International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 4, July-Aug-2022, ISSN (Online): 2395-566X

Neural Network-Based Advanced Cancer Prediction and Classification for Enhanced Diagnosis and Prognosis Accuracy

Valarmathi P, Rubadharshini A K, Subashini P, Arullakshmi A
Department of Computer Science and Engineering
Kongunadu College of Engineering and Technology Tamilnadu, India

Abstract- One of the main areas of contemporary machine learning and data mining research is medical diagnostics. Since single nucleotide polymorphisms (SNPs) contribute significantly to the variability of the human genome, they have been linked to a number of illnesses, including cancer. The most prevalent malignant growth in women, breast cancer, has become much more prevalent during the last 20 years. Several methods have been used on Genetic data to make distinctions between these tumorous and benign data. The large amount of features in SNP data, which makes classification difficult, is one of the main issues.The dimensionality problem for the diagnosis of cancer in women is addressed in this research by an innovative blended intelligence technique based on Association Rules for Harvesting (ARM) and neural network technology (NN) who employs the Evolutionary Computation (EA). While NN is employed to achieve successful classification, ARM optimized by Grammatical Evolution (GE) is used to obtain relationships between SNPs, diminish dimension, which and find the most useful features. The NCBI GEO (Gene Expression Omnibus) website's carcinoma SNP dataset was used to test the suggested NN-GEARM technique. Up to 90% consistency has been achieved by the developed model.

Index Terms- Artificial intelligence, Neural networks, Classification, Filtering

I. INTRODUCTION

The polymorphisms (SNPs) might have important clinical affects, such as associations with complicated disorders, according to a number of studies. The most common genetic variation is an SNP, which is a variation in sequence of DNA brought on by a change in a single nucleotide in the entire genome [1]. The functional and structural effects of SNPs have been extensively studied computationally over the past 20 years [2]. In this context, SNP data had been studied thoroughly with machine learning and data mining tackles [3, 4, 5, 6].

Although typical artificial intelligence algorithms have been successful in analyzing gene expression data, it becomes increasingly recognized that deploying one conventional intelligent strategy for assessing large-scale data is unfeasible. A lack of precision is projected if gene activity is utilized to determine a disease and a typical classifiers is employed to classify an item based on all measurable factors. The phenomenon known as the "curse of multiplicity problem" in the realm of machine learning refers to the vast number of attributes and the comparatively small number of observations (samples) generated as microarray data.

In order to reduce computing costs, find a small fraction of biologically significant genes linked to different diseases, and attain the required prediction accuracy, methods for choosing a limited number of informative features become essential.

Generally speaking, feature selection aims to eliminate redundant and unnecessary features in order to improve classifier performance [7], similar to the diagnostic model [8]. Combining a number of conventional intelligent methodologies has shown beneficial in analyzing large, complex biological data, and as a result, it is gaining popularity.

Recently, there have been several attempts to use ensemble methods for gene selection [9]. Filter, wrapper, and embedding approaches are three distinct groups into which the diverse range of gene methods for selecting are easily divided [10].

In the area of computer intelligence, feature selection [11] is frequently seen as an essential pre-processing data analysis step.

SNP data has wonderful dimension, which and resembles matrix expression information. In the grouping of SNPs, and

choosing features approaches are essential for reducing this vast number of characteristics and enhancing the analysis's accuracy.

In order to determine the best SNP classification of information, the authors in [12] integrated a number of currently used approaches. Three steps made up the analysis: first, informative SNPs were chosen; next, an artificial feature was created using the chosen Polymorphisms; and last, the classification task was carried out. The most successful feature selection methodology had been CBFS (an Algorithm Based on Feature Clearness [13]), which achieved a 94% precision rate with a classifier using an SVM when contrasted to the entirety of the top 100 SNPs from cancer patient data.
Several hybridization tests have prior to effectively employed evolutionary methods of learning [14], [15], and [16], and rule-based classification techniques have recently been evaluated on high-dimensional gene array data [17], [18]. The authors of [19] achieved accuracies of 90% via applying rule-based evolutionary artificial intelligence systems to three openly accessible transgenic oncology datasets. By contrasting it with other benchmark microarray sample classifications based on three other feature selection techniques, they proposed that this evolve method of training might be competing with other approaches.

To find SNPs linked to different illness, combinations of techniques utilizing neural networks proved effective in addition to gene selecting. Because hybrid approaches based on evolutionary algorithms are so common in bioinformatics, we concentrate on them.

