IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 25, 2019, accepted February 14, 2019, date of publication February 19, 2019, date of current version March 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900118

Robust Feature Selection With LSTM Recurrent Neural Networks for Artificial Immune Recognition System

CANAN BATUR ŞAHİN[1] **AND BANU DİRİ**[2]
[1]Computer Engineering Department, Siirt University, 56100 Siirt, Turkey
[2]Computer Engineering Department, Yildiz Technical University, 34220 Istanbul, Turkey
Corresponding author: Canan Batur Şahin (cananbatur@siirt.edu.tr)

ABSTRACT Stability and robustness of feature selection techniques have great importance in the high dimensional and small sample data. The neglected subject in the feature selection is solving the instability problem. Therefore, an ensemble gene selection framework is used in order to provide stable and accurate results of feature selection algorithms. Sequence modeling from high-dimensional data is an important research area for the discovery of biomarkers. Identifying biomarkers requires robust gene selection methods, which makes it possible to find important tumor-related genes with high accuracy. The main issue of this paper is creating a model in order to learn long sequences with the artificial immune recognition system (AIRS) for robust feature selection. Long short-term memory (LSTM) recurrent neural networks are trained with the AIRS in order to obtain the long-lived unit cells for use in the feature selection process. LSTM was used to be better understanding the mechanisms involving the ‘‘remember’’ feature of the immunological behavior of the immune response. We tried to apply a theory suggested by immunologists in order to develop stable associative memory, which capable of solving robustness and optimization tasks. We examined the initial gene selection step based on the different types of group formation algorithm for analysis of the most informative selected features. Microarray datasets are showing remarkable increases in their robustness and classification accuracy. The suggested framework is evaluated on six commonly used microarray datasets.

INDEX TERMS Artificial intelligence, classification algorithms, computational intelligence, neural networks.

I. INTRODUCTION

The early diagnosis of tumors has great benefits in the treatment of cancer, but current techniques for discovering biomarkers that can diagnose cancer are inadequate. Selecting informative gene subsets and accurately classifying tumor samples can help detect important cancer biomarkers. In biology, the main goal of microarrays and mass spectrometry is to detect marker genes or proteins from high output experiments instead of creating models to predict disease susceptibility or phenotypes from original specimens. Although many feature selection algorithms are suitable for selecting feature subsets, they are inadequate in making reliable candidate attribute definitions to validate expensive biological experiments. Discovery of biomarkers requires a robust feature selection method for such datasets. Feature selection for gene expression datasets is usually suboptimal due to the high-dimensionality and small sample characteristics. A popular option for defining attributes is verifying biological experiments through the best classifier accuracy. Nevertheless, different attribute subsets of the same data can be quite similar or even the same in classifier accuracy results and only a large number of distinct attribute subsets can reveal instability in feature selection algorithms. The learning model should not neglect feature selection stability in applications of characteristic markers or biomarker identification. Given these shortcomings, a viable alternative approach is the analysis of microarray data. By using a qualified gene selection mechanism, stability improves and biomarkers can be identified with greater confidence.

The natural immune system is a highly adaptive, distributed, and parallel system. Through learning, memory,

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Afzal.

and relational information acquisition it successfully solves classification and pattern recognition problems. There is no system that can model long-sequence learning in Artificial Immune Recognition Systems (AIRS). The advantage of sequence modeling in AIRS is that it can be converted into an intelligent network system in which the learning process can be performed.

Therefore, in this work LSTM is used with AIRS in order to keep features which are preserved for a long time. It is possible to define an LSTM block as an intelligent network cell since it can remember a value for a time in random length. The sequence learning was performed with an LSTM recurrent neural network, which improves the stability of the consensus gene sequences to find the potential long-term dependency. Intuitively, if an LSTM unit detects an important attribute in an early stage sequence, it will acquire a potential long-term association since this information can be easily transferred. LSTM mechanism trained based on AIRS in order to detect long-lived unit cells for using in the feature selection process. This study attempts to create an ideal infrastructure for stable feature selection using the memory cells involved in immunity acquisition. In AIRS, there is no mechanism to ensure that relational immune memory develops in a way that allows for long-sequence learning. In the long-term preservation of immunological memory, stable feature groups are obtained based on optimal biological gene sequences.

