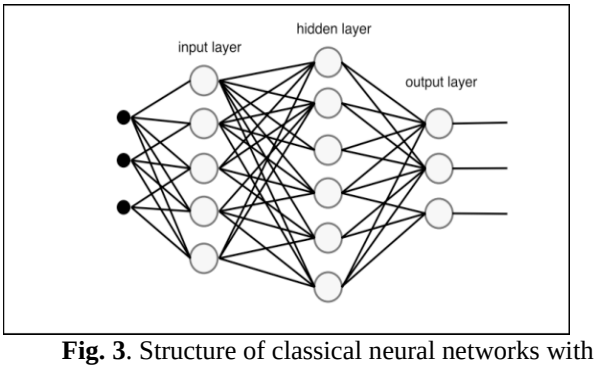


**Fig. 3**. Structure of classical neural networks with one hidden layer.

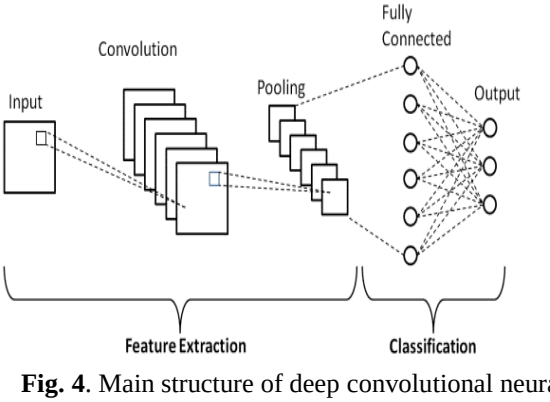


**Fig. 4**. Main structure of deep convolutional neural networks.

### 5.1. CNN Application in RNA-ISH

The term RNA ISH is a technique that can localize and visualize gene expression in a group of cells, in a specific texture, or in an organism [122]. This technique is useful for demonstrating changes in expression patterns during development [123]. This is usually performed by hand, but in recent years, using deep convolutional neural networks (CNN) [124], they have been able to automate the annotation of gene expression patterns. Deep models require a large number of labeled images for training phase. One way to overcome this limitation is to use the transfer learning approach, in which the network is trained on one data set and then used as a feature extractor in other datasets [125-127]. For example, in [124] transfer learning is performed from natural images to ISH images. They used the OverFeat model and trained on the ImageNet dataset. They then generalized the features and finally used them as feature extractors in ISH images. The experimental results show that by using convolutional networks as feature extractors, the accuracy of annotating gene expression patterns at multiple levels of brain structures can be greatly increased. The results in [124] shows that their proposed method provides an average overall accuracy of 0.894, which is higher than the 0.820 accuracy based on the bag of words approach.

### 5.2. Application of deep learning networks CNNs to identify DNA- and RNA-binding protein

One of the most influential processes in developing regulatory biological models and identifying disease types understands the sequence properties of DNA and RNA-binding proteins. Today, sequence properties are completely determined using position weight matrices (PWM). PWMs are easy to interpret and can be easily scanned into a genomic sequence to identify potential binding sites. Many cliché classification models and shallow neural networks have been proposed for this purpose [128-130]. But these models face different problems, including: data production by different technologies, different formats of received data, huge volumes of data that need to be analyzed and specific types of data noise. Alipanahi et al. [131] adopted an in-depth CNN to predict sequence properties and binding scores to address all of these problems. Their proposed method is known as DeepBind. The proposed network provides three points:

a) Derive directly connection properties from input sequence data

b) Discover new motifs

c) Discover the rules needed to combine them into a predicted connection point.

### 5.3. Alzheimer's Disease Diagnosis based on Deep AEs

Deep learning models with much more power than classical neural networks can also analyze and classify more complex patterns. Therefore, in some studies today, deep networks have been used to recognize biomarkers of neurological disorders [118, 121]. Recently, there has been a growing interest in the application of AEs to extract low-dimensional features from several neuroimaging modalities, often used in a multiview fashion. Liu et al. [82] proposed a multilayer neural network (MLP) consisting of several AEs and a soft-max layer is used for the diagnostic of the Alzheimer's disease. MR and PET imaging modalities are fused by jointly training the AEs with the concatenated MR and PET inputs. To avoid neurons that are activated only by one modality, in a pretraining phase, a series of samples is presented to the network where the inputs of one of the modalities are replaced by zeros. In [132], AEs are used to extract hierarchical nonlinear relations between functionally connected regions of the brain, following the idea that the functional organization of the brain is dynamic rather than static.

### 5.4. Convolutional neural networks to analysis brain network

Deep covolutional neural networks (CNN) are commonly used for image analysis because they use the local features of image in the filtering process at each layer. Examples of CNN applications for brain image analysis are available [133-137]. However, CNNs can be used to analyze connectum data to organize the structural / functional areas of the brain. Kawahara et al. [138] proposed a deep structure called BrainNetCNN that is designed to predict the consequences of clinical neurodevelopment**.** Unlike traditional image-based CNNs, BrainNetCNN uses different brain network topologies to create convolutional filters based on edge-to-edge, edge-to-node, and node-to-graph relationships, thus avoiding the need for full connectivity. Some efficient deep learning methods in bioinformatics data analysis are survived briefly in the Table 2.