One of the most prevalent and dangerous cancers in women is breast carcinoma. Breast cancer trends can be predicted and categorized using a variety of methods [20]. In [21], the authors employed artificial intelligence for segmentation and association principles to minimize the level of detail with the state of Wisconsin cancer data. Nine features were subjected to the procedure; four of these have been selected as the most desirable values by the Apriori Algorithm and sent to NN for classification. The described system's reported classifying rate was 95.6%.

It is widely acknowledged that a variety of multifaceted diseases, including cancer, are caused by a confluence of several genes that frequently interact to produce the characteristics [22].

To find SNP-SNP interactions, several ensemble methods have been put forth [23], [24]. Our objective is to unravel the intricate relationships between Polymorphisms that could raise the risk of the development of the desired condition using SNPs as genetic markers. It is reported that this sorting task is a broad pattern recognition problem. The secret to pattern recognition is feature extraction since improperly selecting characteristics will affect the performance of any powerful classifier.

In the present investigation, we provide NN-GEARM, a new hybrid intelligent technique based on Association Rule Mining (ARM) and Neural Networks (NN) that uses Grammatical Evolution (GE), a learning method. This approach is motivated by the success of combining artificial intelligence for feature selection and classification tasks off biological data, as well as by the high performance of the evolutionary algorithms. We also recognize that traditional being chosen characteristics tend to ignore the interaction between features [25], while their combination may have a significant connection with the target [26], [27].

Our two-stage approach was used on sizable SNP datasets related to breast cancer. The input vector feature dimension is decreased in the very initial step. When used in conjunction with GE, ARM is a feature selection technique that considers the interaction connecting the characteristics rather than treating SNPs independently. This has allowed for the extraction essential rules for the selection to necessary characteristics for the medical model. These particular characteristics can be utilized via a neural network in the second phase to organize the (dimensionality-reduced) breast cancer data.

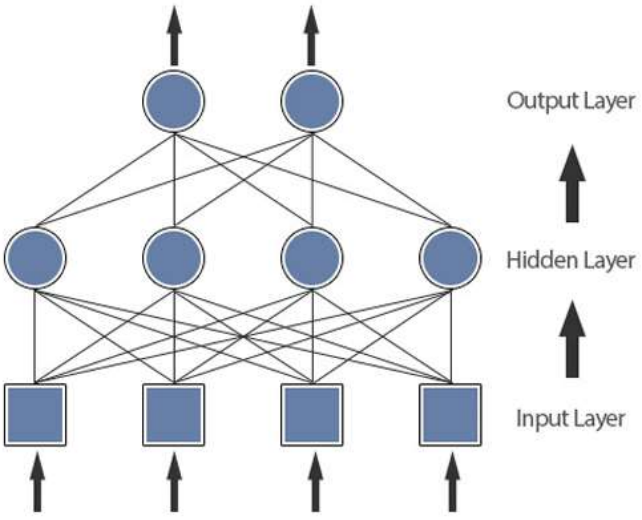

Fig. 1. Neural Network Model

This is how the remainder on the paper is structured. An overview of NNs is provided in Section 2. In Section 3, the ARM procedures are explained, and in the fourth subsection, the GE technique is described. The information that was used presents himself under the heading a five whereas our suggested method, NN-GEARM, is presented in Section 6.

Section 7 presents the findings of the experiment. The article winds down in Part 8.

Artificial Neural Network

Artificial Neural Network Computation, sometimes known as Neural Computation or Network Computation, is a computing method that draws inspiration from the brain's organic neural networks. Its structure is made up of connections between different neurons, which are tiny processing units. The weight of each connection indicates how strongly two neurons exchange signals (Figure 1).

This mathematical model is distinguished by its capacity to employ sequential training processes, learn intricate, nonlinear input-output correlations, and adjust to the data. Building an intelligent module to address a challenging issue, like pattern recognition and classification, is its primary objective [28].

One or more inputs and one or more outputs make up an artificial neural network. There are two stages in the process of creating an ANN: the application (testing) phase and the learning phase, often known as training. In order to learn as closely as possible the input-output relations that define the observed data, the network uses a learning algorithm to modify the strength of its connections based on a sequence of recorded data (inputs and their corresponding outputs) in the training mode. When used to disease data, NNs have demonstrated varying degrees of success [29], [30], and [31]. NNs are utilized for data mining in many different domains. They are quite good at identifying patterns and classifying large amounts of data.

