A Post-Method Condition Analysis of Using Ensemble Machine Learning for Cancer Prognosis and Diagnosis: a systematic review

Kaveh Kavousi (✉ kkavousi@ut.ac.ir)

　Institute of biochemistry and biophysics (IBB)　https://orcid.org/0000-0002-1906-3912

Leila Mirsadeghi

　Department of Biology, Faculty of Science, Payame Noor University, Tehran, Iran

Reza Haji Hosseini

　Department of Biology, Faculty of Science, Payame Noor University, Tehran, Iran

Ali Mohammad Banaei-Moghaddam

　Laboratory of Genomics and Epigenomics (LGE), Department of Biochemistry, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

Seyed Reza Beh-Afarin

　Islamic Azad University, North Tehran Branch, Faculty of Foreign Languages, ELT and Bio-Translation studies Department

---

Research article

Abstract

Background Ensemble methods are supervised learning approaches that integrate different types of data or multiple individual classifiers. It has been shown that these methods can improve professional performance. Methods This study is an attempt to provide an in-depth review on 45 most relevant articles and aims to introduce 42 ensemble classifier (EC) machine learning methods used for the detection of 18 different types of cancer. Compared to other types of cancer, breast cancer, and the 22 ensemble methods introduced for its identification, is extensively investigated. The purpose of this study was to identify, map, and analyze the current academic discourse on EC machine learning methods in order to: 1. identify overarching themes emerging from empirical studies regarding EC methods, 2. determine their input data and decision-making strategies, and 3. evaluate relevant statistical procedures. Results By comparing various approaches, we can introduce Relevance Vector Machine (RVM)-based ensemble learning method that can provide optimal solutions for problems such as curse the dimensionality and high-dimensionality of feature space without missing data values. Conclusions To obtain robust performance and achieve better results, it is tactfully suggested to use multi-omics data integration, which has demonstrated to identify cancers and their subtypes more efficiently.

Background

The prognosis and diagnosis of complex diseases such as cancers are two of the most crucial issues in precision oncology. There are large volumes of bio-molecular, biological and biomedical data in multiple databases including National Human Genome Research Institute (NHGRI), National Cancer Institute, National Institutes of Health (NIH), and The Cancer Genome Atlas (TCGA) project that have developed our ability to screen cancer and have created a lot opportunities for studies on the susceptibility to cancer [1].

Researchers are currently investigating the possibility of using ensemble machine learning tools to help detect cancer. Ensemble methods are supervised learning models that integrate different types of data such as bio-molecular and clinical data with computational function. Committee approaches can unmistakably outperform every one of the single models [2]. The idea also builds a predictive function that integrates two or more models and can enhance gloomy predictions [3]. Accuracy and diversity are two important conceptions in the construction of any classifier ensemble. For designing an accurate ensemble, it is necessary to base classifiers as diverse as possible. Typically, ensembles produce better results when there is greater diversity in the models [4]. Due to recent developments in high-throughput technologies (HTTs), a lot of data associated with cancers has been generated. Previous studies showed that the most recent predictive machines integrate various data types and datasets, including genomic, clinical, histological, imaging, demographic, epidemiological, and proteomic data [5].

Ensemble machine learning theory has a very long history [3,6, 7] Bagging and boosting are among The significant progress in the field of ensemble learning was obtained during the 90s when Bagging and Boosting were created as the most popular ensemble methods [8].

In 1996, and in advancing the research on bagging methods, Breiman noted that this approach would generate better accuracy when the individual base classifiers are unstable [9]. In the same year, bagging nearest neighbor classifiers were used for breast cancer dataset [10]. Thus far, the usage of various ensemble methods for analyzing bio-molecular, biological, and biomedical data has been increased.

Generally, the fusion issue can be investigated at three levels: 1. data and feature integration, 2. decision integration, and 3. model integration (Fig. 1).

Fig. 1) Integration patterns for ensemble learning

The main purpose behind this study, therefore, is to look for the most pre-eminent ensemble method for cancer detection and driver genes prioritization for cancer prognostic and diagnostic. The study is an attempt to answer the following questions:

1. To what extent does the empirical research surrounding the most pre-eminent ensemble method with superior performance (specifically on large or small size datasets) relate to existing typologies?
2. Among the existing studies, which fusion systems have been used to identify cancers?
3. What patterns emerge from empirical studies of ensemble learning addressing the diversity between different decision makers/datasets for improving the performance?
4. Which ensemble model is a more effective approach to process large-scale and high-dimensional problems with massive missing data values?
5. Which ensemble is successful for problems with a small number of imbalanced classes?

The authors will first introduce the most prominent ensemble methods and will determine their base classifiers. Decision-making method for classification will be elaborated, the challenges about each of these methods will be addressed, and the solution for many of these challenges will be worked out.

Methods

In this part, 42 ensemble learning methods used for cancer detection are classified into three famous distinct fusion categories and discussed; include data and feature integration methods, decision integration methods, and on the smaller scale, model integration methods. Then in each case, input data, number of samples, the statistical tools for evaluation of performance will be introduced. And decision-making strategies for decision integration methods will be specifically determined. In this systematic research, ensemble systems regarding 45 most relevant articles in the category of cancer prognosis and diagnosis were studied. The studies that used a valid statistical tool to evaluate their performance were considered and included. In contrast, studies that did not compare the percentage of accuracy of their findings with other methods and also the different aspects of their performance had not clearly assessed, were excluded.

2.1 Data and Feature Integration

When a couple or more biological heterogeneous input data such as clinical data, mutation data, expression data, proteomics data, or gene ontology (GO) database data are combined, a kind of fusion system is created that is called data integration. The higher level of this category called multi-omics data integration. It means that various levels and scales of data, including genomics, epigenomics, transcriptomics, proteomics, and metagenomics data are integrated. It has been mentioned combining multi-Omics data can improve predictive performance [11]. Also, in some studies, various features are selected or extracted from homogeneous or heterogeneous data. When the combination of these features is used to implement an algorithm, another type of integration is created, which is called feature fusion. In most cases, the integration of data and features are simultaneously used as an ensemble system.

2.1.1 Bayesian Network Classifiers (1,2)

Bayesian network classifiers are methods that can integrate heterogeneous data from multiple sources to reveal the mechanisms of complex diseases such as cancers. In a study on hepatocellular carcinoma (HCC), also called malignant hepatoma, carried out by previous researchers., microarray and clinical data were integrated from biological databases and literature. Some liver cancer protein biomarkers were predicted, functional modules that show progression mechanism of liver cancer were identified, and performance was evaluated with 10-fold cross-validation on the testing set. For selecting the test set, the training data was split into ten approximately equal size sets, and then one of them was used for testing. For choosing a different testing set, this process was repeated ten times. Results showed that compared to Bayesian network (BN), Naïve Bayes (NB), full Bayesian network (FBN), and Support Vector Machine (SVM) classifiers, their proposed method gives maximum the area under the curve (AUC) [12].

In other studies, the Bayesian network was used for integrating four types of data consist of GO database data, MicroArray (MA) co-expression data, orthologous Protein-Protein Interactions (PPI) human data and True Positive (TP) data as two networks including (GO+MA+PPI) network or (GO+MA+PPI+TP) network. This approach uses dissimilar data sets and can deal with missing data. This Bayesian network has been applied for prioritizing candidate genes in related to breast cancer such as PIK3CA, CHEK2, BARD1, and TP53 that were predicted with (GO+MA+PPI) network. This data integration method was called Prioritizer. It has been evaluated the performance of the various gene networks using cross-validating on all data sets, ten times. Performance of Prioritizer is good when genes on the basis of their functional interactions are ranked. Prioritizer approach can help to the diagnosis of disorders by introducing driver genes. Accuracy of (GO+MA+PPI) network is significant (AUC = 90%). Also, results showed that the proposed method has a far better performance for the prioritization of genes in Mendelian diseases in comparison with complex disorders [13].

2.1.2 stSVM by Data Integration

In an investigation that has been carried out by in 2013, a new fusion method was introduced. This approach was called smoothed t-statistic SVM (stSVM). It integrates features obtained from experimental

data such as mRNA and miRNA expression data into one SVM classifier. It has been applied for prognosis and diagnosis of cancers including breast, prostate and ovarian cancers and for gene prioritization. In this study, four datasets were used [14]. One breast cancer dataset (GSE4922) [15] two prostate cancer datasets (GSE25136 [16] and GSE21032 [17] and one ovarian cancer dataset (TCGA) [18] from various data repositories. The stSVM was evaluated via 10 times repeated 10-fold cross-validation method [14]. Finally, stSVM were compared with saliency guidedSVM (sgSVM) as a meta-classifier, an SVM machine trained with significant differentially expressed genes with FDR cutoff 5%, selected by Significant Analysis for Microarrays (SAM) [19]. It has been shown that stSVM approach has high predictive power for introducing novel gene lists for mentioned cancers [14].

2.1.3FSCOX-SVM

Feature selection with Cox proportional hazard regression model (FSCOX) is a novel method based on feature selection with Cox proportional hazard regression model that integrates data from different datasets into one SVM classifier. The proposed method was called FSCOX-SVM. This fusion model carried out data integration between miRNA and mRNA features. It has been applied for improving the prediction power of various cancers survival time, especially ovarian cancer and Glioblastoma Multiforme (GBM). In this study, two computational methods were suggested for the prediction of target genes, including TargetScan and miRanda. After computing optimal sequence complementarities between a mature miRNA and mRNA, TargetScan identifies miRNA targets. While a weighted sum of the match and mismatch scores for base pairs and gap penalties was computed by miRanda. The proposed approach predicts the class of each test sample, via the Leave-One-Out Cross-Validation (LOOCV) procedure. Finally, the approach was compared with three classifiers, including RF, SVM, and FSCOX median. Their findings demonstrated that FSCOX-SVM approach showed the highest performance and accuracy, among others. In fact, data integration between miRNA and mRNA features Led to better achievements [20].

2.1.4 Multiple RFE Selection Methods (1,2)

Multiple Recursive Feature Elimination (RFE) is an ensemble feature selection method. This strategy has been applied for identifying metastatic breast cancer core module biomarkers. First, 100 features, including gene expression based on DNA microarray technology and activity vectors features, were the candidate for classification and divided into 500 random splits (with the possibility of overlapping). Then, 500 classifiers were constructed and recorded their AUCs as well as their weight vectors. Third, the features were ranked by the average square weight of each feature among 500 different splits. The lowest ranked feature was eliminated recursively until the maximum average AUC obtained. this procedure was repeated 100 times for selecting a final marker gene set [21]. The consistency of the proposed approach was evaluated with a multi-level reproducibility validation framework [22]. It is a kind of level-by-level validation method [23]. The algorithm related to this strategy identifies the highly reproducible marker. It means that it generates highly reproducible results across multiple experiments.

The results show that this method has improved accuracy and biomarker reproducibility by as much as 15% and 30%, respectively. This method computes an average weight from 500 classifiers and uses algebraic combiners for decision making. Multiple RFE was applied for a classification tool called COre Module Biomarker Identification with Network ExploRation (COMBINER) in the feature selection step [21]. COMBINER software was implemented on three independent breast cancer datasets including the Netherlands [24], the USA [25], and Belgium [26] and identified 13 driver genes as reproducible discriminative biomarkers. Also, a robust regulatory network was constructed [21].

In other studies, RFE selection method has also been applied for biomarker discovery with colon, leukemia, lymphoma, and prostate cancer. Finally, the outputs of all selectors are aggregated, and the ensemble result is computed. In general, this method generated a diverse set of feature selections [27]. The proposed approach was assessed on four microarray datasets, including leukemia dataset [28], colon dataset [29], lymphoma dataset [30], and prostate dataset [31]. For selecting a training set, subsampling was done, and each time, 10% of the data was used as an independent validation set to evaluate classifier performance. The results showed that the robustness of the selected driver genes was increased up to almost 30% and performance improved up to ~15% in the classification method [27].

2.1.5 Feature Subsets Method

In a study aimed for predicting survival in breast cancer patients, the researchers designed an ensemble method to learn models using feature subsets then combined their predictions [32]. Two breast data set were used, including Dataset 1 [24, 33] and Dataset 2 [25]. Data achieved through microarray experiments, and their results were compared with clinical criteria. Feature subsets were obtained from three different methods, including splitting feature selection method, sliding window feature selection method, and random subsets feature selection method. These three feature-subset-selection methods were integrated to construct the proposed ensemble model. The performance of the proposed method was evaluated for 100 different training/test sets from dataset 2, and the results of the method showed a high performance and confidence interval. Compared to the Amsterdam signature and clinical criteria, the proposed method generated a high sensitivity and Negative Predictive Value (NPV). The performance of the approach was enhanced. When splitting feature subsets were used, sensitivity and accuracy were improved[32].

2.1.6 Multimodal Data Fusion of Separate Datasets

Multimodal Data Fusion is a fusion model that integrates clinical and bio-molecular data such as image and microarray data. In fact, in this study, data were heterogeneous. This approach has been applied to the diagnosis of melanoma. There are two different types of multimodal data fusion for fusing separate datasets. Two approaches including a combination of data (COD) and the combination of interpretations (COI) were exploited for feature integration. COD is applied before classification, and it aggregates features from each source for producing a single feature vector, but in COI, independent classifications

are done based on the individual feature subsets using a proper voting mechanism [34], and involve aggregating outputs, so it uses algebraic combiners for decision-making strategy. In another study that was done in related to prostate cancer has been told that COD methods are more optimal [35]. It should be noted that in this method and feature selection step, SBE (sequential backward elimination) with RF (random forest) algorithm was used that utilizes integrations of decision trees. This kind of feature selection techniques leads to dimensionality reduction. Finally, feature selection and dimensionality reduction methods use algorithms for extracting of better biomarkers. The performance of this classification method was evaluated using 10-fold cross-validation procedure with 50 repetitions on different datasets. Results demonstrated a random forest approach that was used in classifying of bootstrapped samples gained an AUC score of high. In contrast, obtained performance with other linear methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) was not high [34].

2.1.7 Meta-classifier Ensemble Learning based on Genetic Programming Technique

In other studies, that were conducted by researchers, an ensemble meta-classifier combining five classifiers was used for feature integration. The features were produced by Genetic Programming (GP) technique which used a kind of evolutionary algorithm for generating thousands of classifiers as features. Input data for the GP system is gene expression data. Finally, the top five classifiers were chosen as the individual classifier. GP classifiers often including five or fewer genes as biomarkers and predict cancer class, successfully. Then, the proposed ensemble method integrates these features for achieving better results. This method is applied for the diagnosis of some cancer types such as prostate and Lung Cancers. Also, it can classify their subtypes, such as Metastatic Prostate Cancer (MPC) and Primary Prostate Cancer (PPC). Results demonstrated GP is a robust method in feature selection and accurately suggests genes for prognostic and diagnostic targets. Also, misclassification Error Rates of GP is very slight. The performance evaluation of the GP system was done using five-fold cross-validation on the training set. The maximal accuracy obtained with the proposed method and average prediction rate was very high when meta-classifier ensemble learning based on GP compared with other classification methods such as 3-nearest neighbors, nearest centroid, covariate predictor, SVM, and Diagonal Linear Discriminant Analysis (DLDA) [36].

2.1.8 Ensembles of BioHEL Rule Sets

Bioinformatics-Oriented Hierarchical Learning (BioHEL) is an evolutionary machine learning approach that integrates microarray data from different datasets. It uses Random Forest based Feature Selection (RFS), Correlation-Based Feature Selection (CFS) and Partial-Least-Squares Based Feature Selection (PLSS) in the feature selection phase. This method has been applied for gene prioritization as diagnostic markers related to prostate, lymphoma, and breast cancers. The proposed method uses algebraic

combiners for feature ranking. For each training set, BioHEL was run 100 times, separately. The main procedure that was used in this classification method is a kind of cross-validation scheme called as two-level external cross-validation. Results compared with some other machine learning tools. Genetic Algorithms based classifier system (GAssist), SVM, RF, and Prediction Analysis of Microarrays (PAM) are among them. The accuracy obtained from BioHEL was highest. BioHEL showed better performance on large datasets in comparison to its nearest rival, the GAssist [37].

2.1.9 Kernel-based Data Fusion Method for Gene Prioritization

This fusion method combines all kernel matrices on human genes. A kernel matrix is used in kernel machines such as SVM. So, all instances are represented by a kernel matrix, which is an n × n positive semidefinite matrix. Each element (ai,j) shows the similarity between ith and jth instances using a pairwise kernel function k(x_i, x_j). By using this approach, we can enhance learning methods by not explicitly making a feature space. In fact, the kernel matrix implicitly represents the inner product between all Paris of instances in an embedded feature space yielded by an explicit feature mapping. Since the resulting feature space may be high-dimensional or even infinite-dimensional, kernel matrix helps to have tractable and efficient computation in the original space without explicit mapping [38]. They are integrated through Log-Euclidean Mean (LogE), Arithmetic Mean (AM) and weighted-version of LogE (W-LogE). Input data were 12,000 human genes. By this approach, 24 novel driver genes were a candidate for 13 diseases such as breast and ovarian cancers. The method uses GO, Swiss-prot (SW) annotation, PPI network based on STRING database, and literature as annotated data sources. For these cancers, kernels performance was evaluated using leave-one-out cross-validation on the training genes. Average True-Positive Rate (TPR) results obtained from proposed Kernel-based data fusion tools were compared with ENDEAVOUR. Results showed that various approaches of Kernel-based data fusion including LogE, W-LogE, and AM performed better than ENDEAVOUR, and the best was W-LogE [39].

2.2 Decision Integration

These methods have been constructed from several base classifiers. Actually, the critical component of any ensemble system is the strategy employed in combining classifiers. The module of outputs combination based-decision making methods is another major issue in this kind of ensemble method. There is no unique naming procedure for the same decision making strategy that has been used in different articles/books.

The terminology we have used to outputs combination and decision-making for final decision integration is according to the following pattern (Fig. 1):

The pattern is divided into two main categories, such as combining class labels (CCL) and combining continuous outputs (CCO). CCL is divided into four sub-types that include majority voting (MV), weighted

majority voting (WMV), behavior knowledge space (BKS) and Borda Count. CCO is divided into three sub-types, which includes algebraic combiners, decision templates, and Dempster-Shafer based combination [3].

2.2.1 Homogeneous Ensemble Methods

Homogeneous ensemble methods refer to the fact that all of the base classifiers are the same and from a single type, but they are different at the data used for training phase or model parameters (e.g., linear combination fusion function model) or a combination of the two categories [40].

2.2.1.1 SVM Classifiers Fusion (three SVM)

The kind of homogeneous ensemble method is SVM Classifiers Fusion. It is a Multi-classification system (MCS) that combines three SVM classifiers. This computational method has been used for breast cancer detection. Combination of three classifiers minimized the classification error in the training phase. For every base SVM, training data and testing data achieved from Digital Database for Screening Mammography (DDSM) mammographic images database [41, 42]. In this study, 300 images were used at the training phase and 100 images for the testing phase. It uses simple Majority Voting for decision making. The cross-validation technique was used for these multiple classifier system evaluations. The results showed that fusion of SVM classifiers improves the performance of the system. It is better than applying of all features in one features vector. Also, results of MCS with voting compared to each single SVM classifier showed that accuracy was increased because of the quality of the decision is improved [43].

