ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

Hybrid Optimized Imperialist Competitive Algorithm with Ensemble Learning for Cancer Subtypes Diagnosis

Received: 24 October 2022, **Revised:** 25 November 2022, **Accepted:** 28 December 2022

P. Avila Clemenshia[1]*, Dr. C. Deepa[2]

[1]PhD Research Scholar
Department of Computer Science
Sri Ramakrishna College of Arts & Science, India. Avilajsph@gmail.com
[2]Associate Professor & Head
Department of CS (AI & DS)
Sri Ramakrishna College of Arts & Science, India. Deepapkd@gmail.com

Keywords

Cancer subtype, Whale based Optimization (WO), Ensemble Learning (EL) and gene expression data

Abstract

Cancer disease diagnosis and treatment now place a high priority on categorization of cancer subtypes. Pathological cancer subgroups have proven difficult to accurately identify molecularly. Numerous supervised learning techniques have been used to classify cancer subtypes in recent years in order to address these demands. For the categorization of cancer subtypes, the prior system used Deep Fuzzy Flexible Neural Forest (DFFNForest) with Binomial Probability Distribution-based Principal Component Analysis (BDPCA).BDPCA is used in this study to conduct dimensional reductions. The Imperialist Competitive Algorithm (ICA) algorithm will next be used to execute the feature selection in order to lower the classifier's miss rate. The DFFNForest is used to classify cancer subtypes in accordance with the chosen characteristic. However, the single classifier does not always produce better results. In order to enhance prediction performance, ensemble learning is necessary. A Hybrid Optimized Imperialist Competitive Algorithm (HOICA) with Ensemble Learning (EL) for Cancer Subtypes Diagnosis was developed by the suggested approach to address this issue. Dimensional reduction, feature choice, and classification processes make up the proposed cancer diagnostic system. Utilizing Improved Independent Component Analysis (IICA), dimensional reductions are carried out in the first step. To reduce the classifier's miss rate in the second stage, features are chosen using HOICA. Finally, the categorization is carried out using EL, which combines methodologies from the Support Vector Machine (SVM), Weighted Activation Function based Convolutional Neural Network (WAFCNN), and DFFN Forest. In terms of accuracy, precision, recall, f-measure, and error, the experimental findings demonstrate that the suggested system outperforms the current system.

1. Introduction

An ongoing development in cancer research has been carried out during the last several decades [1]. Large volumes of cancer data have been gathered and are accessible to the medical research community thanks to the development of new technologies in the area of medicine. Cancer is a heterogeneous disease, that is, each cancer may have a

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

number of subtypes [2-3]. In order to diagnose cancer and find new treatments, cancer subtype identification is crucial. Accurate cancer subtype prediction may help doctors effectively treat patients while reducing side effects. The solution to the basic issues surrounding cancer diagnosis and medication development is understood to lie in cancer subtype categorization utilizing gene expression data [4]. Comparing multiple strategies for categorizing the cancer subgroup with gene expression data reveals significant differences. Occurrence of cancer alters genetic expression of a group of cells or collectively known as a tissue and there are not many techniques for measuring genetic expression of cells.

In recent years, accurate and early cancer detection has been achieved using microarray gene expression data [5]. The categorization of cancer subtypes based on gene expression provides greater benefits than the conventional classification, according to recent study. When the appropriate genes are solely utilized for classification, the combination of gene selection and cancer classification may provide more accuracy. In the area of analyzing micro array data, gene selection is one of the trickiest problems. Due to the singularity problem, the curse of dimensionality, and other issues, traditional data mining approaches cannot be used to analyze gene expression data directly since the data often includes more genes (variables) than there are samples. To prevent these issues, it is necessary to undertake a dimension reduction procedure during the analysis of gene expression data, known as gene selection, which involves removing the genes that are most strongly connected with the patterns of each kind of illness.

For the classification of cancer, a variety of classifiers are utilized, including Rule-based, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbor (NN), Artificial Neural Network (ANN) [6], Rough set, Fuzzy Logic (FL), Genetic Algorithm (GA), and SVM [7].Because microarray experiments are time-consuming, costly, and constrained by sample availability, the number of samples used in microarray-based cancer investigations is often low.Traditional supervised learning methods, which can only use labelled data for learning, are often used for cancer classification utilizing gene expression data. The development of classification models using gene expression data is not without its difficulties. The effectiveness of the classifier may be diminished and the possibility of over-fitting increased due to the often limited number of cancer samples available to train the model in comparison to the number of characteristics present (such as genes). Selecting a subset of features (or genes)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

to develop a model helps prevent a classification model from being over-fit.

The remainder of this study's sections have been organized as follows: The many methods for classifying cancer subtypes in current research studies are condensed in Section 2 of this article. The suggested plan is explained in Section 3. The dataset used in Section 4 is mentioned, and the empirical results are shown. The value of this study is finally determined in Section 5.

2. Literature Review

Guo et al (2018) designed a new method called as Cancer Subtype Prediction using RV_2 (CSPRV)by merging data from several sources on transcriptome expression and diverse biological networks, the prediction of cancer subtypes will be improved. To anticipate the similarities between samples in each view of the expression data, the generalized matrix correlation approach (RV2) is used in this study to extract several expressed characteristics from each genomic element based on the regulation relationships in the heterogeneous biological networks. The information on the similarities should then be combined across several data views in accordance with varying integration weights.Cluster the samples into several subtype groups in accordance with the integrated similarities across samples.The developed strategy may detect more clinically significant cancer subgroups, as shown by extensive studies using the Cancer Genome Atlas (TCGA) cancer datasets [8].

Xu et al (2019) designed based on gene expression data, a new DFNForest for classifying cancer subtypes. Because it divides a multi-classification issue into several binary classification problems for each forest, the developed DFNForest model varies from the traditional Flexible Neural Tree (FNT) model. The flexible neural tree model is explored here to study the DFNForest cascade structure to deepen the model without adding more parameters. For dimensionality reduction of gene expression data in order to improve classification performance, this study included the fisher ratio and neighborhood rough set to the DFNForest model. Studies using RNA-seq gene expression data reveal that the created DFNForest model performs better for classifying different kinds of cancer and the designed gene selection approach has greater accuracy with fewer genes [9].

Halder et al introduced a Cancer categorization using Active Learning using Fuzzy K-Nearest Neighbor (ALFKNN). In order to label the perplexing data in a way that optimizes classification accuracy, the active learning approach is utilized to pick the microarray gene expression patterns that are either the most puzzling or

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

informative.Experts personally name the chosen genes that are the most perplexing or instructive.With the aid of several microarray gene expression cancer datasets, the developed approach is assessed.Experimental findings reveal that the active learning method (ALFKNN) delivers more accurate results for cancer prediction from microarray gene expression data than classic supervised k-Nearest Neighbor (k-NN) and fuzzy k-Nearest Neighbor (fuzzy k-NN) methods [10].

Guo et al (2018) introduced to solve cancer subtype classification on small-scale biology datasets, we developed the BCDForest deep learning model.The following two primary contributions, the BCDForest differs from the traditional deep forest model: In order to foster variation in the ensemble, a named multi-class-grained scanning approach is first developed.While this is happening, representation learning takes into account each classifier's accuracy.In order to spread the advantages of discriminative features throughout cascade layers and increase classification performance, a boosting approach was developed to highlight more significant features in cascade forests.The suggested technique consistently excels in the application of cancer subtype classification, as shown by systematic comparative studies on datasets for both microarray and RNA-Seq gene expression [11].

Kar et al (2015) designed a limited collection of relevant genes that are adequate for the intended classification purpose may be identified using the Particle Swarm Optimization (PSO) approach and an adaptive K-nearest neighborhood (KNN)-based gene selection strategy.To create the right number of neighborhoods to explore and, ultimately, appropriately categorize the database, a suitable value of K would be helpful.So, a heuristic that is motivated by classification accuracy and meant to pick the best values of K efficiently. The small round blue cell tumor (SRBCT), acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), and mixed-lineage leukemia (MLL) datasets are used as benchmark microarray datasets for this method, which is meant to identify the smallest relevant collection of genes that can be found.In terms of classification accuracy on blind test samples, the number of informative genes, and computational time, the results show the method's utility [12].