The SNPs are represented by the NN's input nodes for the genetic application, and the interactions between the elements are captured by the arcs [29]. The output node symbolizes thesample condition, either case or control. Between the input and output layers are one or more hidden layers with a number of interconnected internal nodes that process the information received and generate the final decision based on the weights that have been changed [30].The neural network's goal is to identify matching patterns in actual data by using patterns found in training data

II. ASSOCIATION RULE MINING

Although various approaches have shown encouraging results for genetic data, it is currently unknown how well some of them will handle data with up to 500,000+ SNPs in genome-wide association studies (GWAS). Here, filtering techniques are used to choose a subset of data before the machine learning method looks for interactions in all possible ways [32].

Association Rule Mining (ARM) is a well-researched and significant unsupervised learning technique in data mining. It facilitates the identification and characterization of correlations among several elements (variables) in a sizable data set. This method yields association rules that are generally simple to comprehend and analyze, and they show intriguing relationships between variables.GWAS stands for association studies. In order to choose an area of information before its machine learning methodology thoroughly looks for interactions, filtering techniques are used in this situation [32]. One of the significant unsupervised learning and extensively studied data mining methods is Association Rule Mining (ARM). In a big data set, it aids in identifying and characterizing the connections between various elements (variables). In general, the association rules generated by this method are simple to learn and analyze, and they show intriguing relationships between variables. Numerous domains, including telecommunication networks, market analysis, risk management, inventory control, online usage mining, intrusion detection, bioinformatics, etc., heavily rely on association rules.

Let's give some formal definitions to some terms. Let I be a collection of things, or in our instance, features and their values, such that I = fi1; i2; :::; img Let D be a collection of transactional data.D is a subset of I, and D = fd1; d2; :::;dmg. An implication of the kind X ⸦! Y, where X { I, Y } I, and X \ Y = ;, is called an association rule. An itemset called X is the rule's antecedent, and another itemset called Y is its consequent. The percentage of transactions in the data set that contain an itemset X is known as its support sup(X).

To indicate how strong the rule implication is, two association rule quality metrics—support and confidence—are employed. where count(D) is the total number of transactions in D and count(X [Y]) is the number of transactions that contain every item that appears in X [Y].

If an item set's support is equal to or higher than a minimal support level (Minsup), as determined by the user or expert, it is referred to as frequent. In other words, the confidence denotes the intensity of an implication, while the support suggests the frequency of an occurrence pattern. It is said that a rule (X! Y) is valid if

The extraction of association rules is based on the Apriori algorithm suggested by Agrawal et al. [33] which works in two runs:

- The algorithm starts by generating all the frequent items whose supports are greater than or equal to the specified minimum support.
- It then extracts all the valid rules that satisfy the minimum confidence constraint, which means that the confidence of the generated rule is greater than or equal to the specified minimum confidence.

A limitation of the Apriori algorithm is that it works in two runs and has a very high computational cost.

When there are multiple solutions to a problem, evolutionary algorithms are frequently utilized to discover the best or nearly best one. The two most popular methods for extracting association rules are genetic programming (GP) and genetic algorithms (GA). A methodology for employing GE to extract ARs for the classification of simulated genomic data was put out in [34]. The writers documented the combination's success based on their findings.

III. GRAMMATICAL EVOLUTION

Algorithms for evolutionary computation (EC) are categorized as machine learning in general. These methods have been successfully used in GWAS in the past and are commonly used in the bioinformatics sector. The most popular evolutionary algorithms for optimizing a variety of classifiers (Neural Networks, Naive Bayes classifiers, Decision Trees, etc.) to identify intricate genotype/phenotype relationships include GA, GP, and GE.

The biological process of creating a protein (phenotype) from genetic material (DNA genotype) serves as the inspiration for grammatical evolution, a type of evolutionary computation. It converts populations composed of linear genomes into a computer program using a particular syntax [35].

The binary string (genome) of every member of the population dictates which production rules in a Backus Naur Form (BNF) grammar definition are employed in the mapping process, which is carried out using the rules of a grammar to translate the genotype to the phenotype.

Evolutionary operators are then used at the chromosomal level (strings) to produce following generations after the resultant individual has been evaluated for fitness in the population [35].