**Table 2.** A summary of deep learners applied to bioinformatics data.

| Ref. | Technique | Data | Accuracy |
|---|---|---|---|
| [139] | DeepNetii | KIPAN (RNA Seq.) | 75 % |
| [139] | DeepNetii | COAD (RNA Seq.) | 57 % |
| [139] | DeepNetii | BRCA (RNA Seq.) | 65 % |
| [140] | Shallow DNN | Swiss-Prot (Protein classification) | 54.1 % |
| [140] | LSTM | Swiss-Prot (Protein classification) | 78.4 % |
| [141] | Convolutional neural network | TCGA (20 types of cancers) | 78 % |
| [142] | Recurrent neural network(RNN) | GEO (gene expression) | 63.9 % |
| [143] | Integrate CNN | TCGA (mRNA and | 60 % |

| | | | |
|---|---|---|---|
| | | Methylation) | |
| [144] | ResNet | Alzheimer diagnosis (fMRI data) | 90 % |
| [144] | ACNN | Alzheimer diagnosis (fMRI data) | 91 % |
| [145] | Deep CNN (DCNN) | Breast cancer (gene expression + copy number alternatives) | 96 % |
| [146] | SOM + VGG33 | Breast cancer | 89.7 % |
| [146] | SOM + ResNet-112 | Breast cancer | 97 % |
| [146] | t-SNE + ResNet-112 | Breast cancer | 96 % |

## 6. CONCLUSION

Machine learning is one of the most practical branches of artificial intelligence. The main goal in all machine learning techniques is to analyze the data in different formats and identify the structures in the data based on the problem form .The articles reviewed in this article showed that machine learning techniques can be widely used to solve various problems in bioinformatics. Firstly, the advances in high-throughput technologies for the acquisition of biomedical data have created the need of sophisticated methods able to cope with the complexity of big data. Data on genes, proteins, genomes and brain structure are vast. Therefore, various techniques of dimensional reduction and feature selection can be applied to reduce the computational load on this data.In many bioinformatics and neuroscience data, data labeling is not possible. Therefore, a variety of clustering methods can be widely used to identify data sets with common properties .Categorization means attributing data to one of the predefined groups. There is a classification challenge in many branches of bioinformatics, such as identifying patient subtypes, identifying disease types, identifying gene structures, and identifying RNA binding sites to proteins. Therefore, this group of machine learning techniques can be used .Deep learning is one of the new and versatile branches of machine learning that can be used for a variety of tasks such as categorization, feature extraction and regression. For this purpose, deep learning in neuroscience and bioinformatics is used. Despite some of the limitations and disadvantages mentioned in this article, the results presented in relation to numerous articles showed that machine learning approaches can be much more effective in this scope than traditional and statistical computational methods. This has led to the use of new research areas such as multi-faceted learning and deep learning in the field of biomedicine.

## REFERENCES


[1] N. M. Luscombe, D. Greenbaum, M. Gerstein, **"What is bioinformatics? A proposed definition and overview of the field,"** *Methods of Information in Medicine*, Vol. 40, No. 4, pp. 346–358, 2001.

[2] J. C. Venter, M. D. Adams, E. W. Myers, P. Li, R. J. Mural, G. G. Sutton, J. D. Gocayne, **"The sequence of the human genome,"** *Science*, Vol. 291, pp. 1304–1351, 2001.

[3] A. Ansel, J. P. Rosenzweig, P. D. Zisman, M. Melamed, B. Gesundheit, **"Variation in gene expression in autism spectrum disorders: An extensive review of transcriptomic studies,"** *Frontiers in Neuroscience*, Vol. 10, pp. 601-610, 2016.

[4] I. Hunt, **"From gene to protein: A review of new and enabling technologies for multi-parallel protein expression,"** *Protein Expression and Purification*, Vol. 40, No. 1, pp. 1–22, 2005.

[5] T. Ideker, T. Galitski, L. Hood, **"A new approach to decoding life: Systems biology,"** *Annual Review of Genomics and Human Genetics*, Vol. 2, No. 1, pp. 343–372, 2001.

[6] H. Kitano, **"Systems biology: A brief overview,"** *Science*, Vol. 295, No. 5560, pp. 1662–1664, 2002.

[7] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "**Supervised machine learning: A review of classification techniques,"** *In Proc. of the conference on Emerging artificial intelligence applications in computer engineering: real word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pp. 3-24, 2007.

[8] X. Wang, H. Li, Q. Zhang, R. Wang, **"Predicting subcellular localization of apoptosis proteins combining GO features of homologous proteins and distance weighted knn classifier,"** *BioMed Research International*, pp. 1-12, 2016.

[9] M. Hajighorbani, S. M. R. Hashemi, B. Minaei-Bidgoli, S. Safari, **"A review of some semi-supervised learning methods,"** *In proc. of IEEE First International Conference on New Research Achievements in Electrical and Computer Engineering*, Tehran, 2016.

[10] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos, **"A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages,"** *NeuroImage*, Vol. 155, pp. 530–548, 2017.