Xia et al (2017) introduced a Cancer classification using gene expression data using the Self-Training Subspace Clustering method under Low-Rank Representation (SSC-LRR). From the high-dimensional gene expression data, discriminative features are first extracted using the Low-rank representation (LRR), and the cancer classification predictions are then produced using the self-training

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

subspace clustering (SSC) approach.Four state-of-the-art classification techniques were used to assess the SSC-LRR on two distinct benchmark datasets under control.A general correlation of 0.920 and an accuracy rate of 89.7%, which are 18.9 and 24.4 percent higher than the best control approach, correspondingly, were produced as forecasts for cancer categorization.Additionally, the SSC-LRR program discovered numerous genes (RNF114, HLA-DRB5, USP9Y, and PTPN20) that are novel cancer markers and need additional clinical research.In general, the research showed that there is a novel and sensitive way to distinguish different types of cancer using large-scale gene expression data [13].

3. Proposed Methodology

The proposed system designed a Hybrid Optimized Imperialist Competitive Algorithm (HOICA) with Ensemble Learning (EL) for cancer subtypes diagnosis. Dimension reduction, feature selection, and classification are the phases that make up the suggested cancer diagnostic method. Figure 1 displays the flowchart for the suggested task.

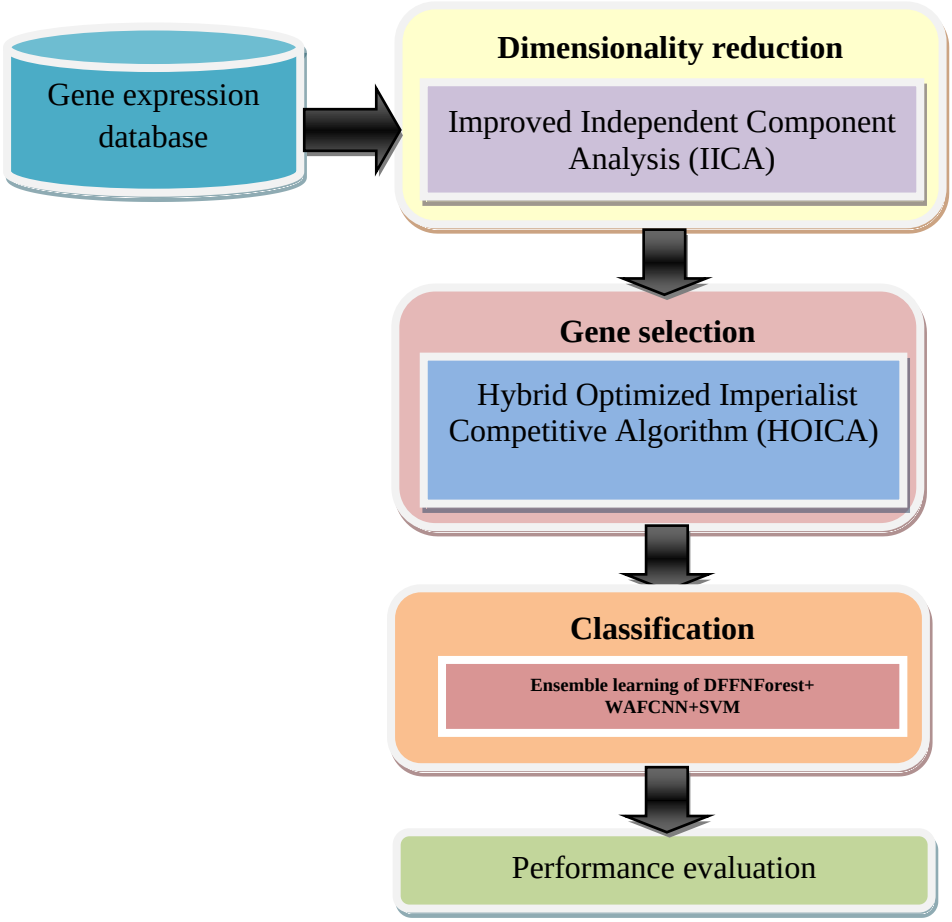

Figure 1: Flow diagram of the proposed research work

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

3.1 Input Gene expression data

The number of samples available is often low, despite the fact that gene expression data typically includes thousands of genes.Only a few number of genes are truly linked to certain cancer subgroups among the hundreds of characteristics in gene expression data; the remainder may be duplicated or noisy features.In order to pick significant genes while keeping the classification accuracy of the original genes, gene selection may be seen as a dimensional reduction challenge.

3.2 Dimensionality reduction

Improvements to IICA are used in the first step to reduce dimensions. Information on higher order statistics is the only thing lacking since PCA considers the second order moments.The independent source components are found from their linear mixes using ICA, which also takes into consideration higher order statistics.Because its objective is to give independent rather than uncorrelated degradation and visualization, ICA offers a more potent data representation than PCA.The main issue is that the sample covariance matrix's eigen values and eigenvectors are poor estimates of the genuine ones, particularly in high dimensions with few data.A Nystrom Method for computing covariance matrices was developed by the suggested system to address this issue.

In order to reduce the statistical dependency between its components, the ICA of a random vector looks for a linear transformation [14–15]. with especial, let X$\in R^N$be a random vector that represents the database, with N being the number of dimensions that the space has.The gene expression database's columns or rows are combined to create the vector, which may then be stabilized to have an equalized histogram or a unit norm.The definition of the covariance matrix is clear.

$$\sum_X = E\{[X - E(X)][X - E(X)]^t\} \quad (1)$$

$E(.)$is the expectation operator, t denotes the transpose operation, $\sum_X \in R^{N\times N}$.

The Nystrom covariance estimator's structural formula for Σ is defined by

$$\widehat{\sum_X} = \frac{1}{n} \text{X P(X) } X^T (2)$$

$\hat{\Sigma}_{(X)}$is an index set function X $\subseteq$ {1, . . . , P} resulting in conditional error performance based on I. an orthogonal projection onto the subspace of where P(X)R^n.

The covariance matrix is factorized using the ICA $\hat{\Sigma}_{(X)}$ into the following form

$$\hat{\Sigma}_{(X)} = F\Delta F^t (3)$$

Where, ΔF converts the initial data X into Z if it diagonal real positive.

$$X=FZ \quad (4)$$

such that, or "the most independent feasible," the elements of

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

the new data Z are independent.A three-step technique, including whitening, rotation, and normalization, is used to create the ICA transformation F.The whitening procedure begins by changing a random vector, denoted by X, into a different vector, denoted by U, which has a unit covariance matrix.

$$X = \varphi \Lambda^{\frac{1}{2}} U \quad (5)$$

Where, φ and $\Lambda$are obtained by resolving the eigen value equation below.

$$\hat{\Sigma}_{(X)} = \varphi \Lambda \varphi^{t} \quad (6)$$

Where,$\varphi = [\varphi_1, \varphi_2, \dots, \varphi_N]$*is an orthonormal eigen vector matrix and* $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_N\}$is a matrix with diagonal eigenvalues of$\hat{\Sigma}_{(X)}$. One observes that the Mean Square Error (MSE), which preferentially favors low frequencies, is offset by whitening, an essential ICA component.By reducing the mutual information, which is approximated using higher order prior phases, the rotation then conducts source separation.The point cloud matrix P is aligned along its primary axis by utilizing the matrix X of eigenvectors produced by the ICA of the covariance matrix, P' = X(P -μ).

3.3 Feature selection using Hybrid Optimized Imperialist Competitive Algorithm (HOICA)

For gene selection in the proposed research project, HOICA is used. ICA and the Enhanced Factor Based Whale Optimization Algorithm (EFWOA) approach are both used in the hybrid model.

3.3.1 Imperialist Competitive Algorithm (ICA)

A novel evolutionary optimization method called The ICA is motivated by imperialist competing. It is important to note that ICA is a reliable strategy founded on imperialism, which is the practice of a government expanding its authority and jurisdiction outside of its own boundaries. In this algorithm, initial populations are considered as number of genes. It is decided that the imperialists' genes are among the finest in the community.The remaining population is split amongst the aforementioned imperialist countries as colonies.Following it, all of the empires start competing imperialistically. The competition will end for the weakest empire, which is unable to get stronger and is unable to prevail.Due to the conflict between empires, all colonies (genes) travel in the direction of their appropriate imperialists.Finally, it is hoped that the collapse mechanism would bring about a condition in which there is only one empire in the globe and all other nations are its colonists.The powerful empire might provide the answer for us.

