Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

ENHANCED ARTIFICIAL BEE COLONY BASED FLEXIBLE NEURAL FOREST (EABC-FNT) AND ENSEMBLE GENE SELECTION (EGS) FOR CANCER SUBTYPES CLASSIFICATION ON GENE EXPRESSION DATA

Dr. N.Jayashri[1]
Assistant Professor,
Department of Computer Applications,
Dr. M.G.R Educatioal and Research Institute,
Maduravayal, Chennai-95
jayashrichandrasekar@yahoo.co.in

Dr. M.Deepika[2]
Assistant Professor,
Department of Computer Applications,
B. S. Abdur Rahman Crescent Institute Of Science and Technlogy,
Vandalur, Chennai-48
deepika.rbp@gmail.com

ABSTRACT: The categorization of cancer subtypes is critical for cancer analysis and diagnosis. Deep learning algorithms have centred substantial recognition for cancer subtype detection in recent years; but, It's difficult to create a Neural Network (NN), and the results of deep learning approaches aren't always predictable are mostly reliant on their structure. In gene expression datasets, a learning strategy takes a long time and the output of the representation diminishes due to duplicated genes and the curse of dimensionality. To increase the classifier's performance and address these difficulties, Enhanced Artificial Bee Colony based Flexible Neural Forest (EABC-FNT) and Ensemble Gene Selection (EGS) is proposed in this paper. The EABC algorithm is used to optimise the parameters of the EABC-FNT classifier to help to utilize the cancer subtypes classification. In the EABC algorithm, new modified onlooker bee behaviour is used with the better fitness food source as the middle. FNT is a particular NN through the benefit of formation and parameter tuning that may be utilised for multi-class classification. The Fisher Ratio (FR), Neighbourhood Rough Set (NRS), Correlation Based Gene Selection (CFS) and Greedy Hill climbing method are combined with the EGS algorithm. It is used to select primarily beneficial genes with data on the expression of a known breast cancer gene. The FR is used for the elimination of useless genes and NRS is then introduced to eliminate redundant genes. Experimentation on Ribonucleic Acid (RNA)-seq gene expression data of Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM), Lung Cancer (LUNG) show with the purpose of proposed EABC- FNT classifier gives higher accuracy with selected genes for cancer subtypes classification when compared to other methods such as Deep Flexible Neural Forest (DFNForest) and FNT classifier concerning the metrics like precision, recall, f-measure, accuracy, and error rate.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

INDEX TERMS: Ensemble Gene Selection (EGS), cascade forest, Cancer subtypes, gene selection, machine learning, Enhanced Artificial Bee Colony (EABC), classification, and Flexible Neural Forest (FNT).

1. INTRODUCTION

Breast cancer is the broadest disease analysed in ladies, and it is additionally the primary reason of death behind lung cancer [1]. The individuals who are diagnosed with breast cancer is growing all the time, and create personal of those diagnosed are ladies under 40 years old [2]. Furthermore, young women Triple-negative or HER2-positive breast cancers are more common, and women who acquire them are more likely to be diagnosed at an advanced stage of the disease [1]. A illness having a great deal of heterogeneity at the time is breast cancer growth, and it is contained unmistakable organic subtypes which current a diverse range of clinical, pathologic, and numerous prognostic and therapeutic propositions on sub-atomic genes [3]. Breast cancer genotyping research is important for determining therapy options and predicting prognosis [4]. In any case, those investigations won't center on the subtype classification.

Cancer subtyping is critical for determining which patients are most likely to benefit from new medicines and programs developed by specialists in their field. It is difficult to reliably classify cancer based on histology and clinical criteria because of this because clinician expertise is often a key factor [5,6]. Only a small portion of the genes in gene expression data are associated to malignant growth subtypes, with the remainder being repetitive or noisy genes. This results in the classification of cancer kinds based on gene expression data into a dimensionality reduction issue.

Using gene expression data to extract information is a prominent research area that has sparked a lot of interest lots of new medical applications [7], [8]. Gene selection approaches, which are split into three types: filter, wrapper, and embedded strategies [9], may be applied. The use of a specific indicator to assess the tight relationship between each variable is central to Filter Strategies and test classification and arranging them as indicated by the level of relationship from enormous excessively little. At long last, a present limit or the top k genes are acquainted with the structure of a gene subset. In bioinformatics, Essential genes/genes were chosen using rough set theory. [10], [11]. Nonetheless, it has a conspicuous issue that makes it unsatisfactory for continuous gene expression processing to fix this issue, Hu et al. [12], [13] proposed a to combine discrete and continuous gene expression data, the Neighborhood Rough Set (NRS) model was used. NRS model is utilized to remove redundant genes, but this results in higher computation time.

Neural networks have newly gained importance in classification issues together with cancer subtype classification [14]. The complexity of biological data is large, and sample sizes are small, which causes a large issue to usual classification algorithms. Using the fast improvement in machine learning algorithms, mostly in deep learning which approved straight dealing out of such high-dimensional biological data is difficult to comprehend without prior knowledge. Chen [15] proposed the Flexible Neural Tree (FNT), a kind of NN with a regular structure and

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

parameter customization. There are also several issues with FNT's multi-class classification. It is not ideal for multi-class categorization to begin with, a single root is an output node. Second, it is critical to continue expanding the FNT model in order to improve classification accuracy results. Moreover, the cost of the parameter optimization method improves considerably using improving the depth of the model.

To solve this issue, Enhanced Artificial Bee Colony based Flexible Neural Forest (EABC-FNT) classifier is proposed in this paper. Ensemble Gene Selection (EGS) It is used as an the top k genes in the data are identified using an algorithm for gene selection for gene expression afterward. Using the EGS algorithm, irrelevant or noisy gene genes are selected from thousands of gene genes in gene expression data. Using cancer subtypes, few genes are ultimately related. EABC-FNT parameter optimization is performed by using EABC algorithm in the direction of aid using the classification of cancer subtypes. In the EABC algorithm, new modified onlooker bee behaviour is implementing using the optimal fitness food source as the middle. The result of the proposed EABC-FNT classifier is assessed with another traditional classifier in terms of traditional metrics.

2. LITERATURE REVIEW

Tao et al [16] proposed a the Sequential Minimal Optimization- Multiple Kernel Learning (SMO-MKL) The estrogen receptor, human epidermal growth factor receptor 2 (HEGF-R2) (ER), and the progesterone receptor (PR) were all used to identify breast cancer subtypes (HER2). The SMO-MKL classifier is recommended for usage with this omics data. In experiments; the proposed SMO-MKL classifier performs better than the other modern algorithms and has rich biological analysis.

Wesolowski and Ramaswamy [17] used with cDNA microarray expression technologies methodologies and Quantitative real-time may now be used to diagnose cancers with precise gene expression patterns (RT-PCR). This technique is reshaping people's perceptions about breast cancer classification and treatment right now. Breast cancer classification based on gene expression is simply establishment towards making its way into the everyday experimental training and likely will balance, however not change, the predictable algorithms of classification.

Gao et al [18] proposed Based on functional spectrum deep learning, DeepCC (DeepCC) classifies cancer subtypes. which quantifies the ways in which genetic pathways work. DeepCC's ability to handle missing data was demonstrated in a Random gene subsampling is used in this simulation. In a nutshell, DeepCC is a new cancer classifier that is platform independent, powerful against missing data, and capable of sample prediction assisting in cancer molecular subtyping clinical performance.

Xu et al [19] The HI- DFNForest framework was utilized to categorize cancer subtypes utilizing multi-omics data using a hierarchical combined deep Flexible Neural Forest framework. Each omics dataset is investigated using the stacked autoencoder (SAE), and then sophisticated representations are developed by merging each learned representation with a layer of the autoencoder. Patients addicted to transformed cancer subtypes may be identified

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

using the Deep Flexible Neural Forest (DFNForest) model. The findings show that combining several the proposed system outperforms previous strategies and the use of the accuracy of cancer subtype classification increases with the use of omics data.

Chen et al [20] The Neighborhood Rough Set (NRS) model is used to develop an entropy-based gene selection strategy was developed to cope with real-valued data while yet preserving the inspired gene classification results. This metric's use might lead to the identification of compressed gene subsets. Finally, the entropy measure and adjacent granules are used to indicate a gene selection method. Several studies employing two types of gene expression data reveal that the recommended gene selection technique is a good way to improve tumor classification results.

Kong and Yu [21] introduced to incorporate of external relational data on genes addicted to the One of the GEDFN's main features is that it is a deep neural network design. Simulated tests and actual data study were carried out using the Cancer Genome Atlas datasets of breast aggressive carcinoma RNA-seq, and renal clear cell carcinoma. The proposed system provides higher results and straightforwardly interpretable gene selection results when compared to existing graph-guided classifiers and gene selection methods. Xu et al [22] DFN Forest (DFNForest) is an ensemble of Flexible Neural Trees (FNT) that was presented models that may help in cancer subtype categorization. Investigate the cascade structure of DFNForest in the direction of deepening the FNT model consequently with the purpose of the depth of the model is improved not including establishing extra parameters. As well as the DFNForest model, Fisher Ratio (FR) and NRS are proposed for Improved categorization results may be achieved by the gene for which genes have been expressed an improved gene selection strategy has been shown utilizing experiments based on RNA-seq gene expression data. When compared to traditional classifiers, the suggested DFNForest model provides enhanced results for the categorization of cancer subtypes.