The production rules that make up the grammar are described in terms of terminals and non-terminals. The production rule's right side can have any combination of terminals and/or non-terminals, whereas the left side can only include non-terminals. Formally, the four sets S, N, T, and P define the grammar.

S: The initial symbol
Non-terminal symbols (N)
T: Symbols for terminals.
P: Set of production rules.

When all non-terminals are changed using the appropriate production rules, the parts that remain in the language are known as terminals.

The following procedures are followed in Grammatical Evolution prior to each individual's evaluation:

The start symbol and non-terminal symbols of the BNF grammar definition are mapped into terminals using the genotype (integer string). The lawful phenotypes are specified by the grammar.

A mapping function then uses the grammar to convert the integer values (from the integer string) into the proper production rule.

The formula shown in (3) is used to choose the production rule, where nb-al is the number of alternatives (i.e., rules) defining the current non-terminal and P-rule is the index of the chosen production rule.

$$\text{Prule} = (\text{V alue}) \text{MOD}(\text{nbal}):$$

The alternative chosen by the MOD operation is used to replace the non-terminals, and this process is repeated until only terminal elements are left. In order to get the next integer, the method can go back to the beginning of the chromosome if the end of the chromosomes is reached and the program still contains non-terminal components. A predetermined T number of times can be allotted for this wrapping procedure to take place [35]. Section VI provides a thorough example that demonstrates how to use grammatical evolution to extract association rules.

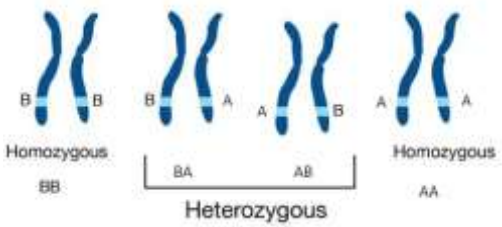

Fig. 2. Homozygous and Heterozygous genotype

IV. THE DATA SET.

A SNP dataset obtained from the NCBI Gene Expression Omnibus (GEO) repository was implemented in this research [36].High-throughput functional genomic data is candidly archived and spread by the GEO project, an available repository.GEO offers an open and adaptable architecture that makes it easier to submit, store, and retrieve diverse data sets from high-throughput genomic hybridized and gene function experiments [36].

With over 500,000 SNPs and two types of designations (case as well as the control), GSE16619 [37], which is linked to breast cancer, was the subject of our investigation. Of the 111

samples we used, 42 were controls whereas 69 comprised cases.

One of the most frequent discrepancies is that some persons may receive an A at specific locations, while others may receive a G. For instance, an individual nucleotide differences between two sequenced DNA pieces from distinct people, AAGCGTA and AAGCATA. SNPs, also known as single nucleotide poly are the name given to all of these spots. A mutation represents a variant of one of the two elements that can appear at the SNP. According to convention, each SNP's highest-frequency genotype is designated A, and its smallest-known allele is designated B.

The genotype at particular loci represents each sample in our data. As seen in Figure 2, a genotype can have one of four values for a bi-allelic SNP: AA, AB, BB, or No Call, depending on the presence of mutation in the copies of the mother's and father's genes.

The homogeneous genotype can be seen by Ab and BB. When an organism is identical, it suggests that its genome possesses two copies of the same gene allele.

- Heterozygous genotype is represented by AB. An organism is said to be heterozygous if it possesses two distinct alleles of a gene.
- No Call: denotes a value that is missing.

The alphabetical format of this dataset must be changed to a numbers format if one is to used feature choice and classification. There are other methods for doing this, but in our instance, we establish the notation as (4) illustrates:
AA = 11, BB = 10, AB = 01, NoCall = 00. (4)

V. THE NN-GEARM METHOD

The aim of this research is to develop a breast cancer model that can distinguish between healthy and impacted samples using a lot of data. To put it another way, we want to investigate the connection between SNPs and breast cancer.

Since feature extraction is essential to pattern recognition, Figure 3 illustrates the two components of the NN-GEARM algorithm:

- GEARM selection of SNPs.
- Using NN for classification.

GEARM's Feature Selection
One feature selection technique is GEARM. The distinction. The difference between our suggested method and the other feature selection techniques is that the ARM strategy considers the interaction between the various SNPs and their influence on one another rather than the features as standalone.When paired with other SNPs, they may exhibit a high link with the illness state, even when the SNP alone may have a particular correlation with the target. Additionally, some SNPs are regarded as dominating; their existence might eradicate other SNPs as the latter's function is a result of the former.