2.2.1.2 enSVM (200 SVM)

In a study, a fusion approach was used and called ensemble SVM (enSVM) that is included in three steps. Step 1 is sub-sampling of genes that generate Gene Subsets and then constructs 200 diverse classifiers. Input data for this step are gene microarray data related to 97 patient samples. In step 2, SVM came into operation and generated 25 candidate classifiers. In this phase, SVM is suitable because it solves variable and high dimensionality problem of training data. In step 3, final decision making with majority voting strategy mechanism was done. The proposed method has been applied for microarray data classification and accurate diagnosis of breast cancer, cancers of the central nervous system, colon tumor, leukemia, and prostate cancer. In this study, LOOCV was used to evaluate the performance of SVM as the base classifier. Results showed that the proposed gene sub-sampling-based ensemble learning that is called enSVM outperforms single SVM and re-sampling ensemble learning methods such as bagging and boosting and enSVM showed relatively the best classification accuracy [44,45] .

2.2.1.3 Three neural networks fusion

In previous studies, the researchers had constructed a combinational feature selection method concerning ensemble three neural networks (NNs). This method has both levels (feature selection integration level and decisions integration level). It has been applied for several cancers diagnosis and treatment via discovering of marker genes of the diseases by gene expression data. These cancers are adult acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), malignant pleural mesothelioma (MPM), adenocarcinoma (ADCA) of the lung, and prostate cancer. At the first step, bagging generates 100 individual classifiers by resampling 100 times on microarray data. Then each classifier as an input will be given to 3 neural networks. The ensemble three neural networks method uses algebraic combiners for decision making, but there are 100 ensemble networks methods with 100 different outputs. Majority voting was used for combining their results. Also, final decision making was done by majority voting. In comparison with other methods, the proposed method effectively improved results. This method can provide more information in microarray data to increase accuracy and introduce the driver genes of the diseases for diagnosing and treatment. The accuracy of this method for ALL, AML, lung cancer, and prostate cancer was 100%, 100%, and 97.06%, respectively. In this method, classification performance and accuracy value were evaluated through 10-fold cross-validation and LOOCV. The accuracy of bagged decision trees for ALL, AML, lung cancer, and prostate cancer were 91.18%, 93.29%, and 73.53%, respectively. The accuracy of the best methods considering ALL, AML, lung cancer, and prostate cancer were 97.06%, 97.99 %, and 73.53 %, respectively [46].

2.2.1.4 NED method (five artificial neural networks fusion)

An ensemble method called Neural Ensemble based Detection (NED) is a learning method that combines five Artificial Neural Networks (ANNs) [48]. The neural algorithm of each network is Fast Adaptive Neural Network Classifier (FANNC) [47]. FANNC has both high performance and speed. Also, FANNC is an automatic algorithm, and it requires no manual set up. The proposed ensemble method has been applied for the lung cancer diagnosis. In this study, images of the specimens of needle biopsies are used as input data containing 552 cell images from biopsies of subjects. This ensemble system has a two-level ensemble structure. At the first level, outputs of each individual neural network place in two classes: normal cell or cancer cell. Then, NED method uses full voting for decision making. In this decision making strategy, the cell is considered normal when all of the individual networks vote it is normal. At the second-level ensemble, each network has five outputs, including adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and normal. This method uses plurality voting for decision making. In this way, the identification rate of NED is high, and its false negative rate is low. This method helps to miss less positive cancer patients. The accuracy of the method was evaluated by 5-fold cross-validation on the data set and was demonstrated that confidence of the first-level ensemble is high, especially. The proposed approach consists of five FANNC networks were compared with a single artificial neural network. Results showed that NED outperforms the single FANNC [48].

2.2.1.5 Clinical decision support system

The Clinical Decision Support System (CDSS) is an ensemble method that is included four different Weighted Random Forest (WRFs). Each WRF has been constructed with 80 trees. This ensemble method combines the results of clinical techniques such as classic and ancillary techniques. Crucial clinical data were included visit dates, patient age, Human Papilloma Virus (HPV) genetic examinations, cytological diagnoses, and histological examination of biopsies. Under the project, 740 cases were studied. With this method, more accurate results were produced. It has been applied for Cervical Cancer (CxCa) diagnosis and uses majority voting for decision making strategy. The performance of the proposed system was estimated using 10-fold cross-validation. Results showed that performance obtained from the proposed method (CDSS consists of four different WRFs) is better than single classifiers approaches including k-Nearest Neighbors (KNN), NB, Classification and Regression Tree (CART), Multi-Layer Perceptron Network (MLP), Radial Basis Function (RBF) Network and Probabilistic Neural Network (PNN), but it showed a bit worse performance than the integrated CDSS consists of two ANNs [49].

2.2.1.6 Bagging subgroup identification trees

Bagging subgroup identification trees is a tree-based ensemble method which combines binary trees. Generation of bootstrap samples is done by resampling the training data with replacement, and then several trees are constructed as diverse classifiers. In the next step, each tree converted to a binary classifier. Finally, "binary trees" are constructed, and the final prediction is done with a simple majority vote strategy. In this study, clinical data such as gender, age, surg, etc., was used. About colon cancer, 929 cases participated [9]. The dataset associated with this cancer can be downloaded from R package survival [50]. Also, the GSE14814 dataset [51] related to lung cancer, including 133 patients, was applied. These two datasets were extracted from the R package survival and GEO database related to colon cancer and lung cancer, respectively. As mentioned, this proposed method uses a majority voting for decision making strategy. The selection of biomarkers and building of classifiers were done with Leave one out cross-validation. Sensitivity, specificity, and accuracy rate of the proposed method was compared with a multivariate Cox model. Results showed that the proposed ensemble bagging (a novel tree-based) method is better, especially when data is imbalanced. For balanced data, the Cox model was slightly better [9].

2.2.1.7 CAD system

Computer-Aided Diagnostic (CAD) system is an ensemble method that combines Bayes classifiers. It was applied for the tissue classification and diagnosis of focal liver lesions. In this study, input data were Computed Tomography (CT) contrast-enhanced images of 20 cases of liver cancer. This classification process has two phases. In the first phase, CT images were classified using the Bayes classifier. In the second phase, the combination of classifier outputs and the decision-making process were done using

majority voting strategy. Also, classification success rates were evaluated by leave-one-out technique. This approach for classifier combination generated better performance. Findings showed that the best performance of this method was obtained by majority voting in data. And CAD-based Bayes generates relatively high accuracy value [52].

2.2.2 Heterogeneous Ensemble Methods

Heterogeneous ensemble methods incorporate different base classifiers in spite of they usually use the same training dataset and input data for running different learning algorithms [53].

2.2.2.1 BNCE method

BNCE is an ensemble approach for training neural network fusions. This ensemble method combines boosting with negative correlation (NC). It has been used for breast cancer detection to the classification of the tumor as either benign or malignant [54]. Data were well-known benchmarks related to breast cancer that were downloaded from the datasets of the UCI machine learning benchmarks repository [55, 56]. The proposed approach uses majority voting for decision making. In order to performance evaluation, classification error in the percentage of BNCE was estimated. The results of this method were compared with the results of other methods, including Evolutionary Programming Network (EPNet), a single NN, a simple NN ensemble, bagging, Ada-boosting, and arc boosting. They used BNCE on breast cancer well-known benchmarks. Generally, a comparison in terms of the classification error rate for benchmark datasets showed the proposed method has the best performance [54].

2.2.2.2 The meta-learning method

In another study, a meta-classification tool was used for prostate cancer detection. Data for this ensemble strategy are mass spectrum data (MS data), and it combines the results of several machine learning approaches [57]. Individual classifiers are ANN, KNN, SVM, LOGISTIC-REGRESSION, and CART [58]. It uses weighted majority voting for decision making. The proposed method combines multiple error independent base classifiers into a meta-classifier. Validation of meta-classifier was done with k-folding validation (leave-one-out) experiments on the training set. This ensemble method improves prediction accuracy over individual classifiers. In comparison with individual classifiers such as ANN, KNN, SVM, LOGISTIC-REGRESSION, CART, results showed that their proposed method accuracy is better. Also, sensitivity and specificity were high (respectively, 91.30% and 98.81%). By the way, 11 biomarkers associated with prostate cancer was diagnosed [57].

2.2.2.3 Heterogeneous ensemble (KNN-SVM- DT-LDA)

Generally, heterogeneous ensemble methods combine the outputs of several base classifiers. They train some learner machines with different learning strategies using a single common training dataset. This definition is in contrast with the methods that use different datasets for training a single learner machine. In one of the heterogeneous ensemble methods proposed by previous studies, five base classification algorithms such as KNN (K=3), KNN (K=5), SVMs, DT (Decision Trees) and LDA (Linear Discriminant Analysis) were used. This method has been designed for increasing the chance of early prostate cancer diagnosis [59]. Data were proteomic prostate cancer data obtained from protein mass spectrometry available in JNCI Data 7-3-02 [60]. The statistical population was 322 patients that data obtained from their sera. 63 people were with normal prostate, 190 patients were with benign prostate tumors, 26 patients with prostate cancer and Prostate-Specific Antigen (PSA) level in the range 4–10, and 43 patients with prostate cancer and PSA levels above 10. The proposed approach uses simple majority voting for decision making. Also, for performance validation of the method, 10-fold cross validation was used. The results showed that accuracy and sensitivity were increased, but specificity slightly decreased after using ensemble method. This simple fusion strategy improved prostate cancer mass spectroscopy dataset-based methods which are faced with High Dimensionality Small Sample (HDSS) problem. By this approach, overall performance is boosted. Diagnosis using protein mass spectrometry technique is a new solution. Many of the learning algorithms use it to increase the chances of prognosis of cancer in the early stages. However, the problem of small samples with high dimensions concerning the proteomic data in cancers requires more sophisticated solutions to improve classification accuracy. In this method, five classification algorithms were used. Applying this simple strategy in making the final decision, it yields a more promising performance for the use of mass spectroscopy data related to prostate cancer [59].

2.2.2.4 MRS method

The mixture of Rough set and SVM (MRS) is a mixture classification model based on clinical markers that are made by combining rough set and SVM classification tools in serial form. This model is a serial multi-sensor system that integrates several methods with different sources and characteristics for breast cancer prognosis. In this fusion method, rough set classifier acts as the first layer for identification of some singular samples in data, and the SVM classifier comes into operation as the second layer for the classification of remaining samples. The upper layer is called shrinking classifier, too and uses a voting strategy for decision making. For each sample, the rough set tries to assign a class type to it. In this step, if the class type is unknown, the second layer comes into operation for assigning a class type to its sample. This two-layer construction without voting is a suitable way for better clinical prognosis. MRS has used two open breast cancer datasets for prediction [61]. One dataset called BRC-1 hereafter that is included both clinical data and gene expression data from 97 breast cancer tumors of lymph node-negative patients [62]. The other dataset is BRC-2 hereafter that uses baseline human primary breast tumor clinical data from Lawrence Berkeley Laboratory (LBL) breast cancer cell collection containing 174 samples [63]. This approach gives higher accuracy, specificity, sensitivity, and Matthew's correlation

coefficient (MCC) than previous prognostic methods such as NB, SVM, J48, random forest, and attribute selected classifier. Also, the higher accuracy of the method was validated by 5-fold cross-validation [61].

2.2.2.5 Bagging (bootstrap aggregating) method

In a study was done in 2016, two different meta-learning algorithms were used. They applied Bagging- RF, Bagging- NB and Bagging-K* instance (K*) and compared their results with individual classifiers including RF, NB, K* and vote ensemble classifiers (RF-NB-K*) too. All of the methods have been used for melanoma skin cancer detection. Data were clinical images of skin lesions. These ensemble methods use simple majority voting for decision making. Because of the Bagging reduces variance and helps to avoid overfitting, so Bagging aggregation improves the accuracy and stability of the other selected tools. In this study, 10 – fold cross-validation test was used for estimation of its accuracy. In comparison with other methods, the results show that when the number of positive cases is insufficient, using Bagging with the Random Forest is suitable. Using this approach, sensitivity and AUC have been meaningfully improved [64].

2.2.2.6 Artificial intelligence based hybrid ensemble technique

Researchers have also designed a novel artificial intelligence based hybrid ensemble technique for screening of cervical cancer. They used smear images data for diagnosis as clinical data. The hybrid ensemble system combines fifteen different classifiers [65] including Bagging, decorate, decision table, Ensemble of Nested Dichotomies (END) [66], filtered classifier, J48 graft [67], Projective Adaptive Resonance Theory (PART) [68], multiple backpropagation ANN, multiclass classifier, NB, random subset space, radial basis function network [69], rotation forest, random forest and random committee. The method uses voting for decision making. Validation is done on multiple training and testing datasets, and 10-folds cross-validation is applied for evaluation of this algorithm. This approach provides high performance for classification of complex datasets. The hybrid ensemble technique is a promising method for classification of pap-smear images and can be used to detect cancer cervical cancer. receiver operating characteristic (ROC) Area of the proposed novel hybrid ensemble system was increased, and the overall performance of the ensemble approach was improved. Also, in comparison with individual classifiers, results were better for both multi-class and two-class problems [65].

2.2.2.7 Boosting-TWSVM method

In other studies, researchers used boosting with SVM together for MicroCalcifications (MCs) clusters detection in digital mammograms. MCs clusters are an important sign of breast cancer diagnosis. This ensemble method uses algebraic combiners for decision making because the aggregation is computed by the weighted averaging in this method [70]. In comparison, they showed that their proposed method

outperformed the other methods such as twin SVM (TWSVM) [71]. In this study, there were 650 positive and 3567 negative samples. Samples were split into two subsets, the first part was used for the training set and validation, while the second part was applied for the testing set. Also, the TWSVM classifier was trained using 10-fold cross-validation technique for evaluation of method performance. Since the TWSVM is sensitive to the training samples, it is inconsistent, but when Bagging was integrated into TWSVM, the inconsistent problem in the training set will be solved. Result related to BOOSTING-TWSVM showed that Sensitivity and specificity were increased. Also, they demonstrated that Boosted-TWSVM is a promising approach for MCs detection [70].

2.2.2.8 Bagging and boosting-based TWSVM

Bagging and boosting based twin support vector machine (BBTWSVM) is yet another ensemble method. The structure of the algorithm of this ensemble method consists of three modules: the image preprocessing, the feature extraction component, and the BBTWSVM modules. Also, BBTWSVM modules composed of 2 two algorithms: bagging TWSVM and boosting-TWSVM Combining these algorithms results in a more efficient solution that composed of several classifiers: BBTWSVM. This method has been applied for clustered MCs detection and so is called MCs detection approach too. Breast cancer can be diagnosed by MCs detection approach. This fusion method uses algebraic combiners for decision making because it finds the maximum score from all the base classifiers, or computes a weighted scoring scheme from among the base learners. Data for validation were chosen through the training set and like of Boosting-TWSVM method, 10-fold cross-validation technique was used in the training phase for evaluation of method performance. BB-TWSVM outperforms TWSVM. The sensitivity of BB-TWSVM classifier was increased, and ROC curves showed that in comparison with TW-SVM, the performance of the proposed approach is improved [72,73].

2.2.2.9 Ensemble multi-class learning

Ensemble multi-class learning algorithm is an ensemble approach that combines error-correcting output coding (ECOC) scheme and one-against-one pairwise coupling (PWC) scheme. This method has been used for finding biomarkers in liver cancer. The method uses an algorithm called extended Markov blanket (EMB). Also, a liver cancer matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) dataset was used as a training dataset. In this method, redundancy and relevance were two aspects of biomarkers that were considered for feature selection. By this ensemble method, identification of proteomic biomarkers for liver cancer was possible. It uses voting for decision making [74]. Samples for liver cancer data were 201 spectra of MALDI-TOF MS belonging to HCC patients, cirrhosis patients, and healthy participants [75]. Then all samples were divided into 10 exclusive folds, randomly. In this study, the error rate was estimated, and 10-fold cross-validation was selected for evaluation of experimental results such as accuracy value. The proposed method was compared with

random forest, NB, classical ECOC, and J48 approaches, and the results showed accuracy was increased to 88.71% [74].

2.2.2.10 RSS-SCS method

This approach has combined Random Subspace (RSS) and Static Classifiers Selection (SCS) Paradigms. Proposed ensemble method has been used for Breast Cancer diagnosis by CAD. The approach used from a real database, including 300 mammograms as clinical data. These mammograms collected from the DDSM. In this research, it has been shown that CAD is an effective approach for breast cancer detection in the initial stages. At first, RSS constructs diverse classifiers by using different subsets of features that are used in the training phase. The second, SCS selects diverse classifiers. Then, the outputs of classifiers are combined. These diverse classifiers use majority voting for decision making. For estimating the final accuracy of feature subsets, cross-validation was used. Results demonstrated that in comparison with the three best ensemble methods such as Bagging, AdaBoost, and Random Subspace, proposed approach generated higher rate in three metrics including sensitivity, specificity, and accuracy [76].

2.2.2.11 REIS-based ensemble method

The Resonance-frequency Electrical Impedance Spectroscopy (REIS) − based ensemble method has fused five classifiers. It is a kind of heterogeneous ensemble method. These classifiers are ANN, SVM, Gaussian mixture models (GMM), CART, and LDA. This fusion method has been applied for suspicious breast lesions detection. The lesions are the sign of the risk of having or developing breast cancer. In this investigation, 174 cases were examined. Imaging-based examinations such as mammography, additional views, ultrasound, and magnetic resonance imaging were used as clinical data. For feature selection stage, a genetic algorithm was applied. The Reis-based method uses algebraic combiners for decision making. Actually, this method combines the results of classifiers via three rules, including sum rule, Weighted Sum Fusion Rule (WSFR), and Weighted Median Fusion Rule (WMFR). The performance was evaluated using a leave-one-case-out cross-validation technique. In this study, ROC curves were compared among the ANN, SVM, GMM, CART, and LDA individual classifiers. Without fusion, ANN had a higher rate. Also, Comparison of ROC curves for the single best classifier (ANN) and proposed fusion model with three rules showed that WSFR and WMFR are better than ANN and Sum Rule. So, the weighted median fusion rule is the best fusion approach in this study [77].

2.2.2.12 MV-ACE method

Multi-view based AdaBoost classifier ensemble (MV-ACE) framework is an ensemble method that has integrated multiple views in a straight forward manner, such as the linear combination of different views and AdaBoost algorithm. AdaBoost produced the base classifiers and optimized them. In this study, gene expression datasets were used. This ensemble method has been applied for class prediction from several

cancers gene expression profiles, including blood, bladder, liver, prostate, brain, endometrium, and bone marrow. MV-ACE works well for cancer classification by gene expression profiles. This ensemble method uses algebraic combiners for decision making. In this way, the algorithm was run 20 times separately, and the average value was calculated. Also, prediction accuracies were evaluated using 3-fold cross-validation. In this investigation, an accuracy value of the proposed method (MV-ACE) was compared with other best classifier ensemble methods, such as Bagging, MultiBoosting (MB), RF, RSS, and AdaBoost The results showed this approach achieved relatively better performance for most of the data sets [78].