[11] X. Rui, D. C. Wunsch, **"Clustering algorithms in biomedical research: A review,"** *IEEE Reviews in Biomedical Engineering*, Vol. 3, pp. 120–154, 2010.

[12] Y. Saeys, I. Inza, P. Larrañaga, **"A review of feature selection techniques in bioinformatics,"** *Bioinformatics*, Vol. 23, No. 19, pp. 2507–2517, 2007

[13] A. Hyvärinen, E. Oja, **"Independent component analysis: Algorithms and applications,"** *Neural Networks*, Vol. *13*, No. 4, pp. 411–430, 2000.

[14] I. K. Fodor, **"A survey of dimension reduction techniques,"** *Lawrence Livermore National Lab, Livermore, CA*. pp.1-34, 2002.

[15] Y. Garcia-Chimeno, B. Garcia-Zapirain, M. Gomez-Beldarrain, B. Fernandez-Ruanova, J. C. Garcia-Monco, **"Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data,"** *BMC Medical Informatics and Decision Making*, Vol. 17, No. 1, pp. 38-42, 2017.

[16] N. Shuke, **"Voxel-based control database generated from clinical fdg pet data for statistical analysis of brain fdg pet: Comparison with subject-based normal database,"** *Journal of Nuclear Medicine*, Vol. 58, pp. 1257–1257, 2017.

[17] A. Wang, N. An, J. Yang, G. Chen, L. Li, G. Alterovitz, **"Wrapper-based gene selection with markov blanket,"** *Computers in Biology and Medicine*, Vol. 81, pp. 11–23, 2017.

[18] Z. Zhu, Y. S. Ong, M. Dash, **"Markov blanket-embedded genetic algorithm for gene selection,"** *Pattern Recognition*, Vol. 40, No. 11, pp. 3236–3248, 2007.

[19] D. Heider, N. Genze, U. Neumann, **"Efs: An ensemble feature selection tool implemented as r-package and web-application,"** *BioData Mining*, Vol. 10, No. 1, pp. 21-30, 2017.

[20] H. Polat, H. D. Mehr, A. Cetin, **"Diagnosis of chronic kidney disease based on support vector machine by feature selection methods,"** *Journal of Medical Systems*, Vol. 41, No. 4, pp. 55-59, 2017.

[21] M. S. Hannah-Immanuel, S. G. Jacob, **"Feature selection techniques for Alzheimer's disease: A review,"** *International Journal of Engineering Technology Science and Research*, Vol. 4, No. 7, pp. 494–49, 2017.

[22] V. Fortino, P. Kinaret, N. Fyhrquist, H. Alenius, D. Greco, **"A robust and accurate method for feature selection and prioritization from multi-class omics data,"** *PLoS One*, Vol. 9, pp. e107801, 2014.

[23] R. Díaz-Uriarte, S. A. De Andres, **"Gene selection and classification of microarray data using random forest,"** *BMC Bioinformatics*, Vol. 7, No. 1, pp. 3-12, 2006.

[24] H. Yu, S. Hong, X. Yang, J. Ni, Y. Dan, B. Qin, **"Recognition of multiple imbalanced cancer types based on DNA microarray data using ensemble classifiers,"** *BioMed Research International*, Vol. 2013, pp. 1–13, 2013.

[25] L. Kruglyak, D. A. Nickerson, **"Variation is the spice of life,"** *Nature Genetics*, Vol. 27, No. 3, pp. 234–236, 2001

[26] T. Charlon, M. Martínez-Bueno, L. Bossini-Castillo, F. David Carmona, A. Di Cara, J. Wojcik, M. E. Alarcón-Riquelme, **"Single nucleotide polymorphism clustering in systemic autoimmune diseases,"** *PloS One*, Vol. 11, No. 8, pp. e0160270, 2016

[27] F. De Martino, F. Gentile, F. Esposito, M. Balsi, F. Di Salle, R. Goebel, E. Formisano, **"Classification of fmri independent components using ic-fingerprints and support vector machine classifiers,"** *NeuroImage*, Vol. 34, No. 1, pp. 177–194, 2007

[28] E. Formisano, F. Esposito, N. Kriegeskorte,

[29] M. J. McKeown, S. Makeig, G. G. Brown, T. Jung, S. S. Kindermann, A. J. Bell, T. J. Sejnowski, **"Analysis of fMRI data by blind separation into independent spatial components,"** *Naval Health Research Center, San Diego, CA*, 1997.

[30] G. Tedeschi, F. Di-Salle, R. Goebel, **"Spatial independent component analysis of functional magnetic reso- nance imaging time-series: Characterization of the cortical components,"** *Neurocomputing*, Vol. 49, No. 1, pp. 241–254, 2002.

[31] H. Shen, L. Wang, Y. Liu, D. Hu, **"Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embed- ding of fMRI,"** *NeuroImage*, Vol. 49, No 4, pp. 3110–3121, 2020.