The unique aspect of this work is the development of the immune memory of the AIRS with the mechanism which is created by LSTM memory blocks. LSTM will function as a kind of internal stimulation mechanism in this paper. Suggested associative memory mechanism was modeled based on the LSTM in order to create a robust immune memory and realize association of the learned pattern quickly and effectively. The attributes that will be selected (cell survival in the resting state) during the long-term preservation of the immune memory will be referred to as long-lived unit cells attribute groups.

The dense group finder (DGF), correlation-based feature group (CFG), and information gain based feature group (IGFG) algorithms are used to obtain associative feature subsets groups. Each associative set of feature groups were developed with the meta-dynamics of the suggested LSTM-AIRS versions which are, long-short-term memory based artificial immune recognition system version-1 (LSTM-AIRS1), long-short-term memory based parallel artificial immune recognition system version-1 (LSTM-PAIRS1), long-short-term memory based artificial immune recognition system version-2 (LSTM-AIRS2), long-short-term memory based parallel artificial immune recognition system version-2 (LSTM-PAIRS2), Genetic algorithm (GA), and Artificial neural network with Genetic algorithm (ANN+GA) algorithms like a single cell in order to find the optimal sequences. Then algorithms compared with fast correlation-based feature selection (FCBF), consensus group stable feature selection (CGS), sequential forward feature selection (SFS), and sequential backward elimination (BES) algorithms.

The other contents of this paper are as followed: the second section includes the related work and the third section includes the ensemble feature selection. The fourth section includes the methods and the suggested method represented in the fifth section. Experiment setup is explained in the sixth section. The seventh section includes the results and discussions and finally, the conclusions are given in the eighth section.

II. RELEATED WORK

In this study [3], the proposed artificial immune systems and local attribute selection approach were created within the scope of the epitope and paratope regions formed by amino acids residues generated during the binding of an antigen and an antibody in artificial immune systems. The local selection mechanism proposed within the scope of the study is the original aspect of the study. It was observed that the proposed local attribute selection mechanism eliminated the multidimensionality complexity.

Standard artificial immune recognition systems developed by using the Opposite Sign Test (OST) approach [7] were used in attribute selection. The classification accuracy obtained as a result of the attribute selection was increased by combining the local search technique, OST, with the efficient aspects of artificial immune recognition systems.

This study [10] includes the selection of attribute groups with a high correlation by the FAST (fast clustering-based feature selection) algorithm. The most distinctive attribute groups were defined by the Markov blanket approach. The stable attribute groups were obtained by the Greedy Variable neighborhood search approach.

Fast correlation-based filter (FCBF) used within the scope of this study [1] as a preprocessing data and Artificial Immune Recognition System (AIRS) algorithm used as a model prediction. Experiment results show the classification of the dataset Breast-Cancer-Wisconsin reaches correctness 100% based on k-NN.

The distance-based feature selection method is investigated in this study [19] for a two-category classification problem. The marker genes selected by implementing the Bhattacharyya distance to measure the dissimilarity in gene expression levels between groups. A novel method is proposed for marker gene selection and cancer classification based on SVM. The proposed algorithm is very promising in gene selection and cancer classification.

In this study [13], the two-stage filter-wrapper model suggested for feature selection in microarray datasets. An ensemble of filter method developed based on the Relief-F, chi-square and symmetrical uncertainty. The ensemble learning was applied by the union and intersection of the three ranking. Multi-objective GA used in order to improve the quality of the selected features. Experimental results show the success of the suggested model in cancerous gene identification.

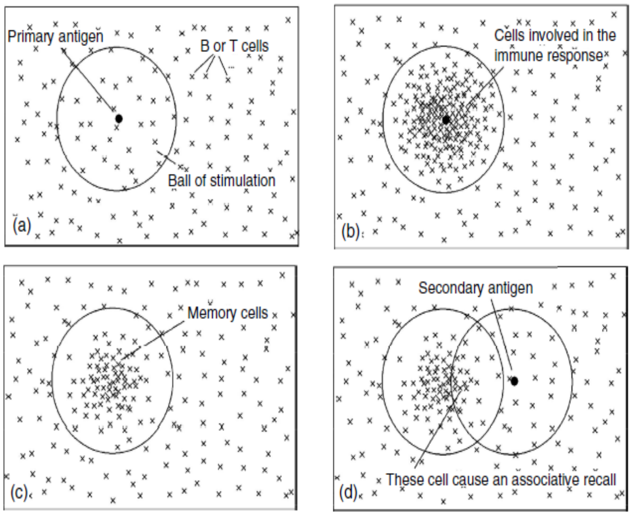

FIGURE 1. Modeling Immune Memory as Associative Memory [2].