1) Generating initial empires

The objective of optimization is to identify the best genes. In this

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

proposed research work, number of genes as population (country). The number of genes is represented by a 1Narray in an N-dimensional issue as follows:

Number of genes =(x_1,x_2,...,x_n),
$x_i \in R, 1 \le i \le N$ (7)

The precision of the genes' categorization is what determines a gene's cost value. To classify data, a variety of classifiers may be utilized. In this proposed research work, for categorization, we use the DFFN Forest. The following formula is used to calculate the cost function:

$$\text{Cost} = \frac{Numberofcorrectlyclassifiedgenes}{Totalnumberofgenes} \quad (8)$$

The designed system should generate N_{pop} of them. Each gene's price is calculated using the variables' f(x) (x_1,x_2,...,x_n). Then

cost=f(genes)=f(x1,x2,...,xn)
(9)

To build our empires, we choose N_{imp}, one of the strongest genes. It will be the colonies that make up the remainder N_{col} of the population.Here is how an imperialist's normalized cost is determined:

$$c_n = c_n - {}^{max}_{i}\{c_i\}$$
(10)

where c_n n^{th} value is imperialist and$c_n$is the cost normalized.Each imperialist's normalized power is determined by

$$P_n = \left\{ \frac{c_n}{\sum_{i=1}^{N_{imp}} c_i} \right\} \quad (11)$$

So, an empire will begin with the following number of colonies:

No.C_n=round(p_n.N_{col}) (12)

where No.C_n, where N_{col} is the total number of colonies, and is the nth empire's beginning colony count. For the imperialists' division of the colonies, choose a No.$C_n$to the n^{th} empire in terms of colonies.

2)*Imperialist movement of a country's colonies*

The vector from colony to imperialist is composed of each colony that progresses x-units in the direction of the imperialist. With a uniform distribution, the random variable x will be. Then

$x \sim U(0, \beta \times D),\ \beta > 1$ (13)

Where did the imperialist and genetic distance come from. β represents the colony and imperialist closer.

3) Revolution

An empire's genes are replaced with an equal number of newly created nations in each iteration.This has been accomplished by the algorithm by randomly producing some new nations and then replacing them with colonies of the respective empire.A similar amount of newly created genes are expected to replace each colony in the empire's population, which is

N.R.C=round{RevolutionRate ×No.(The colonies of $empire_n$)} (14)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Where, the number of revolutionary colonies is abbreviated as N.R.C. By doing this, the ICA's global convergence will be enhanced and it won't continue to operate at a local minimum.

4)A colony and an imperialist exchanging positions

During the process of relocating, a colony could obtain a better position than the one held by the imperialist. Therefore, the stance of the imperialist shifts to that of the colony, and vice versa.

5) Total power of an empire

The following are the ways in which an empire's whole power is dependent on all of its colonies:

$$T\cdot Cn=cost(imperialist_n)+\xi\cdot mean(cost(colonies\ of\ empire_n)) \quad (15)$$

Where ξ is a position coefficient.

6) Imperialistic competition

To seize and maintain control over the colonies of rival empires, all empires compete with one another.As a consequence, the stronger empires gain ground while the weaker ones eventually lose strength.Find the likelihood that each empire will be possessed depending on its overall power to achieve this aim. The averaged total cost is

$$N\cdot T\cdot C_n=T\cdot C_n-\max\{T\cdot C_i\} \quad (16)$$

Where, $T\cdot C_n$and $N\cdot T\cdot C_n$are the nth empire's normalized total cost and total cost overall, respectively. The possession likelihood of any empire might now be determined by

$$P_{p_n}=\left\{\frac{N\cdot T\cdot C_n}{\sum_{i=1}^{N_{imp}} N\cdot T\cdot C_i}\right\} \quad (17)$$

Based on the likelihood that each empire would eventually hold the mentioned colonies, divide them among them. the formation of the vector P is

$$P=[p_{p1},p_{p2},p_{p3},\ldots,p_{p_{Nimp}}] \quad (18)$$

and also the vector Rwith uniformly distributed elements

$$R=[r_1,r_2,r_3,\ldots,r_{N_{imp}}]r_1,r_2,r_3,\ldots,r_{N_{imp}}p\sim U(0,1) \quad (19)$$

Finally, we have vector D by

$$D=P-R=[p_{p1}-r_1,p_{p2}-r_2,p_{p3}-r_3,\ldots,p_{p_{Nimp}}-r_{N_{imp}}] \quad (20)$$

An empire with the highest applicable index in D will get the aforementioned colonies from the components of D.

7) The eliminated empire

An empire will fall apart and become one of the remaining colonies when it loses every colony.

8) Convergence

When everything is said and done, there will be just one mighty empire, with no rivals, and it will rule over all of the colonies.Therefore, all of the colonies will be charged the same prices as the special empire. Therefore, there is no distinction between their distinctive empire and colonies (genes).

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

3.3.2 Gene Selection using EFWOA

The suggested research endeavor employs the EFWOA to increase the classifier's accuracy.Optimizations for swarm intelligence imitate the behavior of whales as predators. Once it has found its food, the whale moves upstream to create a bubble net along the spiral route.The essential premise of whale bubble-net foraging is this. Encircling the prey, launching a bubble net assault, and pursuing the victim are the three stages of this predation behavior.

1) Surrounding Prey Stage

In the whale population, multiply Xi by the amount of genes it contains (i=1, 2,..., n). An objective function is thought to be the categorization's accuracy.However, in actuality, whales are unable to anticipate exact locations of praise. In WOAs, whales find preys before surrounding them.The other members of the group will all go to the location where the intended prey is, if that location is the gene's current optimum position.It is possible to express the enclosing stage mathematically using the model shown below:

$$D = |C.X^*(t)X(t)| \quad (21)$$

$$X(t+1)=X^*(t)\text{-}A.D \quad (22)$$

Where t represents the number of iterations most recently performed, X*(t) represents the prey's (solution) current ideal location, X(t) symbolizes the prey's positional vector, A. D denotes surround step sizes, and

$$A=2\vec{a}.rand\text{-}\vec{a} \quad (23)$$

$$C=2.rand \quad (24)$$

Within the formula above, $\vec{a}$ is a control parameter that drops by a constant amount from 2 to 0 iterations. Between [0, 1] is the range of the random number rand. To make sure that the algorithm converges more fast and avoids the local optimum, modify the convergence factor using the formula below.

$$\vec{a}= 2*(e^{-\left(1+\frac{t}{Maximumiteration}\right)} - 1) \quad (25)$$

2) Bubble-Net Attack Stage

Within the confines of its bubble, the humpback whale follows a spiral route to get close to its prey. foraging technique with nets. In order to describe the predation behavior of whales, two methodsthe spiral update position and the shrinking and encircling mechanismhave been created by WOA.

In order to accomplish shrinking and surrounding processes, lower the convergence factor.

Spiral update position: Determine the distance between each gene and its present optimal location before attempting to replicate the whale's spiral feeding technique.The mathematical model may be stated in the following ways:

$$X(t+1)=D'.e^{bl}.\cos(2\pi l) + X^*(t) \quad (26)$$

$$D' = |X^*(t) - X(t)| \quad (27)$$

Where D'b and an are the coefficient constants that produce logarithmic spiral forms, and l is a

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

random value between [1, 1], which are all variables in the logarithmic spiral form. This suggests distances between i^{th} genes and current ideal places. While circling their prey during predation stages, whales also need to limit the area surrounding them. Therefore, for the purpose of creating synchronous models, spiral envelops and contractions are carried out with the same probability.

3) Hunting Prey Stage

If |A|≥ 1 , The algorithm's global explorations are improved by preventing whales from reaching the existing reference targets. Additionally, they must come up with better replacements for the existing reference genes, which are chosen at random for the best possible outcomes at this time and are mathematically represented as:

$$X(t+1)=X_{rand}\text{-}A.D \quad (28)$$

$$D=|C.X_{rand}\text{-}X(t)| \quad (29)$$

where X_{rand}implies choosing at random the gene's location vector.