[32] S. T. Roweis, L. K. Saul, **"Nonlinear dimensionality reduction by locally linear embedding,"** *Science*, Vol. 290, No. 5500, pp. 2323–2326, 2000

[33] A. Delcher, **"Improved microbial gene identification with glimmer,"** *Nucleic Acids Research*, Vol. 27, pp. 4636–4641, 1999

[34] N. Chuzhanova, **"Feature selection for genetic sequence classification,"** *Bioinformatics*, Vol. 14, pp. 139–143, 1998

[35] N. Zavaljevsky, **"Support vector machines with selective kernel scaling for protein classification and identification of key amino acid positions,"** *Bioinformatics*, Vol. 18, pp. 689–696, 2002

[36] P. Conilione, D. Wang, **"A comparative study on feature selection for E.coli promoter recognition,"** *Int. J. Inf. Technol.*, Vol. 11, pp. 54–66, 2005

[37] S. Kim, **"miTarget:micro RNA target gene prediction using a support vector machine,"** *BMC Bioinformatics*, Vol. 7, pp. 411, 2006

[38] G. Tedeschi, F. Esposito, **"Neuronal networks observed with resting state functional magnetic resonance imaging in clinical populations,"** *In Bright P. Edition, 2012*

[39] S. Sinha, **"Discriminative motifs,"** *Journal Computer Biology*, Vol. 10, pp. 599–615, 2019

[40] A. Ben-Dor, **"Tissue classification with gene expression profiles,'** *Journal of Computer Biology*, Vol. 7, pp. 559–584, 2000

[41] S. Degroeve, **"Feature sub set selection for splice site prediction,"** *Bioinformatics*, Vol. 18, No. 2, pp. 75–83, 2002

[42] J. A. Hartigan, M. A. Wong, **"Algorithm AS 136: A K-means clustering algorithm,"** *Journal of the Royal Statistical Society*, Vol. 28, pp. 100–108, 1979

[43] R. Xu, D. Wunsch-II, **"Clustering Hoboken, NJ:,"** *IEEE access*, 2009

[44] T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, **"Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications,"** *In Proc. of the National Academy of Sciences*, Vol. 98, No. 19, pp. 10869–10874, 2001

[45] D. Cordes, V. Haughton, J. D. Carew, K. Arfanakis, K. Maravilla, **"Hierarchical clustering to measure connectivity in fmri resting-state data,"** *Magnetic Resonance Imaging*, Vol. 20, No. 4, pp. 305–317, 2002

[46] T. Blumensath, S. Jbabdi, M. Glasser, D. C. Van Essen, K. Ugurbil, T. E. Behrens, S. M. Smith, **"Spatially constrained hierarchical parcellation of the brain with resting-state fmri,"** *NeuroImage*, Vol. 76, pp. 313–324, 2013

[47] Z. Quan, L. Gang, J. Xingpeng, L. Xiangrong Z. Xiangxiang, **"Sequence clustering in bioinformatics:**

**an empirical study,"** *Briefings in Bioinformatics*, Vol. 21, No. 1, pp. 1–10, 2020

[48] H. Y. Lin, **"Gene discretization based on em clustering and adaptive sequential forward gene selection for molecular classification,"** *Applied Soft Computing*, Vol. 48, pp. 683–690, 2016

[49] E. Segal, H. Wang, D. Koller, **"Discovering molecular pathways from protein interaction and gene expression data,"** *Bioinformatics*, Vol. 19, pp. 264–272, 2003

[50] Maddah, M., L. Grimson, S. K. Warfield, W. M. Wells, **"A unified framework for clustering and quantitative analysis of white matter fiber tracts,"** *Medical Image Analysis*, Vol. 12, No. 2, pp. 191–202, 2008

[51] S. Rani, G. Sikka, **"Recent techniques of clustering of time series data: A survey,'** *International Journal of Computer Applications*, Vol. 52, No. 15, pp. 1–9, 2012

[52] M. Ester, H. P. Kriegel, J. Sander, X. Xu, **"A density-based algorithm for discovering clusters in large spatial databases with noise,"** *Knowledge Discovery and Data Mining*, Vol. 96, pp. 226–231, 1996

[53] P. Galdi, M. Fratello, F. Trojsi, A. Russo, G. Tedeschi, R. Tagliaferri, F. Esposito, **"Consensus-based feature extraction in rs-fmri data analysis,"** *Soft Computing*, pp. 1–11, 2017

[54] W. Poole, K. Leinonen, I. Shmulevich, T. Knijnenburg, B. Bernard, **"Multiscale mutation clustering algorithm identifies pan-cancer mutational clusters associated with pathway-level changes in gene expression,"** *PLoS Computational Biology*, Vol. 13, No. 2, pp. e1005347, 2017

[55] F. Pennacchietti, S. Vascon, T. Nieus, C. Rosillo, S. Das, S. K. Tyagarajan, F. Cella-Zanacchi, **"Nanoscale molecular reorganization of the inhibitory postsynaptic density is a determinant of gabaergic synaptic potentiation,"** *Journal of Neuroscience*, Vol. 37, No. 7, pp. 1747–1756, 2017

[56] M. Meila, J. Shi, **"Learning segmentation by random walks,"** *Advances in neural information processing systems*, pp. 873–879, 2001

[57] W. Elhefnawy, M. Li, J. Wang, Y. Li, **"Construction of protein backbone fragments libraries on large protein sets using a randomized spectral clustering algorithm,"** *In Proc. international symposium on bioinformatics research and applications*, Switzerland: pp. 108–119, 2017.