The study [22] focuses on the analysis of the co-expressed genes in the microarray data. Gene clustering performed with developing the multi-objective clonal selection optimization algorithm (MCSOA). Two conflicting objective function used based on the clustering validity indexes in order to utilize a robust MCSOA. The first objective function formed by the combination of the IDE, homogeneity, DB, and XB. I index and separation used together as a second objective function. Results show the superiority of the used gene clustering method.

III. PREMINILARIES

A. BIOLOGICAL INSPIRATION

The artificial immune system (AIS) is a computation system and supervised learning algorithm inspired by biological immune system metaphors. A major feature of the adaptive immune system response is its ability to "remember" when the same antigen encountered later time. We try to understand the mechanisms involved in this behavior. The modeling immune memory as an associative memory is illustrated in Figure 1.

The formation of the immune memory is visualized with the density leveling of the various memory cells during the primary and secondary responses. When the immune system first encounters an antigen, a primary response is induced: a number of lymphocytes are produced. Some of these lymphocytes are kept as memory cells. The next time the same antigen is detected, memory cells generate a faster and more intense response (the secondary response). In this way, memory cells work like associative memories.

Various strategies are suggested for understanding the immune system's learning and memory. According to one hypothesis, when the ARB cells are activated by an antigen, they have a time period in which they expect to encounter the same antigen. ARB cells stay in memory for weeks or even months. The lymphocytes are able to maintain a simple dormant state for many years in this time interval. The antigen itself or a partially deformed state of it (a small variation) accumulates in the lymph nodes or organs, and the memory of the immune system which is exposed to the antigen is fed at a periodic interval. However, survivors are not known in case ARB cells have not been activated by antigen for years. A basic theory [2] behind this idea depends on the immunological theory. According to this theory, some immunologists suggest that some kind of internal restimulation mechanism protects the immune memory for a long time.

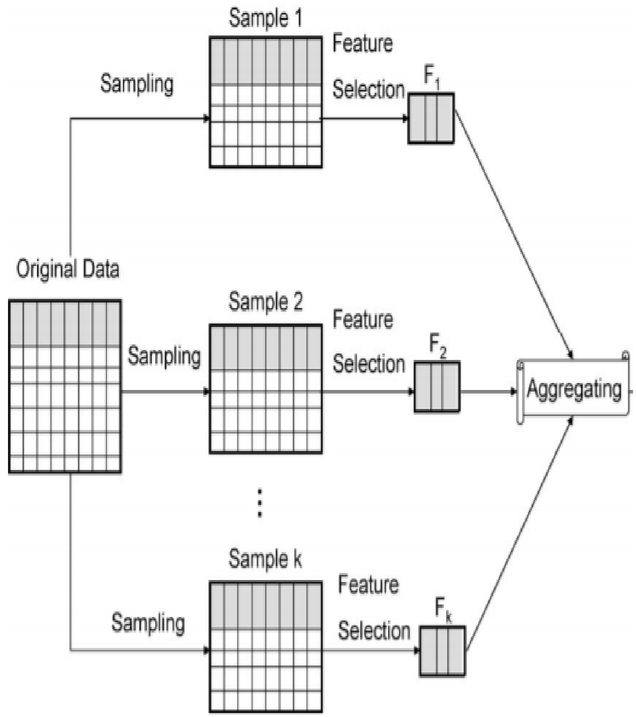

FIGURE 2. Ensemble feature selection framework [21].

The immune memory of the AIS was developed in order to investigate a system for obtaining associative immune memory with an intelligent network system for sequences could be effectively memorized as robust patterns. Associative memory remembers the patterns being learned and enables effective recall of learned patterns.

B. ENSEMBLE FEATURE SELECTION

In this paper, the ensemble feature selection framework is used with the random subsampling of the dataset. Figure 2 shows the ensemble feature selection framework. The relevance and redundancy are the main concepts in feature selection. Therefore, in this work, ensemble gene selection framework is used in order to provide stable and accurate results of the feature selection algorithm. In this framework, the different types of associative feature groups which are created based on the principles of group-based learning method are used. The stability is achieved by the resistance of the associative feature groups to the variations of the training sample. Each type of associative feature groups is created based on the subsampling of the training samples. Creating associative feature groups based on group relevance less prone to overfitting on small samples and reduce the complexity of multidimensionality [16]. Learning was performed at the group level. The purpose of using this idea is to handle selection instability with sample selection and approximate intrinsic properties of the original dataset.