Exploration and exploitation are two essential parts of an optimization algorithm, and the proper trade-off between the two helps the system break out of local optimum conditions so that it can find correct answers.With each iteration of WOA, the step size of a search agent steadily decreases. The parameter A may be used to control this step size. Additionally, it demonstrated how weak divergence affects WOA's capacity to constrain traps within local optimum at later iterations.

The enhanced factor for whale optimization technique used in this study, which adds the levy flight function to modify the value A, addresses those problems. It enhances concurrent WOA capabilities inquiry and use.

A function called the Levy probability distribution function, which is based on a power-law, is used to calculate the leap size in a Levy flight. Following is the mathematical formula for the Levy distribution:

$$L(s,\gamma,\mu)=\begin{cases}\sqrt{\frac{\gamma}{2\pi}}\exp\left[-\frac{\gamma}{2(s-\mu)}\right]\frac{1}{(s-\mu)^{3/2}} & if\ 0<\mu<\infty \\ 0 & if s\le 0\end{cases}$$

(30)

Where μ, γ, and $s$are the location parameter, scale parameter, and sample collection parameter, which govern the distribution's size, respectively.

Algorithm1: EFWOA

Setinitialcountsof genesin$X_i(i=1,2,\ldots,n)GES_d$atabase.

Set Step 2's starting values for a, A, C, l, and p.A search agent's fitness should be determined such that X*=the search agent

while (it<Maxiter)foreachsearchagent

if(p<0.5)

if(|A|<1)

current search agents' positions using (21)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Elseif(|A|≥1)
Chooserandomsearch agents
current search agents' positions using (28)end
Elseif(p≥0.5)
Update search results positions based on (26)End
End
Any search agents that go beyond the search region should be checked and adjusted.
Determine the fitness of each search agent (classification accuracy)
UpdateX_(best)Whenever a better option developst=t+1
end-while
Step 3: return X_(best)

In HOICA,the length of the solution is shortened by each empire's growth in strength and the advancement of its colonists and imperialists.Consequently, the HOICA's convergence rate is larger than the ICA's and the EFWOA's.

Algorithm 2: Hybrid Optimized Imperialist Competitive Algorithm (HOICA)

Input: Number of genes in Gene expression data

Output: Optimal genes

Step 1: Determine the goal function of classification accuracy: f(x), x=$(x_1, x_2, .., x_d)$

Step 2: Set the initial number of genes

Step 3: Create early empires by generating a random solution in the search space.

Step 4: Assimilation: Different paths in which genes migrate in the direction of imperialist nations.

Step 5: Revolution: Random changes occur in the characteristics of some countries.

Step 6: An Imperialist gene and a gene switch places. By dislodging the imperialist genes now in place, a gene with a superior position has the potential to seize leadership of the empire.

Step 7: Imperialistic competition: For control of one another's colonies, all imperialists compete.

Step 8: If there is an empire with no colonies.

Step 9: Take out the weak empires. Weak empires eventually lose their influence before being destroyed.

Step 10: Else

Step 11: Perform EFWOA

Step 12: Determine each search agent's fitness

Step 13: Update X_(best)if a better option becomes available. t=t+1

Step 14: Else

Step 15: Stop the condition

Step 16: End

3.4 Classification using Ensemble Learning (EL)

The utilization of numerous individual students and a specific combination approach in ensemble

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

learning is a novel method that produces outcomes superior to those of any one student working alone.The categorization in this study is carried by employing ensemble learning which consists of Deep Fuzzy Flexible Neural Forest (DFFNForest), Weighted Activation Function based Convolutional Neural Network (WAFCNN) and SVM approaches.

3.4.1 DFFNForest based classification

Then, in order to improve the outputs of the DFNForest classifier, the DFFNForest has been conducted with the selected feature, in which the fuzzy function has been supplied.While predicting a cancer subgroup, the DFFNForest technique, on the other hand, uses the fuzzy to update the classifier's weight values.

Flexible Neural Tree

The function set F and terminal instruction set T have been used to create the FNT model, as stated by

$$S=F \cup T =\{+_2, +_3, \ldots\ldots,,, +_N\} \cup \{x_1, \ldots\ldots, x_n\} \quad (31)$$

Wherein non-leaf node instructions with i parameters specified by$+_i$(i= 2, 3, 4,…,N), as the leaf nodes without parameters, which are represented by x_1, x_2 …, x_n. The non-terminal directive should be$+_i$(i=2, 3, 4,….,N) in order to generate a flexible neural tree, the weights of the children must be concatenated and any random generation of $i$values must be handled for non-leaf nodes. The flexible neural tree may be considered for the following flexible activation function,

$$f(x)=(1 + e^{-x})^{-1} \quad (32)$$

The outcome of a flexible neuron is produced by the following statement $+_n$

$$\sum_n = \sum_{j=1}^{n} w_j * x_j \quad (33)$$

which inputs are specified byx_j(j = 1, 2, …,n). The output of node $+_n$ expressed by,

$$out_n = f(sum_n) = ((1 + e^{-sumn})^{-1} \quad (34)$$

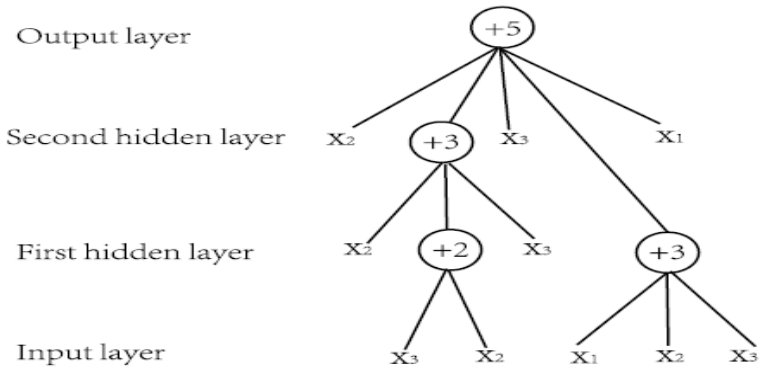

Figure 2:Typical FNT with function instruction set representation F={$+_2, +_3, +_4, +_5$}, and terminal instruction set T ={x_1, x_2, x_3}.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Figure 2depicts the FNT in accordance with a common representation. The flexibility neural tree's total output may also be iterated from left to right using the depth-first approach.Through consenting to the over-layer correlations and recognizing the structure spontaneously, the FNT is recognized to be a sparse model that ensures excellent generalization performance.The optimization of the tree structure is followed by the optimization of the parameter, and these two stages may be separated into two main stages in the FNT's optimization process.

Deep Fuzzy Flexible Neural Forest Model

It is well recognized that the flexible neural tree is the only kind of neural network that can automatically optimize both structure and parameters. But this approach is not without its difficulties. First of all, since it only has a single root node as an output node, it may not be the greatest solution to deal with the issues with multi-classification. Additionally, for the model to operate at its best, more work must be done. However, using this approach leads to an increase in the number of parameters that the parameter optimization technique needs, which ultimately raises the performance cost. The novel, flexible neural tree-related technique, known as DFFNForest, has successfully overcome the aforementioned difficulties.

To allow deep neural networks to achieve the remarkable success on the tasks of voice and image recognition, representation learning and model complexity are crucial. The layer-by-layer method of features is how representation learning is defined. Concerning the generation of in-depth FNT without the inclusion of extra parameters, the established framework has adopted the cascade forest structure. In order to process it as an input to the subsequent layer, the layer-by-layer process of features could supply the novel features that have combined with original features.