[58] Y. Zhang, H. Xiaohua, X. Jiang, **"Multi-view clustering of microbiome samples by robust similarity network fusion and spectral clustering,"** *ACM Transactions on Computational Biology and Bioinformatics*, Vol. 14, No. 2, pp. 264–271, 2017

[59] N. C. Sheffield, G. Pierron, J. Klughammer, P. Datlinger, A. Schönegger, M. Schuster, E. M. Tomazou, **"DNA methylation heterogeneity defines a disease spectrum in ewing sarcoma,"** *Nature Medicine*, Vol. 23, No. 3, pp. 386–395, 2017

[60] I. Cribben, Y. Yi, **"Estimating whole-brain dynamics by using spectral clustering,"** *Journal of the Royal Statistical Society: Series C: Applied Statistics*, Vol. 66, No. 3, pp. 607–627, 2017

[61] A. K. Alok, S. Saha, A. Ekbal, **"Semi-supervised clustering for gene-expression data in multiobjective optimization framework,"** *International Journal of Machine Learning and Cybernetics*, Vol. 8, No. 2, pp. 421–439, 2008

[62] Q. Wang, G. Chen, **"Fuzzy soft subspace clustering method for gene co-expression network analysis,"** *International Journal of Machine Learning and Cybernetics*, Vol. 8, No. 4, pp. 1157–1165, 2017

[63] P. Manocha, S. Bhasme, T. Gupta, B. K. Panigrahi, T. K. Gandhi, **"Automated tumor segmentation and brain mapping for the tumor area,"** *arXiv preprint:*1710.11121, pp. 1-14, 2017

[64] V. Kinani, J. Marie, A. J., Silva, F. G. Funes, D. M. Vargas, E. R. Díaz, A. Arellano, **"Medical imaging lesion detection based on unified gravitational fuzzy clustering,"** *Journal of Healthcare Engineering*, Vol. 2017, pp. 1-16, 2017

[65] U. Vercelli, M. Diano, T. Costa, A. Nani, S. Duca, G. Geminiani, F. Cauda, **"Node detection using high-dimensional fuzzy parcellation applied to the insular cortex,"** *Neural Plasticity*, Vol. 2016, pp. 1–8, 2016

[66] R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan, **"Automatic subspace clustering of high dimensional data,"** *Data Mining and Knowledge Discovery*, Vol. 11, No. 1, pp. 5–33. 2005

[67] J. Quackenbush, **"Computational analysis of microarray data,"** *Nature Reviews Genetics*, Vol. 2, No. 6, pp. 418-427, 2001

[68] Z. Wang, M. Gerstein, M. Snyder, **"RNA-Seq.: A revolutionary tool for transcripttomics,"** *Nature Reviews Genetics*, Vol. 10, No. 1, pp. 57–63, 2009

[69] A. A. Alizadeh, M. B. Eisen, R. Eric-Davis, C. Ma, I. S. Lossos, A. Rosenwald, **"Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling,"** Nature, Vol. 403, No. 6769, pp. 503–511, 2000

[70] R. Suzuki, H. Shimodaira, **"Pvclust: An R package for assessing the uncertainty in hierarchical clustering,"** *Bioinformatics*, Vol. 22, No. 12, pp. 1540–1542, 2006

[71] N. J. Borg, M. Mitchell, P. Lukacs, C. Mack, L. P. Waits, P. R. Krausman, **"Behavioral connectivity among bighorn sheep suggests potential for disease spread,"** *The Journal of Wildlife Management*, Vol. 81, No. 1, pp. 38–45, 2017

[72] A. H. Ellebedy, K. J. Jackson, H. T. Kissick, H. Nakaya, C. Davis, K. Roskin, R. Ahmed, **"Defining antigen-specific plasmablast and memory b cell subsets in human blood after viral infection or vaccination,"** *Nature Immunology*, Vol. 17, No. 10, pp. 1226–1234, 2016

[73] C. Esnault, F. Gualdrini, S. Horswell, G. Kelly, A. Stewart, P. East, R. Treisman, **"Erk-induced activation of tcf family of srf cofactors initiates a chromatin modification cascade associated with transcription,"** *Molecular Cell*, Vol. 65, No. 6, pp. 1081–1095, 2016

[74] P. Galdi, F. Napolitano, R. Tagliaferri, **"Consensus clustering in gene expression,"** *In Proc. Of International meeting on computational intelligence methods for bioinformatics and biostatistics*, pp. 57–67, Switzerland, 2014

[75] S. K. Pal, S. S. Ray, A. Ganivada, **"Gene function analysis,"** *In Proc. of Granular neural networks, pattern recognition and bioinformatics*, pp. 163–193). Cham, Switzerland, 2017

[76] R. C. Rennert, R. Schäfer, T. Bliss, M. Januszyk, **"High-resolution microfluidic single-cell transcriptional profiling reveals clinically relevant subtypes among human stem cell populations commonly utilized in cell-based therapies,"** *Frontiers in Neurology*, Vol. 7, pp. 41, 2016