GEARM creates a set of SNP association rules based on the phases shown in Figure 3's first section. A collection of ideal feature selection rule solutions is the result of this section. The following are the several steps:

- As vectors of integers, an initial population of N random solutions is produced.
- Using the specified grammar, the MOD (modulus) operation maps each vector to an association rule.
- New rules are created in place of the vectors that provide contradictory or inaccurate ones.
- Every generated rule is assessed based on its fitness (5) for each confidence and support that are higher than the minimal confidence and minimum support, respectively.

$$Fitness(R) = (a*sup(R)) + (b*conf(R))$$

where a and b are weights set by experimentation and their sum is equal to 1.The best solutions (best rules) are selected for crossover and mutation which are performed at the chromosomal level.

The new generation, which is equal in size to the original population and contains the best rules, is used in the cycle time and again until the maximum number of generations is reached, after which GEARM stops.

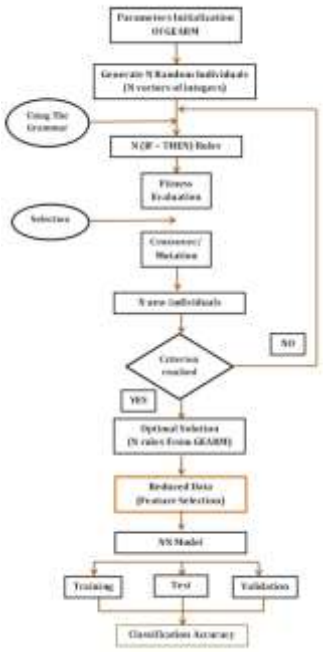

Fig. 3. NN-GEARM Algorithm

Each generation is followed by the identification of an ideal solution.The best overall solution is chosen as the ideal AR set at the conclusion of the GEARM evolution.

An relationship between SNPs is generically defined by the grammar employed in this study, and each SNP may have one of four values (AA, BB, AB, or NC) (see Section V). The following is the grammar:

We provide an example to demonstrate how grammatical evolution is used to derive association rules. Let f55, 9, 22, 10, 29, 44, 96, 7, 25g be a vector of integers representing a randomly created individual. The grammar mentioned above is used to map this vector to an association rule. < Side > Sep < Side > is the result of the start symbol < Rule >. The grammar demonstrates two different definitions (rules) for the initial nonterminal < Side >. We choose alternative number 1 (beginning at 0) to replace the current nonterminal using the Mod operation and the first integer of the vector 55Mod2 = 1.Thus, < SNP >< V al >< Side > Sep < Side > is obtained. The vector's next integer is 9, and its next non-terminal is < SNP >.

Assuming that there are three SNPs in the data, 9Mod3 = 0 and SNP1 < Val >< Side > Sep < Side > are the results of selecting SNP1.The next non-terminal is < Val >, which contains four options, and the following integer is 22. Since 22Mod4 = 2, the alternative AB is chosen to generate SNP1 AB. Sep < Side > < Side >.The vector is used to replace all of the non-terminals in the mapping process. Only terminal symbols remain at the conclusion. If SNP1 = AB and SNP3 = AA, then SNP2 = BB will be the final rule that is obtained.

The rule is then expressed using the number notation.
SNP2 = 10 if SNP1 = 01 and SNP3 = 11. We provide two distinct methods for choosing characteristics.
When there is a circular dependency,

GEARM-OE: This technique aims to extract the set of rules from the data by using all of the samples. We refer to technique as Overall Extraction (OE) because we look for relationships between SNPs based on their values for the tumor and the healthy samples, without differentiating between case and control people. The SNPs in the antecedent portion are regarded as the dominant and more informative elements if a good rule is retrieved with sufficient support and high confidence.The rule "if SNP1= AA and SNP7= AB then SNP3= BB," for instance, suggests that SNP1 and SNP7 work together to affect SNP3, meaning that SNP3 is dependent on the other two SNPs. Thus, SNP1 and SNP7 are selected to be used as inputs to NN.

GEARM-PE: In this second approach, the idea is to divide the data into different parts and separately search for the association rules in each part. We call it Parallel Extraction (PE). Since our application involves diagnosing issues, the data was separated into two sections according to the sample class: all of the case samples are in one section, and all of the control samples are in another.