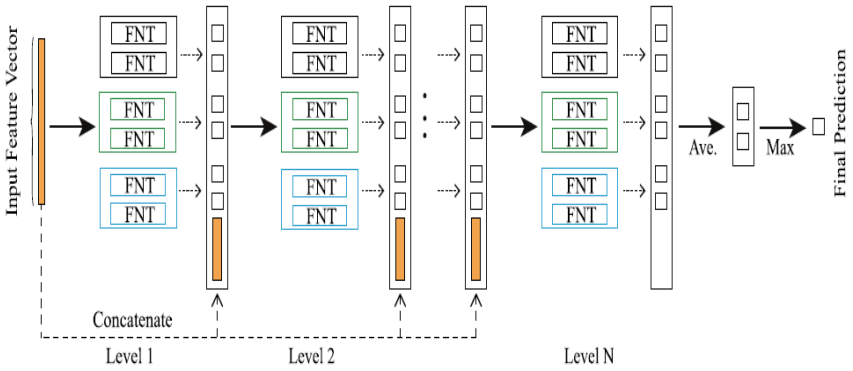

Figure 3: Cascade forest structure illustration

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

FNT has been used to construct each level of the proposed framework. Although decision trees are used in multi-grain cascade forests, they do have one disadvantage: they cannot be applied directly to continuous data; instead, they must be split first, which might result in information being lost. Given that the gene expression data in this case is known to be continuous data, FNT has been thought to be the best option as a base classifier. The suggested strategy has maintained the following characteristics of FNT, i.e. i) In order to avoid overfitting and get the best generalization performance, the FNT, which is a sparse model, allows for cross-layer association; ii) Additionally, the various FNTs allow the suggested ensemble learning to improve the overall performance since the structures and parameters have been automatically adjusted by FNT. In order to improve diversified ensemble learning, the system has produced a variety of FNT structures employing a variety of grammars. Examine that each of the three forests has two FNTs for more understanding. Then, consider the function set F of $\{+_2, +_3, +_4\}$ first forest will utilize it, and then $\{+_2, +_4, +_5\}$ lastly, second forest will make use of $\{+_3, +_4, +_5\}$ the final forest will use (see. Figure 3). Then, by reducing the multi-classification problem into several two-class problems across a forest, the M-ary technique assists in fixing the FNT's multi-classification issues. Another simple way of putting it is to say that, in the case of the four-class issue, each individual forest was required to incorporate k = log24 = 2 FNTs. Determining the number of trees in the forest was therefore made possible by the categorization issue.

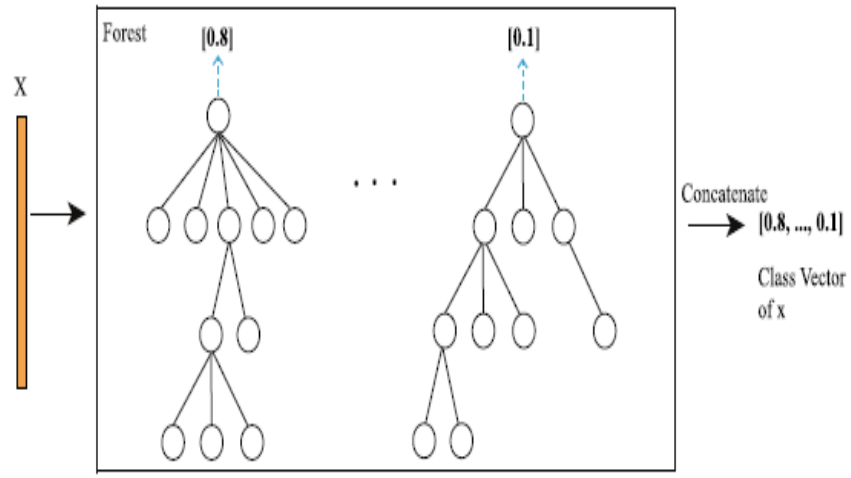

Figure 4: Class vector generating illustration. Each FNT will provide a guessed value, which will then be added together.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

To organize a class vector that was concatenated to the first input feature vector as the source to be processed in a subsequent phase, Figure 4 depicts an example where the approximation value has been created by each FNT. Consider the following scenario: If there are four classes, each forest has formed a two-dimensional class vector as a result, and the subsequent level of cascade has obtained six (2 x 3) augmented features. Additionally, in order to allocate them to the training and validation procedures, the training set would be divided into two portions. At the moment of including a new level, the validation set validates the total cascade. When there is no improvement in accuracy, the increment of further levels will stop. Thus, cascade levels may be used on varied datasets and with small-scale gene expression data since they are automatically determined.

Weight value updating using fuzzy function.

When predicting the cancer subtype, the DFFNForest approach makes use of the fuzzy to update the weight values of the classifier. As a reduced version of a fuzzy rule-based classifier, similar to the one used in fuzzy control, a fuzzy if-then technique has been suggested. Considering that there are three classes in the instance, the following weight values for the gene (feature) are established by the designated classification rules:

IF x_i is medium AND x_j is small THEN class is 1

IF x_i is medium AND x_j is large THEN class is 2

IF x_i is large ANDx_j is small THEN class is 2

IF x_i is small ANDx_j is large THEN class is 3

The two attributes are numerical, and$x_j$in addition to the language values utilized by the rules, includes weight values. The number of alternative if-then rules of this conjunction type that may be used in the issue is assuming each feature has M potential linguistic values and n characteristicsM^n. Every linguistic value is shown through a membership function.

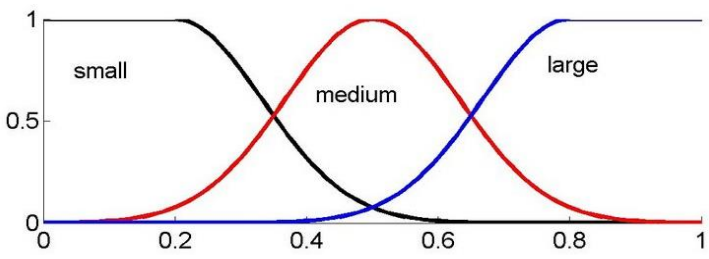

Figure 5: Membership functions for weight values for $gene\ x_i$

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Deep neural networks may be replaced with an innovative deep learning method called DFFNForest. The FNT structure is automatically selected by the tree structure optimization method in each unique forest, and the levels of cascade are determined deceptively. Although the FNT is included into the DFFNForest method, this defect in FNT is addressed and corrected so that it may be used to deal with multi-classification difficulties, which have evolved into multiple binary classification concerns throughout every forest. Additionally, without the use of any additional parameters, the cascade structure contributes to increase the model depth.

3.4.2Weighted Activation Function based Convolutional Neural Network (WAFCNN)

Cancer subtype categorization is carried out in this suggested research project utilizing WAFCNN. Convolution layers, sub sampling or pooling layers, and fully linked layers are the three primary layers that make up these networks [16-17]. The network comprises three layers: an input layer that accepts chosen characteristics (genes) as input; an output layer from which the system receives trained output; and intermediate layers known as hidden layers, as illustrated in figure 6.

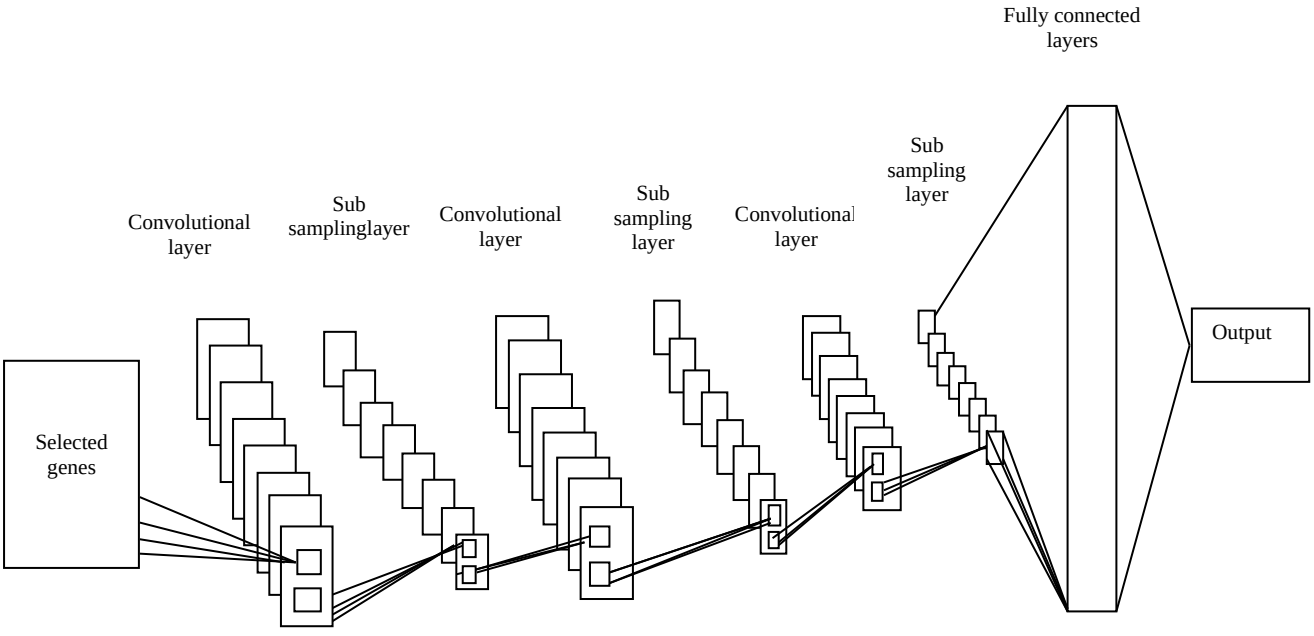

Figure 6: Convolutional Neural Network (CNN)

Convolution layer

Selected genes are used as input in the research that is being suggested. The CNN network's initial layer is the convolution layer.The input features are convolved using a kernel in this stage of

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

convolution (filter). The kernel convolution individually with each feature of the input matrix to produce the output. It produces n outputs using the result of convolution of the input and kernel. In general, the output produced by convolving the kernel and the input is referred to as feature maps of size i*i, while the convolution matrix's kernel is known as a filter.