2.2.2.13 DECORATE method

Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE) is another ensemble classifier method that can be categorized as a member of decision integration level class. Decorate is an ensemble method that combines four base classifiers including NB, Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier, C4.5 DT, and forest of random trees. It has been applied for introducing prognostic biomarkers in related to breast cancer and ovarian cancer. For feature selection step, 10 different methods were used as individual classifiers, and then 10 feature vectors were constructed, respectively [79]. Input data for these classifiers is somatic mutation data that is available in TCGA data set [18, 80]. Five individual classifiers which rank candidate genes based on p-values are including OncodriveFM, OncodriveCLUST, MutSig, ActiveDriver, Simon. The five remaining individual classifiers, including FLN, NetBox, MEMo, Dendrix, and FLNP, choose driver genes based on linkage weights. This method also has a layer of feature integration. Input test data is 20,624 genes annotated as protein-coding that was downloaded from NCBI database. Finally, supervised classification is done and DECORATE uses the average posterior probabilities of four above mentioned base classifiers. It is necessary to note when data is limited, and this ensemble method is effective because DECORATE is an approach that creates diverse artificial data. This method has been applied for introducing cancer driver genes, including breast cancer and ovarian cancer. Although DECORATE is grouped in decision integration methods with algebraic combination rule, in a distinct layer, it employs some kind of feature fusion mechanism. During training, the method was run 50 times, and performance evaluation was estimated using 10-fold cross-validation. Results showed that when the training set is small, DECORATE gained higher accuracy than other best ensemble approaches such as Bagging or boosting [79].

2.2.2.14 HyDRA method

Hybrid Distance-score Rank Aggregation (HyDRA) is an ensemble approach that has combined advantages of score and distance methods [81]. The score and distance methods [82] are aggregation techniques. The predictive potency of this aggregation approach is evaluated very high. HyDRA aggregates genomic data based on mutation data and has been applied on several gene sets related to diseases such as autism, breast cancer, colorectal cancer, endometriosis, glioblastoma, meningioma,

ischaemic stroke, leukemia, lymphoma, and osteoarthritis. By this method, driver genes for these diseases were prioritized. The proposed approach uses decision templates for decision making strategy because this method ranks driver genes based on different similarity criteria that are combined with statistical tools. In each disease and for disease gene discovery, the performance of HyDRA method was evaluated by Cross-validation. Results showed that performance obtained from HyDRA was higher than other methods such as Endeavor [83] and ToppGene [84] for a majority of quality criteria. The analyses also show that each method has specialized advantages in prioritization for some diseases [81].

2.2.2.15 Stacking IB3-NBS-RF-SVM method

This ensemble approach has combined four well-known individual classifier types, including Instance Based3 (IB3), Naïve Bayes Simple (NBS), RF, and SVM. This method is grouped as a decision integration level tool. It classifies DNA microarray data using biological gene sets such as KEGG gene sets. This ensemble approach has been applied for breast cancer and Leukemia diagnosis. It uses weighted majority voting for decision-making strategy. In this study, the kappa value is calculated instead of the accuracy because it is better criteria for the classification of unbalanced data; then, evaluation of individual classifiers was estimated using a 10-fold cross-validation schema. In this study, the proposed approach gives a better performance in comparison of the various integration methods, including AdaBoost hybrid, Bagging hybrid, Stacking-IB3, Stacking-NBS, Stacking-RF, and Stacking-SVM. The proposed approach is able to generate a ranked list of genes that can be effective for cancer diagnosis and shows meaningful improvement in cancer classification results such as accuracy and kappa values [85].

2.2.2.16 GenEnsemble method (NBS-IB3-SVM-C4.5 DT)

Similar to the previous method is GenEnsemble method. This ensemble method has combined biological knowledge in the form of gene sets for the microarray data classification process. Four base classifiers for these approaches are NBS, IB3, SVM, and C4.5 DT. Clinically, GenEnsemble model has been applied for cancer diagnosis such as breast cancer in bi-class classification issue and leukemia in multi-class classification issue. In this study and in the training phase, each gene set was used as the informed feature selection subsets to train base classifiers and then determine their accuracy. Similar to the previous case in 2.2.17 section, this approach uses weighted majority voting for decision-making strategy too. An internal k-fold cross validation strategy was used for each data set, and GenEnsemble was evaluated over the training data. Although, the Naïve Bayes algorithm as base classifier of Bagging or AdaBoost ensembles gave the best results for the three breast cancer datasets but other evidence showed that the proposed approach achieved better performance compared with other popular ensemble algorithms, such as Bagging, Boosting, IB3, SVM, J48, AdaBoost-IB3, AdaBoost-J48, Stacking-IB3 and Stacking-SVM [86].

2.2.2.17 ADASVM method

ADASVM is an ensemble method that has incorporated AdaBoost with linear SVM classifier. It classifies cancers based on microarray gene expression data using Support Vector Machines Ensemble [87]. In this study, the cancer dataset that is selected as a benchmark was leukemia dataset [28]. Leukemia has two sub-types, including AML and ALL. ADASVM is a suitable algorithm for two class problem. This algorithm resolves defects and dilemmas of AdaBoost and SVM. This fusion method has dealt with the diversity of the AdaBoost algorithm. Also, the boosting mechanism was caused to reduce misclassification rate to improve accuracy. It uses weighted majority voting for decision-making strategy. The main measure for evaluation of AdaBoost is weight error of the component, and if it was higher than 0.5, the process is stopped. In this study, researchers showed that the proposed method outperforms than SVM and KNN classifiers. Results showed ADASVM accuracy was 100%. In return, SVM and KNN accuracy were lower, respectively [87].

2.2.2.18 NB (Naïve Bayes) combiner method

The previous investigations showed that NB combiner could be introduced as a fusion strategy in decision integration level. It combines 100 decision tree classifiers [88]. This integration model has been used with 73 benchmark datasets such as breast cancer [55, 56], Arrhythmia [89, 90], Hypothyroid [91, 92]. These datasets belonged to the UCI Machine Learning Repository. The UCI Machine Learning Repository is an open-access database of machine learning problems. In this study, the NB combiner was compared with other three combination methods, including Majority Vote combiner, Weighted Majority Vote combiner, and Recall Combiner (REC). In this study, 10-fold cross-validation was applied. For each cross-validation fold, the training set was divided into two equal parts, including "proper" training and validation. All base classifiers were validated on the proper training part, but combiners were evaluated on the validation part. Of course, estimation of prior probabilities was validated on the whole training data. Results have shown among mentioned methods; NB combiner was the best. Accuracy value of NB combiner was high. This approach has solved problems with a large number of fairly balanced classes while WMV combiner is successful for problems with a small number of imbalanced classes. The previous studies with simulation had demonstrated that NB combiner estimates are inaccurate, but these results did not show such anomalies when data were real and sufficient [88].

2.2.2.19 Collective approach (correlation, color palette, color proportion, and SVM)

The collective ensemble method operates at the decision integration level. It has combined several methods, including correlation method, color palette approach, color proportion method, and SVM classifier. This fusion method determines CEN17 and HER2 biomarkers status by Fluorescence in Situ Hybridization (FISH) images that are important in breast cancer detection as clinical data. This method

uses weighted majority voting for decision-making strategy. Performance of the ensemble approach was confirmed using the statistical evaluation of the mentioned spot recognition system. It was demonstrated that the main advantage of this method is the absolute repeatability of scores in several independent runs. This property is in contrast with human expert decisions which are dependent on mental and physical condition. In this study, the average sensitivity, specificity, and the mean of summed sensitivity and specificity of the proposed fusion approach were compared with different individual methods. Results showed the fusion method gives better efficiency than other methods [93].

2.2.2.20 Rankboost_W

Rankboost _weighting function (Rankboost_W) is a Rankboost algorithm that a heuristic weighting function has been added to it [94]. It is an ensemble method that uses boosting learning techniques for combining different computational approaches as a set of weak features to improve overall performance [95]. Similar to DECORATE method, this approach also has a layer of features integration. It has been applied for gene prioritization related to prostate cancer. In this study that was carried out in 2013, training and test data were genomic data based on mutations for prostate cancer detection. Driver genes and protein-coding genes data were downloaded from Online Mendelian Inheritance in Man (OMIM) and HUGO Gene Nomenclature Committee (HGNC) databases, respectively. It uses algebraic combiners for final decision-making strategy with a novel weighting function. They used the LOOCV method for determining confidence interval estimation. In comparison with other approaches, including ToppGene and ToppNet [87], the performance of the proposed model (Rankboost_W) was better. AUC and mean average precision (MAP) as two performance indicators showed better results for Rankboost_W in comparison with ToppGene method [94].

2.2.2.21 RVM-based ensemble learning

Relevance Vector Machine (RVM) is an ensemble approach that is a combination of AdaBoost and reduced-feature. The RVM has been applied to classify and diagnose different cancers by the construction of a human genetic network. Input data were heterogeneous genomics data such as microarray data. There are three major problems related to heterogeneous genomics data in the construction of a human genetic network. Lack of gold-standard negative set, large-scale learning, and massive missing data values are these problems. This ensemble method has addressed two problems by using kernel-based techniques. AdaBoost helped to solve the problem of large-scale learning, and the reduced-feature model resolved the problem of massive missing data values, which both caused meaningful improvement in performance. 10-fold cross- validation testing was used to evaluate the performance of models. Generally, RVM is an effective approach for encountering the large dimensionality of feature space and the existence of massive missing values. The proposed ensemble method uses algebraic combiners for decision making strategy. In comparison with a robust ensemble approach such as Naïve Bayes Baseline, the proposed model is preferred. Its performance, even with

massive missing data values, is high, and this method can be used for classification tasks in biological datasets [96].

2.2.2.22 PSO–ANN ensemble

particle swarm optimization (PSO)–ANN ensemble is an ensemble method that was used in microarray data classification. The critical point of microarray data analysis is related to the fact that only a few numbers of thousands of genes affect the classification results. This fusion approach has been applied for cancer diagnosis, including leukemia, colon cancer, ovarian cancer, and lung cancer by microarray data classification. PSO-ANN approach has four steps. In step 1 and for gene selection, Fisher-ratio is used. Also, for feature selection and dimension reduction, correlation analysis is employed. In step 2, feature subsets are re-sampled with PSO algorithm, and several base classifiers are trained. In step 3, appropriate base classifiers are chosen. In the step 4, selected base classifiers are combined using an Estimation of Distribution Algorithms (EDAs). In this study, the ANN was used as the base classifiers and was trained with the PSO algorithm. This intelligent ensemble method uses algebraic combiners for decision making. For each data set and evaluation of classification, leave-one-out cross-validation was used. In this investigation, the proposed method was compared with single PSO–ANN, SVM, C4.5, Neuro-fuzzy, and KNN. On the basis of this comparison, the accuracy of classification was improved. Results showed the PSO–ANN ensemble model offers the best overall classification accuracy [97].

2.2.2.23 MF-GE system

The multi-filter enhanced genetic ensemble (MF-GE), hybrid ensemble model, includes two sequentially phases. The first phase consists of a filtering process, and the second phase includes the wrapper process. In phase 1, genes in the microarray dataset were scored using multiple filtering (MF) algorithm and obtained scores were integrated. In the wrapper process, genes were selected with genetic ensemble (GE) algorithm [98]. This approach has been applied to four benchmark microarray datasets for gene selection related to leukemia [31], colon cancer [32], liver cancer [99] and mixed-lineage leukemia (MLL) [100]. This fusion method can be effective for binary-class and multi-class classification problems. Also, the hybrid system overcame the overfitting problem of the GE algorithm. It was used both majority voting and algebraic combiners for decision making, but majority voting generated better classification results. In this study MF-GE system compared with the original GE system and the GA/KNN hybrid. In this study, the double cross-validation process was applied, including internal cross-validation and external cross-validation. The internal cross-validation was done in gene selection phase while the external cross-validation was used for evaluation of selection results. Results showed the proposed approach (MF-GE system) achieved higher classification accuracy value, generated more compact gene subset, and led to the election results more quickly [98].

2.2.2.24 Evolutionary Ensemble Model

In a study, an ensemble model was designed that integrated results of three modules of evolutionary Multilayer Perceptron Neural Networks (MLPNNs). This approach is a parallel ensemble method. Four techniques including polling, maximum, minimum, and weighed average were used for integration, separately. An evolutionary ensemble model is suitable for breast cancer correct diagnosis [101]. Data were taken from Wisconsin Diagnostic Breast Cancer dataset [102, 103] from the UCI Machine Learning Repository that is contained data vectors from 569 patients. About 70% of the total dataset were used as training data using a genetic algorithm, and 30% of data were used as testing data. Each module used algebraic combiner for decision making strategy. Then, voting came into operation between modules. The results demonstrated that Maximum fusion operator generates the best performance when was compared with other fusion technique. The authors of the mentioned work proposed that considering the type of their training method, validation data is not necessary. Accuracy value obtained using the maximum integration operator showed the best performance. Also, sensitivity, specificity, False Positive Rate (FPR), and False Negative Rate (FNR) values were reported [101].

2.2.2.25 Optimized naïve-Bayes model

This classification fusion system is a heuristic algorithm that improves the performance of the naïve-Bayes classifier. It integrates different heterogeneous data, including clinical, laboratory, and flow cytometry. The dataset was included of 112 cases of B-Cell Chronic Lymphocytic Leukemia (B-CLL) patients. The mentioned dataset was obtained from clinical, general laboratory (hematological) examinations and flow cytometry analysis. The proposed method uses algebraic combiner for decision-making strategy. In this study, data classification was done using naïve-Bayes, and performance was evaluated by 10-fold cross-validation. The proposed optimized naïve-Bayes model showed high classification accuracy values. Results demonstrated that including the flow cytometry parameters can improve performance [104].

2.3 Model Integration

Model integration implies when we construct the model, the integration should be done at the model level. In this approach, each model transforms the input data into the format required, and then models are combined. By linking models, a single model comes to decisions to be made based on it. This method can be developed using different tools [105]. One of the tools is based on Bayesian networks that are mentioned in the literature.

2.3.1 Bayesian networks-based model integration (1,2)

In a study experimented in 2006, Bayesian networks have also been used concerning breast cancer. Researchers used three models for integration. In the first model named full integration, they integrated two data sources and then built a Bayesian network based on integrated data and two data sources,

including the clinical and microarray data, were combined. So, at this step, data integration was just done. In the second modal in this approach named decision integration model, an independent model was built for each data source, and then the outcome decision from these models was combined based on weighting policy. In the last modal named partial integration, similar to the second one, and the independent model was developed for each data source, and then the models were linked and integrated for building a single combined model and final decision making. This method used model integration for decision making. The models mentioned above were used for predicting metastatic state in breast cancer. The training set was selected 100 times, randomly and proposed methods performance was evaluated using ROC (Receiver Operator Characteristic) curves analysis. The obtained results revealed that partial integration achieved higher performance and proved to be the best method for data integration [106].

In another study, a novel Bayesian hierarchical model-based method has been proposed. This approach uses single-nucleotide variants (SNVs) and insertions and deletions (InDels) in whole genome sequence data as mutation data [107] obtained from sequencing of the breast cancer cell lines dataset that are available in TCGA [108] and data can be downloaded from https://gdc.cancer.gov/files/public/file/TCGA_mutation_calling_benchmark_files.zip. It first generates two models include of the tumor model and error model by setting partition rules on paired-end reads and datasets, and then this framework integrates these models for mutation calling associated with breast cancer through input data partitioning. So, it is confirmed that the proposed method can improve performance using incorporating heterozygous single nucleotide polymorphisms (SNPs) and strand bias information comparison with other Bayesian network classifiers [107].

Results

There are numerous ensemble methods which yield various research results. In this review, 45 works of literature related to 42 ensemble machine learning methods (Table 1) associated with various cancers were investigated. These 42 approaches can be used for diagnosis, prognosis, or predicting 18 cancer types. We have presented these relations in the following bipartite graph (Fig. 2).

Table 1) 42 ensemble machine learning methods associated with 18 various cancers

Fig. 2) The identified cancers by ensemble learning methods. Note that items marked by an asterisk are a complex disease and they are not a type of cancer

Among these 18 cancer types, breast cancer has been studied more than others, and 22 methods have been introduced for detection purposes. These approaches have been presented in the following pattern (Fig. 3):

Fig. 3) The introduced methods for breast cancer detection

In the previous section, it was mentioned that we have classified learning methods into three famous fusion categories including data and feature integration methods and decision integration methods. On a smaller scale, there is another method called model integration. In most of these methods, we attempted

to introduce data type, number of samples, and decision-making strategies which are used for them. Also, we determined the statistical tools and methods used for their accuracy validation.

Within the category of fusion and level of data and feature integration, nine methods (21.43% of the total) were introduced. These methods were presented in the following pattern (Fig. 4):

Fig. 4) The methods based on data and feature integration. They are BN Classifiers, multiple RFE selection, feature subsets, multimodal data fusion of separate datasets, meta-classifier ensemble learning based on genetic programming technique, ensembles of BioHEL rule sets, Kernel-based data fusion for gene prioritization, stSVM by data integration, and FSCOX-SVM methods

In the next level, decision integration, 32 methods (76.19% of the total) were introduced. They are listed in the 2.2 section, including homogeneous ensemble (seven methods) and heterogeneous ensemble methods (25 methods). These methods are presented in the following pattern [Fig. 5]:

Fig. 5) The methods based on decision integration. Homogeneous ensemble methods are SVM classifiers fusion, three neural networks fusion, CDSS, enSVM, Bagging subgroup identification trees, and NED approaches. Heterogeneous ensemble methods are BNCE, meta-learning, KNNs-SVMs-DT-LDA, MRS, Bagging, artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, BBTWSVM, Ensemble multi-class learning, RSS-SCS, REIS-based ensemble method, MV-ACE, DECORATE, HyDRA, ADASVM, NB combiner, Collective approach (correlation, color palette, color proportion and SVM), Rankboost_W, PSO−ANN ensemble, MF-GE system, Evolutionary Ensemble, CAD system, optimized NB approaches, Stacking IB3-NBS-RF-SVM, and GenEnsemble that is similar to Stacking IB3-NBS-RF-SVM

In the third level, model integration, two studies, one method (2.38 % of the total), have been introduced on the topic of Bayesian networks-based model integration.

The following diagram shows the percentage of studies used at each level of fusion (Fig. 6).

Fig. 6) The number of ensemble learning methods (%) used at each level of fusion

In the majority of these 45 studies, about 200 cases had participated. In two studies, 740, 929, and even 4217 (650 positive and 3567 negative) cases have participated. We also introduced the input data for each machine that have been mentioned in the following pattern (Fig. 7):

In these studies, various data were applied as input data. Input data includes genomic data such as mutation data, gene expression data based on microarray technology, clinical data, GO database data, PPI data, literature data, proteomic data based on protein mass spectrometry technology and multivariate data based on UCI Machine Learning Repository.

Fig.7) Various input data that have been used in different studies

Clinical data is various, including different images, flow cytometry, age, gender, surgery, and cytological diagnoses and histological examination of biopsies (Fig. 8).