High-dimensional data have many features and few samples, which poses a classification problem. Therefore, it is very important to obtain feature groups from high dimensional data. In this study, feature groups were obtained based on density, correlation, and knowledge acquisition. By clustering feature groups based on high-dimensional data, group-level learning is achieved, as opposed to simply bringing together the members of each feature group though random relationship variation. Thus, there is more potential to converge on the intrinsic feature groups of the dataset.

Data-driven feature group formation used to pre-select the different type associated gene subset groups. The DGF (Dense Group Finder), CFG (Correlation Based Feature Group) and IGFG (Information Gain Based Feature Group) algorithms are used to obtain an associative set of feature groups. The formation algorithms of the feature groups were used by the ensemble gene selection framework in the initial stage of feature selection tasks. The improvement of each attribute group and achieving group-level learning is more effective than in bringing each attribute group based on the random relationship variations of group members.

In (1), the kernel density function is used in order to obtain the density based feature groups. The sequence of consecutive locations of the kernel function determined the calculation of C_{j+1} equation. The parameter h used for kernel bandwidth refers to the nearest neighbor number, the parameter p used for the total number of attributes in the datasets, any attribute represented by f_i parameter, K represents the kernel function.

$$c_{j+1} = \frac{\sum_{i=1}^{p} fiK(\frac{c_j - fi}{h})}{\sum_{i=1}^{p} K(\frac{c_j - fi}{h})}, \quad j = 1, 2, \cdots \tag{1}$$

In (2), the usefulness of a subset of attributes determined based on the correlation based feature group function. Evaluation of the attributes determined based on the value of the heuristic evaluation function. The intuitive usability of a subset of S with k attributes is represented by merits equation. The parameter rcf represents the mean attribute-class correlation and rff parameter represents the correlation between the mean attributes.

$$\mathrm{merits}_S = \frac{\mathrm{k} * \mathrm{rcf}}{\sqrt{\mathrm{k} + (\mathrm{k}-1) * \mathrm{rff}}} \tag{2}$$

The importance of a given attribute of the original feature set obtained based on information gain function. The knowledge in the feature is evaluated using the entropy criteria. The entropy presented by of the f_t attribute with M data can be found by the (3).

$$\mathrm{E} = -\sum_{i=1}^{M} (\mathrm{f_t}(\mathrm{i}) \log (\mathrm{f_t}(\mathrm{i})) \tag{3}$$

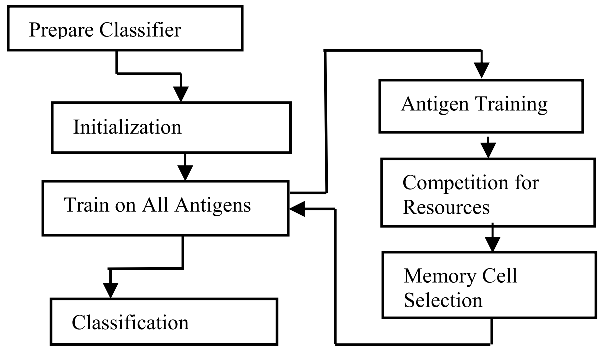

FIGURE 3. Standard AIRS schema [5], [6].

IV. METHODS

A. ARTIFICIAL IMMUNE RECOGNITION SYSTEM (AIRS)

Artificial Immune Recognition System (AIRS) is a supervised learning algorithm which depends on the metaphors of the biological immune system.

Figure 3 presents the standard AIRS schema. At the initialization stage, the data set is normalized to the range of [0, 1]. After normalization, the affinity threshold is calculated by (4). At the next stage, antigens are presented to the storage pool with antigen training. At the memory cell recognition stage, a stimulation value is assigned to these cells by stimulating the recognition cells in the memory cell pool. Affinity is calculated by (5), the stimulation value is calculated by (6). Best match memory cell ($M_{cmatch)}$ is calculated by (7). A number of clones calculated based on the (8). Affinity describes the degree of similarity between a recognition cell and an antigen. Stimulation is an inverted affinity. Numclones is the clone number, which utilizes local searching during training. A clonal rate is the rate of the cloned cell for the best matching memory cell. Within AIRS, ag_i and ag_j are the sequential antigens in the input training data set (antigens). Where n is the number of antigens in the input training data set and m_c is the memory cell.