Analyzing the relationship between the SNPs in the same environment—whether healthy or impacted—is the primary motivation behind this division. The greatest characteristics of this class were chosen since a set of association rules was created specifically for the SNPs of healthy individuals. The SNPs from the illness samples were used to create an additional set of guidelines. Additionally, the greatest features for this class were chosen.

Together, the two sets of the best features were fed into a neural network.
if SNP1= AB and SNP4= AB then SNP2= AA
if SNP4= AA then SNP3= BB and SNP7= AB
if SNP8= AA and SNP5= BB then SNP6= AA

The selected features from this part will be: SNP1, SNP4, SNP3, SNP8 and SNP5. Consider also the following rules extracted from the control samples:

if SNP6= AA then SNP2= AA
if SNP10= AB and SNP4= BB then SNP3= AA
if SNP15= BB then SNP12= AA and SNP3= BB

The selected features from this part will be SNP6, SNP10, SNP4 and SNP15.

Combining the two sets of selected features, the best SNPs that are considered as input for NN will be: SNP

Classification Based on Neural Networks
The reduced collection of SNPs, as shown in the second section of Figure 3, was used to do the classification using a neural network.With a sigmoid transfer function applied to both the hidden layer and the output layer, NN is a two-layer feedforward network. The network's input is represented by the chosen SNPs from the earlier step. Three subdata sets—training, validation, and test—have been created from the reduced data (Figure 3). For each set of input features, a number of tests were conducted by increasing the number of neurons from 20 to 55. The results are shown in the section that follows.

VI. EXPERIMENTATION AND RESULTS

111 samples and more than 500,000 SNPs for breast cancer were used in this investigation. The study demonstrates the effectiveness of GEARM-OE in classifying features with varying numbers of SNPs. The best classification performance was achieved with 32 SNPs, achieving almost 75% accuracy. The study also reveals that parallel feature extraction provides

better performance than overall extraction, with GEARM-PE accuracy always better for any number of neurons except for 50 neurons. The study also shows the average fitness for each set of association rules and execution time for GEARM-PE and GEARM-OE.

To select the best features, the data set was subjected to both GEARM-OE and GEARM-PE.

VII. CONCLUSION

This study proposes a hybrid intelligent approach for classifying SNP data for breast cancer, using over 500,000 SNPs against 111 samples. The approach uses GEARM for input vector selection and NN for classification, with two ideas for reducing dimensionality: GEARM-OE and GEARM-PE. GEARM-PE outperforms GEARM-OE in feature selection, but both methods offer promising results. The hybrid intelligent model of NNGEARM is considered a promising technique for SNP data classification.