Fig. 8) Kinds of clinical data in these studies

It was found that most of the methods have used gene expression data based on microarray technology. 19 methods associated with 21 studies that have used gene expression data have been shown in the following pattern (Fig. 9):

Fig. 9) 19 methods have used gene expression data as input data. The star-labeled methods have used several input data types in addition to the gene expression data. 15 of these 19 methods (33.32% of the 45 studies) have used only gene expression data. These methods are Multiple RFE selection, Feature subsets, Meta-classifier ensemble learning based on Genetic programming technique, Ensembles of BioHEL rule sets, stSVM, FSCOX-SVM, three neural networks fusion, MV-ACE, Stacking IB3-NBS-RF-SVM, GenEnsemble, ADASVM, RVM-based ensemble learning, enSVM, PSO−ANN ensemble, and MF-GE system. Besides, four methods labeled by an asterisk including Bayesian network, Multimodal Data Fusion of Separate Datasets, MRS, and Bayesian networks-based model integration are approaches that used several types of data as multiple input data, at least one of them is gene expression data

Meanwhile, 17 methods associated with 17 studies, which have used clinical data are presented below (Fig. 10):

Fig. 10) 17 mentioned methods in this review have used clinical data as input data. The star-labeled methods have used several input data types in addition to the clinical data. 13 of these 17 methods (28.89% of the 45 studies) including of SVM Classifiers Fusion, Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, Bagging and Boosting-TWSVM, RSS-SCS, REIS-based ensemble, Collective approach (correlation, color palette, color proportion and SVM), Bagging subgroup identification trees, CAD system, optimized naïve-Bayes model, and Neural Ensemble based Detection (NED) have just applied clinical data as input data. Also, four methods marked by an asterisk such as BN Classifiers, Multimodal Data Fusion of Separate Datasets, MRS, and Bayesian networks-based model integration1 have used another data in addition to the clinical data as multiple input data

In the third rank, five methods related to seven studies have used several types of data as multiple input data. These methods have been shown in the following diagram (Fig. 11). One of these methods called Kernel-based data fusion has used GO, SW, PPI network based on STRING database, and literature data as input data sources.

Fig. 11) Five methods (15.56% of the 45 studies) used several data as multiple input data. They are BN Classifiers, Multimodal Data Fusion of Separate Datasets, MRS, Kernel-based data fusion, and Bayesian networks-based model integration1

Note that Bayesian network, Multimodal Data Fusion of Separate Datasets, MRS, and Bayesian networks-based model integration1 are common between methods that have used expression data, clinical data, and multiple data as input data.

In next rank, three methods have used proteomic data based on protein mass spectrometry technology, and approaches that have been done on datasets belonged to the UCI (University of California, Irvine) Machine Learning Repository (Figure 12).

Fig.12) The methods (6.67% of the 45 studies) that have used proteomic data as input data. They are meta-learning, KNNs-SVMs-DT-LDA, and Ensemble multi-class learning

Also, there are three methods that have used data belonged to datasets, such as the UCI Machine Learning Repository (Fig. 13).

Fig. 13) Three methods (6.67% of the 45 studies) have used various data. This data is belonged to different datasets, such as the UCI Machine Learning Repository. They include BNCE, NB combiner, and Evolutionary Ensemble Model

Finally, four methods have used mutation data as input data that have been shown below (Fig. 14):

Fig. 14) The methods (8.89% of the 45 studies) that have used somatic mutation data. They are HyDRA and DECORATE, Rankboost_W, and Bayesian networks-based model integration 2

The following diagram shows the kinds of input data used in these 45 studies by percentage (Fig. 15).

Fig. 15) The number of ensemble learning methods (%) that have used different kinds of input data

Furthermore, there are several decision-making strategies for 32 decision integration-based ensemble methods. These strategies are simple majority voting, full voting, weighted majority voting, algebraic combiners, and decision templates. At this review, the number of methods that have been decided with the majority voting strategy roughly equal to those methods that have been decided with algebraic combiner strategy. 13 methods (40.63% of the 32 studies) have made a decision with algebraic combiner strategy (Fig. 16).

Fig. 16) The methods that have been decided based on majority voting strategy. These methods are: MRS, Boosting-TWSVM, BBTWSVM, REIS-based ensemble, MV-ACE, DECORATE, NB combiner, Rankboost_W, RVM-based ensemble learning, PSO–ANN ensemble, MF-GE system, Evolutionary Ensemble Model, and optimized NB model

And 12 methods (37.50% of the 32 studies) have been decided with the majority voting strategy (Fig. 17).

Fig. 17) The methods that have been decided based on majority voting strategy. They are BNCE, KNNs-SVMs-DT-LDA, SVM Classifiers Fusion (three SVM), Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Ensemble multi-class learning, three neural networks fusion, RSS-SCS, CDSS, Bagging subgroup identification trees, enSVM, and CAD system

Also, five methods (15.63% of the 32 studies) called meta-learning, Stacking IB3-NBS-RF-SVM, GenEnsemble, ADASVM, and Collective approach (correlation, color palette, color proportion, and SVM)

have been decided with weighted majority voting strategy (Fig. 18).

Fig. 18) The methods that have been decided based on the weighted majority voting strategy

Full voting and decision templates strategies were only used for NED (3.13% of the 32 studies) and HyDRA (3.13% of the 32 studies) methods, respectively.

The number of methods (%) that have used different kinds of decision-making strategy has been shown in the following diagram (Fig. 19):

Fig. 19) The number of methods (%) that have decided with different strategies

Evaluation of performance for these mentioned ensemble methods were carried out by several statistical validation techniques such as kinds of cross-validation forms, ROC curves analysis, level-by-level (multi-level reproducibility) validation method, subsampling as an independent validation (repeatability of scores in several independent runs), determination of error in percentage (percentage error) and calculation of weight error. Kinds of cross-validation forms include: The k-fold cross-validation, leave-one-out cross-validation, two-level external cross-validation and double cross-validation including internal cross-validation and external cross-validation. According to the findings of these studies, the accuracy of results for 35 methods associated with 36 studies (80% of the 45 studies) has been evaluated using a variety of cross-validation methods (Fig. 20).

Fig. 20) The methods that have been evaluated using a variety of cross-validation methods (k-fold cross-validation). These methods include: Two of three Bayesian networks (Bayesian network 1 and Bayesian network 2), Multimodal Data Fusion of Separate Datasets, meta-classifier ensemble learning based on Genetic programming technique, Ensembles of BioHEL rule sets, Kernel-based data fusion, stSVM, FSCOX-SVM, meta-learning, KNNs-SVMs-DT-LDA, SVM Classifiers Fusion (three SVM), MRS, Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, Bagging and Boosting-TWSVM, Ensemble multi-class learning, three neural networks fusion, RSS-SCS, REIS-based ensemble, MV-ACE, DECORATE, HyDRA, Stacking IB3-NBS-RF-SVM, GenEnsemble, NB combiner, CDSS, Rankboost_W, Bagging subgroup identification trees, RVM based ensemble learning, enSVM, PSO−ANN ensemble, MF-GE system, CAD system, optimized NB model, Ensemble multi-class learning, and NED

In parallel, three methods (6.67% of the 45 studies) including Multiple RFE selection 2, Feature subsets and Collective approach (correlation, color palette, color proportion, and SVM) were validated by the statistical test of repeatability of scores and results in several independent runs. Respectively, Bayesian networks-based model integration1,2(4.44% of the 45 studies) and Multiple RFE selection1 (2.22% of the 45 studies) use ROC curves analysis and multi-level reproducibility (level-by-level) validation framework for statistical validation. Also, weight error and percentage error formula have been used for ADASVM (2.22% of the 45 studies) and BNCE (2.22% of the 45 studies) methods, respectively. These classifications have been shown in the following graph (Fig. 21):

Fig. 21) Different statistical validation methods for evaluation of the performance of six ensemble learning methods

About one of the methods (2.22% of the 45 studies) called Evolutionary Ensemble model has been emphasized that considering the type of their training method, validation data is not necessary.

The number of methods (%) that use any kind of statistical method for validation has been shown in the following diagram (Fig. 22).

Fig. 22) The number of ensemble learning methods (%) validated with kinds of statistical methods

Discussion

In this study, the trail of genesis and the use of ensemble learning methods in medical studies and cancer detection were researched. It was mentioned that in 1996, bagging approach was applied on a breast cancer dataset to help identify and classify benign and malignant samples. Since then, the use of ensemble systems has dramatically expanded in medical studies and diagnosis of cancers. After all these studies and classification of ensemble learning methods associated with detection of cancers, one key question remains to be answered: which proposed fusion method is the best?

Results demonstrate that there is no best and superior approach for all classification problems, and the optimum solution depends on the kind of problem, the structure of the available data and prior knowledge about the related algorithm. So, this review can help researchers choose the most suitable ensemble method in the field of molecular biology and complex diseases diagnosis. However, in this study, it was found that Ensembles of BioHEL rule sets outperform on large datasets. Besides, the fusion of KNNs-SVMs-DT-LDA as heterogeneous ensemble method can provide superior performance specifically on small size datasets. RVM-based ensemble learning is a more effective approach to solve high-dimensionality problems of feature space without missing data values. As compared to other methods, ADASVM method seems to be the most appropriate method for improving the classification performance by applying diversity. ADASVM is based on the AdaBoost algorithm. Finally, NB combiner approach is successful for problems with a large number of fairly balanced classes, and WMV combiner method can solve problems with a small number of imbalanced classes.

Meanwhile, a careful look at Figures 9 and 10, it can be inferred that clinical data and microarray data have been used over time. However, due to advances in equipment and imaging methods, the trend of using clinical data has changed. For example, digital images data were used as clinical data in recent research. Also, by the development of bio-molecular technique such as Next Generation Sequencing (NGS), the massive genomic data, including somatic and germline mutation data, has been generated. The use of this data has revolutionized the diagnosis of complex diseases such as cancers. These aspects of genomic diagnostics contribute to the prosperity of modern sciences, including personalized medicine and precision oncology [109].

However, one of our limitations in classifying and introducing cancers-based ensemble methods was the low number of approaches based on model integration, and so these categories of methods were studied on a much smaller scale.

Conclusions

The authors hope that in subsequent publications, they will further introduce the applications of multi-omics data integration in precision oncology. In this paper, due to the complexity and challenges involved in integrating omics data, these types of fusion system have not been thoroughly investigated. In summary, multi-omics approaches have been used for identifying cancers and their subtypes [110]. These methods integrate multi-omics data to diagnose cancers effectively.

Keyword search results, such as "multi-omics data" and "cancer" in Scopus show that the trend of using the multi-omics data as input data in cancer researches has increased significantly from 2006 to 2018, especially in 2017 and 2018.

The first multi-omics database called "LinkedOmics database" includes more than a billion data points. This database contains clinical data and multi-omics (genomics, epigenomics, metagenomics, transcriptomics, and proteomics) data. This data consists of 32 cancer types in 11158 patients from TCGA project [111]. It can help researchers who want to use this data for solving medical problems.

So, for robust performance and to get better results, we especially suggest using multi-omics approaches for integration of different omics data types such as genomics data, epigenomics, transcriptomics data, proteomics data, metabolomics, and microbiomics in different layer levels [112]. This kind of data integration is our next plan for developing a novel and robust predictive fusion method for identifying driver genes as biomarkers, especially, for breast invasive carcinoma (BRCA).

Finally, using machine learning methods is less expensive than bio-molecular testing. It can help reduce the magnitude of the search space. There is hope that this study can provide a roadmap to help researchers in using appropriate ensemble methods in the field of molecular biology and diagnosis of complex diseases. However, for better evaluation of the accuracy of results, it is suggested that proper bio-molecular techniques such as quantitative reverse transcription-polymerase chain reaction (qRT-PCR) and Whole Exome Sequencing (WES) based on NGS experiments are applied.

Abbreviations

ADCA: Adenocarcinoma; ALL: Adult acute lymphoblastic leukemia; AM: Arithmetic Mean; AML: Acute myeloid leukemia; ANNs: Artificial Neural Networks; AUC: Area under the curve; BBTWSVM: Bagging and boosting based twin support vector machine; B-CLL: B-Cell Chronic Lymphocytic Leukemia; BioHEL: Bioinformatics-Oriented Hierarchical Learning; BKS: Behavior knowledge space; BN: Bayesian network; BRCA: Breast invasive carcinoma; CAD: Computer-Aided Diagnostic; CART: Classification and Regression

Tree; CCL: Combining class labels; CCO: Combining continuous outputs; CDSS: Clinical Decision Support System; CFS: Correlation-Based Feature Selection; CFS: Correlation-Based Feature Selection; COD: Combination of data; COI: Combination of interpretations; COMBINER: COre Module Biomarker Identification with Network ExploRation; CT: Computed Tomography; CxCa: Cervical Cancer; DDSM: Digital Database for Screening Mammography; DECORATE: Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples; DLDA: Diagonal Linear Discriminant Analysis; ECOC: Error-correcting output coding; EDAs: Estimation of Distribution Algorithms; EMB: Extended Markov blanket; END: Ensemble of Nested Dichotomies; enSVM: Ensemble SVM; EPNet: Evolutionary Programming Network; FANNC: Fast Adaptive Neural Network Classifier; FBN: Full Bayesian network; FISH: Fluorescence in Situ Hybridization; FNR: False Negative Rate; FPR: False Positive Rate; FSCOX: Feature selection with Cox proportional hazard regression model; GAssist: Genetic Algorithms based classifier system; GBM: Glioblastoma Multiforme; GE: genetic ensemble; GMM: Gaussian mixture models; GO: Gene Ontology; GP: Genetic Programming; HCC: Hepatocellular carcinoma; HDSS: High Dimensionality Small Sample; HGNC: HUGO Gene Nomenclature Committee; HPV: Human Papilloma Virus; HTTs: High-throughput technologies; HyDRA: Hybrid Distance-score Rank Aggregation; IB3: Instance Based3; K*: K* instance; KNN: K-Nearest Neighbors; LBL: Lawrence Berkeley Laboratory; LDA: Linear discriminant analysis; LogE: Log-Euclidean Mean; LOOCV: Leave-One-Out Cross-Validation; MA: MicroArray; MALDI-TOF: Matrix-assisted laser desorption/ionization time-of-flight; MAP: Mean average precision; MB: MultiBoosting; MCC: Matthew's correlation coefficient; MCs: MicroCalcifications; MCS: Multi-classification system; MF: Multiple filtering; MF-GE: Multi-filter enhanced genetic ensemble; MLL: Mixed-lineage leukemia; MLP: Multi-Layer Perceptron Network; MLPNNs: Multilayer Perceptron Neural Networks; MPC: Metastatic Prostate Cancer; MPM: Malignant pleural mesothelioma; MRS: Mixture of Rough set and SVM; MS: Mass spectrometry; MV: Majority voting; MV-ACE: Multi-view based AdaBoost classifier ensemble; NB: Naïve Bayes; NBS: Naïve Bayes Simple; NC: Negative correlation; NED: Neural Ensemble based Detection; NGS: Next Generation Sequencing; NHGRI: National Human Genome Research Institute; NIH: National Cancer Institute, National Institutes of Health; NNs: Neural networks; NPV: Negative Predictive Value; OMIM: Online Mendelian Inheritance in Man; PAM: Prediction Analysis of Microarrays; PART: Projective Adaptive Resonance Theory; PCA: Principal component analysis; PLSS: Partial-Least-Squares Based Feature Selection; PNN: Probabilistic Neural Network; PPC: Primary Prostate Cancer; PPI: Protein-Protein Interactions; PSA: Prostate-Specific Antigen; PSO: Particle swarm optimization; PWC: Pairwise coupling; qRT-PCR: quantitative reverse transcription-polymerase chain reaction; Rankboost_W: Rankboost _weighting function; RBF: Radial Basis Function; REC: Recall Combiner; REIS: Resonance-frequency Electrical Impedance Spectroscopy; RF: Random forest; RFE: Recursive Feature Elimination; RFS: Random Forest based Feature Selection; ROC: Receiver operating characteristic; RSS: Random Subspace; RVM: Relevance Vector Machine; SAM: Significant Analysis for Microarrays; SBE: Sequential backward elimination; SCS: Static Classifiers Selection; sgSVM: Saliency guidedSVM; SMO: Sequential Minimal Optimization; stSVM: Smoothed t-statistic SVM; SVM: Support Vector machine; SW: Swiss-prot; TCGA: The Cancer Genome Atlas; TP: True Positive; TPR: True-Positive Rate; TWSVM: Twin SVM; WES: Whole Exome Sequencing; W-LogE: weighted-version of LogE; WMFR: Weighted Median Fusion Rule; WMV: Weighted majority voting; WRFs: Weighted Random Forest; WSFR: Weighted Sum Fusion Rule

Declarations

Additional files

Not applicable

Acknowledgements

Not applicable

Funding

Availability of data and materials

Please contact the author for data requests.

Authors' contributions

KK developed the concept, designed the study and participated in project supervision. LM performed the comprehensive survey of related publications. KK, SRBA, and LM reviewed the articles. LM wrote the manuscript and contributed to visualize results. LM and KK contributed to the interpretation of the data and discussion. KK, RHH, AMBM, and SRBA gave final approval for publication.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1] Department of Biology, Faculty of Science, Payame Noor University, Tehran, Iran. [2] Laboratory of Genomics and Epigenomics (LGE), Department of Biochemistry, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran. [3] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran,

Tehran, Iran. [4] Islamic Azad University, North Tehran Branch, Faculty of Foreign Languages, ELT and Bio-Translation Studies Department.

References

1. Yang MQ, Yang JY. An investigation into the feasibility of detecting microscopic disease using machine learning. In: Bioinformatics and Biomedicine Workshops, 2007 BIBMW 2007 IEEE International Conference on. IEEE; 2007. p. 1–2.

2. Baronti F, Micheli A, Passaro A, Starita A. Machine learning contribution to solve prognostic medical problems. Outcome Predict Cancer. 2006;261.

3. Rokach L. Ensemble-based classifiers. Artif Intell Rev. 2010;33(1–2):1–39.

4. Kuncheva LI, Whitaker CJ. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach Learn. 2003;51(2):181–207.

5. Kourou K, Exarchos TP, Exarchos KP, Karamouzis M V, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8–17.

6. Polikar R. Ensemble based systems in decision making. Circuits Syst Mag IEEE. 2006;6(3):21–45.

7. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. Neural Comput. 1991;3(1):79–87.

8. Dietterich TG. Ensemble methods in machine learning. In: Multiple classifier systems. Springer; 2000. p. 1–15.

9. Chen Y, Chen JJ. Ensemble survival trees for identifying subpopulations in personalized medicine. Biometrical J. 2016;58(5):1151-63.

10. Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123–40.

11. Ma S, Ren J, Fenyö D. Breast cancer prognostics using multi-omics data. AMIA Summits Transl Sci Proc. 2016;2016:52.

12. Wang J, Zuo Y, Liu L, Man Y, Tadesse MG, Ressom HW. Identification of functional modules by integration of multiple data sources using a bayesian network classifier. Circ Cardiovasc Genet. 2014;7(2):206–17.

13. Franke L, Bakel H Van, Fokkens L, Jong ED De, Egmont-petersen M, Wijmenga C. Reconstruction of a Functional Human Gene Network, with an Application for Prioritizing Positional Candidate Genes. 2006;78(June):1011–25.

14. Cun Y, Fröhlich H. Network and data integration for biomarker signature discovery via network smoothed t-statistics. PLoS One. 2013;8(9):e73074.

15. Ivshina A V, George J, Senko O, Mow B, Putti TC, Smeds J, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res. 2006;66(21):10292–301.

16. Sun Y, Goodison S. Optimizing molecular signatures for predicting prostate cancer recurrence. Prostate. 2009;69(10):1119–27.

17. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18(1):11–22.

18. Network CGAR. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609.

19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci. 2001;98(9):5116–21.