$$\text{Affinity threshold} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \times \left(\frac{\text{Affinity}\left(ag_i, ag_j\right)}{\frac{n(n+1)}{2}} \right) \tag{4}$$

$$\text{Affinity}\left(ag_i, ag_j\right) = 1 - \text{Euclidean distance}\left(ag_i, ag_j\right) \tag{5}$$

$$\text{Stimulation} = \begin{cases} \text{Affinity}\,(m_c, ag) \text{ if } m_c.\text{class} = ag.\text{class} \\ 1 - \text{affinity otherwise} \end{cases} \tag{6}$$

$$\mathrm{Mc_{match}} = \operatorname{argmax}\left(\text{Stimulation}\,(m_c, ag)\right) \tag{7}$$

$$\text{NumClones} = \text{Stimulaion} * \text{clonalRate} \tag{8}$$

Figure 4 shows the standard AIRS resource competition schema. The crucial point of artificial immune systems is the evolutionary process that memory cells spend in the Artificial Recognition Ball (ARB) population. The process of evolution involves certain concepts. First, the ARB population

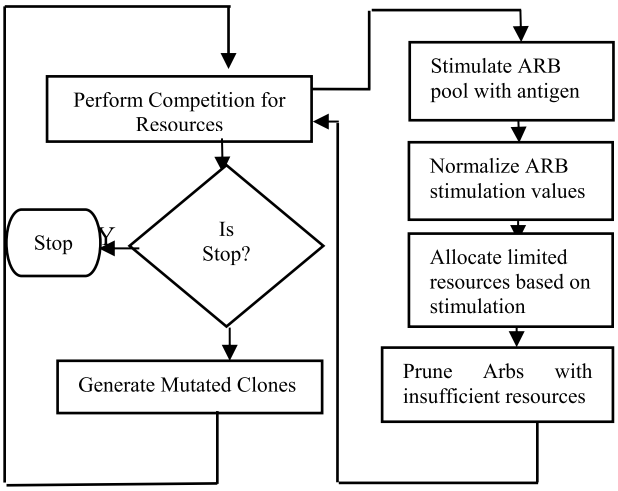

FIGURE 4. **Standard AIRS resource competition schema [5], [6].**

is evolved by the competition created by memory cells for large resources of the evolutionary constraint system required for the development of memory cells. In this way, which memory cell survivor is determined, they are transformed into qualified classifier cells. The goal of resource competition is to develop the most appropriate individuals, as in genetic algorithms. Within the artificial immune recognition systems, the suitability of the cells is measured by their strong activation of the ARB cells by antigen (antigen-antibody binding). Firstly, the value of activation between a memory cell and antigen is calculated and is processed by taking into account the memory cell and antigen class values. Finally, the reward system allocates large system resources to the cells. Memory cells (close to anti-cells), which have a high activation value (antigen-antibody binding strength) and have the same class label, and cells with low activation value and the same class label as the antigen (far from the antigen) are the most rewarded. Cells with low activation values and memory cells with different class labels with antigen, cells with high activation value and different class labels with antigen are not only seen as weak for classification but also are least rewarded as potentially harmful cells. Since these ARB cells have insufficient ability to obtain the rewards, they are removed from the system and the system is purified. Within the search space, there is a serious challenge that evolves each ARB cell with the most award acquisition. After the evolutionary process of this ARB population is completed, the original memory cell may have the potential to displace with the best memory cell evolved. This potential for displacement is only possible when the presented antigen is closer to the original memory cell previously located in the memory pool and when both the original memory cell evolves and the subsequent memory cells are in proximity to each other.

The main difference between AIRS1version and AIRS2 version are that the ARB pool used as a permanent resource in the AIRS1 algorithm, vice versa temporary resource in the AIRS2 algorithm. In AIRS1, the ARBs remaining from previous steps cause the algorithm to spend more time competing for limited resources. This makes AIRS1 a complex algorithm. While AIRS1 uses the mutation parameter that can be defined by the user, AIRS2 uses the somatic hypermutation where the mutation ratio of a clone is proportional to the affinity [10]. While the classes of clones may change after the mutation process in the AIRS1 algorithm, classes are not allowed to change in the AIRS2 algorithm. Parallel-AIRS1 and Parallel-AIRS2 versions demonstrate the distributed nature of the immune systems and their parallel processing qualities. In these versions, initially, each part of the training data set is assigned to np number of processes. Thus, it ensured that np number of the memory pool is created by running the AIRS algorithm on each process. As a result, the memory pools obtained are merged [18].