3. PROPOSED METHODOLOGY

For optimal gene selection, four gene subsets such as Fisher Ratio (FR), Neighbourhood Rough Set (NRS), Correlation-Based Gene Selection (CFS), and greedy hill-climbing are combined in the Ensemble-Based Gene Selection (EGS) approach. To achieve greater precision, it is used to select genes for breast cancer gene expression data. To extract useless genes and consequently reduce redundant genes, an EGS algorithm is used. Enhanced Artificial Bee Colony based Flexible Neural Forest (EABC-FNT) classifier is used for classification of cancer subtypes; In EABC-FNT classifier, parameter optimization is performed using EABC algorithm. In the EABC algorithm, new modified onlooker bee behavior is used with the better fitness food source as the middle. FNT is a particular NN using the benefit of Automatic formation and parameter optimization that may be utilised for multi-class categorization. The suggested EABC-FNT classifier is shown to provide better outcomes than existing classifiers when analysing RNA-seq gene expression data. The suggested system's flow diagram is shown in Figure 1.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

3.1.DATASET DETAILS

Cancer subtype predictions were made using Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM), and Lung Cancer (LUNG) RNA-Seq gene expression datasets from The Cancer Genome Atlas (TCGA) [23]. BRCA data is divided into four groups in 514 samples: basal-like, HER2-enriched, Luminal-A, and Luminal-B. 164 GBM samples have Classical, Mesenchymal, Neural, and Proneural subtypes, whereas 275 LUNG samples have Bronchioid, Magnoid, and Squamoid subtypes. The characteristics of the three cancer types are shown in Table 1. These datasets were used to compare the proposed EABC - DFNForest model to Deep Flexible Neural Forest (DFNForest) and FNT approaches in terms of gene selection and classification performance.

TABLE 1. THE DETAIL OF THE THREE CANCER TYPES

Dataset	Sample	Gene	Class
RCA	514	4247	4
GBM	164	3398	4
LUNG	275	4596	3

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

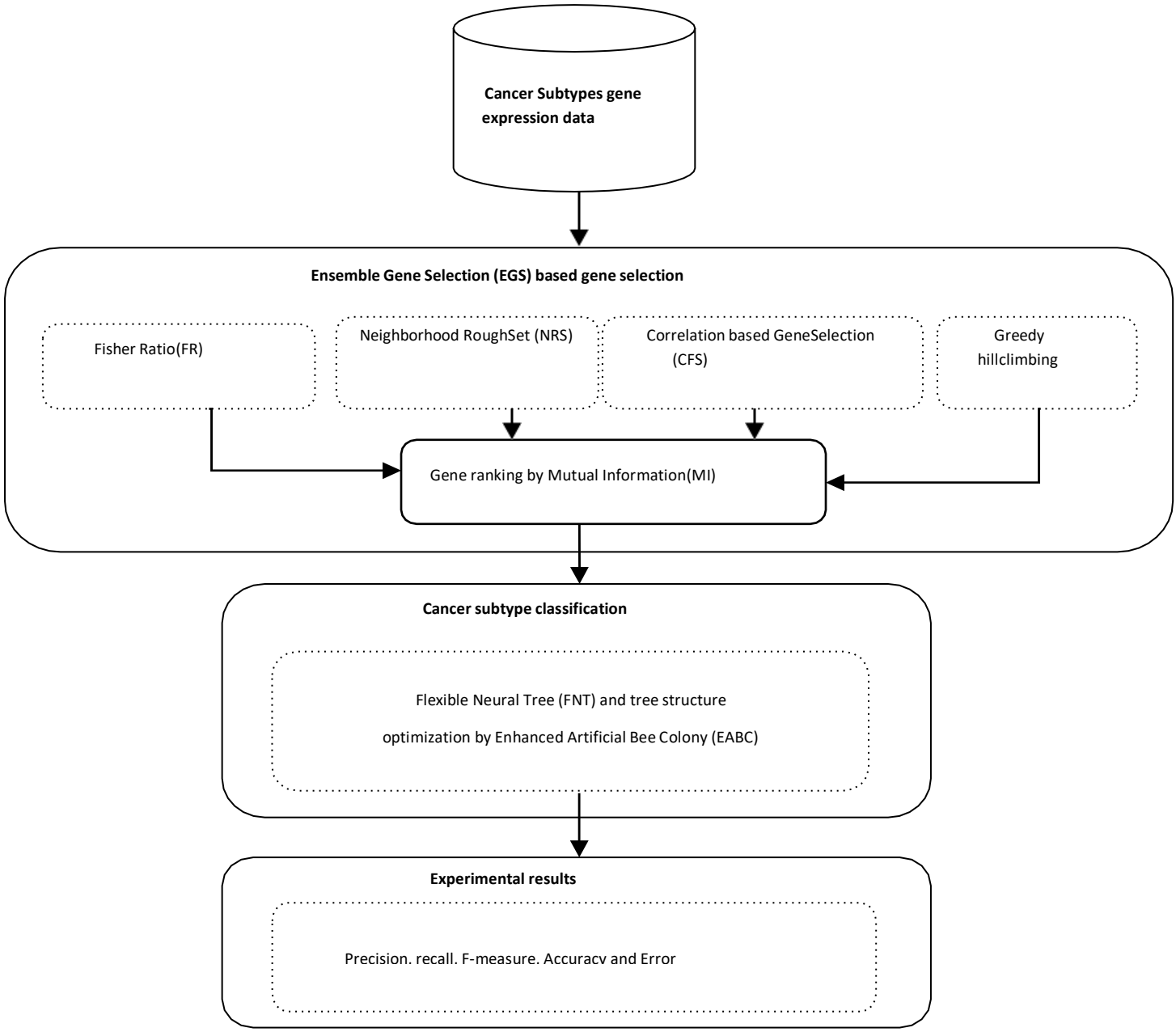


FIGURE 1. FLOW DIAGRAM OF THE PROPOSED SYSTEM

3.2.GENE SELECTION METHODS

Thousands of genes are usually represented in gene expression data; but, the number of samples available is frequently minimal. Only a few genes are fundamentally related with cancer subtypes among hundreds of genes in gene expression data, while the remainder might be termed redundant or noisy genes. As a result, gene selection may be defined as the challenge of reducing the dimensionality while attempting to choose key genes while maintaining unique gene classification accuracy [24]. The selected subsets of genes are aggregated in the second stage,

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

which employs aggregation.

3.2.1. FISHER RATIO(FR)

The Fisher Ratio reflects lengths across classes to distances within classes. Each sample in a dataset with two classes must be able to be labeled as $Y \in \{+1, -1\}$ and gene express vector i be = { , . . , } meant for each gene i, the standard deviation σ_i^+ (resp., σ_i^-) and the mean μ_i^+ (resp., μ_i^-) are considered and the fisher ratio FR_i is computed as follows in equation (1):

$$FR_i = \frac{(\mu_i^+ - \mu_i^-)^2}{(\sigma_i^+)^2 + (\sigma_i^-)^2} \quad (1)$$

Gene using maximum FR_i value is mostly about giving information, and the levels are different mainly on standard in the two classes at the same time as moreover supporting those through little difference between the classes [25]. So, the genes with high FR_i values are chosen as the best ones.

3.2.2. NEIGHBOUR ROUGH SET(NRS)

A Neighborhood Rough Set (NRS) [12], [13] classifier has been developed for handling together distinct and continuous gene expression data along with keeping information fundamental towards classifying data accurately. Let us consider an individual a set of samples U = {x1, … xn} An individual a set of real-type genes describing U, and D a finish gene. If A generate a family of neighborhoods on the region, it is called NDT = {U, A, D} a neighborhood classifier. If D splits U into N sameness classes:X1, X2, … , XN , ∀B ⊆ A, consequently the lower and upper determination of assessment D relating to B be able to be described as in the equation (2-3):

$$N_B^L D = \bigcup_{i=1}^{N} N_B^L X_i \quad (2)$$

$$N_B^U D = \bigcup^{\cdot\cdot} N_B^U X_i \quad (3)$$

NLBX = {xi|δB(xi) ⊆ X, xi ∈ U}, NUBX = {xi|δB(xi) ∩ X ≠ ϕ, xi ∈ U}, δB(xi) is indicated as the piece of data about the area made with gene B and measure. The decision positive area is also known as the bottom target of decision D, or POSB(D) [13]. The size of the positive area shows how well the classification problem can be split up in known gene space. The more positive an area is, the less likely it is that something related to class boundaries will happen. This ensures a more accurate explanation of an issue with categorizing the genes selected. As a result, the necessity for the decision gene D on the situation gene B is reflected in the equation as (4)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

$$\gamma B(D) = Card(NLBD)/Card(U)(4)$$

Especially a local decision system NDT= {U, A, D}, B ⊆ A, ∀ ∈ A − B, the importance of a capable in the direction of B be capable of being explained as in the equation (5)

$$SIG(a, B, D) = \gamma BUa(D) - \gamma B(D)(5)$$

A greedy attribute reduction method is used, and it is chosen based on the gene importance index. The method starts with an empty set, figures out the importance of each remaining gene's gene each time, and joins the reduction set with the gene having the greatest gene significance value. This method is repeated until all of the remaining genes have a value of 0. This implies that no matter how many additional genes are introduced, the way the addiction system functions remains same. The algorithm for the forward search guarantees with the purpose of significant genes are further in the direction of the reduction set first, consequently, the purpose of significant genes is not absent.