[77] D. T. Pham, S. S. Dimov, C. D. Nguyen, **"Selection of k in k-means clustering,"** *Journal of Mechanical Engineering Science*, Vol. 219, No. 1, pp. 103–119, 2005

[78] R. Sharan, R. Shamir, **"Click: a clustering algorithm with applications to gene expression analysis,"** *In Proc. of the 8th International Conference on Intelligent Systems for Molecular Biology*, Vol. 16, 2000

[79] S. Tavazoie, J. D. Hughes, M. Campbell, R. Cho, G. M. Church, **"Systematic determination of genetic network architecture,"** *Nature Genetics*, Vol. 22, No. 3, pp. 281–285, 1999

[80] S. Saria, A. Goldenberg, **"Subtyping: What it is and its role in precision medicine,"** *IEEE Intelligent Systems*, Vol. 30, No. 4, pp. 70–75, 2015

[81] R. Mirnezami, J. Nicholson, A. Darzi, **"Preparing for precision medicine,"** *New England Journal of Medicine*, Vol. 366, No. 6, pp. 489–491, 2012

[82] G. Liu, C. Dong, L. Liu, **"Integrated multiple omics data reveal subtypes of hepatocellular carcinoma,"** *PLoS One*, Vol. 11, No 11, pp. e0165457, 2016

[83] E. Taskesen, S. M. Huisman, A. Mahfouz, J. Krijthe, M. Reinders, **"Pan-Cancer subtyping in a 2D-map shows substructures that are driven by specific combinations of molecular characteristics,"** *Scientific Reports*, Vol. 6, pp. 109–128, 2016.

[84] R. Higdon, R. K. Earl, L. Stanberry, C. Hudac, E. Montague, E. Stewart, **"The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders,"** *Omics: A Journal of Integrative Biology*, Vol. 19, No. 4, pp. 197–208, 2015

[85] K. V. Nielsen, B. Ejlertsen, S. Møller, H. T. Mouridsen, **"The value of top2a gene copy number variation as a biomarker in breast cancer: Update of dbcg trial 89d,"** *Acta Oncologica*, Vol. 47, No. 4, 725–734, 2008

[86] C. Wang, S. M. Armasu, K. R. Kalli,G. L. Keeney, E. L. Goode, **"Pooled clustering of high-grade serous ovarian cancer gene expression leads to novel consensus subtypes associated with survival and surgical outcomes,"** *Clinical Cancer Research*, Vol. 23, No. 15, pp. 4077–4085, 2017

[87] C. Goutte, P. Toft, E. Rostrup, F. Nielsen, L. K. Hansen, **"On clustering fMRI time series,"** *NeuroImage*, Vol. 9, No. 3, pp. 298–310, 1999

[88] B. Thirion, O. Faugeras, **"Feature characterization in fMRI data: The information bottleneck approach,"** *Medical Image Analysis*, Vol. 8, No. 4, pp. 403–419, 2004

[89] K. Duan, S. S. Keerthi, **"Which is the best multiclass svm method? An empirical study,"** *Multiple Classifier Systems*, Vol. 3541, pp. 278–285, 2005

[90] C. W. Hsu, C. J. Lin, **"A comparison of methods for multiclass support vector machines,"** *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, pp. 415–425, 2002

[91] Y. Zou W. Hongjie G. Xiaoyi L. Peng, **"MK-FSVM-SVDD: A Multiple Kernel-based Fuzzy SVM Model for Predicting DNA-binding Proteins via Support Vector Data Description, Current,"** *Bioinformatics*, Vol. 16, No. 2, 2021, pp. 274-283, 2021

[92] L. Zhou, L. Wang, L. Liu, P. Ogunbona, D. Shen, **"Support vector machines for neuroimage analysis: Interpretation from discrimination,"** *In Proc. of International conference on Support Vector Machines Applications*, Switzerland, pp. 191–220, 2014

[93] E. Byvatov, G. Schneider, **"Support vector machine applications in bioinformatics,'** *Applied Bioinformatics*, Vol. 2, No. 2, pp. 67–77, 2003

[94] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, T. Eichele, **"Regularized linear discriminant analysis of eeg features in dementia patients,"** *Frontiers in Aging Neuroscience*, Vol. 8, pp. 273-284, 2016

[95] Y. Guo, T. Hastie, R. Tibshirani, **"Regularized linear discriminant analysis and its application in microarrays,"** *Biostatistics*, Vol. 8, No. 1, pp. 86–100, 2006

[96] D. Huang, Y. Quan, M. He, B. Zhou **"Comparison of linear discriminant analysis methods for the classification of cancer based on gene expression data,"** *Journal of Experimental & Clinical Cancer Research*, Vol. 28, No. 1, pp. 149-153, 2009

[97] L. Li, J. Meng, J. Zhang, Y. S. Luan, X. He, Y. F. Zhu, **"Parallel gene selection and dynamic ensemble pruning based on affinity propagation,"** *Computers in Biology and Medicine*, Vol. 87, pp. 8–21, 2017

[98] Y. Qi, **"Random forest for bioinformatics,"** *In Proc. Ensemble machine learning*, pp. 307–323, 2012

[99] K. C. Chou, H. B. Shen, **"Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic k-nearest neighbor classifiers,"** *Journal of Proteome Research*, Vol. 5, No. 8, pp. 1888–1897, 2006