B. GENETIC ALGORITHM (GA)

Genetic Algorithm (GA) is a class of the evolutionary algorithms uses Darwin's principle of the "survival of the fittest" and techniques inspired by evolutionary biology. In the genetic algorithm, each possible candidate solution is represented in the initial population. The evaluation of each candidate solution is performed by genetic operators which are crossover and mutation. Each candidate solution has a fitness value in order to show goodness which computed based on the fitness function determined for the problem. Evolving each candidate solution is performed by the fitness function. Based on the "survival of the fittest" principle, the individuals can propagate throughout the population and yield more "fitter" individuals.

C. ARTIICIAL NEURAL NETWORK WITH GENETIC ALGORITHM (ANN+GA)

Artificial Neural Network (ANN) is a biological inspired computational intelligence technique which inspired from human brain. In this paper, we try to implement the genetic algorithm for optimization of the artificial neural network algorithm for the learning behavior [23]. Designing of the ANN and setting the number of hidden node (HN) and connection weights parameters is difficult. Optimization of the HN parameter is performed by GA. Fewer numbers of HN causes the under-fitting problem and big numbers of the HN causes the over-fitting problem. DGF, CFG and IGFG associative set of feature groups used as the feature subsets for ANN.

In this paper, we try to construct a hybrid system in order to observe the generalization ability of the ANN based on the associative feature groups. The fitness function is calculated based on validation mean square error (MSE). Small MSE has a better chance to be selected as a parent for mating and propagate throughout the population [4].

D. STANDARD FEATURE SELECTION ALGORITHMS

The Fast Correlation-based Feature Selection (FCBF) algorithm is finding a set of predominant features. FCBF algorithm consists of two parts. In the first part, for each feature,

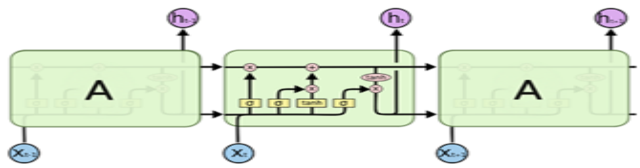

FIGURE 5. Structure of the LSTM block [21].

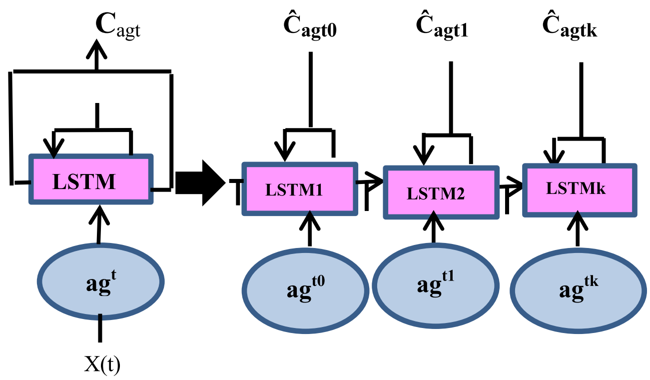

FIGURE 6. A LSTM block with LSTM units.

the symmetrical uncertainty (SU) value computed and then based on the predefined threshold δ relevant features selected in the feature list. The feature list is ordered in descending order according to their SU values. In the second part, remove redundant features from the ordered list and among all the selected relevant features keep the only predominant ones [8].

The Consensus Group Stable Feature Selection (CGS) algorithm identifies consensus feature groups by sub- sampling training samples. The major aim of the algorithm is to converge to the intrinsic feature groups by an aggregating set of consensus feature groups from ensemble learning with generating the DGF groups on each sample.

The Sequential forward feature selection (SFS) and backward elimination feature selection techniques (BES) are simple and effective methods for selecting attributes. While these methods create a set of attributes, they perform extraction or addition of an attribute from the attribute subset according to the method selected at each step. The selection criterion here is the performance ratio of the classifier algorithm. According to the performance status of the specified classifier algorithm, the discriminative attributes are determined at each step.