3.2.3. CFS based Gene Selection

The CFS-based FS method figures out how accurate a subset of genes is by looking at how well each gene can predict on its own and how much overlap there is between them. We prefer subsets with attributes that have a high correlation with the class gene but a low correlation with each other. People say that a good gene set has genes that have the most in common with the class and the least in common with each other. A gene is called "relevant" if and only if there are vi and c for which the gene is needed p(Vi = vi) > 0 such that

$$(C = c|Vi = vi) \neq (C = c)(6)$$

CFS only looks at how tightly two nominal genes are linked, so the process starts by making numeric genes into nominal ones. The generalised correlation-based gene selection, on the other hand, doesn't need to change the data in any way. Instead, it just measures the correlation between any two variables. So, the method can be used to solve a wide range of problems, even ones with numbers. CFS is an algorithm that does everything by itself and does not need to be supervised in terms of threshold limits. It works in the space of the original genes, so it can be understood in terms of those genes. Due to the repeated use of the learning algorithm, the CFS filtering method does not have a high computational cost. If the correlation between the parts is known and the correlation between the parts is given, then equation can be used to figure out the correlation. (7)

$$r_{zc} = \frac{k\overline{r_{ic}}}{\sqrt{k+(k-1)\overline{r_{ii}}}} \quad (7)$$

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

A component's average correlation to an outside gene is called rzi, and it may be calculated by multiplying the total number of component-outside gene correlations by rzc, and rii is the average component-to-component inter-correlation [30-31].

3.2.4. Attribute filtering using greedy hill-climbing

If a gene selection algorithm is to function on data containing numerous genes, it must search the space of gene subsets within realistic time restrictions. Local changes to the present gene subset are taken into account by one basic search approach dubbed greedy hill-climbing [32]. A local alteration is sometimes as simple as adding or removing a single gene from a gene subset. Best initially travels across the search space by making local adjustments to the current gene subset, similar to greedy hill-climbing. The evaluation approach employed here is Best first, which uses a greedy hill-climbing algorithm 3.1 supplemented with a backtracking capability to search the space of attribute subsets.

Algorithm 1. Greedy Hill Climbing Algorithm
1. Let s ←start state
2. Expand s by making each possible local change
3. Evaluate each child t of s
4. Let s′ ←child t with highest fitness evaluation e(t)
5. If e(s′) ≥ e(s) then s ← s′ , go to step 2
6. Return s

3.2.5. Combiner

Based on gene–class and gene–gene mutual information, it combines a subset of genes. First, the combiner evaluates the top-ranked genes from all of the chosen subsets, and if all of the top-ranked genes are the same, we choose that common gene as the best gene without calculating gene-class and gene-gene mutual information. However, if the genes are different, we calculate gene–class mutual information for each gene and then choose the gene with the greatest gene–class mutual information. A user-defined threshold is used to determine whether or not a gene is picked based on its gene-gene mutual information with all other genes that have been previously identified as optimal. A non-selected gene's gene-relevance with chosen genes is measured using gene-gene mutual information. Based on an extensive experimental analysis, introduce an effective value for α (α 0.75). A 'combiner' plays a key part in ensemble multiple gene selection approaches in an ensemble approach. The suggested ensemble gene selection method's combiner, on the other hand, Emphasizes the removal of redundant information within the selected subset of genes by merging information about genes within classes and genes within individual genes.

In information theory, the mutual information I(X, Y) measures the degree of uncertainty in X as a result of knowing Y. (8)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

$$(X, Y) = \sum_{x,} (x, y) \log \quad 2 \frac{p(x,y)}{(x)(y)} \quad (8)$$

The equation, p(x) and p(y) are the marginal probability distribution functions for X and Y, and p(x) and p(y) are the joint probability distribution functions of X and Y. (9)

$$(X, Y) = (X) - (X|Y) \ (9)$$

H(X) is the marginal entropy, H(X|Y) is the conditional entropy, and H(X,Y) is the combined entropy (X; Y). Assuming H(X) measures random variable uncertainty, then H(X) quantifies the information that Y does not provide about X. Knowing one gene about another validates the intuitive idea of reciprocal information, which is the amount of information supplied by knowing one gene about the other. To establish how much information is obtained between genes and between genes and class genes, it is necessary to do a comparative analysis the technique employs a mutual information measure.

3.3.CLASSIFICATION

In this section, a brief explanation of the Flexible Neural Tree (FNT) go after with the proposed Enhanced Artificial Bee Colony based Flexible Neural Forest (EABC-FNT) into a classifier is discussed.

3.3.1. Flexible Neural Tree (FNT)

Combining the function set F with the terminal instruction set T, the FNT model is formed as shown in equation (10):

$$S = F \cup T = \{+2, +3, \ldots +N\} \cup \{x1, \ldots xn\} \ (10)$$

where +i(i =2, 3, 4,,..,N) By using I parameters, non-leaf node instructions are denoted. x1, x2,..., xn are leaf node commands that do not include arguments. To make a FNT, use the nonterminal instruction +i(i =2, 3, 4,...,N), in which the I values for non-leaf nodes and between weights linking children are generated at random [15], [25]. For the FNT in the equation, the following flexible activation function might be assessed. (11),

$$f(x) = (1 + e{-}X){-}1 \ (11)$$

The results of a flexible neuron +n can be created as tracked in the equation (12),

$$\sum n = \sum n \quad wj * xj (12)$$

where xj(j =1, 2,…,n) are the inputs. The result of the node +n is calculated as in the equation (13),

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

$$outn = f(sumn) = (1 + e{-}sumn){-}1 (13)$$

The depth-first technique [26] may be used to recursively calculate the whole output of the FNT from left to right. This FNT format allows for over-layer connections and automatically selects a sparse model that provides improved classification performance. two major stages are involved while developing FNT optimization: tree structure optimization followed by parameter optimization.

3.3.2. TREE STRUCTURE OPTIMIZATION BY Grammar Guided Genetic Programming (GGGP)

To advance the development of FNT, Grammar Guided Genetic Programming (GGGP) is presented [27]. GGGP helps to prevent establishing an invalid tree as a result of crossover or mutation. A grammar that is not dependent on context G is represented by the four-tuple G = {N, T, P, ∑ }where N stands for nonterminal characters and T stands for terminal characters. The portions of P are referred to as grammatical rules. ∑ is a portion of N and the start symbol. The grammatical rules are written as x y, with x referring to N and y referring to NUT.

3.3.3. PARAMETER OPTIMIZATION BY ENHANCED ARTIFICIALBEE COLONY(EABC) ALGORITHM

Artificial bees may be divided into three types using the standard ABC method for migratory behaviour: employed bees, observer bees, and scout bees. Onlooker bees settle on a certain food supply after seeing the dance of employed bees, while scout bees opt to look for food everywhere they can. The detailed bee colony behaviour and associated model may be found in the literature [28]. Each of the food sources is initialized with equation at the startup step (14). Predefined parameters are used to determine the quantity of food sources available.

$$tj, = li + (0,1) * (ui - li) \ (14)$$

where tj,i is the jth food source's ith dimensional information. The parameters tj, li, and ui grow in value as we go closer to their bottom and upper boundaries. Tj is the jth food source. When she is in the employed bee stage, her memory is searched at a certain rate j,i. The rate of change of the tj,i food supply (as mentioned in equation (15) affects the convergence speed to a certain extent. If equation (14)'s result is superior than the bee's, a greedy selection technique is used to simplify the memory.

$$vj, = tj, + \varphi j,(tj,i - tr,i) \ \ (15)$$

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

an unpredictably selected food source (tr), an unpredictably chosen index (I), and $\varphi j,i$ represents a random integer in the range [−1, 1]. Fitness of the solution is computed via the equation (16). The improved fitness value denotes an enhanced objective function value; consequently, make best use of the fitness function which is able towards attain the optimal thresholds.

$$fit(t_j) = \begin{cases} \frac{1}{(1+f(t_j))}, f(t_j) \geq 0 \\ 1 + abs\,(f(t_j)), f(t_j) < 0 \end{cases} \quad (16)$$

where $f(tj)$ be able to be computed via the use of classification accuracy. An employed bee informs an observer bee on the health of the food supply. The observer bee will most likely pick one source to investigate. The equation may be used to represent the probability function. (17)

$$prob_j = \frac{fit(t_j)}{\sum_{j=1}^{SN} fit(t_j)} \quad (17)$$

where SN is denoted as the amount of food sources. In the classic ABC, the employed bee and the spectator bee use equation (15) to find food sources. To put it another way, the employed bee and the spectator bee utilize the same tactics to find a new and better food source in the same area. Aside from genuine honey bee colonies, the employed bee and the spectator bee use distinct algorithms in the direction of new food source discovery [29]. As a result, it is fairly practical for spectator bees when they seek for the finest meal since the pattern varies from the search strategy. The optimum fitness food source as the middle is used to apply modified observer bee behaviour in the literature. The equation may be used to represent the created hunt for new food sources. (18)

$$vbest = tbest + \varphi j, * (tbest - tr,) \quad (18)$$

$$Nj,i \qquad Nj,i$$

where $tbest$ is as the best outcomes from each of the nearby food sources in relation to the present food source and tj , Nj described each one the neighborhood food sources together with tj; The remaining parameters are identical to those in equation (15). It's possible to be visibly experimental with the key being how the local food supply is portrayed. The centre of this pattern has a distortion in how the nearby food supply is characterised, according to equation (14).