REFERENCES

1. F. Collins, L. Brooks, and A. Chakravarti, "A dna polymorphism discovery resource for research on human genetic variation," Genome Res, vol. 8(12), pp. 1229–1231, 1998.
2. S. Mooney, "Bioinformatics approaches and resources for single nu cleotide polymorphism functional analysis," Briefings in Bioinformatics, vol. 6 (1), pp. 44–56, 2005.
3. H. Schwender and K. Ickstadt, "Identification of snp interactions using logic regression," Biostatistics, vol. 9, pp. 187–198, 2008.
4. S. Chen, J. Sun, L. Dimitrov, A. Turner, T. Adams, and D. M. et al, "A support vector machine approach for detecting gene-gene interaction," Genetic Epidemiology, 32, vol. 32(2), pp. 152–167, 2008.
5. H. He, W. Oetting, M. Brott, and S. Basu, "Pair-wise multifactor dimensionality reduction method to detect gene-gene interactions in a case-control study," Hum Hered, vol. 69(1), pp. 60–70, 2010.
6. K. Chen, K. Wang, and M. T. et al, "Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm," BMC Bioinformatics, vol. 15(1), p. 49, 2014.
7. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," The Journal of Machine Learning Research, vol. 3, pp. 1157 1182, 2003.
8. C. Christin, H. C. Hoefsloot, A. K. Smilde, B. Hoekman, F. Suits, R. Bischoff, and P. Horvatovich, "A critical assessment of feature selection methods for biomarker discovery in clinical proteomics," Molecular and Cellular Proteomics, vol. 12(1), pp. 263–276, 2013.
9. H. Liu, L. Liu, and H. Zhang, "Ensemble gene selection by grouping for microarray data classification," J Biomed Inform, vol. 43(1), pp. 81–87, 2010.
10. Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23(19), pp. 2507 2517, 2007.
11. M. S. Mohamad, S. Omatu, M. Yosioka, and S. Deris, "A cyclic hybrid method to select a smaller subset of informative genes for cancer classification," Int. J. Innov. Comput. Inf. Control, vol. 5(8), pp. 2189 2202, 2009.
12. N. Batnyam, A. Gantulga, and S. Oh, "An efficient classification for single nucleotide polymorphism (snp) dataset," Studies in Computational Intelligence, vol. 493, pp. 171–185, 2013.
13. M. Seo and S. Oh, "Cbfs:high performance feature selection algorithm based on feature clearness," PLoS ONE, vol. 7(7), p. 1, 2012 and S. W′ ojtowicz, "Finding differential paths in arx ciphers through nested monte-carlo search," International Journal of electronics and telecommunications, vol. 64, no. 2, pp. 147 150, 2018
14. T. Jirapech-Umpai and S. Aitken, "Feature selection and classification for microarray data analysis: evolutionary methods for identifying pre dictive genes," BMC Bioinformatics, vol. 6(148), 2005.
15. R. Blanco, P. Larranaga, I. Inza, and B. Sierra, "Gene selection for cancer classification using wrapper approaches," Intern J Pattern Recognit Artif Intell, vol. 18, pp. 1373–1390, 2004.
16. J. Deutsch, "Evolutionary algorithms for finding optimal gene sets in microarray prediction," Bioinformatics, vol. 19, pp. 45–52, 2003.
17. V. Popovici, E. Budinska, and M. Delorenzi, "Rgtsp: a generalized top scoring pairs package for class prediction," Bioinformatics, vol. 27, pp. 1729–1730, 2003.
18. A. Tan, D. Naiman, L. Xu, R. Winslow, and D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," Bioinformatics, vol. 21, pp. 3896–3904, 2005.
19. E. Glaab, J. Bacardit, J. Garibaldi, and N. Krasnogor, "Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data," PLoS ONE, vol. 7(7), 2012.
20. T. Sridevi and A. Murugan, "A novel feature selection method for effective breast cancer diagnosis and prognosis," International Journal of Computer Applications, vol. 88(11), 2014.
21. M. Karabataka and M. C. Inceb, "An expert system for detection of breast cancer based on association rules and neural network," Expert Systems with Applications, vol. 36(2), pp. 3465–3469, 2009.
22. J. Cordell, "Detecting gene-gene interactions that underlie human dis eases," Nature Reviews Genetics, vol. 10, pp. 392–404, 2009.

23. Z. Zhang, S. Zhang, M. Wong, N. Wareham, and Q. Sha, “An ensemble learning approach jointly modeling main and interaction effects in genetic association studies,” Genetic Epidemiology, vol. 32(4), pp. 285 300, 2008.
24. R. Upstill-Goddard, D. Eccles, J. Reige, and A. Collins, “Machine learning approaches for the discovery of gene-gene interactions in disease data,” Briefing in Bioinformatics, vol. 14(2), p. 251, 2013.
25. A.Jakulin and I. Bratko, “Testing the significance of attribute inter actions,” Proceedings of the twenty-first international conference on Machine learning. ACM, New York, USA, p. 52, 2004.
26. Z. Zhao and H. Liu, “Searching for interacting features in subset selection,” Intelligent Data Analysis, vol. 13(2), pp. 207–228, 2009.
27. R. Shen, Y. Yang, and F. Shao, “Intelligent breast cancer prediction model using data mining techniques,” Intelligent Human-Machine Sys tems and Cybernetics (IHMSC), pp. 384–387, 2014.
28. C. M. Bishop, “Pattern recognition and feed-forward networks,” The MIT Encyclopedia of the Cognitive Sciences, Wilson and F. C. Keil (editors), MIT Press, vol. 13(2), 1999.
29. A.Motsinger-Reif, S. Dudek, and L. H. et al, “Comparison of ap proaches for machine-learning optimization of neural networks for detecting gene-gene interactions in genetic epidemiology,” Genet Epi demiol, vol. 32, pp. 325–340, 2008.
30. P. Lucek and J. Ott, “Neural network analysis of complex traits,” Genet Epidemiol, vol. 14, pp. 1101–1106, 1997.