V. THE SUGGESTED METHOD

The underlying principle of the suggested method is the modeling of the potential long-lived unit cells based on the LSTM. The LSTM is a recurrent neural network (RNN) architecture that is a kind of artificial neural network (ANN). An LSTM network is a fairly reasonable approach to predicting the time period of the process when learning from the experience for a classifier is not known when long delays between significant events are unknown. The overall effect is that the network makes it possible to store and retrieve information over a long period. Figure 5 illustrates the structure of the LSTM block. Each of the LSTM units has an input gate that decides which new information should be stored for the cell state, a forget gate that remembers which information should be discarded and an output gate that allows new information to enter the cell. The sigmoid layer lets activation out to potentially affect other cells or the network's output.

Where X_t represents an input at time t and h(t-1) represents a hidden state at time t-1. The sigmoid layer decides whether the information is discarded or retained. The output gate is represented by h(t). σ and *tanh* are respectively a sigmoid function and hyperbolic tangent function. V is the diagonal matrix.

The key of the LSTM is the state of the cell. The cell states in LSTM blocks, behave a kind of conveyor belt. Linear interactions in LSTM chain preserve the long-lived unit cells. Figure 6 presents an LSTM block with LSTM units. Each X (t) represents an antigen during the training process. The antigen training set (ags) is introduced into the network as an LSTM block series to compute linear mappings from hidden state outputs.

Each j. LSTM unit has a memory amount of C_t^j at time t. At time t, the activation of the j. LSTM output is h_t^j. LSTM unit;

$$h_t^j = \sigma_t^j \tanh\left(c_t^j\right) \tag{9}$$

σ_t^j, is an output gate that modulates the amount of memory content exposure.

Output gate;

$$\sigma_t^j = \sigma(W_0X_t + U_0h_{t-1} + V_0C_t)^j \tag{10}$$

The new memory contents of the memory unit Ĉjt are updated by partially forgetting the current memory and adding the new memory contents to Cjt.

$$C_t^j = f_t^j C_{t-1}^j + i_t^j \hat{C}_t^j \tag{11}$$

The new memory contents,

$$\hat{C}_t^j = \tanh\left(W_cX_t + U_ch_{t-1}\right)^j \tag{12}$$

To what extent the current memory is forgotten, a forgetting gate is modulated by f_t^j and the degree of addition of the new memory content to the memory cell is modulated by an input port i_t^j.

$$f_t^j = \sigma(W_fX_t + U_fh_{t-1} + V_fC_{t-1})^j \tag{13}$$
$$i_t^j = \sigma(W_iX_t + ih_{t-1} + V_iC_{t-1})^j \tag{14}$$

Figure 7 shows LSTM blocks along time series. During training, the LSTM evolves an antigenic pattern by remembering the state of each evolution of the cell and treats them as network inputs. The time-series information for each memory cell was transferred with a sequential time as a long-term association.

Figure 8 represents the LSTM-AIRS algorithm. Each memory cell will be used as an LSTM block in evolutionary

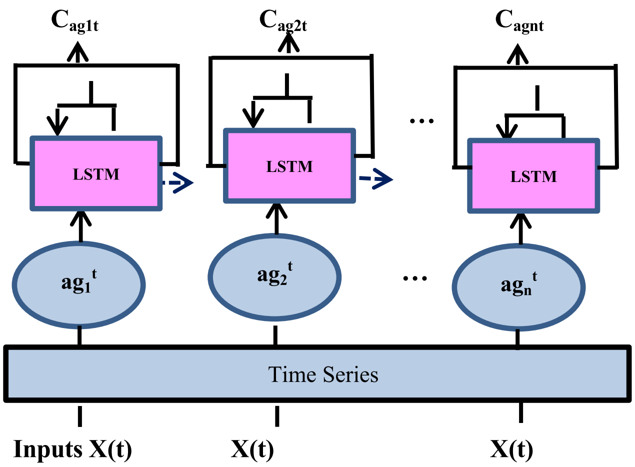

FIGURE 7. LSTM blocks along time series.