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

Karaboga and Gorkemli [29] introduced an explanation of neighboring food sources. Regarding different ways of describing problem solution, it requires in the direction of describe diverse capacity for similarity. In this work, the neighboring food source distance is described as go behinds with equation (19).

$$d_{j,i} = \frac{\sum_{x=1}^{S} t_{x,i}}{SN}, \; i = 1 \ldots SL \quad (19)$$

where $dj,i$is the scope of the food source tj's research in the neighborhood. The entire amount of food sources is marked by SN, while the dimension of the food source is given by SL. If a solution's Euclidean distance in the direction of tj is smaller than dj,i, it is considered the neighborhood of solution tj, which differs from the traditional ABC approach. If the spectator bee approaches the food source tj, she will first inspect all of the nearby food sources before picking the best food source tbest and expanding her search using the equation (18). In Nj, the best food source is described via the equation (20)

$$fit\left(t_{N_i}^{best}\right) = \max\left(fit\left(t_{N_i}^{1}\right), \ldots, fit\left(t_{N_i}^{s}\right)\right) \quad (20)$$

Proposed classifier has some of advantages which are described as follows: 1) FNT is a sparse model that lets cross-layer connections happen. This prevents overfitting and gives better results for classification. 2) FNT optimises the structure and settings automatically. Additionally, with the use of numerous FNTs, The overall accuracy of the classifier is improved by the use of ensemble learning methods. It is utilized to choose the shape of the FNT in each forest, and the cascade stages are specified adaptively.

4. RESULTS AND DISCUSSION

In this section, we utilized the Cancer Genome Atlas' RNA-Seq gene expression datasets to predict cancer subtypes in three cancer types: BRCA, Glioblastoma Multiforme (GBM), and Lung Cancer (LUNG). The experimental results of classification results with respect to three datasets are shown in Table 2- 4. Overall results comparison of various classifiers with respect to the precision, recall, f- measure, accuracy and error results and their chosen genes of cancer subtypes are also discussed in Table 2, 3, and 4. Highly important model evaluation metrics are recall and precision. Percentage of relevant results is referred using precision (equation (21)) and percentage of total relevant results which are classified correctly is referred as recall (equation (22)).

$$Prec = \frac{T_P}{T_P + F_P} \quad (21)$$

$$Rec = \frac{T_P}{T_P + F_N} \quad (22)$$

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

Recall and precision's harmonic mean produces F-measure and it is expressed as in equation (23),

$$F - Measure = \frac{2*Prec*Rec}{Prec+Rec} (23)$$

Accuracy is simply the sum of the diagonals divided by the total discussed in equation (24):

$$Accuracy = \frac{T_P+T_N}{T_P+F_P+T_N+F_N} (24)$$

Table 2 provides a full explanation of the identified important genes of BRCA subtypes, along with their metrics data.

TABLE 2. RESULTS COMPARISON OF THE CLASSIFIERS WITH RESPECT TO SELECTED BRCA GENES

METRIC	Method	Genes	FNT	DFN Forest	EABC-FNT
Precision (%)	Original data	4247	87.6647	87.6647	92.3515
	NRS	8	89.3188	90.0944	94.1886
	Fisher Ratio(FR)	30	92.4257	92.5385	95.1975
	FR +NRS	5	92.8458	92.7014	96.8208
	Proposed EGS	3	94.4664	96.0726	97.3624
Recall(%)	Original data	4247	87.6592	90.0833	92.3510
	NRS	8	89.3058	92.5333	94.1816
	Fisher Ratio(FR)	30	92.4146	92.6996	95.1956
	FR +NRS	5	92.8395	94.5619	96.8219
	Proposed EGS	3	94.4666	96.0653	97.3633
F-measure (%)	Original data	4247	87.6619	88.874	92.3512
	NRS	8	89.3123	91.3138	94.1851
	Fisher Ratio(FR)	30	92.4201	92.6190	95.1965
	FR +NRS	5	92.8426	93.6316	96.8213
	Proposed EGS	3	94.4665	96.0689	97.3628
Accuracy (%)	Original data	4247	87.6619	90.0871	92.3475
	NRS	8	89.3101	92.5359	94.1841

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

	FR	30	92.4182	92.7007	95.1966
	FR +NRS	5	92.8420	94.5609	96.8213
	Proposed EGS	3	94.4667	96.0678	97.3628
Error (%)	Original data	4247	12.3381	9.9129	7.6525
	NRS	8	10.6899	7.4641	5.8159
	FR	30	7.5818	7.2993	4.8034
	FR +NRS	5	7.1580	5.4391	3.1787
	Proposed EGS	3	5.5333	3.9322	2.6372

Table 3 shows a full explanation of the identified relevant genes of GBM subtypes, along with their metrics data.

TABLE 3. CLASSIFICATION RESULTS FOR THE SELECTED GBM GENES

METRIC	Method	Gene	FNT	DFN Forest	EABC-FNT
Precision (%)	Original data	3398	82.7827	82.7827	87.9333
	NRS	6	83.5862	85.9039	90.0227
	Fisher Ratio(FR)	287	86.5205	87.5224	90.9937
	FR +NRS	6	88.6132	89.0914	92.9070
	Proposed EGS	5	90.8564	92.6764	95.6153
Recall(%)	Original data	3398	82.7835	85.9008	87.9333
	NRS	6	83.5716	87.5198	90.0231
	Fisher Ratio(FR)	27	86.5225	89.0755	90.9955
	FR +NRS	6	88.6075	90.6132	92.9074
	Proposed EGS	5	90.8418	92.6685	95.6184
F-measure (%)	Original data	3398	82.7831	84.3417	87.9333
	NRS	6	83.5789	86.7118	90.0229
	Fisher Ratio(FR)	27	86.5215	88.2989	90.9946
	FR +NRS	6	88.6103	89.8523	92.9072
	Proposed EGS	5	90.8491	92.67245	95.6168
Accuracy(%)	Original data	3398	82.7840	85.9035	87.9341
	NRS	6	83.5786	87.5221	90.0235
	Fisher Ratio(FR)	27	86.5215	89.0818	90.9947
	FR +NRS	6	88.6109	90.6121	92.9076
	Proposed EGS	5	90.8476	92.6722	95.6151
Error(%)	Original data	3398	17.2160	14.0965	12.0659
	NRS	6	16.4214	12.4779	9.9765
	Fisher Ratio(FR)	27	13.4785	10.9182	9.0053
	FR +NRS	6	11.3891	9.3879	7.0924
	Proposed EGS	5	9.1524	7.3278	4.3849

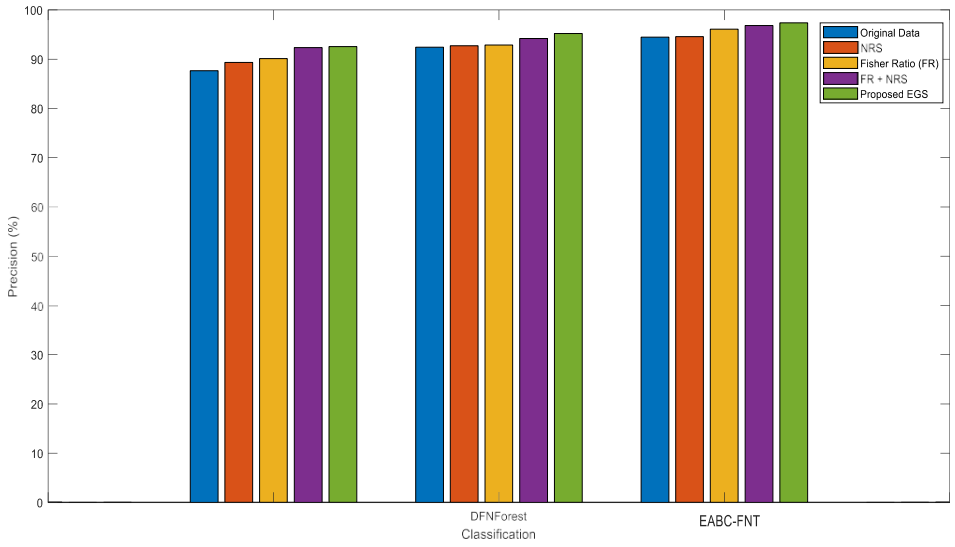
Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

Table 4 shows a full explanation of the identified important genes of LUNG subtypes, together with their metrics data.