Convolutional layers may be included in the CNN in multiples, with feature vectors serving as their inputs and outputs in subsequent levels. Each convolution layer has a number of n filters. The depth of the resulting maps (n*) is equal to the number of filters used in the convolution process after these filters are convolved with the input. It is important to keep in mind that each filter map is regarded as a unique feature at a particular point of the input.

The l-th convolution layer's output is indicated by the symbol$C_j^{(l)}$, is made up of maps. The calculation is

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{i,j}^{(l-1)} * C_j^{(l)} \quad (35)$$

Where, $B_i^{(l)}$ bias matrix is present, and$K_{i,j}^{(l-1)}$j-th feature map in layer (l 1) and i-th feature map in the same layer are connected by a convolution filter. The output $C_i^{(l)}$ feature maps make up the layer. In (37), the first layer of convolution $C_i^{(l-1)}$ a space for input, that is, $C_i^{(0)} = X_i$.

Map of features is produced by the kernel. For nonlinear modification of the convolution layer's outputs after the convolution layer, the activation function may be used:

$$Y_i^{(l)} = Y(C_i^{(l)}) \quad (36)$$

Where,

$Y_i^{(l)}$- output of the activation function

$C_i^{(l)}$- input that it receives.

Sigmoid, tanh, and rectified linear units (ReLUs) are the most often employed activation functions. Weighted ReLUs, which are referred to in this study as$Y_i^{(l)} = w.\max(0, Y_i^{(l)})$ are used. Due to its assistance in decreasing the interaction and nonlinear effects, this function is often utilized in deep learning models. The performance is improved by adding feature weight value in this case. If the ReLU gets a negative input, it will convert the output to 0, but if it receives a positive input, it will return the original input value. Because to the error derivative's reduction in size in the saturating zone, which causes the updates of the weights to nearly cease, this activation function has an advantage over other functions in terms of speedier training. The vanishing gradient issue refers to this.

Sub sampling or pooling Layer

Following the convolutional layer is the sub sampling layer. The primary objective of this layer is to lower the spatial dimensionality of the feature

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

maps that were retrieved from the layer before this one that performed the convolution. It averages the characteristics by dividing them into blocks of size 2x2.The relative information between features, rather than their precise relationship, is preserved by the sub sampling layer. It should be noted that the presence of a sub sampling layer enables the convolution layer to better withstand rotation and translation among the input.

Fully Connected layer

The softmax activation function is used by the output layer:

$$Y_i^{(l)} = f(z_i^{(l)}), \qquad (37)$$

$$\text{where, } z_i^{(l)} = \sum_{i=1}^{m_i^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)}$$

Where, $w_{i,j}^{(l)}$ are the weights that must be adjusted by the whole fully connected layer to provide the representation of each class, and f is the transfer function that symbolizes the nonlinearity.

3.4.3 Support Vector Machine (SVM)

A group of machine learning algorithms known as SVMs may be used to identify patterns in supplied information. The chosen genes are used in this study as a first input to classify the various types of cancer. In cancer subtype classification, the training features belonging to two separate classes, $(x_1, y_1),\ldots,(x_n, y_n)$, where, $x_i \in R^n$ which represents the i-th training features and $y_i \in \{-1, +1\}$ which represents the corresponding class label. In order to perform separation, one wants to find a hyperplane $\omega.x$ + b =0. In this weight vector ω, b, and x are training features, and is a normal hyperplane vector. The formula for categorization is:

$$(w.x) + b \geq if\ y_i = +1 \qquad (38)$$

$$(w.x) + b \leq if\ y_i = -1 \qquad (39)$$

To identify the optimal hyperplane, $||w||^2$ must be reduced to satisfy the limitation $y_i(w.x_i + b) \geq 1, i = 1,2,..,n$. For this reason, it is necessary to find a solution to the optimization issue presented by

$$\min \frac{1}{2} ||w||^2 \qquad (40)$$

$$\text{Subject to } y_i(x_i.w + b) \geq 1$$

The slack variables that are positive now ξ_i are included as a replacement in the optimization problem, allowing the approach to be expanded to include nonlinear decision surfaces. The current optimization issue is presented as

$$\min_{w,\xi} \frac{1}{2} ||w||^2 + C\sum_{i=1}^{N} \xi_i \qquad (41)$$

$$\text{s.t } y_i(\omega.x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad i=1,2,..,n$$

Where C is a penalty parameter that handles the balance between making the most profit possible and making the fewest mistakes possible.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

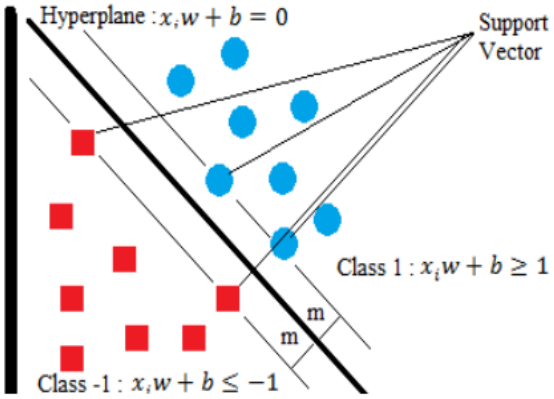

Figure 7: Illustration of classification in SVM

The ensemble learning such as Deep Fuzzy Flexible Neural Forest (DFFNForest), WAFCNN and SVM methods are mentioned in figure 8.

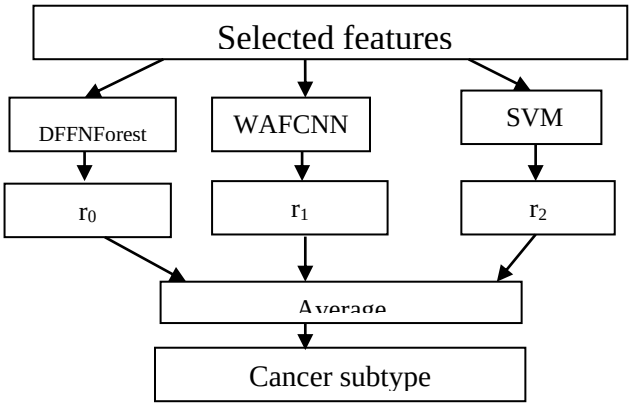

Figure 8. Ensemble Convolutional Neural Networks

When classifying a cancer subtype, all CNN's output probabilities for a particular input are summed. The average output S_i for output i is calculated as follows:

$$S_i = \frac{1}{n}\sum_{j=1}^{n} r_j(i) \qquad (4\ 2)$$

where $r_j(i)$ for a certain input pattern, is network j output i(selectedgenes).

The strategy involves giving each network a distinct weight. When integrating the findings, networks in the validation set with smaller classification errors will be given more weight. All networks, including DFFNForest, WAFCNN, and SVM, multiply their output probabilities by a weight given an input pattern α before the prediction:

$$S_i = \sum_{j=1}^{n} \alpha_j r_j(i)$$

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

The weight is determined in the proposed study project using the mean. The following formula is used to determine weight.

$$\alpha_k = A_k \quad (44)$$

where i traverse the n, A_k is the network k classification accuracy in the validation set, and. The network's maximum categorization is based on its average output; which accuracy is selected to predict the cancersubtype.