20. Kim S, Park T, Kon M. Cancer survival classification using integrated data sets and intermediate information. Artif Intell Med. 2014;62(1):23–31.

21. Yang R, Daigle BJ, Petzold LR, Doyle FJ. Core module biomarker identification with network exploration for breast cancer metastasis. BMC Bioinformatics. 2012;13(1):1.

22. Drosos C, Bisdounis L, Metafas D, Blionas S, Tatsaki A. A multi-level validation methodology for wireless network applications. In: International Workshop on Power and Timing Modeling, Optimization and Simulation. Springer; 2004. p. 332–41.

23. Zhang R, Tian J, Li Z, Su H, Chen S, Tang X. Principles and methods for the validation of quantitative remote sensing products. Sci China Earth Sci. 2010;53(5):741–51.

24. Van De Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347(25):1999–2009.

25. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005;365(9460):671–9.

26. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin cancer Res. 2007;13(11):3207–14.

27. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics. 2010;26(3):392–8.

28. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (80-).

1999;286(5439):531–7.

29. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci. 1999;96(12):6745–50.

30. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503.

31. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell. 2002;1(2):203–9.

32. Djebbari A, Liu Z, Phan S, Famili F. An ensemble machine learning approach to predict survival in breast cancer. Int J Comput Biol Drug Des. 2008;1(3):275–94.

33. Van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530–6.

34. Moutselos K, Maglogiannis I, Chatziioannou A. Integration of High-Volume Molecular and Imaging Data for Composite Biomarker Discovery in the Study of Melanoma. Biomed Res Int. 2014;2014.

35. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. Med Image Anal. 2013;17(2):219–35.

36. Yu J, Yu J, Almal AA, Dhanasekaran SM, Ghosh D, Worzel WP, et al. Feature selection and molecular classification of cancer using genetic programming. Neoplasia. 2007;9(4):292-IN3.

37. Glaab E, Bacardit J, Garibaldi JM, Krasnogor N. Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data. PLoS One. 2012;7(7):e39932.

38. Tsuda K, Uda S, Kin T, Asai K. Minimizing the cross validation error to mix kernel matrices of heterogeneous biological data. Neural Process Lett. 2004;19(1):63–72.

39. Zakeri P, Elshal S, Moreau Y. Gene prioritization through geometric-inspired kernel data fusion. In: Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on. IEEE; 2015. p. 1559–65.

40. Gilpin SA, DUNLAVY DM. Heterogeneous ensemble classification. CSRI SUMMER Proc 2008. 2008;90.

41. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. The digital database for screening mammography. In: Proceedings of the 5th international workshop on digital mammography. Medical Physics Publishing; 2000. p. 212–8.

42. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, et al. Current status of the digital database for screening mammography. In: Digital mammography. Springer; 1998. p. 457–60.

43. Azizi N, Tlili-Guiassa Y, Zemmal N. A computer-aided diagnosis system for breast cancer combining features complementarily and new scheme of SVM classifiers fusion. Int J Multimed Ubiquitous Eng. 2013;8(4):45–58.

44. Peng Y. A novel ensemble machine learning for robust microarray data classification. 2006;36:553–73.

45. Peng Y. Integration of gene functional diversity for effective cancer detection. Int J Syst Sci. 2006;37(13):931–8.

46. Liu B, Cui Q, Jiang T, Ma S. A combinational feature selection and ensemble neural network method for classification of gene expression data. BMC Bioinformatics. 2004;5(1):1.

47. Zhou Z, Chen S, Chen Z. FANNC: A fast adaptive neural network classifier. Knowl Inf Syst. 2000;2(1):115–29.

48. Zhou Z-H, Jiang Y, Yang Y-B, Chen S-F. Lung cancer cell identification based on artificial neural network ensembles. Artif Intell Med. 2002;24(1):25–36.

49. Bountris P, Haritou M, Pouliakis A, Karakitsos P, Koutsouris D. A decision support system based on an ensemble of random forests for improving the management of women with abnormal findings at cervical cancer screening. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015. p. 8151–6.

50. Therneau T. A Package for Survival Analysis in S. version 2.38. 2015.

51. Zhu C-Q, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non–small-cell lung cancer. J Clin Oncol. 2010;28(29):4417.

52. Kourou K, Exarchos TP, Exarchos KP, Karamouzis M V, Fotiadis DI. Artificial Intelligence Methods Application in Liver Diseases Classification from CT Images 1 Introduction 2 Methods. Comput Struct Biotechnol J. 2015;13:8–17.

53. Reid S. A review of heterogeneous ensemble methods. Dep Comput Sci Univ Color Boulder. 2007;

54. Alam KMR, Islam MM. Combining boosting with negative correlation learning for training neural network ensembles. In: 2007 International Conference on Information and Communication Technology. IEEE; 2007. p. 68–71.

55. UCI Machine Learning Repository, Breast Cancer Data Set. University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia, Zwitter M, Soklic M. 1986. https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/breast-cancer.names. Accessed 11 Jul 1988.

56. Michalski RS, Mozetic I, Hong J, Lavrac N. The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. Proc AAAI 1986. 1986;1–41.

57. Bhanot G, Alexe G, Venkataraghavan B, Levine AJ. A robust meta-classification strategy for cancer detection from MS data. Proteomics. 2006;6(2):592–604.

58. Loh W. Classification and regression trees. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(1):14–23.

59. Assareh A, Moradi MH, Esmaeili V. A novel ensemble strategy for classification of prostate cancer protein mass spectra. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2007. p. 5987–90.

60. Petricoin III EF, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, et al. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst. 2002;94(20):1576–8.

61. Zeng T, Liu J. Mixture classification model based on clinical markers for breast cancer prognosis. Artif Intell Med. 2010;48(2):129–37.

62. Edén P, Ritz C, Rose C, Fernö Ma, Peterson C. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. Eur J Cancer. 2004;40(12):1837–41.

63. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo W-L, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006;10(6):529–41.

64. Grzesiak-Kopeć K, Ogorzałek M, Nowak L. Computational Classification of Melanocytic Skin Lesions. In: International Conference on Artificial Intelligence and Soft Computing. Springer; 2016. p. 169–78.

65. Sarwar A, Suri J, Ali M, Sharma V. Novel benchmark database of digitized and calibrated cervical cells for artificial intelligence based screening of cervical cancer. J Ambient Intell Humaniz Comput. 2016;1–14.

66. Frank E, Kramer S. Ensembles of nested dichotomies for multi-class problems. In: Proceedings of the twenty-first international conference on Machine learning. ACM; 2004. p. 39.

67. Webb GI. Decision tree grafting from the all-tests-but-one partition. In: Ijcai. 1999. p. 702–707.

68. Cao Y, Wu J. Dynamics of projective adaptive resonance theory model: the foundation of PART algorithm. IEEE Trans Neural Networks. 2004;15(2):245–60.

69. Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans neural networks. 1991;2(2):302–9.

70. Zhang X. Boosting twin support vector machine approach for MCs detection. In: Information Processing, 2009 APCIP 2009 Asia-Pacific Conference on. IEEE; 2009. p. 149–52.

71. Khemchandani R, Chandra S. Twin support vector machines for pattern classification. IEEE Trans Pattern Anal Mach Intell. 2007;29(5):905–10.

72. Zhang X, Gao X, Wang M. MCs detection approach using Bagging and Boosting based twin support vector machine. In: Systems, Man and Cybernetics, 2009 SMC 2009 IEEE International Conference on. IEEE; 2009. p. 5000–505.

73. Zhang X, Chen Y, Luo Z. Microcalcification Clusters Detection with Bagging and Boosting based Twin Support Vector Machine. Int Inf Inst (Tokyo) Inf. 2012;15(1):67.

74. Oh JH, Kim YB, Gurnani P, Rosenblatt KP, Gao JX. Biomarker selection and sample prediction for multi-category disease on MALDI-TOF data. Bioinformatics. 2008;24(16):1812–8.

75. Ressom HW, Varghese RS, Drake SK, Hortin GL, Abdel-Hamid M, Loffredo CA, et al. Peak selection from MALDI-TOF mass spectra using ant colony optimization. Bioinformatics. 2007;23(5):619–26.

76. Cheriguene S, Azizi N, Zemmal N, Dey N, Djellali H, Farah N. Optimized Tumor Breast Cancer Classification Using Combining Random Subspace and Static Classifiers Selection Paradigms. In: Applications of Intelligent Optimization in Biology and Medicine. Springer; 2016. p. 289–307.

77. Lederman D, Wang X, Zheng B, Sumkin JH, Tublin M, Gur D. Fusion of classifiers for REIS-based detection of suspicious breast lesions. In: SPIE Medical Imaging. International Society for Optics and Photonics; 2011. p. 79661C-79661C.

78. Li L, Yu Z, Liu J, You J, Wong H-S, Han G. Multi-view based adaboost classifier ensemble for class prediction from gene expression profiles. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE; 2014. p. 178–83.

79. Liu Y, Tian F, Hu Z, DeLisi C. Evaluation and integration of cancer gene classifiers: identification and ranking of plausible drivers. Sci Rep. 2015;5.

80. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70.

81. Kim M, Farnoud F, Milenkovic O. HyDRA: gene prioritization via hybrid distance-score rank aggregation. Bioinformatics. 2015;31(7):1034–43.

82. Kemeny JG. Mathematics without numbers. Daedalus. 1959;88(4):577–91.

83. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, et al. Gene prioritization through genomic data fusion. Nat Biotechnol. 2006;24(5):537–44.

84. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37(suppl 2):W305–11.

85. Reboiro-Jato M, Glez-Peña D, Díaz F, Fdez-Riverola F. A novel ensemble approach for multicategory classification of DNA microarray data using biological relevant gene sets. Int J Data Min Bioinform. 2012;6(6):602–16.

86. Reboiro-Jato M, Díaz F, Glez-Peña D, Fdez-Riverola F. A novel ensemble of classifiers that use biological relevant gene sets for microarray classification. Appl Soft Comput. 2014;17:117–26.

87. Begum S, Chakraborty D, Sarkar R. Cancer classification from gene expression based microarray data using SVM ensemble. In: 2015 International Conference on Condition Assessment Techniques in Electrical Systems (CATCON). IEEE; 2015. p. 13–6.

88. Kuncheva LI, Rodríguez JJ. A weighted voting framework for classifiers ensembles. Knowl Inf Syst. 2014;38(2):259–75.

89. UCI Machine Learning Repository, Arrhythmia Data Set. Bilkent University, Baskent University, Ankara, Turkey, Guvenir HA, Acar B, Muderrisoglu H. 1997. https://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia/arrhythmia.names. Accessed 01 Jan1998.

90. Guvenir HA, Acar B, Demiroz G, Cekin A. A supervised machine learning algorithm for arrhythmia analysis. In: Computers in Cardiology 1997. IEEE; 1997. p. 433–6.

91. UCI Machine Learning Repository, Thyroid Data Set. Garavan Institute in Sydney, Australia, Quinlan JR. 1986. https://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/HELLO. Accessed 01 Jan 1987.

92. Quinlan JR. Induction of Decision Trees. Machine Learning 1. I. 1986;8(1):106.

93. Les T, Markiewicz T, Osowski S, Kozlowski W, Jesiotr M. Fusion of FISH image analysis methods of HER2 status determination in breast cancer. Expert Syst Appl. 2016;61:78–85.

94. Lee P-F, Soo V-W. An ensemble rank learning approach for gene prioritization. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2013. p. 3507–10.

95. Freund Y, Schapire RE. An Efficient Boosting Algorithm for Combining Preferences. 2003;4:933–69.

96. Wu C-C, Asgharzadeh S, Triche TJ, D'Argenio DZ. Prediction of human functional genetic networks from heterogeneous data using RVM-based ensemble learning. Bioinformatics. 2010;26(6):807–13.

97. Chen Y, Zhao Y. A novel ensemble of classifiers for microarray data classification. 2008;8:1664–9.

98. Yang P, Zhou BB, Zhang Z, Zomaya AY. A multi-filter enhanced genetic ensemble system for gene selection and sample classification of microarray data. BMC Bioinformatics. 2010;11(1):1.

99. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, et al. Gene expression patterns in human liver cancers. Mol Biol Cell. 2002;13(6):1929–39.

100. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet. 2001;30(1):41.

101. Janghel RR, Shukla A, Sharma S, Gnaneswar A V. Evolutionary Ensemble Model for Breast Cancer Classification. In: International Conference in Swarm Intelligence. Springer; 2014. p. 8–16.

102. UCI Machine Learning Repository, Breast Cancer Wisconsin (Diagnostic) Data Set. University of Wisconsin, Wolberg WH, Street WN, Mangasarian OL. 1993. https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.names. Accessed 01 Nov 1995.

103. Street WN, Wolberg WH, Mangasarian OL. Nuclear feature extraction for breast tumor diagnosis. In: Biomedical image processing and biomedical visualization. International Society for Optics and Photonics; 1993. p. 861–71.

104. Lakoumentas J, Drakos J, Karakantza M, Sakellaropoulos G, Megalooikonomou V, Nikiforidis G. Optimizations of the naïve-Bayes classifier for the prognosis of B-Chronic Lymphocytic Leukemia incorporating flow cytometry data. Comput Methods Programs Biomed. 2012;108(1):158–67.

105. Belete GF, Voinov A, Laniak GF. An overview of the model integration process: From pre-integration assessment to testing. Environ Model Softw. 2017;87:49–63.

106. Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics. 2006;22(14):e184–90.

107. Moriyama T, Imoto S, Hayashi S, Shiraishi Y, Miyano S, Yamaguchi R. A Bayesian model integration for mutation calling through data partitioning. Bioinformatics. 2019.

108. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109–12.

109. Sheikine Y, Kuo FC, Lindeman NI. Clinical and technical aspects of genomic diagnostics for precision oncology. J Clin Oncol. 2017;35(9):929–33.

110. Guo Y, Zheng J, Shang X, Li Z. A Similarity Regression Fusion Model for Integrating Multi-Omics Data to Identify Cancer Subtypes. Genes (Basel). 2018;9(7):314.

111. Vasaikar S V, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. 2017;46(D1):D956–63.

112. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18(1):83.

Table

Table 1) 42 ensemble machine learning methods associated with 18 various cancers

Methods number	42 ensemble methods name	Number of studies that used these 45 methods
1	BN classifier (1,2)	2
2	Multiple RFE selection method (1,2)	2
3	Feature subsets method	1
4	Multimodal data fusion	1
5	Meta-classifier ensemble learning based on GP technique	1
6	Ensembles of BioHEL rule sets	1
7	Kernel-based data fusion for gene prioritization	1
8	stSVM (smoothed t-statistic SVM)	1
9	FSCOX-SVM	1
10	enSVM	1
11	SVM classifiers fusion (three SVM)	1
12	CDSS	1
13	Bagging subgroup identification trees	1
14	NED method	1
15	Three NNs fusion	1
16	BNCE	1
17	Meta-learning	1
18	KNNs-SVMs-DT-LDA	1
19	MRS	1
20	Bagging (Bootstrap Aggregation)	1
21	Artificial intelligence-based hybrid ensemble technique	1
22	Boosting-TWSVM	1
23	BBTWSVM	1
24	Ensemble multi-class learning	1
25	RSS-SCS	1
26	REIS-based ensemble method	1
27	MV-ACE	1
28	DECORATE	1
29	HyDRA	1
30	ADASVM	1
31	NB combiner	1
32	Collective approach	1
33	RVM	1
34	Rankboost_W	1
35	PSO–ANN ensemble	1
36	MF-GE system	1
37	Evolutionary Ensemble model	1
38	CAD system	1
39	Optimized NB model	1
40	Stacking-IB3, NBS, RF, SVM	1
41	GenEnsemble (NBS-IB3-SVM-C4.5 DT)	1
42	BNs-based model integration (1,2)	2