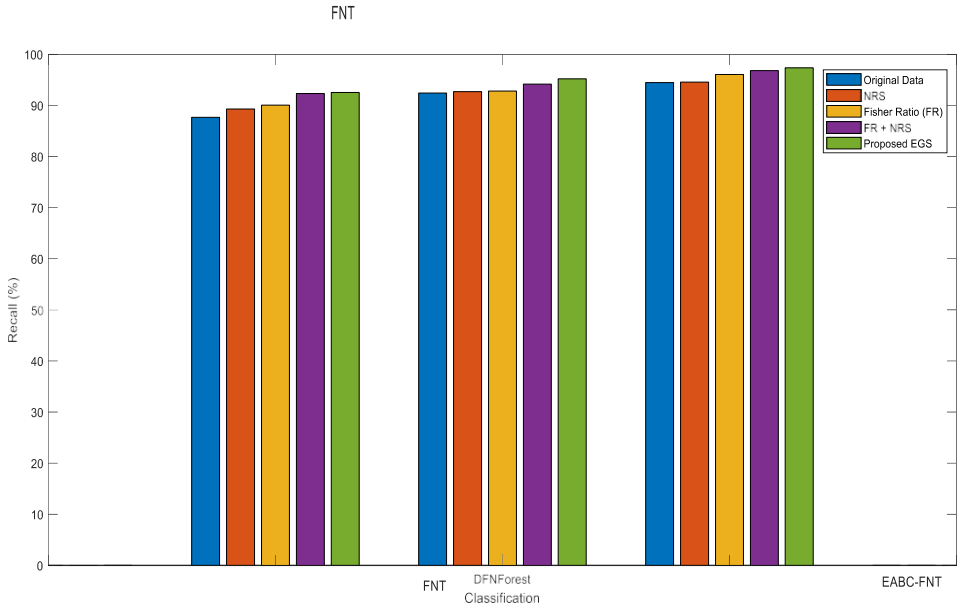
TABLE 4. CLASSIFICATION RESULTS FOR THE SELECTED LUNG GENES

METRIC	Method	Genes	FNT	DFN Forest	EABC-FNT
Precision (%)	Original data	4596	86.4381	86.4381	90.4731
	NRS	6	86.7701	87.9033	92.5156
	Fisher Ratio(FR)	25	89.2726	89.0130	93.1256
	FR +NRS	8	90.2119	90.9511	93.8262
	Proposed EGS	4	91.8623	94.3437	96.6493
Recall(%)	Original data	4596	86.4313	87.9055	90.4745
	NRS	6	86.7713	89.0100	92.5182
	Fisher Ratio(FR)	25	89.2748	90.9457	93.1227
	FR +NRS	8	90.2055	91.5388	93.8169
	Proposed EGS	4	91.8616	94.3463	96.6521
F-measure (%)	Original data	4596	86.4347	87.1718	90.4738
	NRS	6	86.7707	88.4566	92.5169
	Fisher Ratio(FR)	25	89.2737	89.9793	93.1241
	FR +NRS	8	90.2087	91.2450	93.8215
	Proposed EGS	4	91.8620	94.3450	96.6507
Accuracy(%)	Original data	4596	86.4230	87.9025	90.4700
	NRS	6	86.7711	89.0122	92.5152
	Fisher Ratio(FR)	25	89.2733	90.9487	93.1245
	FR +NRS	8	90.2089	91.5361	93.8207
	Proposed EGS	4	91.8625	94.3429	96.6493
Error(%)	Original data	4596	13.5770	12.0975	9.5300
	NRS	6	13.2289	10.9878	7.4848
	Fisher Ratio(FR)	25	10.7267	9.0513	6.8755
	FR +NRS	8	9.7911	8.4639	6.1793
	Proposed EGS	4	8.1375	5.6571	3.3507

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

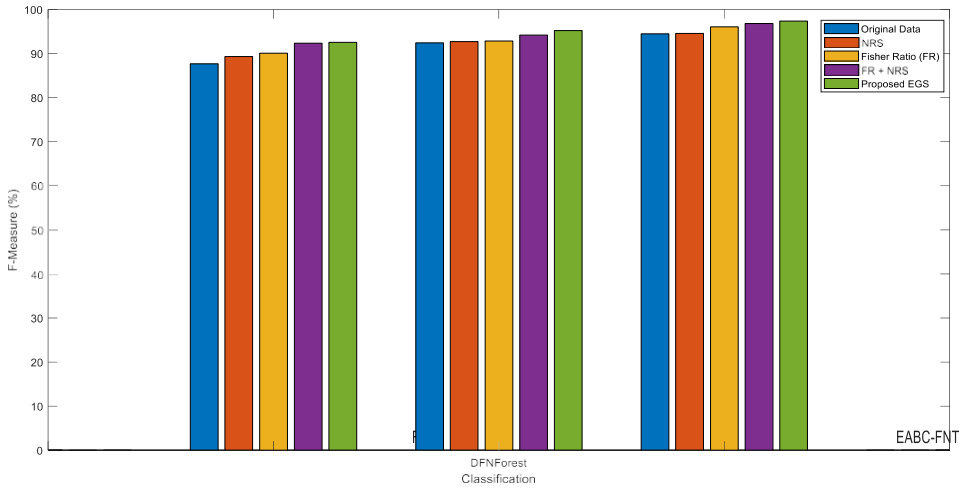


(a) Precision Results Comparison vs. Classifiers (BRCA genes)

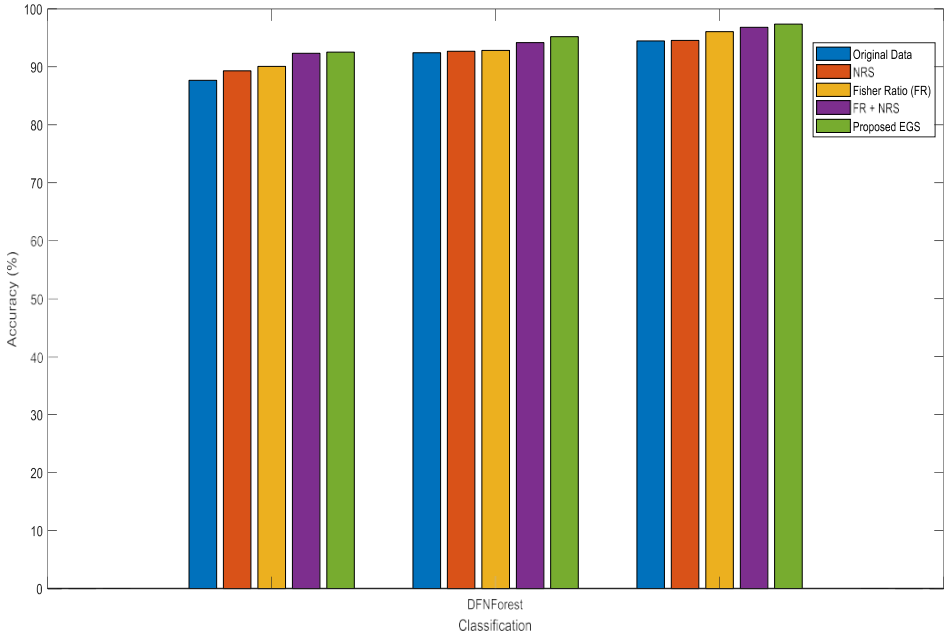


(b) Recall Results Comparison vs. Classifiers (BRCA genes)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

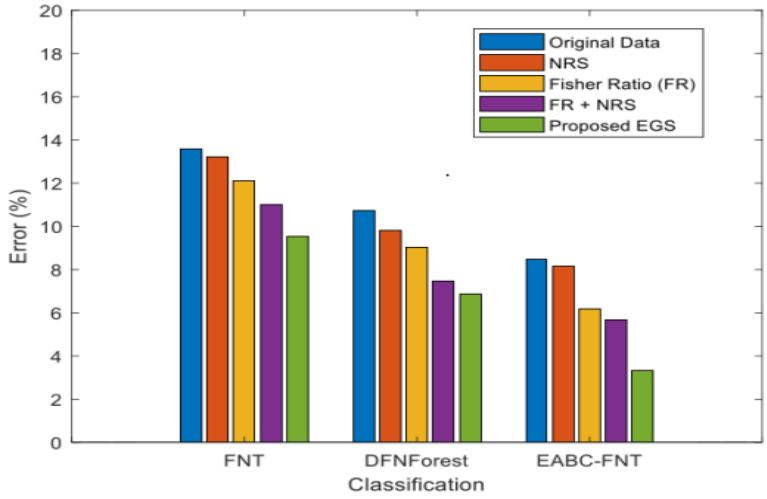


(c)F-measure Results Comparison vs. Classifiers (BRCA genes)



(d) Accuracy Results Comparison vs. Classifiers (BRCA genes)

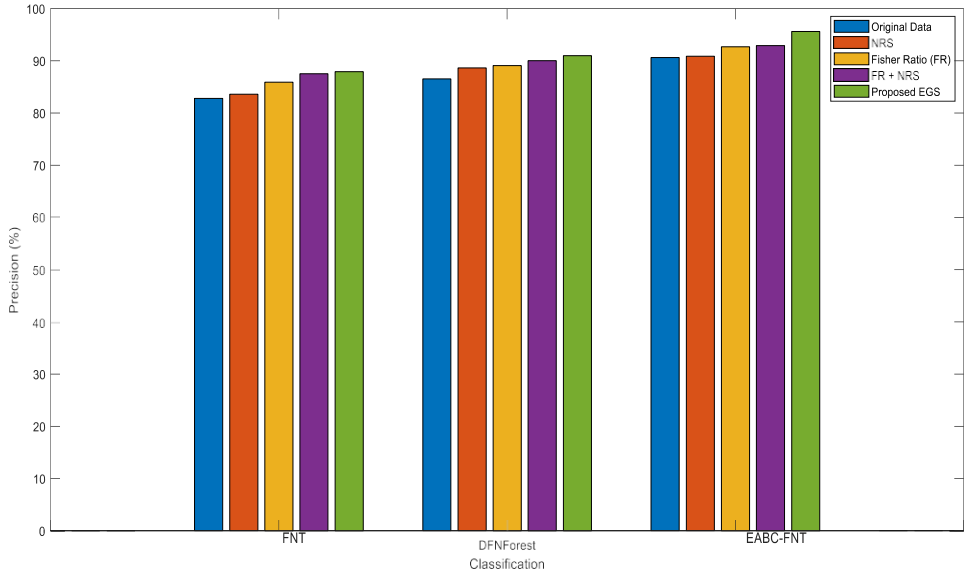
Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G



(e) Error Results Comparison vs. Classifiers (BRCA genes)