[100] L. A. Xijie, A. M. Moses, **"An unsupervised knn method to systematically detect changes in protein localization in high-throughput microscopy images,"** *PloS One*, Vol. 11, No. 7, pp. e0158712, 2016

[101] P. Horton, K. J. Park, T. Obayashi, N. Fujita, H. Hajime, C. J. Adams-Collier, K. Nakai, 2007

[102] H. A. Vrooman, C. Cocosco, F. Van-der-Lijn, R. Stokking, **"Multi-spectral brain tissue segmentation**

**using automatically trained k-nearest-neighbor classification,"** *NeuroImage*, Vol. 37, No. 1, pp. 71–81, 2007

[103] A. Ghayoor, J. S. Paulsen, R. E. Kim, H. J. Johnson, **"Tissue classification of large-scale multi-site mr data using fuzzy k-nearest neighbor method. SPIE Medical Imaging,"** *International Society for Optics and Photonics*, Vol. 97841, 2016

[104] M. Havaei, P. M. Jodoin, H. Larochelle, **"Efficient interactive brain tumor segmentation as within-brain knn classification,"** *In Proc. of 22nd International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, pp. 556–561, 2014

[105] D. H. Tran, K. Satou, H. T. Bao, **"Finding microrna regulatory modules in human genome using rule induction,"** *BMC Bioinformatics*, Vol. 9, No. 12, pp. 5-15, 2008

[106] G. Huang, K. Huang, T. Lee, J. T. Weng, **"An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients,"** *BMC Bioinformatics*, Vol. 16, No. 1, pp. 1-15, 2015

[107] D. Cangelosi, F. Blengio, R. Versteeg, A. Eggert, A. Garaventa, C. Gambini, L. Varesio, **"Logic learning machine creates explicit and stable rules stratifying neuroblastoma patients,'** *BMC Bioinformatics*, Vol. 14, No. 7, pp. 1-12, 2013

[108] A. Gruca, M. Sikora, **"Data-and expert-driven rule induction and filtering framework for functional interpretation and description of gene sets,"** *Journal of Biomedical Semantics*, Vol. 8, No. 1, pp. 1-23, 2017

[109] C. A. Kumar, S. Ramakrishnan, **"Binary classification of cancer microarray gene expression data using extreme learning machines,"** *In Proc. of IEEE International Conference on Computational Intelligence Computer*, pp.1-4, 2014

[110] S. Vural, X. Wang, C. Guda, **"Classification of breast cancer patients using somatic mutation profiles and machine learning approaches,"** *BMC Syst. Biol.*, Vol. 10, pp. 264-276, 2016

[111] P. Danaee, R. Ghaeini, D. Hendrix, **"A deep learning approach for cancer detection and relevant gene identification,"** *In Proc. of Pacific Symp. Bio comput.* pp.219-229, 2017

[112] S. Tarek, E. Abd, M. Shoman, **"Gene expression-based cancer classification,'** *Egyptian Information Journal*, Vol. 18, No. 3, pp. 151-159, 2017

[113] S. Turgut, M. Dagtekin, T. Ensari, **"Microarray breast cancer data classification using machine learning methods,"** *In Proc. of Electric Electron., Compute. Sci. Biomed. Eng. Meeting (EBBT)*, pp.1-3, 2018

[114] R. A. Musheer, C. K. Verma, N. Srivastava, **"Novel machine learning approach for classification of high-dimensional microarray data,"** *Soft Computing*, Vol. 23, No. 24, pp. 13409-13421, 2019\

[115] K. KR, A. R. KV, A. Pillai, **"An Improved Feature Selection and Classification of Gene Expression Profile using SVM,"** *In Proc. of 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Kannur, India, 2019

[116] N. Y. Simsek, B. Haznedar, C. Kuzudisli, **"A comparitie study of different algorithms on RNA-Seq cancer data,'** *New Trends and Issues Proceedings on Advances in Pure and Applied Sciences*. Vol. 12, pp. 24– 35, 2020

[117] J. Wu, C. Hicks, **"Breast cancer type classification using machine learning,"** *Journal of personalized medicine*, Vol. 11, pp. 1-12, 2021

[118] S. M. Plis, D. Hjelm, V. D. Calhoun, **"Deep learning for neuroimaging: A validation study,"** *Frontiers in Neuroscience*, Vol. 8, pp. 229, 2014

[119] C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle, **"Deep learning for computational biology,"** *Molecular Systems Biology*, Vol. 12, No. 7, pp. 878-883, 2016

[120] P. Mamoshina, A. Vieira, E. Putin, A. Zhavoronkov, **"Applications of deep learning in biomedicine,"** *Molecular Pharmaceutics*, Vol. 13, No. 5, pp. 1445–1454, 2016

[121] S. Vieira, W. H. Pinaya, A. Mechelli, **"Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications,"** *Neuroscience & Biobehavioral Reviews*, Vol. 74, pp. 58–75, 2017

[122] J. Bayani, J. A. Squire, **"Fluorescence in situ hybridization (FISH),"** *Current Protocols in Cell Biology*, Vol. 22, pp. 1–52, 2004