Figures
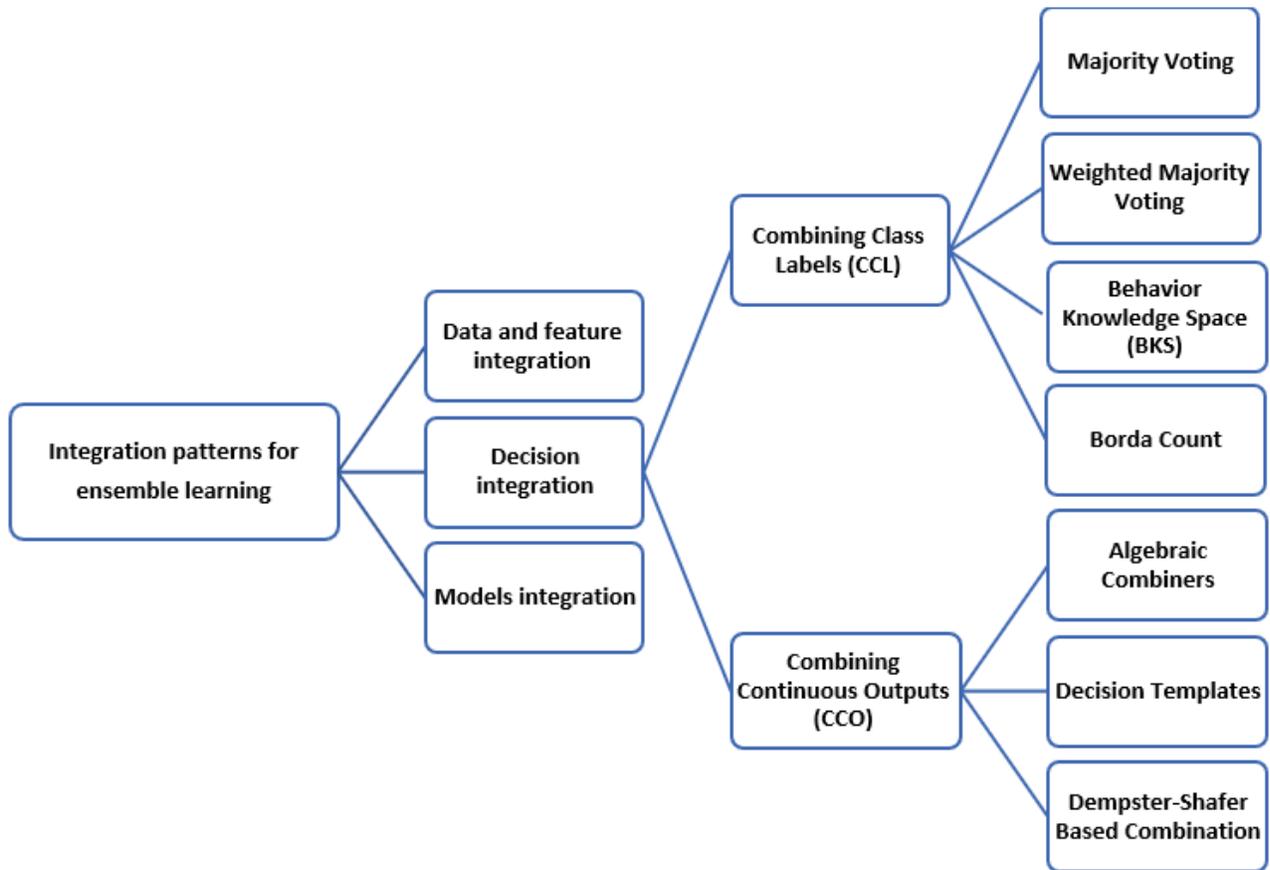
Figure 1

Integration patterns for ensemble learning
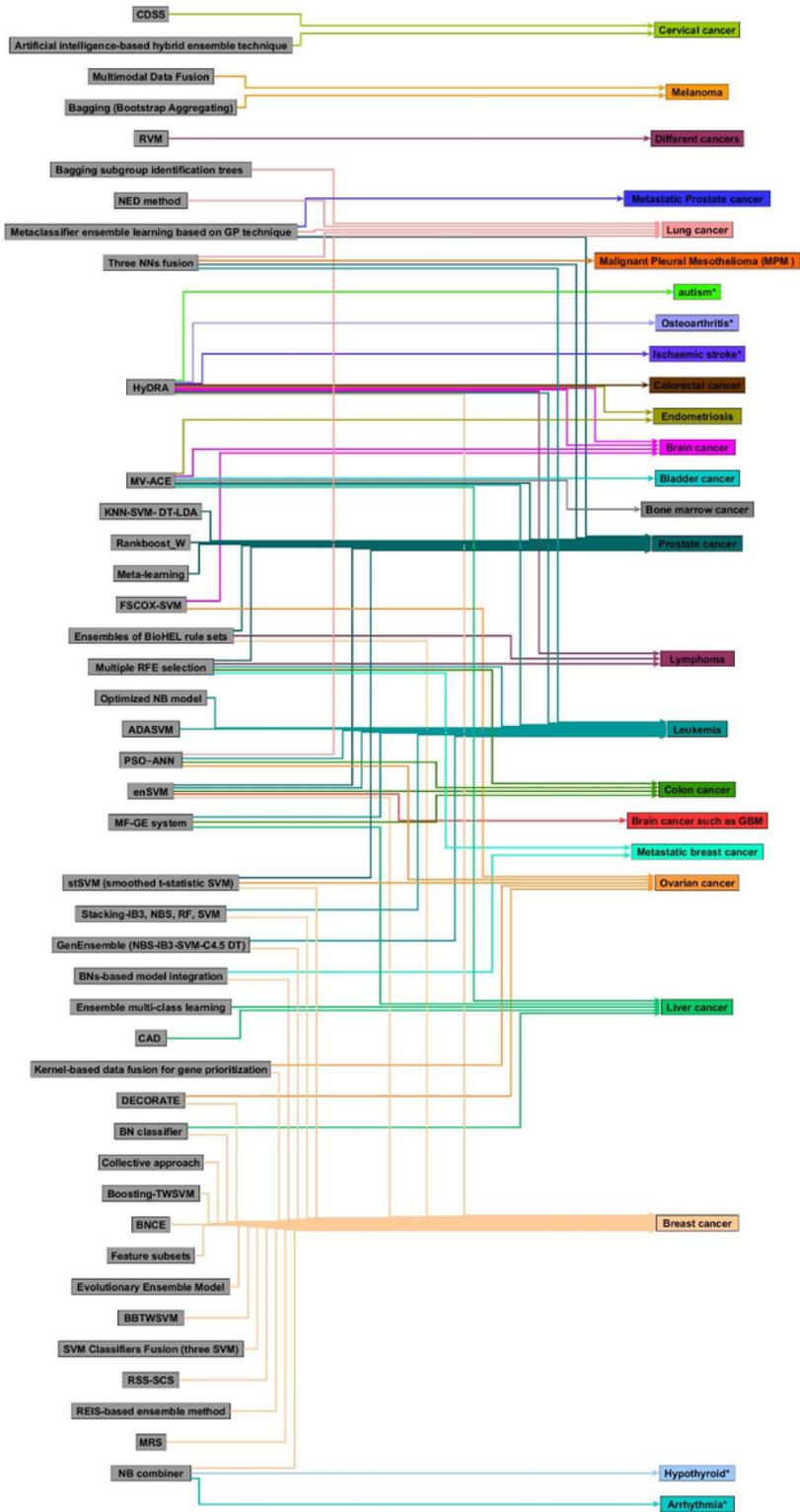
Figure 2

The identified cancers by ensemble learning methods. Note that items marked by an asterisk are a complex disease and they are not a type of cancer
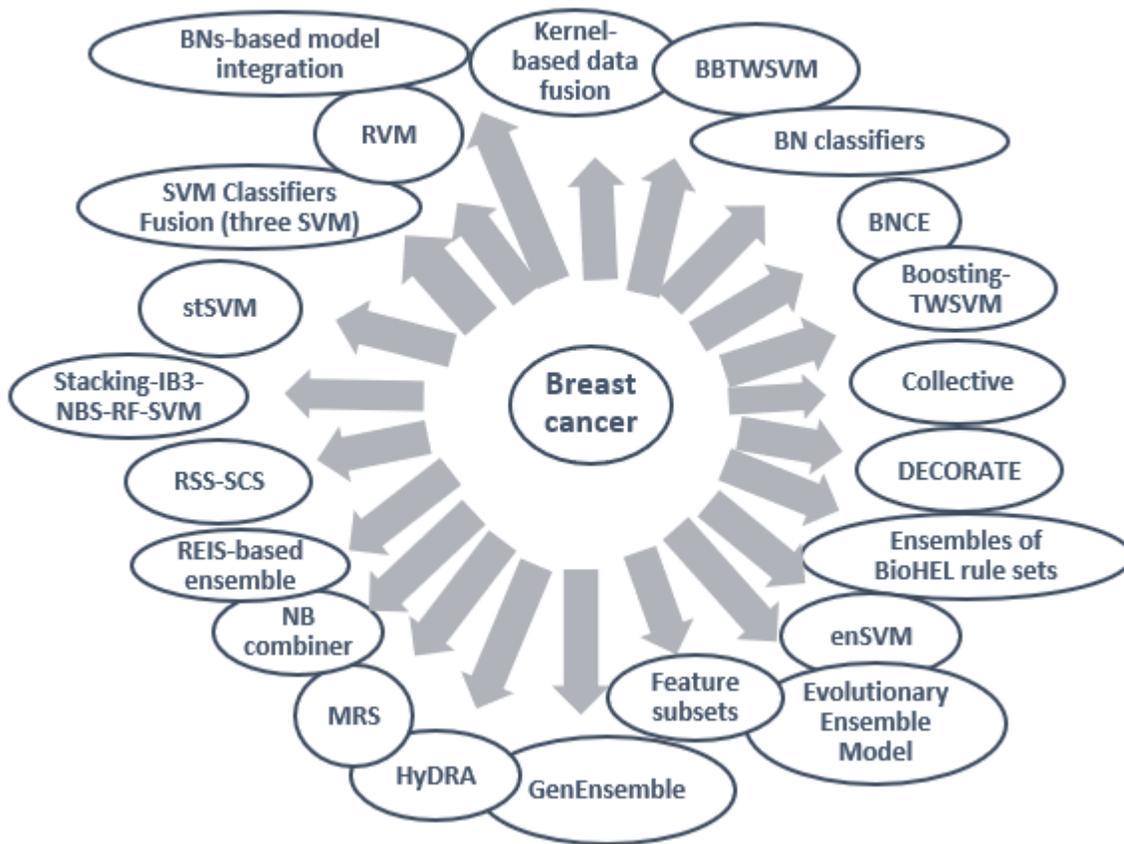
Figure 3

The introduced methods for breast cancer detection
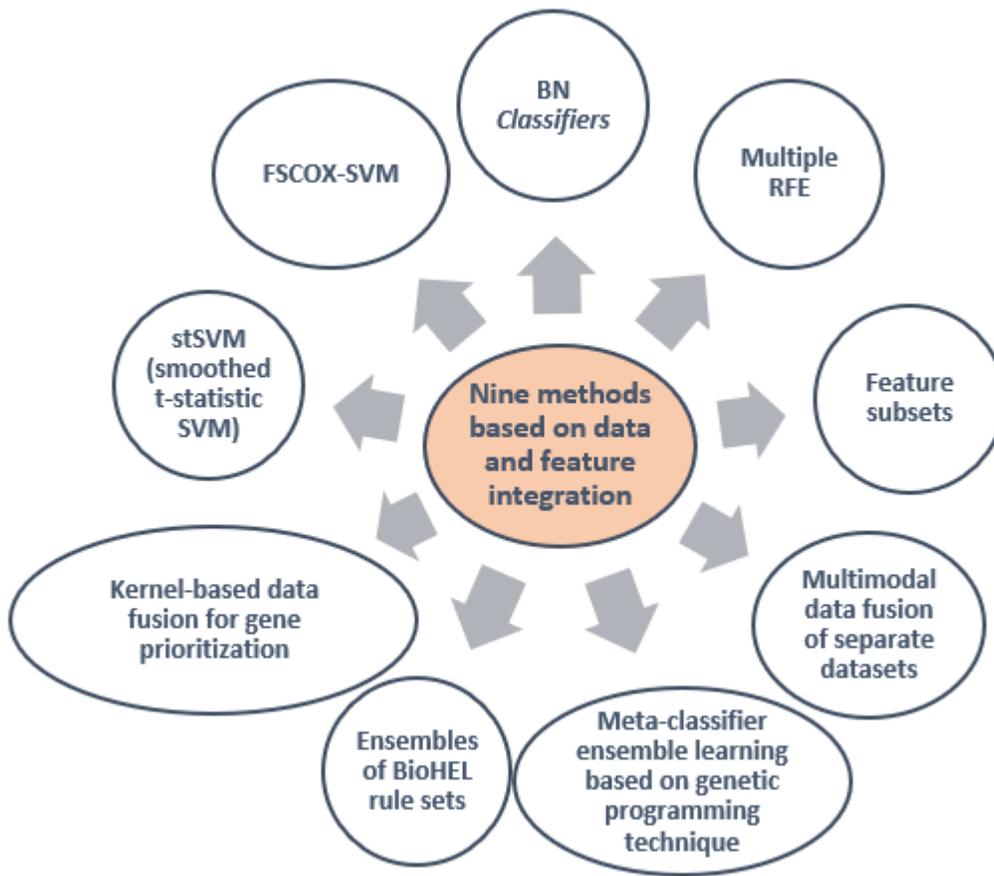
Figure 4

The methods based on data and feature integration. They are BN Classifiers, multiple RFE selection, feature subsets, multimodal data fusion of separate datasets, meta-classifier ensemble learning based on genetic programming technique, ensembles of BioHEL rule sets, Kernel-based data fusion for gene prioritization, stSVM by data integration, and FSCOX-SVM methods
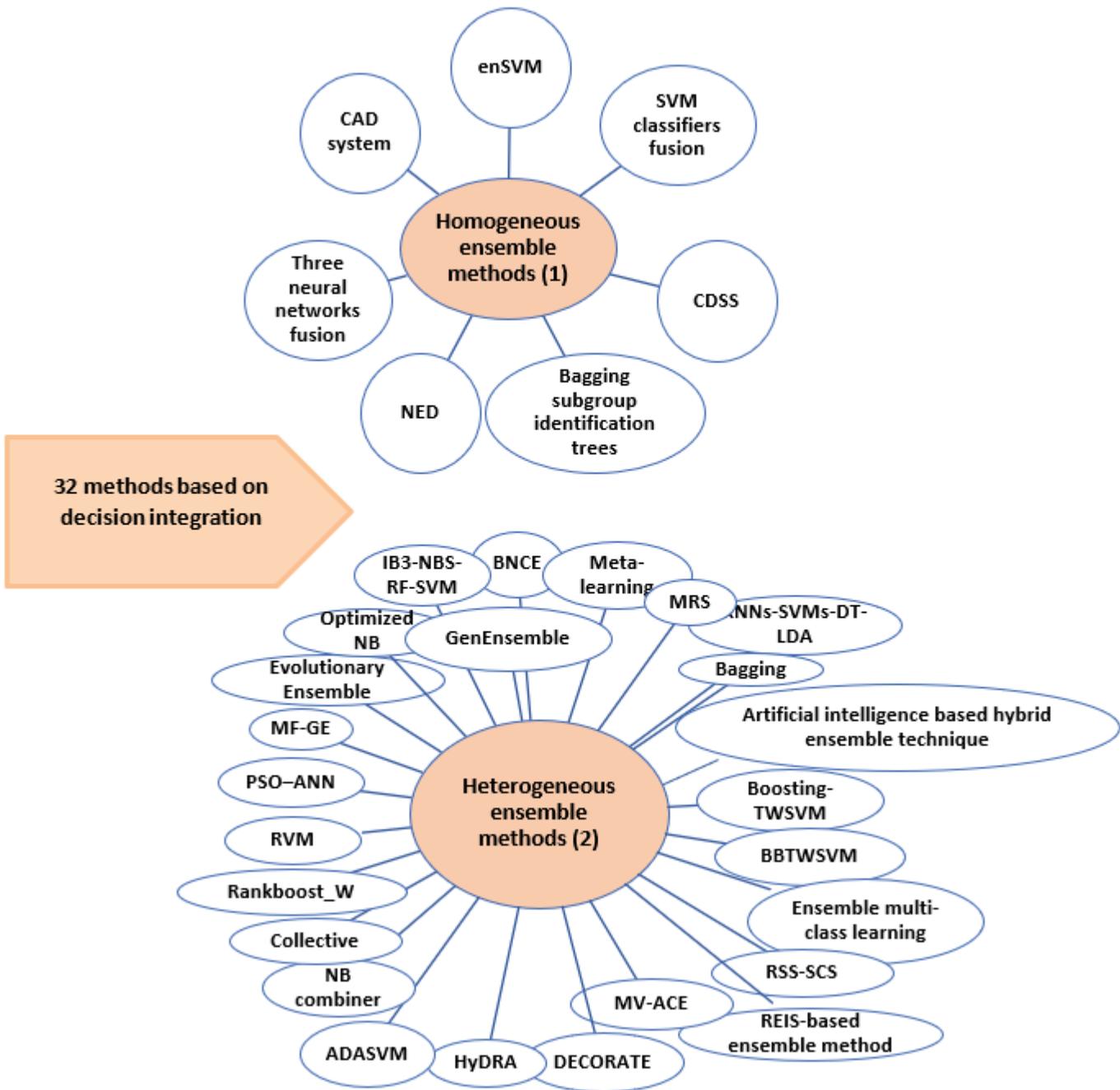
Figure 5

The methods based on decision integration. Homogeneous ensemble methods are SVM classifiers fusion, three neural networks fusion, CDSS, enSVM, Bagging subgroup identification trees, and NED approaches. Heterogeneous ensemble methods are BNCE, meta-learning, KNNs-SVMs-DT-LDA, MRS, Bagging, artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, BBTWSVM, Ensemble multi-class learning, RSS-SCS, REIS-based ensemble method, MV-ACE, DECORATE, HyDRA, ADASVM, NB combiner, Collective approach (correlation, color palette, color proportion and SVM), Rankboost_W, PSO–ANN ensemble, MF-GE system, Evolutionary Ensemble, CAD system, optimized NB approaches, Stacking IB3-NBS-RF-SVM, and GenEnsemble that is similar to Stacking IB3-NBS-RF-SVM

Figure 6

The number of ensemble learning methods (%) used at each level of fusion
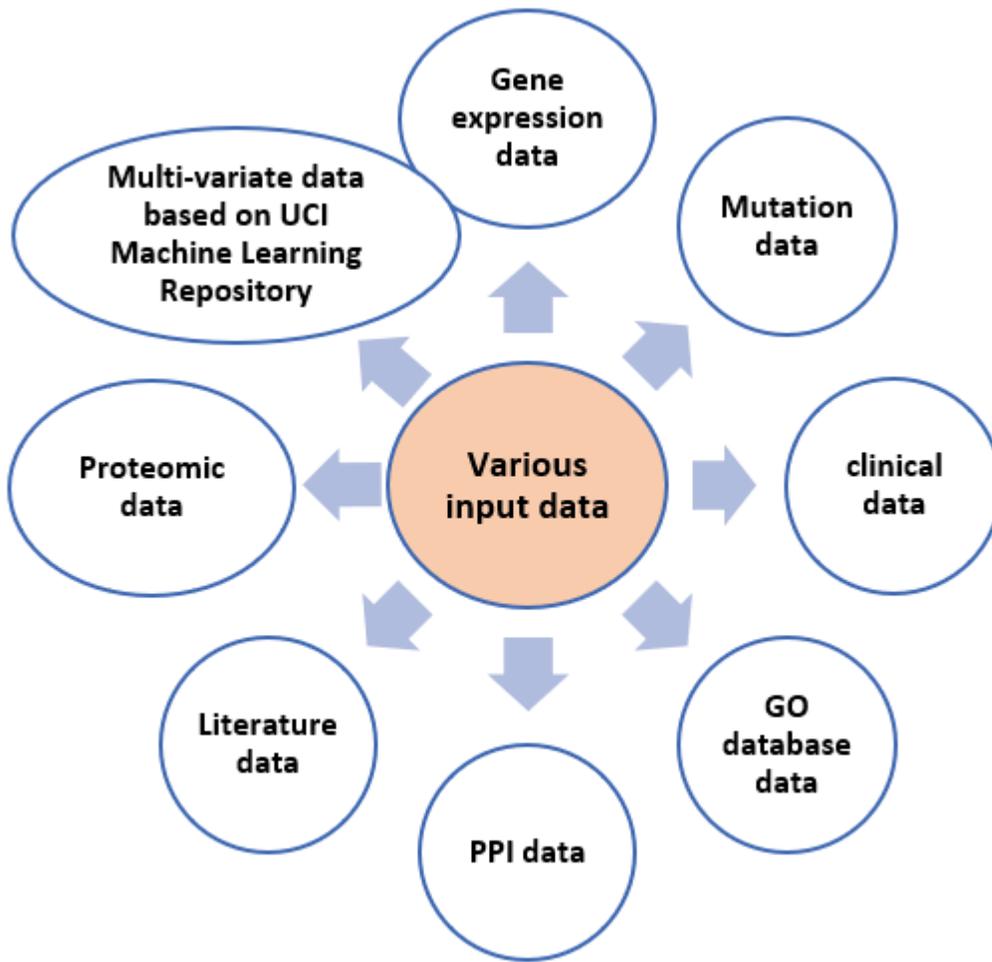
Figure 7

Various input data that have been used in different studies
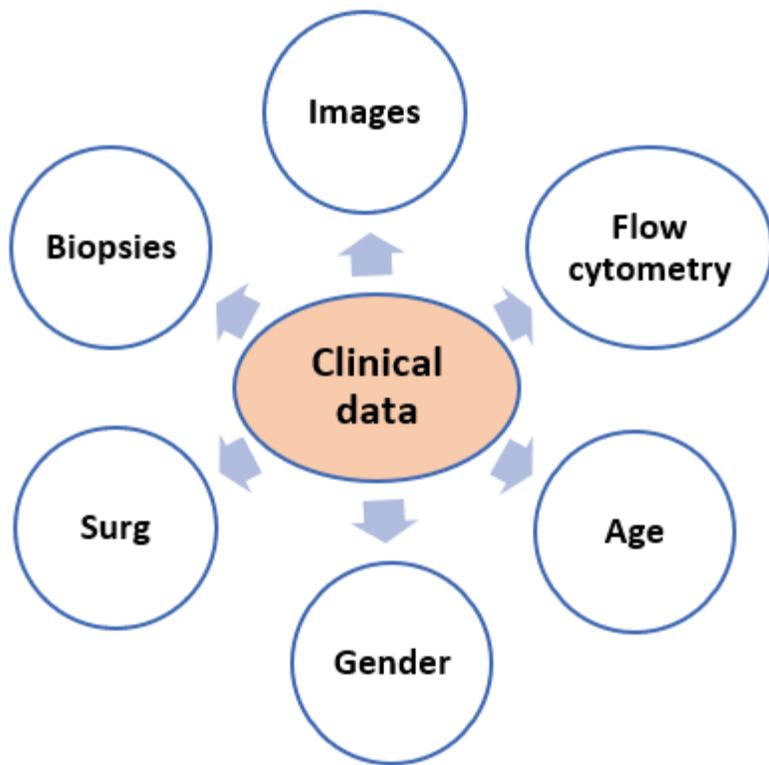
Figure 8