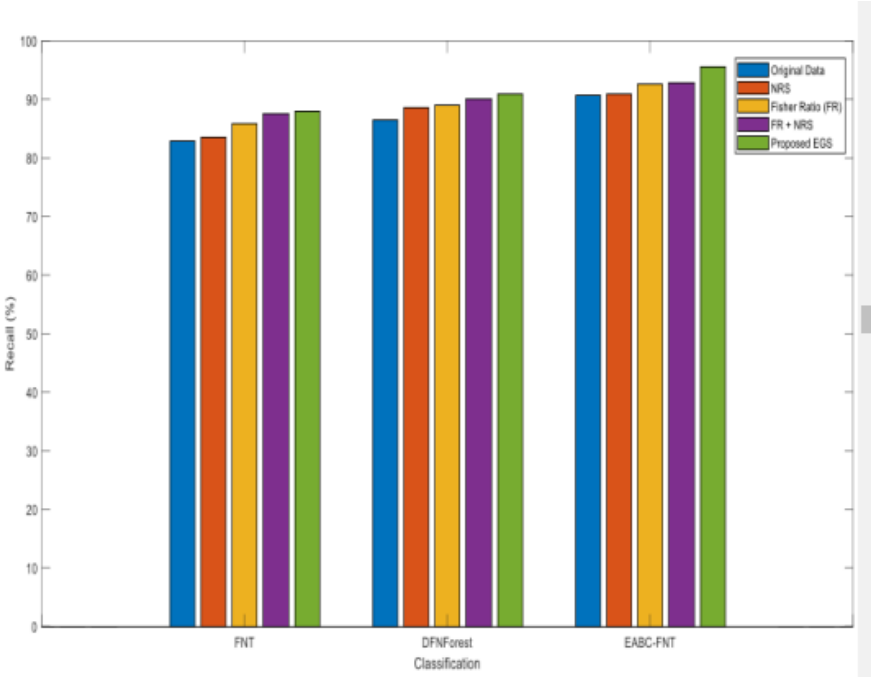
FIGURE 2. FINAL RESULTS COMPARIOSN VS. CLASSIFIERS (BRCA GENES)

Figure 2(a-e) shows the performance comparison results of the various classifiers such as FNT, DFNForest and EABC-FNT classifier with four feature selection methods such as NRS, FR, FR+NRS and proposed EGS algorithm under BRCA genes. The proposed EABC-FNT classifier produces higher accuracy results of 96.6493% for proposed EGS feature selection, the other methods such as FNT, and DFN Forest gives lesser accuracy results of 91.8625% and 94.3429% respectively.

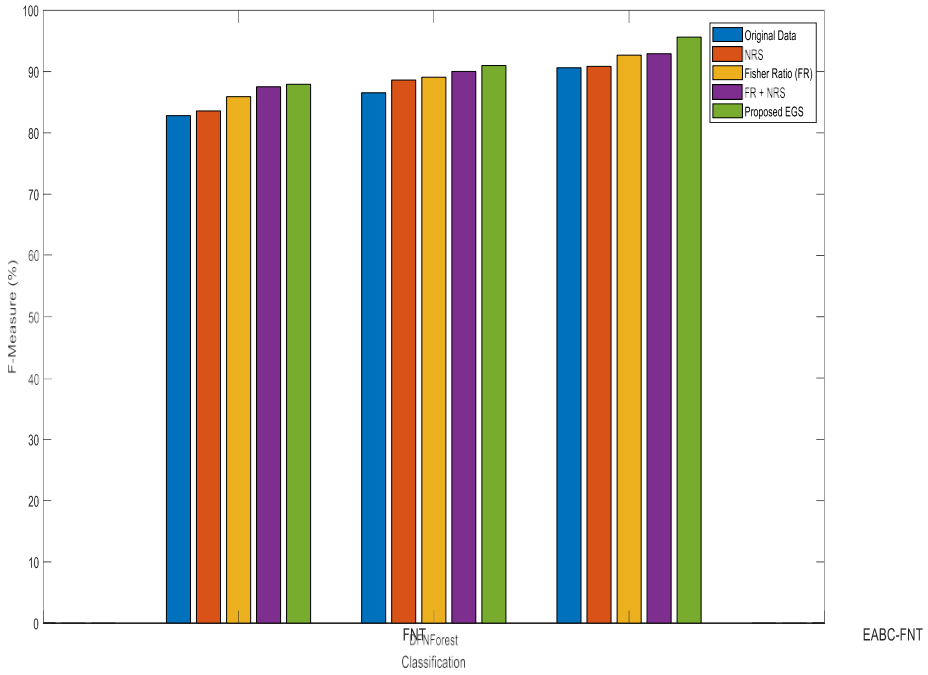


(a) Precision Results Comparison vs. Classifiers (GBM genes)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

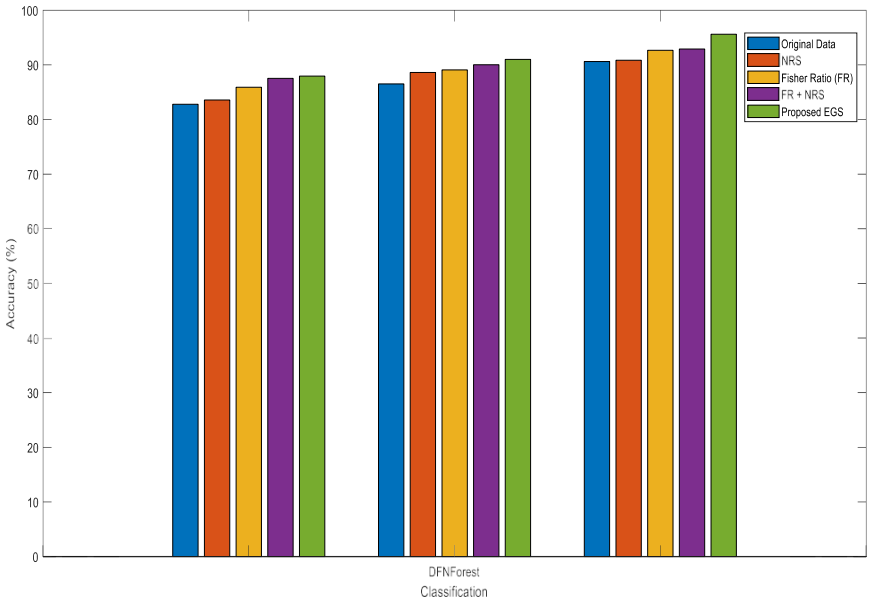


(b) Recall Results Comparison vs. Classifiers (GBM genes)

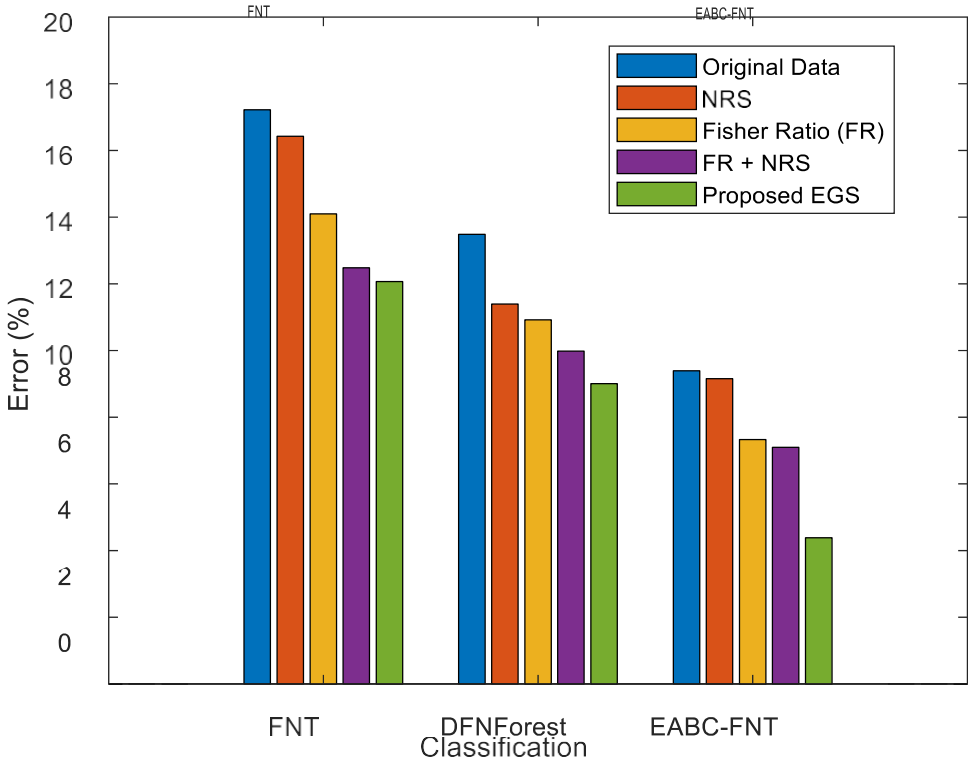


(c) F-measure Results Comparison vs. Classifiers (GBM genes)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G