4. Experimental Results

In MATLAB, the experimental analysis is completed. Prostate cancer dataset, which can be accessed from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15484, was used to perform the cancer subtype predictions in this study. There are 65 samples in all, and there are two grade levels (6 and 8).The performance of the proposed Hybrid - Imperialist Competitive Algorithm with Ensemble Learning (HICA-EL) is compared with previous methods such as Deep Flexible Neural Forest (DFNForest),Imperialist Competitive Algorithm with Deep Fuzzy Flexible Neural Forest (ICA-DFFNForest) and Modified - Imperialist Competitive Algorithm with Deep Fuzzy Flexible Neural Forest (MICA-DFFNForest) approach in terms of accuracy, precision, recall, F-measure and error. The performance comparison is shown in table 1.

Table 1: Performance comparison

Performance metrics in (%)	Methods			
	DFNForest	ICA-DFFNForest	MICA-DFFNForest	HOICA-EL
Accuracy	85	93.25	96.62	98.99
Precision	86.3035	94.48	98.74	98.83
Recall	86.10	90.05	97.92	98.84
F-measure	86.20	92.21	98.33	98.81
Error	15	6.75	3.37	1.05

Performance metrics

4.1 Accuracy

Accuracy is the performance metric that makes the most sense to most people, because all it really is a ratio of properly predicted observations to the total number of observations.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (45)$$

Where,

TP - True Positive

FN - False Negative

FP - False Positive

TN- True Negative

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

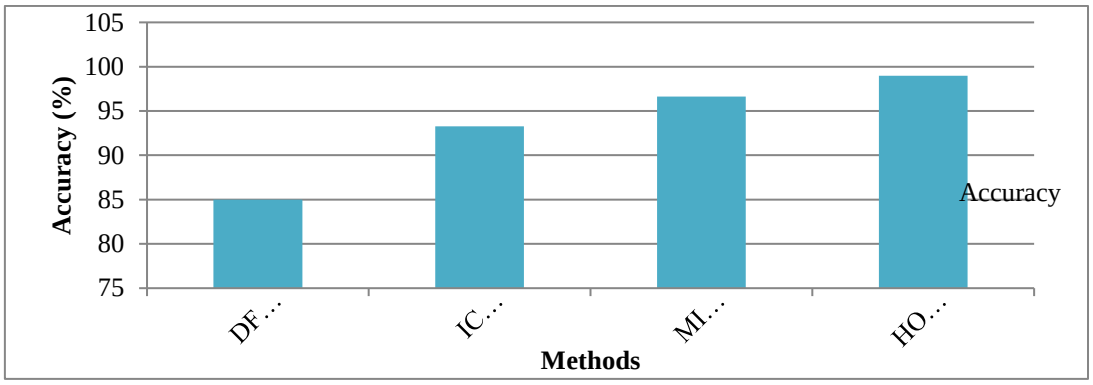

Figure 9: Accuracy comparison

Figure 9 shows the accuracy performance of the proposed HOICA-EL is compared with existing DFNForest, ICA-DFFNForest and MICA-DFFNForest approaches. Selecting the best genes for this planned research project involves Hybrid Optimized Imperialist Competitive Algorithm (HOICA) algorithm. It improves the accuracy rate. From the results, 98.99% accuracy is achieved by the suggested technique, compared to other methods like DFN Forest, which may be concluded, ICA-DFFNForest and MICA- DFFNForest approaches attains 85% ,93.25% and 96.62% respectively.

4.2 Precision

The term "precision" refers to the proportion of accurately anticipated positive observations relative to the overall number of expected positive observations.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (46)$$

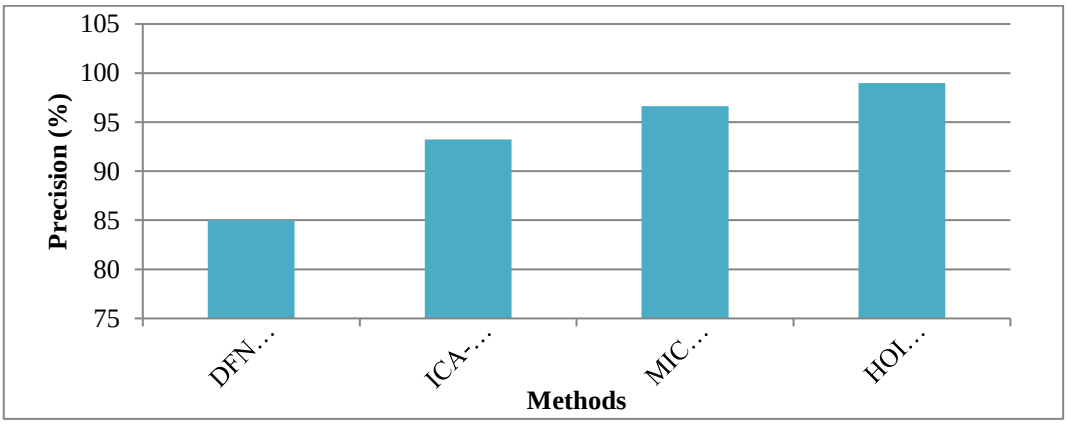

Figure 10: Precision comparison

The proposed HOICA-EL is compared with existing DFNForest, ICA-DFFNForest and MICA- DFFNForest methods in terms of precision. Figure 10 illustrates a comparison of accuracy. The results of the experiment demonstrate

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

that the HOICA-EL achieves 98.83 % of precision when existing DFNForest, ICA-DFFNForest and MICA-DFFNForest methods provides 86.30%, 94.48 % and 98.74 % respectively.

4.3 Recall

The proportion of accurately anticipated positive observations relative to the total number of observations made in the actual class is referred to as recall.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (47)$$

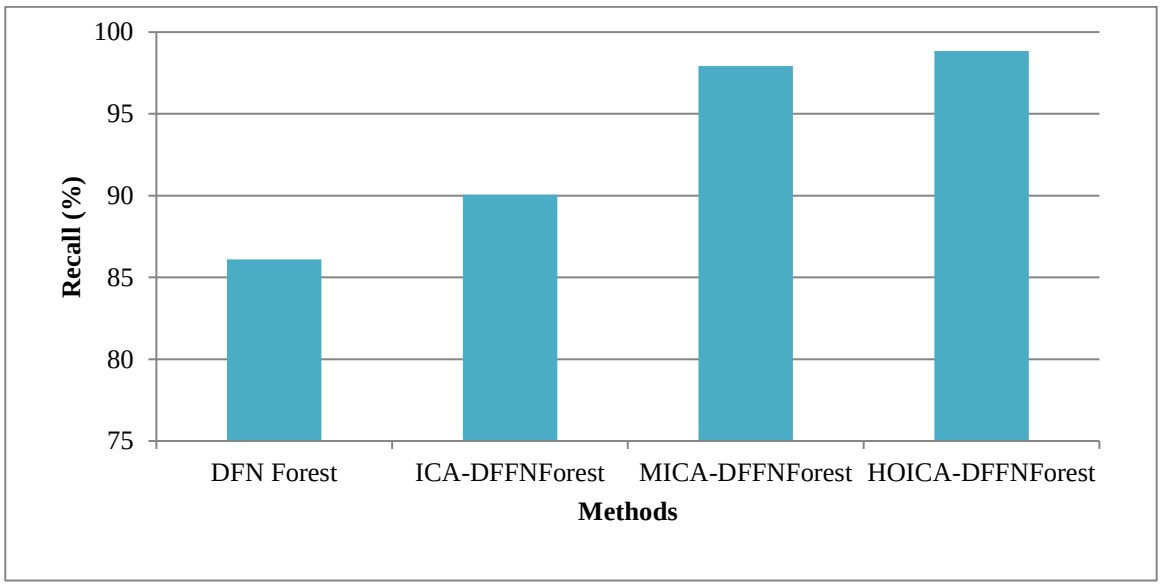

Figure 11: Recall comparison

When compared to the current DFNForest, the recall performance of the proposed HOICA-EL is evaluated, ICA-DFFNForest and MICA- DFFNForest approaches which are shown in figure 11. The dimensionality reduction process used in this suggested research work employs theImproved Independent Component Analysis (IICA). It improves the recall rate. From the graph, it shows that the proposed HOICA-EL attains 98.84% of recall whereasDFNForest, ICA-DFFNForest and MICA- DFFNForestmethods achieves 86.10%,90.05% and 97.92 % respectively.