Kinds of clinical data in these studies
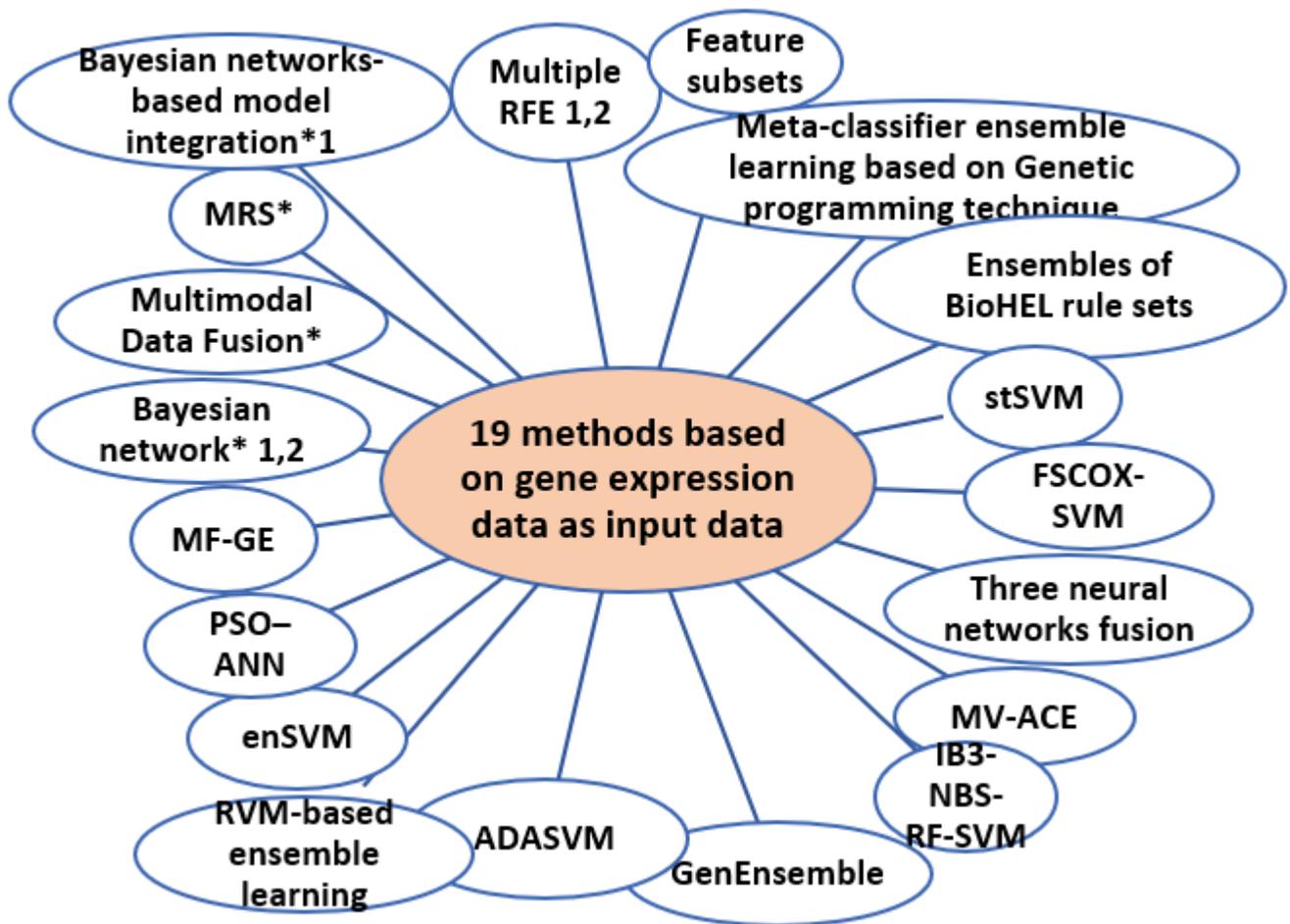
Figure 9

19 methods have used gene expression data as input data. The star-labeled methods have used several input data types in addition to the gene expression data. 15 of these 19 methods (33.32% of the 45 studies) have used only gene expression data. These methods are Multiple RFE selection, Feature subsets, Meta-classifier ensemble learning based on Genetic programming technique, Ensembles of BioHEL rule sets, stSVM, FSCOX-SVM, three neural networks fusion, MV-ACE, Stacking IB3-NBS-RF-SVM, GenEnsemble, ADASVM, RVM-based ensemble learning, enSVM, PSO−ANN ensemble, and MF-GE system. Besides, four methods labeled by an asterisk including Bayesian network, Multimodal Data Fusion of Separate Datasets, MRS, and Bayesian networks-based model integration are approaches that used several types of data as multiple input data, at least one of them is gene expression data

Figure 10

17 mentioned methods in this review have used clinical data as input data. The star-labeled methods have used several input data types in addition to the clinical data. 13 of these 17 methods (28.89% of the 45 studies) including of SVM Classifiers Fusion, Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, Bagging and Boosting-TWSVM, RSS-SCS, REIS-based ensemble, Collective approach (correlation, color palette, color proportion and SVM), Bagging subgroup identification trees, CAD system, optimized naïve-Bayes model, and Neural Ensemble based Detection (NED) have just applied clinical data as input data. Also, four methods marked by an asterisk such as BN Classifiers, Multimodal Data Fusion of Separate Datasets, MRS, and Bayesian networks-based model integration1 have used another data in addition to the clinical data as multiple input data

Figure 11

Five methods (15.56% of the 45 studies) used several data as multiple input data. They are BN Classifiers, Multimodal Data Fusion of Separate Datasets, MRS, Kernel-based data fusion, and Bayesian networks-based model integration1
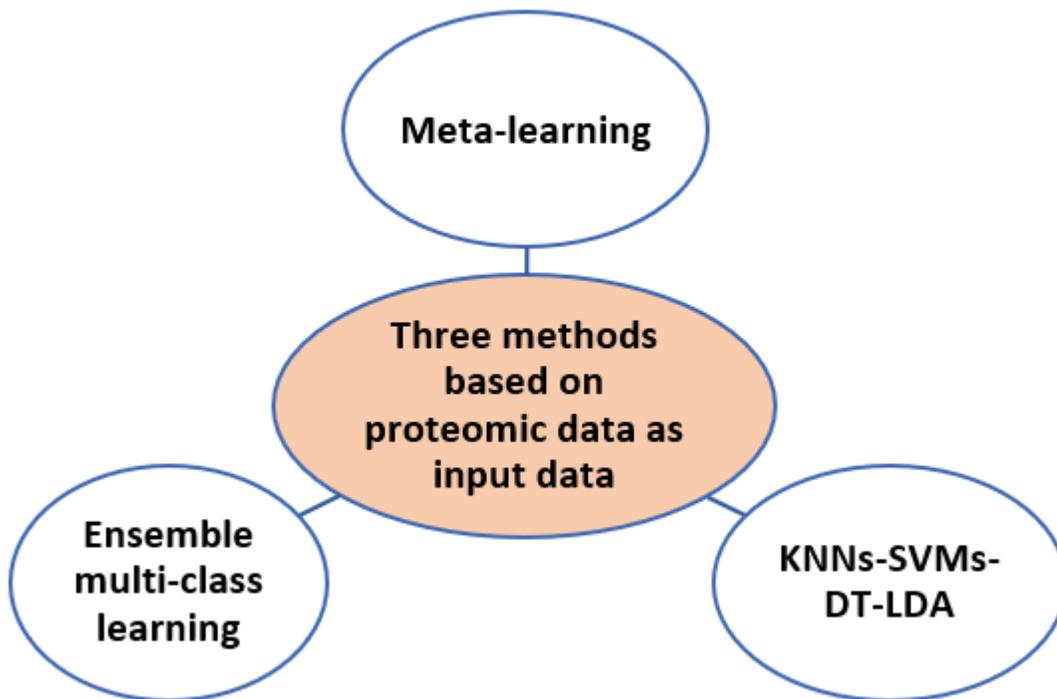
Figure 12

The methods (6.67% of the 45 studies) that have used proteomic data as input data. They are meta-learning, KNNs-SVMs-DT-LDA, and Ensemble multi-class learning
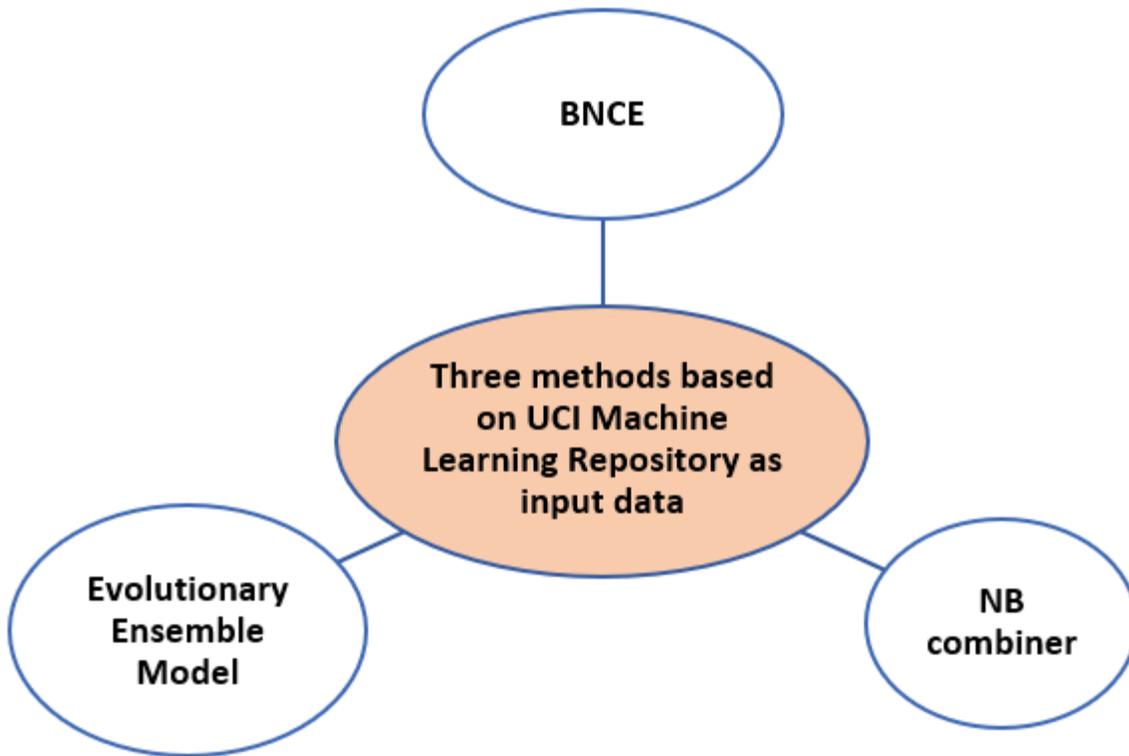
Figure 13

Three methods (6.67% of the 45 studies) have used various data. This data is belonged to different datasets, such as the UCI Machine Learning Repository. They include BNCE, NB combiner, and Evolutionary Ensemble Model
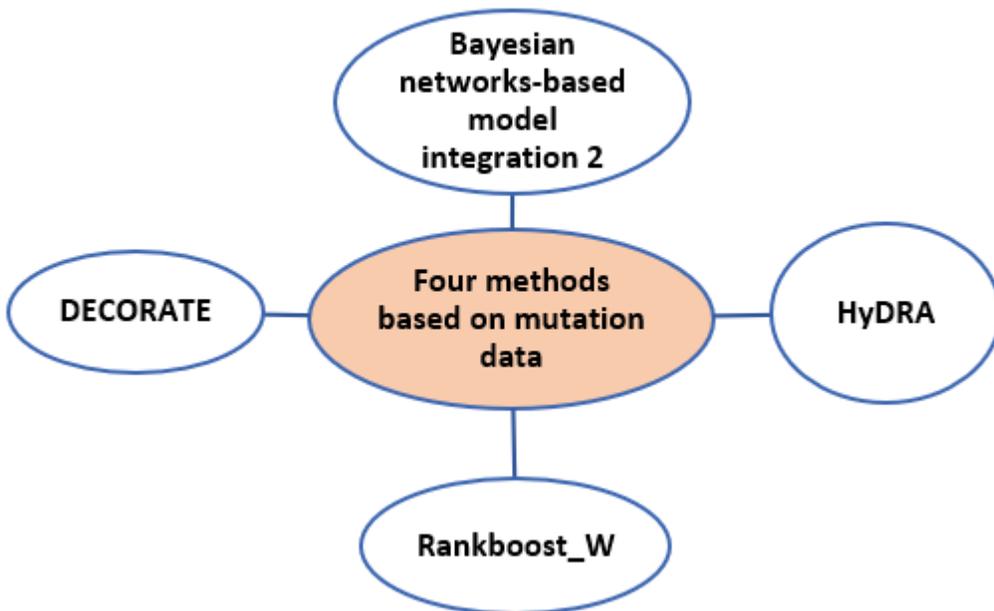
Figure 14

The methods (8.89% of the 45 studies) that have used somatic mutation data. They are HyDRA and DECORATE, Rankboost_W, and Bayesian networks-based model integration 2

Ensemble learning methods (%)
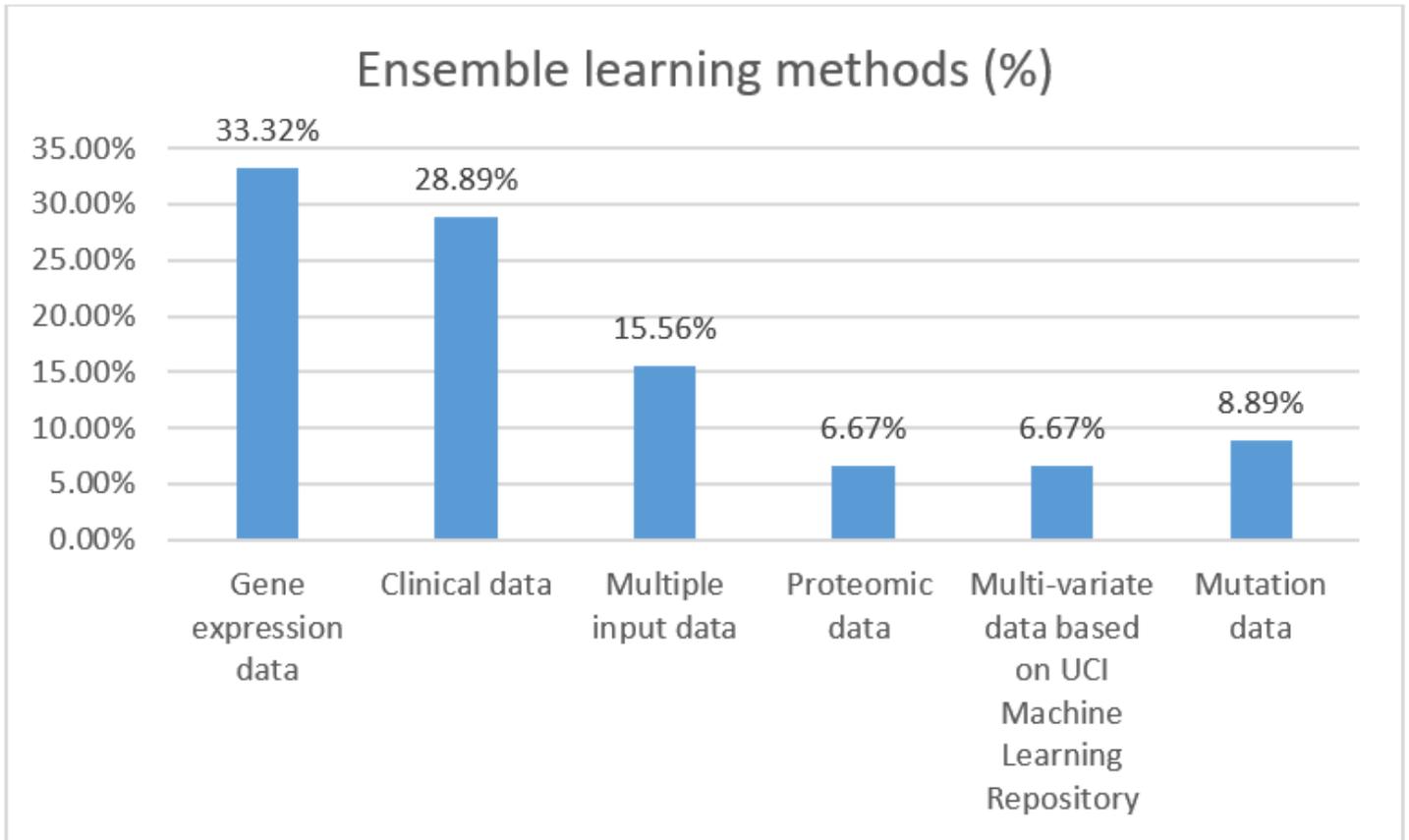
Figure 15

The number of ensemble learning methods (%) that have used different kinds of input data
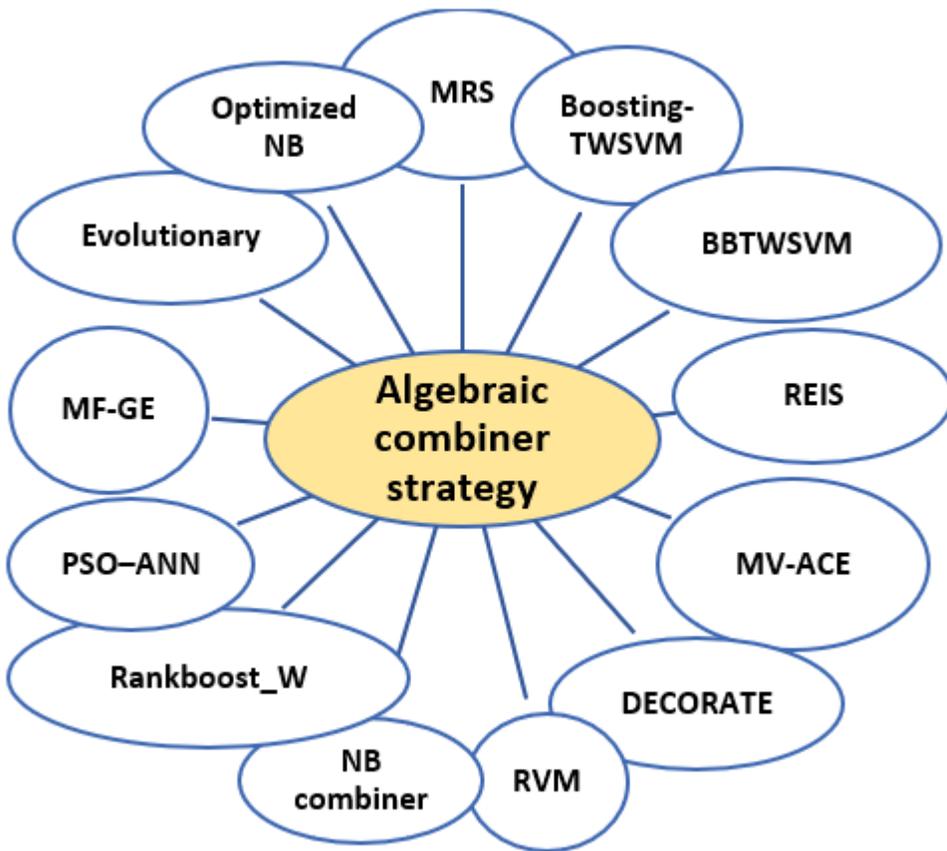
Figure 16