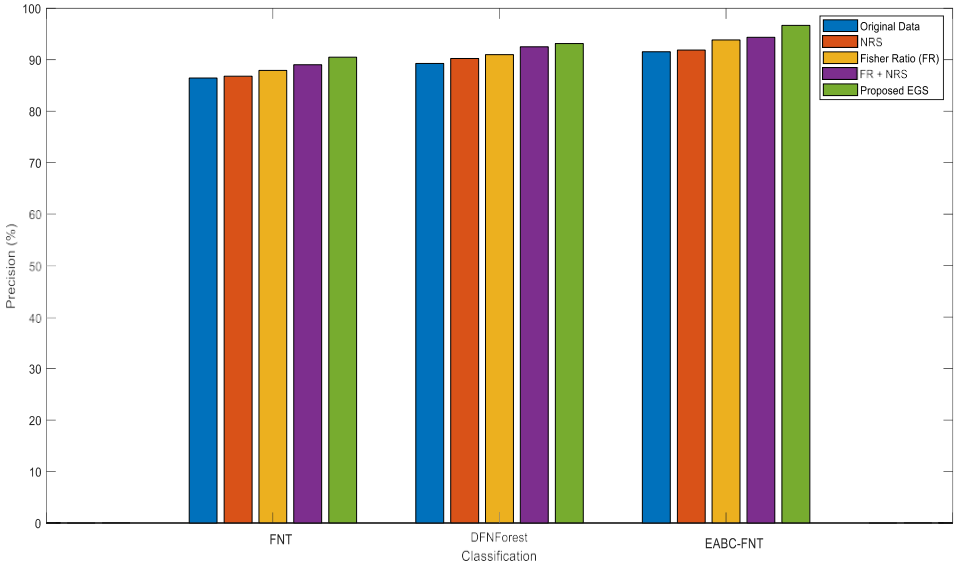
(d) Accuracy Results Comparison vs. Classifiers (GBM genes)



(e) Error Results Comparison vs. Classifiers (GBM genes)

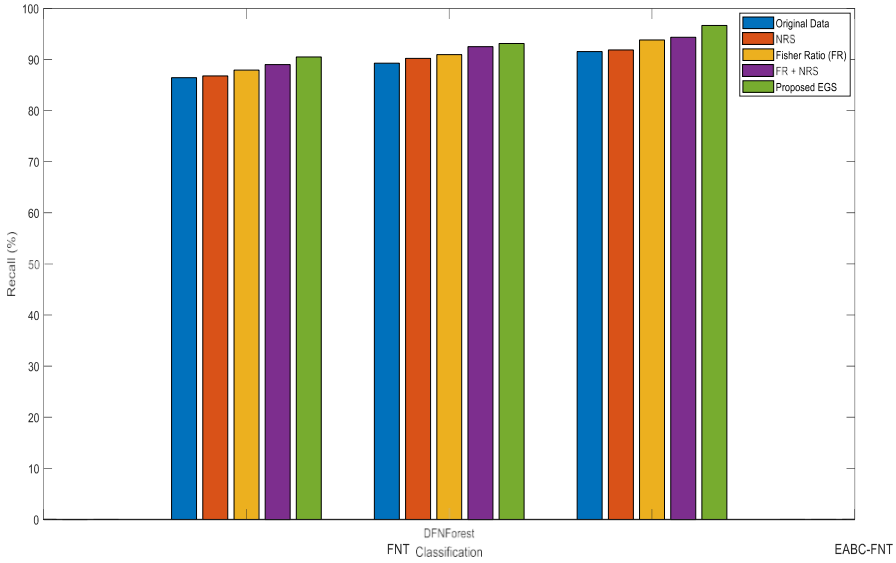
Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

FIGURE 3. FINAL RESULTS COMPARIOSN VS. CLASSIFIERS (GBM GENES)



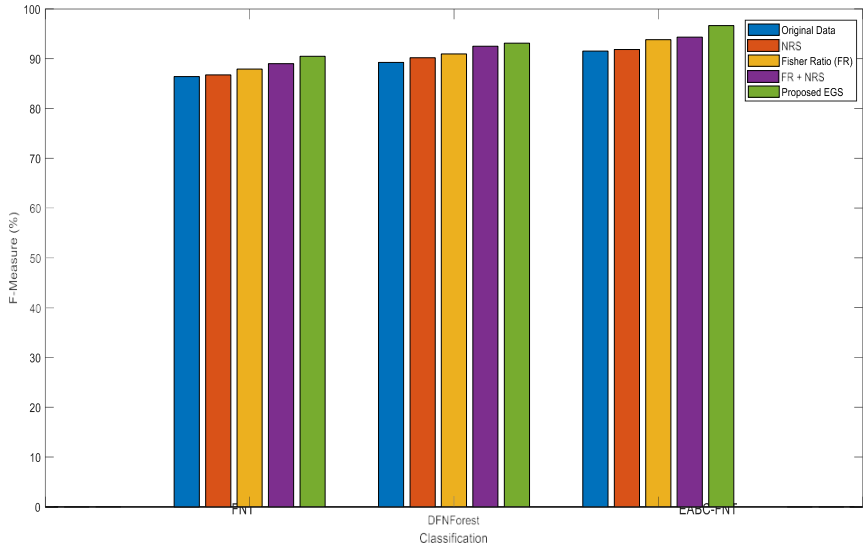
(a) Precision Results Comparison vs. Classifiers (LUNG genes)

Figure 3(a-e) shows the performance comparison results of the various classifiers such as FNT, DFNForest and EABC-FNT classifier with four feature selection methods such as NRS, FR, FR+NRS and proposed EGS algorithm under GBM genes. The proposed EABC-FNT classifier produces higher accuracy results of 95.6151% for proposed EGS feature selection, the other methods such as FNT, and DFN Forest gives lesser accuracy results of 90.8476% and 92.6722% respectively.

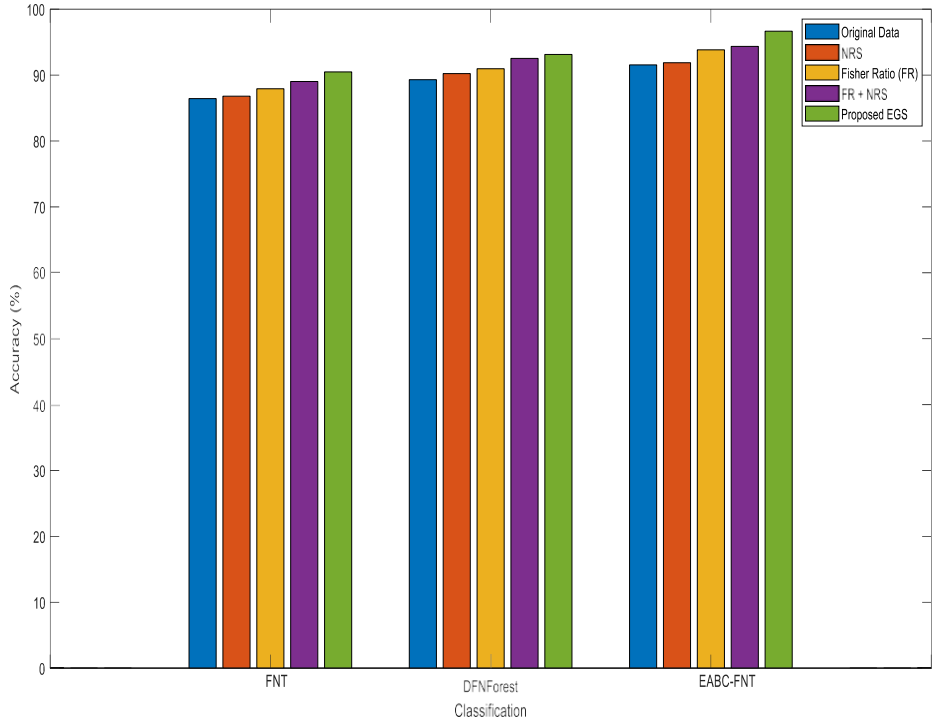


(b) Recall Results Comparison vs. Classifiers (LUNG genes)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

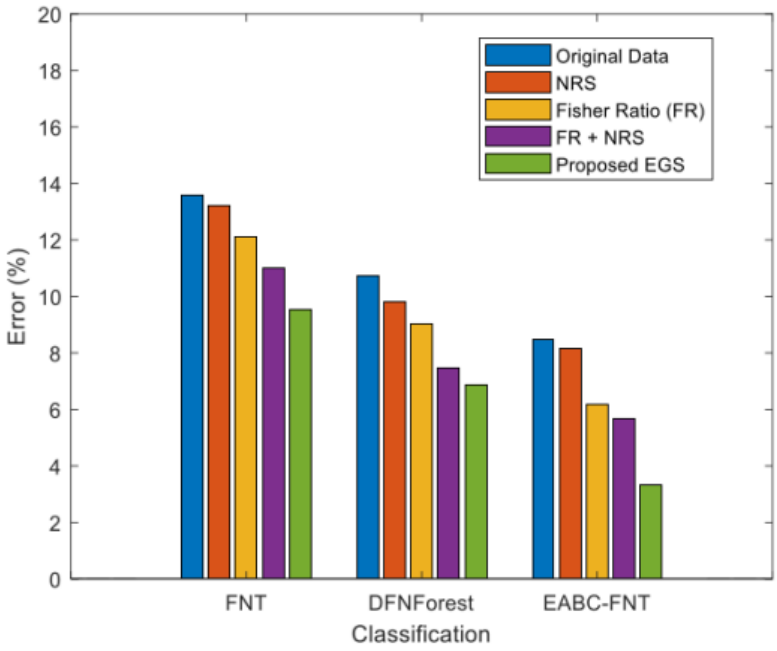


(c) F-Measure Results Comparison vs. Classifiers (LUNG genes)



(d) Accuracy Results Comparison vs. Classifiers (LUNG genes)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G



(e) Error Results Comparison vs. Classifiers (LUNG genes)

FIGURE 4. FINAL RESULTS COMPARIOSN VS. CLASSIFIERS (LUNG GENES)

Figure 4(a-e) shows the performance comparison results of the various classifiers such as FNT, DFNForest and EABC-FNT classifier with four feature selection methods such as NRS, FR, FR+NRS and proposed EGS algorithm under LUNG genes. The proposed EABC-FNT classifier produces higher accuracy results of 96.6493% for proposed EGS feature selection, the other methods such as FNT, and DFNForest gives lesser accuracy results of 91.8625% and 94.3429% respectively.