[123] C. L. Thompson, L. Ng, V. Menon, A. Jones, **"A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain,"** Neuron, Vol. 83, No. 2, pp. 309–323, 2014

[124] T. Zeng, R. Li, R. Mukkamala, J. Ye, S. Ji, **"Deep convolutional neural networks for annotating gene expression patterns in the mouse brain,"** *BMC Bioinformatics*, Vol. 16, No. 1, pp. 147-155, 2015

[125] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, **"Decaf: A deep convolutional activation feature for generic visual recognition,"** *In Proc. International Conference on Machine Learning*, Beijing, China, pp. 647–655, 2014

[126] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, **"Cnn features off-the-shelf: An astounding baseline for recognition,"** *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, USA, pp. 806–813, 2014

[127] M. Oquab, L. Bottou, I. Laptev, J. Sivic, **"Learning and transferring mid-level image representations using convolutional neural networks,"** *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1717–1724, 2014

[128] R. Rohs, X. Jin, S. West, R. Joshi, B. Honig, R. S. Mann, **"Origins of specificity in protein-dna recognition,'** *Annual Review of Biochemistry*, Vol. 79, pp. 233–269, 2010

[129] H. Kazan, D. Ray, E. T. Chan, T. R. Hughes, Q. Morris, **"RNAcontext: A new method for learning the sequence and structure binding preferences of**

**rna-binding proteins,'** *PLoS Computational Biology*, Vol. *6*, No. 7, pp. e1000832, 2010

[130] T. Siggers, R. Gordân, **"Protein–DNA binding: Complexities and multi-protein codes,"** *Nucleic Acids Research*, Vol. *42*, No. 4, pp. 2099–2111, 2013

[131] B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, **"Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning,"** *Nature Biotechnology*, Vol. 33, No. 8, pp. 831–838, 2015

[132] H. Suk, C. Y. Wee, S. W. Lee, D. Shen, **"State-space model with deep learning for functional dynamics estimation in resting-state fMRI,"** *NeuroImage*, Vol. *129*, pp. 292–307, 2016

[133] Y. Bengio, I. J. Goodfellow, A. Courville, **"Deep learning,"** *Nature*, Vol. 521, pp. 436–444, 2015

[134] X. W. Gao, R. Hui, **"A deep learning based approach to classification of ct brain images,"** *In Proc. of SAI Computing Conference (SAI)*, London, United Kingdom, pp. 28–31, 2016

[135] A. Gupta, M. Ayhan, A. Maida, ***"Natural image bases to represent neuroimaging data,"*** *In Proc. of International Conference on Machine Learning*, pp. 987–994, 2013

[136] A. Krizhevsky, I. Sutskever, G. Hinton, ***"Imagenet classification with deep convolutional neural networks,"*** *Advances in neural information processing systems*, pp. 1097–1105, 2012

[137] A. Payan, G. Montana, **"Predicting Alzheimer's disease: A neuroimaging study with 3d convolutional neural networks,"** *arXiv:1502.02506*, pp. 1-14, 2015

[138] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. Grunau, G. Hamarneh, **"Brainnetcnn: Convolutional neural networks for brain net- works: Towards predicting neurodevelopment,"** *NeuroImage*, Vol. *146*, pp. 1038–1049, 2017

[139] D. Urda, F. Moreno, **"Deep learning to analyze RNA-seq gene expression data,"** *In Proc. International Work-Conference on Artificial Neural Networks*, Vol. 3. pp. 50-59, 2017

[140] A. Vazhayil, R. Vinayakumar, K. Soman, **"Deep Proteomics: Protein family classification using shallow and deep networks,"** *arXiv:1809.04461*, pp. 1-17, 2018

[141] A. Cheerla, O. Gevaert, **"Deep learning with multi modal representation for pan cancer prognosis prediction,"** *Bioinformatics*, Vol. 35, No. 14, pp. 446-454, 2019

[142] L. Chen, X. Pan, Y. H. Zhang, M. Liu, T. Huang, Y. D. Cai, **"Classification of widely and rarely expressed genes with recurrent neural network,"** *Computer Structural Biotechnology Journal*, Vol. 17, pp. 49-60, 2019

[143] J. Lim, S. Bang, J. Kim, C. Park, J. Cho, S. Kim, **"Integrative deep learning for identifying differentially expressed (DE) biomarkers,"** *Computational and mathematical methods in medicine*, Vol. 2019, pp. 1-10, 2019

[144] S. Bian, **"Diagnosis of Alzheimer's disease using structural MRI and Convolution neural network,"** *In Proc. of ICEEB*, Vol. 185, pp. 03037-03042, 2020

[145] M. M. Islam, S. Huang, R. Ajwad, C. Chi, Y. Wang, P. Hu, **"An integrative deep learning framework for classifying molecular subtypes of breast cancer,"** *Computational and Structural Biotechnology Journal*, Vol. 18, pp. 2185-2199, 2020

[146] L. Zhou, M. Rueda, A. Alkhateeb, **"Classification of Breast Cancer Nottingham Prognostic Index Using High-Dimensional Embedding and Residual Neural Network,"** *Cancers*, Vol. 14, pp. 1-16, 2022.