The methods that have been decided based on majority voting strategy. These methods are: MRS, Boosting-TWSVM, BBTWSVM, REIS-based ensemble, MV-ACE, DECORATE, NB combiner, Rankboost_W, RVM-based ensemble learning, PSO−ANN ensemble, MF-GE system, Evolutionary Ensemble Model, and optimized NB model
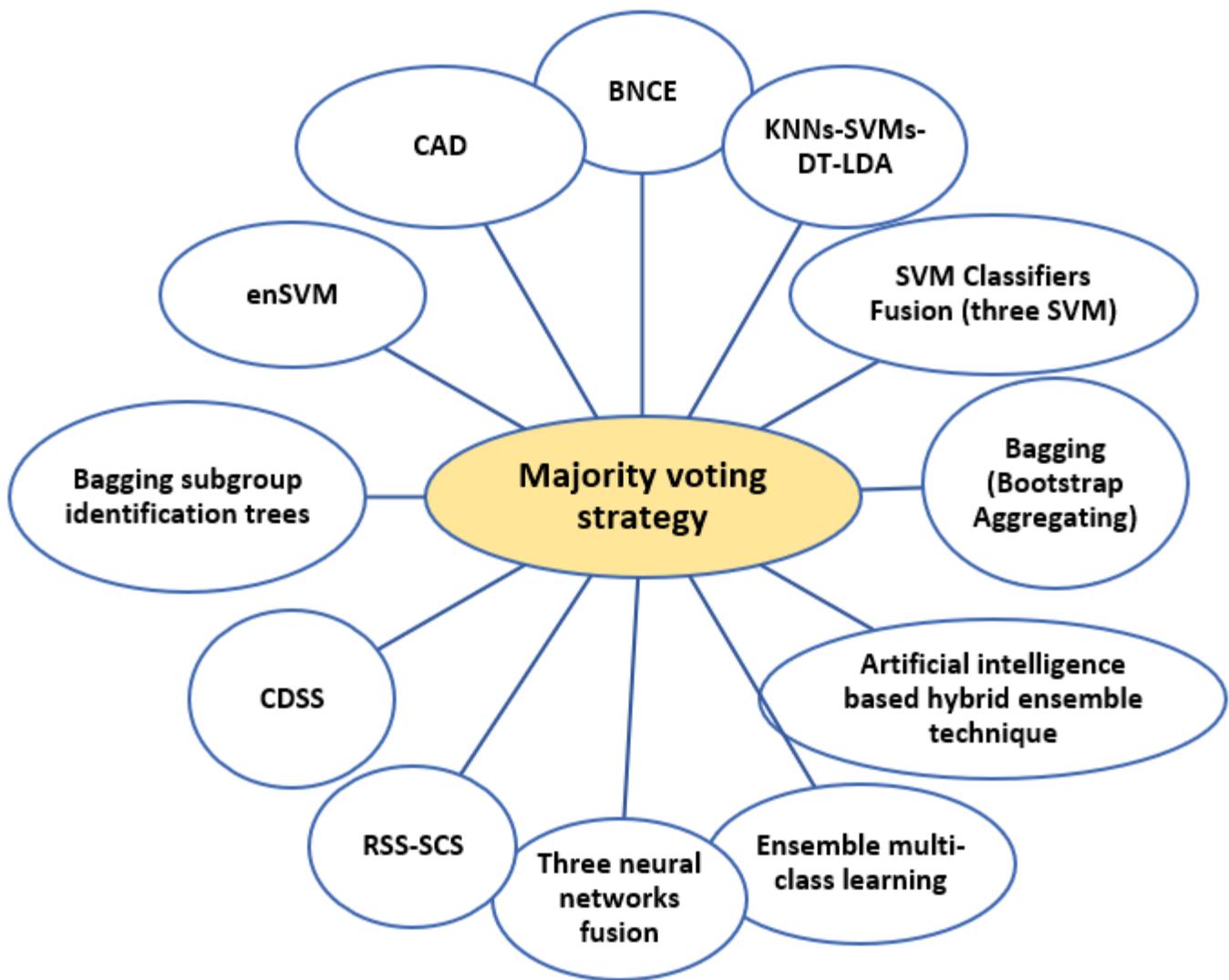
Figure 17

The methods that have been decided based on majority voting strategy. They are BNCE, KNNs-SVMs-DT-LDA, SVM Classifiers Fusion (three SVM), Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Ensemble multi-class learning, three neural networks fusion, RSS-SCS, CDSS, Bagging subgroup identification trees, enSVM, and CAD system
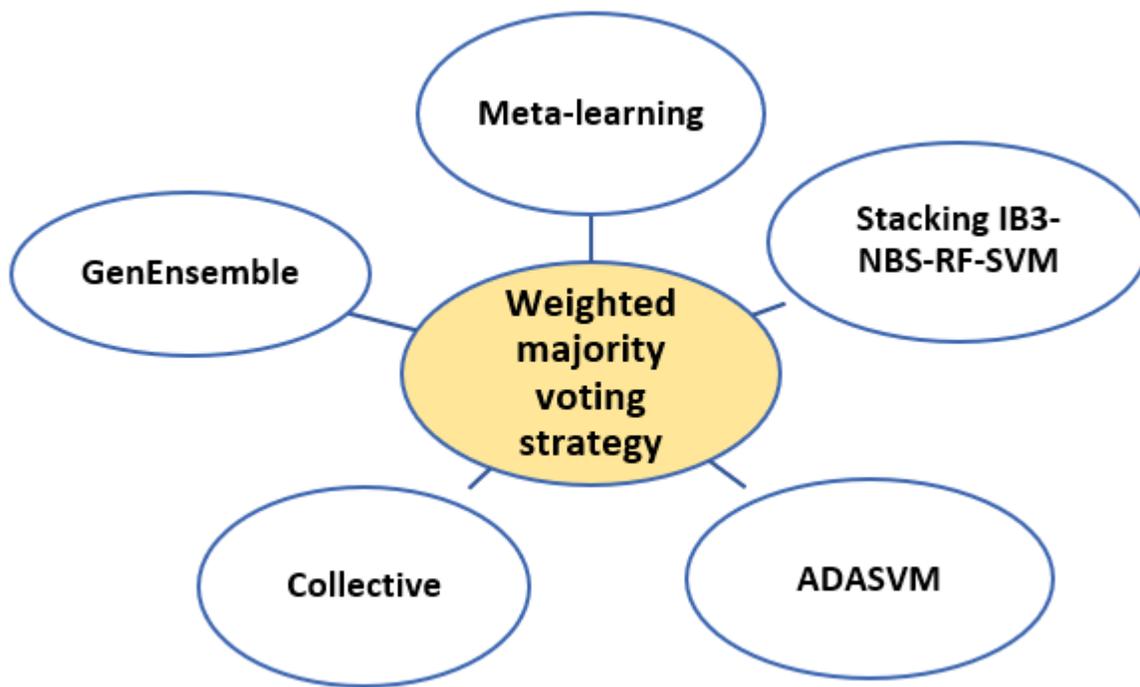
Figure 18

The methods that have been decided based on the weighted majority voting strategy
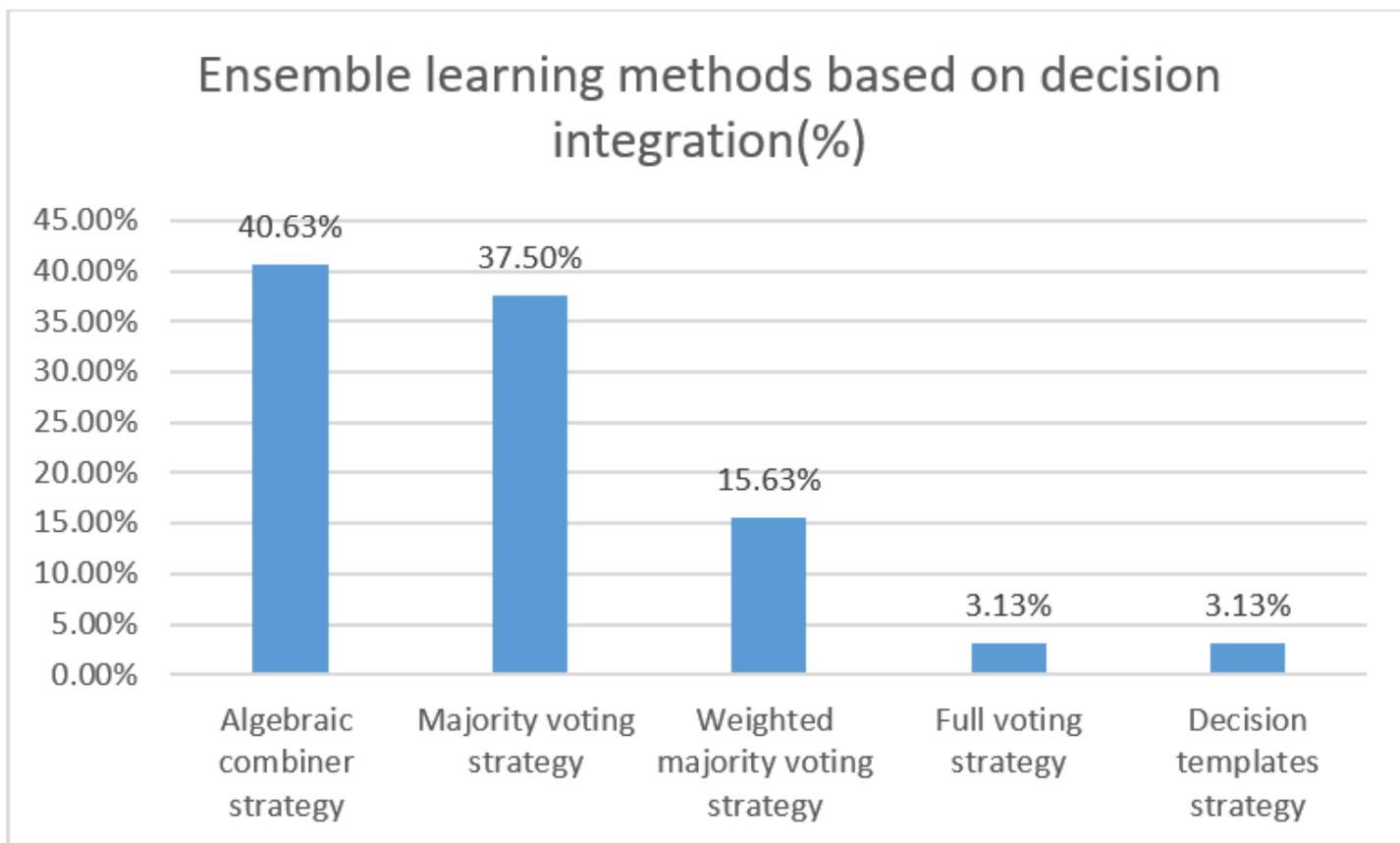
Figure 19

The number of methods (%) that have decided with different strategies
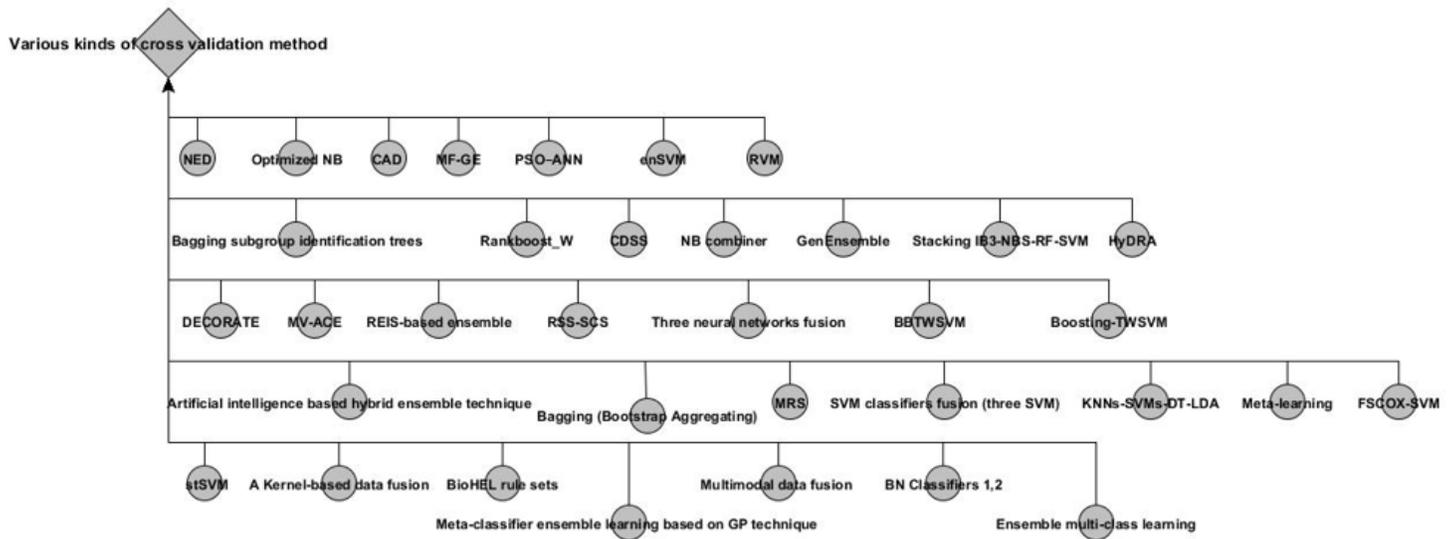

Figure 20

The methods that have been evaluated using a variety of cross-validation methods (k-fold cross-validation). These methods include: Two of three Bayesian networks (Bayesian network 1 and Bayesian network 2), Multimodal Data Fusion of Separate Datasets, meta-classifier ensemble learning based on Genetic programming technique, Ensembles of BioHEL rule sets, Kernel-based data fusion, stSVM, FSCOX-SVM, meta-learning, KNNs-SVMs-DT-LDA, SVM Classifiers Fusion (three SVM), MRS, Bagging (Bootstrap Aggregating), Artificial intelligence based hybrid ensemble technique, Boosting-TWSVM, Bagging and Boosting-TWSVM, Ensemble multi-class learning, three neural networks fusion, RSS-SCS, REIS-based ensemble, MV-ACE, DECORATE, HyDRA, Stacking IB3-NBS-RF-SVM, GenEnsemble, NB combiner, CDSS, Rankboost_W, Bagging subgroup identification trees, RVM based ensemble learning, enSVM, PSO−ANN ensemble, MF-GE system, CAD system, optimized NB model, Ensemble multi-class learning, and NED

Figure 21

Different statistical validation methods for evaluation of the performance of six ensemble learning methods
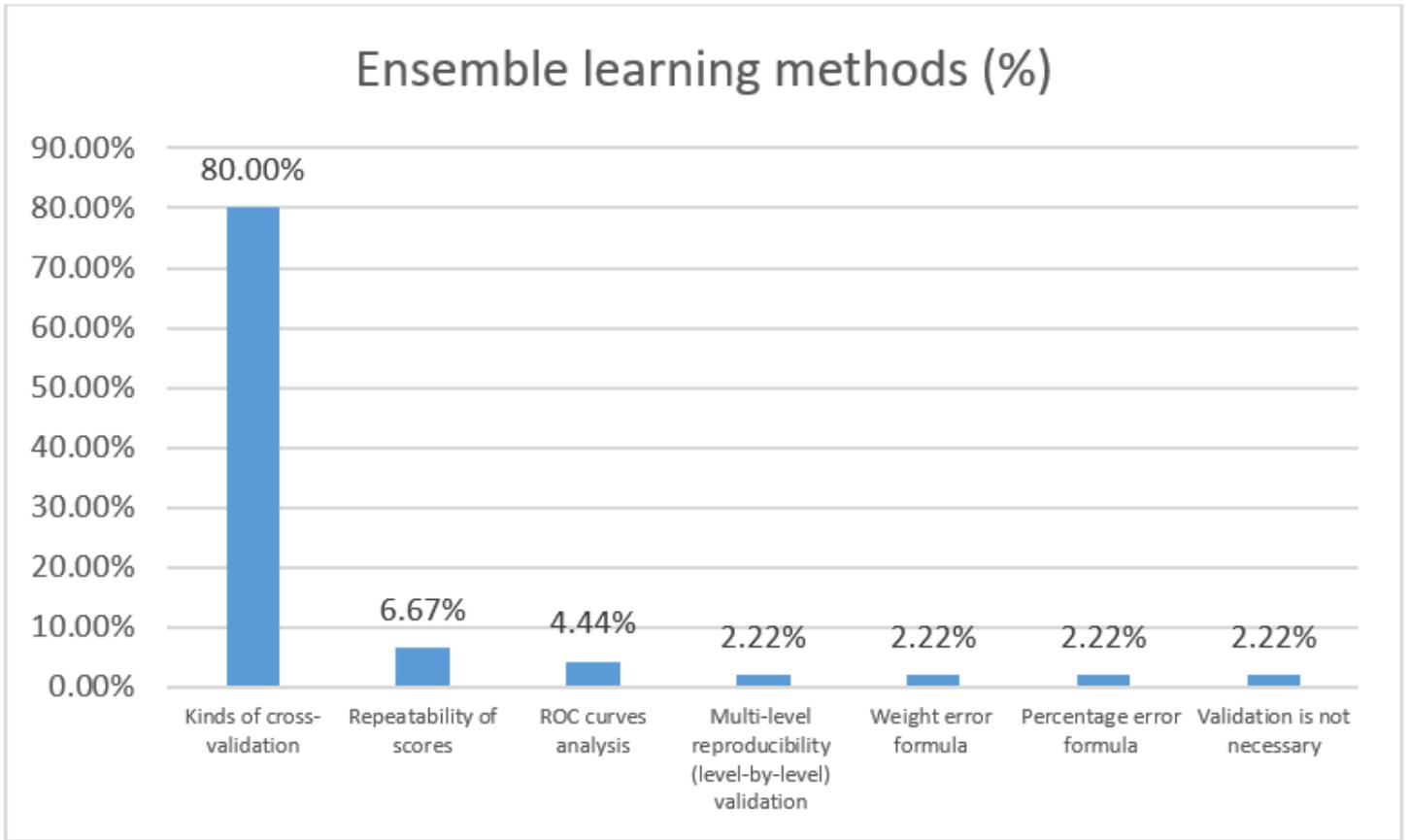
Figure 22

The number of ensemble learning methods (%) validated with kinds of statistical methods