5. CONCLUSION AND FUTURE WORK

The proper classification of breast cancer subtypes has made significant progress in recent years. Despite the fact that data on gene expression is classified, has been extensively established in the direction of is a successful algorithm in the last decade, Stage differences, group effects, and the problem of classifying patient's individual samples have all delayed the comprehensive operation for a long time. For this, figuring out the subtypes of cancer is important for understanding and treating cancer. On the other hand, the RNA-seq gene expression data that has been utilised to classify cancer subtypes has a high complexity and

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

small sample size context. The Ensemble-Based Gene Selection (EGS) approach incorporates four gene subset outcomes in this paper, such as Fisher Ratio (FR), Neighborhood Rough Set (NRS), Correlation-Based Gene Selection (CFS) and Greedy Hill climbing. These outcomes are combined to pick an appropriate gene subset using Mutual Knowledge (MI) that enhances the identification of cancer subtype outcomes. Enhanced Artificial Bee Colony based Flexible Neural Forest (EABC-FNT) is proposed in this work for breast cancer subtype classification, where EABC algorithm is proposed for parameter optimization in order to improve the classification performance of cancer subtypes. The main goal of this EABC-FNT classifier is to turn a multi-classification problem into a lot of binary classification problems in each forest. In the meantime, the strength of the whole EABC-based cascade structure is being tested. The FNT classifier has been improved, but no new parameters have been added.The optimization of parameters via the EABC algorithm thus improves the results of FNT classifier for cancer subtype discovery. Experimentation on Ribonucleic Acid (RNA)-seq gene expression data of Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM), Lung Cancer (LUNG) show with the purpose of proposed EABC-FNT classifier gives higher accuracy with selected genes for cancer subtypes classification when compared to other methods such as Deep Flexible Neural Forest (DFNForest) and Precision, recall, f-measure, accuracy, and error rate are all parameters that the FNT classifier excels at. The findings show that the suggested EABC-FNT classifier is effective gives a choice towards categorize cancer subtypes with using FNT on three gene expression datasets. Future work determination includes other classifiers such as deep learning and ensemble classifiers for other gene expression datasets such as lung and tumor.

REFERENCES

1. Callahan, R.; Hurvitz, S. HER2-Positive Breast Cancer: Current Management of Early, Advanced, and Recurrent Disease. Curr. Opin. Obstet. Gynecol. 2011, 23, pp.37–43.

2. Assi, H.A.; Khoury, K.E.; Dbouk, H.; Khalil, L.E.; Mouhieddine, T.H.; El Saghir, N.S. Epidemiology and prognosis of breast cancer in young women. J. Thorac. Dis. 2013, 5, S2– S8.

3. Almendro, V. and Fuster, G., 2011. Heterogeneity of breast cancer: etiology and clinical relevance. Clinical and Translational Oncology, 13(11), pp.767-773.

4. Blows, F.M.; Driver, K.E.; Schmidt, M.K.; Broeks, A.; van Leeuwen, F.E.; Wesseling, J.; Cheang, M.C.; Gelmon, K.; Nielsen, T.O.; Blomqvist, C.; et al. Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. PLoS Med. 2010, 7, e1000279.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

5. Breugom, A.J., Bastiaannet, E., Boelens, P.G., Iversen, L.H., Martling, A., Johansson, R., Evans, T., Lawton, S., O'Brien, K.M., Van Eycken, E. and Janciauskiene, R., 2016. Adjuvant chemotherapy and relative survival of patients with stage II colon cancer–a EURECCA international comparison between the Netherlands, Denmark, Sweden, England, Ireland, Belgium, and Lithuania. European Journal of Cancer, 63, pp.110-117.

6. Dotan, E. and Cohen, S.J., 2011, Challenges in the management of stage II colon cancer. In Seminars in oncology (Vol. 38, No. 4, pp. 511-520). WB Saunders.

7. Wang S. L., X. Li, S. Zhang, J. Gui, and D. S. Huang, "Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction," Comput. Biol. Med., vol. 40, no. 2, pp. 179-189, 2010.

8. Dai J. and Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification," Appl. Soft Comput., vol. 13, no. 1, pp. 211-221, 2013.

9. Chan, W.H., Mohamad, M.S., Deris, S., Zaki, N., Kasim, S., Omatu, S., Corchado, J.M. and Al Ashwal, H., 2016. Identification of informative genes and pathways using an improved penalized support vector machine with a weighting scheme. Computers in biology and medicine, 77, pp.102-115.

10. Maji P. and S. Paul, "Rough set based maximum relevance-maximum significance criterion and Gene selection from microarray data," Int. J. Approx. Reasoning, vol. 52, no. 3, pp. 408-426, Mar. 2011.

11. Liu T. K., Y. P. Chen, Z. Y. Hou, C. C.Wang, and J. H. Chou, "Noninvasive evaluation of mental stress using by a refined rough set technique based on biomedical signals," Artif. Intell. Med., vol. 61, no. 2, pp. 97-103, Jun. 2014.

12. Hu Q., J. Liu, and D. Yu, "Mixed gene selection based on granulation and approximation," Knowl.Based Syst., vol. 21, no. 4, pp. 294-304, May 2008.

13. Hu Q., D. Yu, J. Liu, and C.Wu, "Neighborhood rough set based heterogeneous gene subset selection," Inf. Sci., vol. 178, no. 18, pp. 3577-3594,2008.

14. Dwivedi A. K., "Artificial neural network model for effective cancer classification using microarray gene expression data," Neural Comput.Appl., vol. 29, no. 12, pp. 1545-1554, Jun. 2016.

15. Chen Y., B. Yang, and A. Abraham, "Flexible neural trees ensemble for stock index modeling," Neurocomputing, vol. 70, nos. 4-6, pp. 697-703, Jan. 2007.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

16. Tao, M., Song, T., Du, W., Han, S., Zuo, C., Li, Y., Wang, Y. and Yang, Z., 2019. Classifying breast cancer subtypes using multiple kernel learning based on omics data. Genes, 10(3), pp.1-14.

17. Wesolowski, R. and Ramaswamy, B., 2011. Gene expression profiling: changing face of breast cancer classification and management. Gene Expression the Journal of Liver Research, 15(3), pp.105-115.

18. Gao, F., Wang, W., Tan, M., Zhu, L., Zhang, Y., Fessler, E., Vermeulen, L. and Wang, X., 2019. DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. Oncogenesis, 8(9), pp.1-12.

19. Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H. and Dawood, H., 2019. A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data. BMC bioinformatics, 20(1), pp.1-11.

20. Chen Y., Z. Zhang, J. Zheng, Y. Ma, and Y. Xue, "Gene selection for tumor classification using neighborhood rough sets and entropy measures," J. Biomed. Inform., vol. 67, pp. 59- 68, 2017.

21. Kong Y. and T. Yu, "A graph-embedded deep feedforward network for disease outcome classification and gene selection using gene expression data," Bioinformatics, vol. 34, no. 21, pp. 3727-3737, 2018.

22. Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H. and Khan, M.M., 2019. A novel deep flexible neural forest model for classification of cancer subtypes based on gene expression data. IEEE Access, 7, pp.22086-22095.

23. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M. and Cancer Genome Atlas Research Network, 2013. The cancer genome atlas pan-cancer analysis project. Nature genetics, 45(10) pp. 1113-1120, 2013.

24. Lu W., Z. Li, and J. Chu, "A novel computer-aided diagnosis system for breast MRI based on gene selection and ensemble learning," Comput. Biol. Med., vol. 83, pp. 157-165, 2017.

25. Chen Y. and Y. Zhao, "A novel ensemble of classifiers for microarray data classification," Appl. Soft Comput., vol. 8, no. 4, pp. 1664-1669, Sep. 2008.

26. Chen Y., L. Peng, and A. Abraham, "Gene expression profiling using flexible neural trees," in Proc. Int. Conf. Intell. Data Eng. Automated Learn., 2006, pp. 1121-1128.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 65 Issue 08 | 2022
DOI 10.17605/OSF.IO/9U82G

27. Wu P. and Y. Chen, "Grammar guided genetic programming for flexible neural trees optimization," in Advances in Knowledge Discovery and Data Mining (Lecture Notes in Computer Science), vol. 4426. Sep. 2007, pp. 964-971.

28. Akay B. and D. Karaboga, "A modified Artificial Bee Colony algorithm for real-parameter optimization," Information Sciences, vol. 192, pp. 120–142, 2012.

29. Karaboga, D. and Gorkemli, B., 2012, "A quick artificial bee colony-qABC-algorithm for optimization problems", In International symposium on innovations in intelligent systems and applications, pp.1-5.

30. Chuanlei, Z., Shanwen, Z., Jucheng, Y., Yancui, S. and Jia, C., 2017. Apple leaf disease identification using genetic algorithm and correlation based gene selection method. International Journal of Agricultural and Biological Engineering, 10(2), pp.74-83.

31. Shahbaz, M.B., Wang, X., Behnad, A. and Samarabandu, J., 2016, On efficiency enhancement of the correlation-based gene selection for intrusion detection systems. In 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) ,pp. 1-7.

32. Wu, H. and Liu, X., 2008, Dynamic bayesian networks modeling for inferring genetic regulatory networks by search strategy: Comparison between greedy hill climbing and mcmc methods. In Proc. of World Academy of Science, Engineering and Technology (Vol. 34, pp. 224-234).