4.4 F-measure

Precision and Recall are combined to create the F1 Score. The result is that this score accounts for both false positives and false negatives.

$$\text{F-measure} = 2*\frac{(Recall*Precision)}{(Recall+Precision)} \quad (48)$$

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

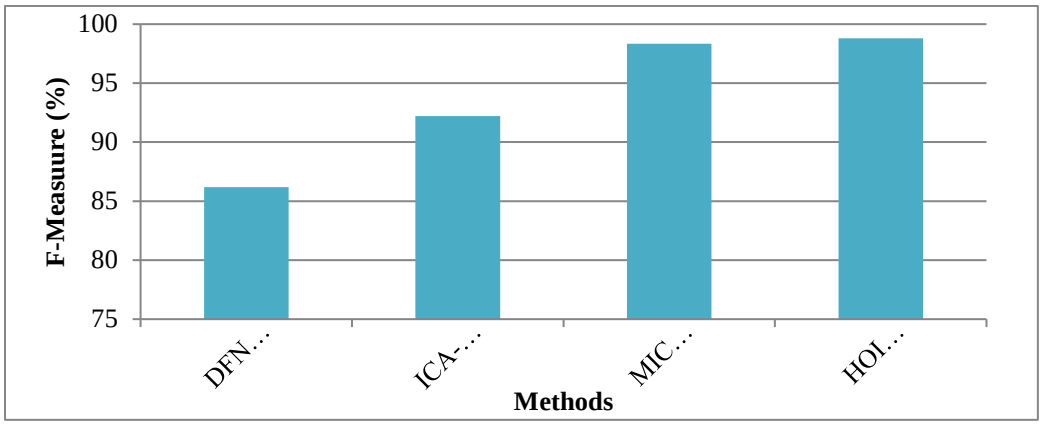

Figure 12: F-measure comparison

The f-measure performance of the proposed HOICA-EL is compared with existing DFNForest, ICA-DFFNForest and MICA- DFFNForest approaches and they are seen in picture 12. The f-measure is used as the y-axis in the x-axis procedures.The results of the experiments demonstrate that the suggested system's f-measure is 98.81% when DFNForest, ICA-DFFNForest and MICA- DFFNForestmethods attains 86.20%, 92.21% and 98.33% respectively.

4.5 Error

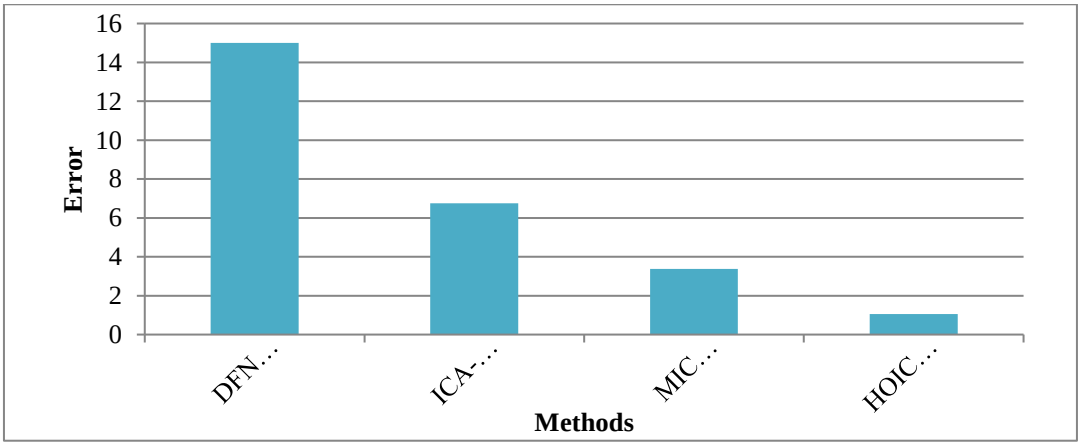

Figure 13: Error comparison

Figure 13 compares the performance of the current DFNForestto that of the planned HOICA-EL, ICA-DFFNForest and MICA- DFFNForest methods interms of error. The error is considered as the y-axis and the x-axis is techniques. This suggested work, an optimal geneis chosen with the help of Hybrid - Imperialist Competitive Algorithm (HOICA). It reduces the error

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

rate. The proposed HICA-EL system achieves 1.05% error rate, according to the graph, whereas DFNForest has a higher error rate, ICA-DFFNForest and MICA- DFFNForest methods achieves 15%,6.75% and 3.37% respectively.

5. Conclusion

The proposed research workdesigned aHybrid Optimized Imperialist Competitive Algorithm (HOICA) with Ensemble Learning (EL) in order to diagnose cancer subgroups. To lessen the difficulties presented by high - dimensional, Improved Independent Component Analysis (IICA) scheme is designed. In order to improve the classification accuracy, Hybrid Optimized Imperialist Competitive Algorithm (HOICA) is utilized. Finally, cancer subtype classification is done by using EL which consists of Deep Fuzzy Flexible Neural Forest (DFFNForest), WAFCNN and SVM approaches. In terms of accuracy, precision, recall, f-measure, and error, the experimental findings demonstrate that the suggested system out performs the current system.

References

[1] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, pp. 8-17.

[2] Fu, L. M., & Fu-Liu, C. S. (2004). Multi-class cancer subtype classification based ongene expression signatures with reliability analysis. FEBS SXEFR65TJNUHREGBTHBGTssss swederletters, 561(1-3), pp.186-190.

[3] Rhee, S., Seo, S., & Kim, S. (2017). Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. arXiv preprint arXiv:1711.05859.

[4] Tung, W. L., & Quek, C. (2005). GenSo-FDSS: a neural-fuzzy decision support system for pediatric ALL cancer subtype identification using gene expression data. Artificial intelligence in medicine, 33(1), pp.61-88.

[5] Franks, J. M., Cai, G., & Whitfield, M. L. (2018). Feature specific quantile normalization enables cross-platform classification of molecular subtypes using gene expression data. Bioinformatics, 34(11), pp.1868-1874.

[6] Joshi, P., Jeong, S., & Park, T. (2019, November). Cancer subtype classification based on superlayered neural network. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1988-1992). IEEE.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

[7] Hung, F. H., & Chiu, H. W. (2017). Cancer subtype prediction from a pathway-level perspective by using a support vector machine based on integrated gene expression and protein network. Computer methods and programs in biomedicine, 141, pp.27-34.

[8] Guo, Y., Qi, Y., Li, Z., & Shang, X. (2018). Improvement of cancer subtype prediction by incorporating transcriptome expression data and heterogeneous biological networks. BMC medical genomics, 11(6), pp.87-98.

[9] Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., & 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.

[10] Halder, A., Dey, S., & Kumar, A. (2015). Active learning using fuzzy k-NN for cancer classification from microarray gene expression data. In Advances in Communication and Computing (pp. 103-113). Springer, New Delhi.

[11] Guo, Y., Liu, S., Li, Z., & Shang, X. (2018). BCDForest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. BMC bioinformatics, 19(5), 118.

[12] Kar, S., Sharma, K. D., & Maitra, M. (2015). Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. Expert Systems with Applications, 42(1),pp. 612-627.

[13] Xia, C. Q., Han, K., Qi, Y., Zhang, Y., & Yu, D. J. (2017). A self-training subspace clustering algorithm under low-rank representation for cancer classification on gene expression data. IEEE/ACM transactions on computational biology and bioinformatics, 15(4), pp.1315-1324.

[14] Spurek, P., Tabor, J., &Śmieja, M. (2018). Fast independent component analysis algorithm with a simple closed-form solution. Knowledge-Based Systems, 161, pp.26-34.

[15] Nordhausen, K., &Oja, H. (2018). Independent component analysis: A statistical perspective. Wiley Interdisciplinary Reviews: Computational Statistics, 10(5), e1440.

[16] Yamashita, R., Nishio, M., Do, R. K. G., &Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), pp.611-629.

[17] Zhou, D. X. (2020). Universality of deep convolutional neural networks. Applied and computational harmonic analysis, 48(2), pp.787-794.