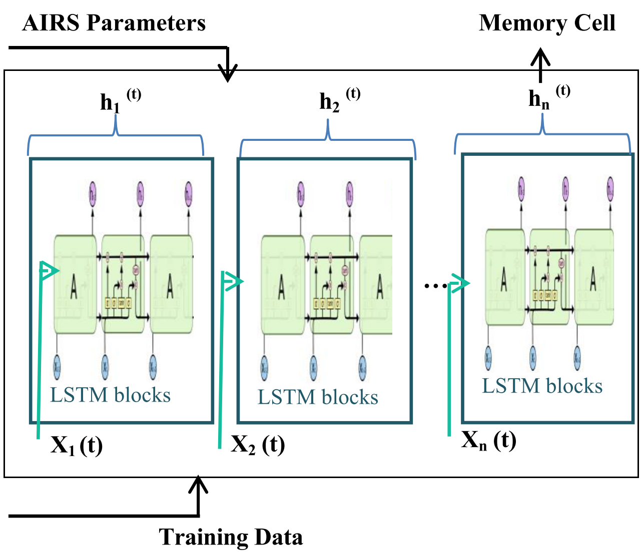

FIGURE 8. LSTM-AIRS algorithm.

terms in order to strongly match the antigen pattern. Each mutation treatment changes the current state of the cell (Cat time t, in order to predict inputs with time series) with an iterative time in terms of the amount of new memory content exposure ($\hat{C}$ at time t).

That the improvements of the state achieved by applying mutation to specific sub-populations allow children to make further exploration along the entry search space based on the LSTM treatment. The updating process of LSTM block's state is enhanced until learning of a presented antigen pattern. Stable attribute groups will be obtained from the output gate by updating the previous state-of-the-art feature matrices (hidden state) at the base of long-short-term memory functions. During the evolution process, the memory cells with desired qualities are presented in output gate as an h (t)n (n. output in time t) will be added to this long sequences in this way based on restimulation mechanism. These memory cells are used for classification later.

Algorithm 1 Pseudocode of the LSTM-AIRS Algorithm

Input: InputPatterns, Clone_{rate}, mutation_{rate}, Stim_{thres}, Resorce_{max}
Output: $\text{Cells}_{memory} \leftarrow$ MemoryPool(InputPatterns)
{*InitializeMemoryPool* (InputPatterns)}
{*InitializeFeatureSet*(Ω)}
[Train] {1...N} $\leftarrow$ (Input Patterns)
For I=1 $\rightarrow$ N **do**
$m_{cbest} \leftarrow$ *GetMostStimulated* (Cells_I)
Fitness $\leftarrow$ *Accuracy* (Ω, m_{cbest})
$\text{Clone}_{num} \leftarrow m_{cbest} * \text{Clone}_{rate} * \text{mutation}_{rate}$
For ($\text{Cells}_i \in \text{Clone}_{num}$)
$\text{Cells}_{clones} \leftarrow$ (CloneandMutate (m_{cbest}))
{*IntroduceARBPool*} $\leftarrow$ (Cells_{clones})
mcandidate $\leftarrow$ Affinity (Cells_{clones}, Cells_{best})
Fitness $\leftarrow$ *Accuracy* (Cells_{clones}, mcandidate)
While (AverageStimulation(Cells_{clones})) $\leq$ Stim_{thres}
CompetitionforLimitedResources(Cells_{clones}, Resorce_{max})
Fitness $\leftarrow$ *Accuracy* (Feature set (Ω), Cells_i)
$\text{Cells}_{LSTM} \leftarrow$ *CreateLSTMMemory* (Cells_i, t, State_id)
$\hat{C}_{LSTM} \leftarrow$ *UpdateLSTMMemory* (Cells_{LSTM}, t, State_id)
State_id+=State_id
$C_{LSTM} \leftarrow$ *UpdateLSTMMemory* ($\hat{C}_{LSTM}$, t+1, State_id)
$\text{Cells}_{LSTM} \leftarrow C_{LSTM}$
END
$\text{Cells}_{memory} \leftarrow \text{Cells}_{LSTM}$
END
$\Omega^* \leftarrow$ NewFeatureSet(Cells_{memory})
END

TABLE 1. Microarray datasets.

Dataset	Gene	Sample	Class
Colon	2000	62	2
Lung	5000	181	2
Prostate	6034	102	2
SRBCT	2308	63	4
Lymphoma	4026	62	3
Leukemia	7129	72	2

The Figure 9 represents the flowchart of the suggested framework.

VI. EXPERIMENT SETUP

The most common six microarray data sets were used in this study [17]. Table 1 includes information on the genes, samples and class numbers.

Experimentally obtained performance values of the algorithms were achieved by dividing the data set as 70% training,

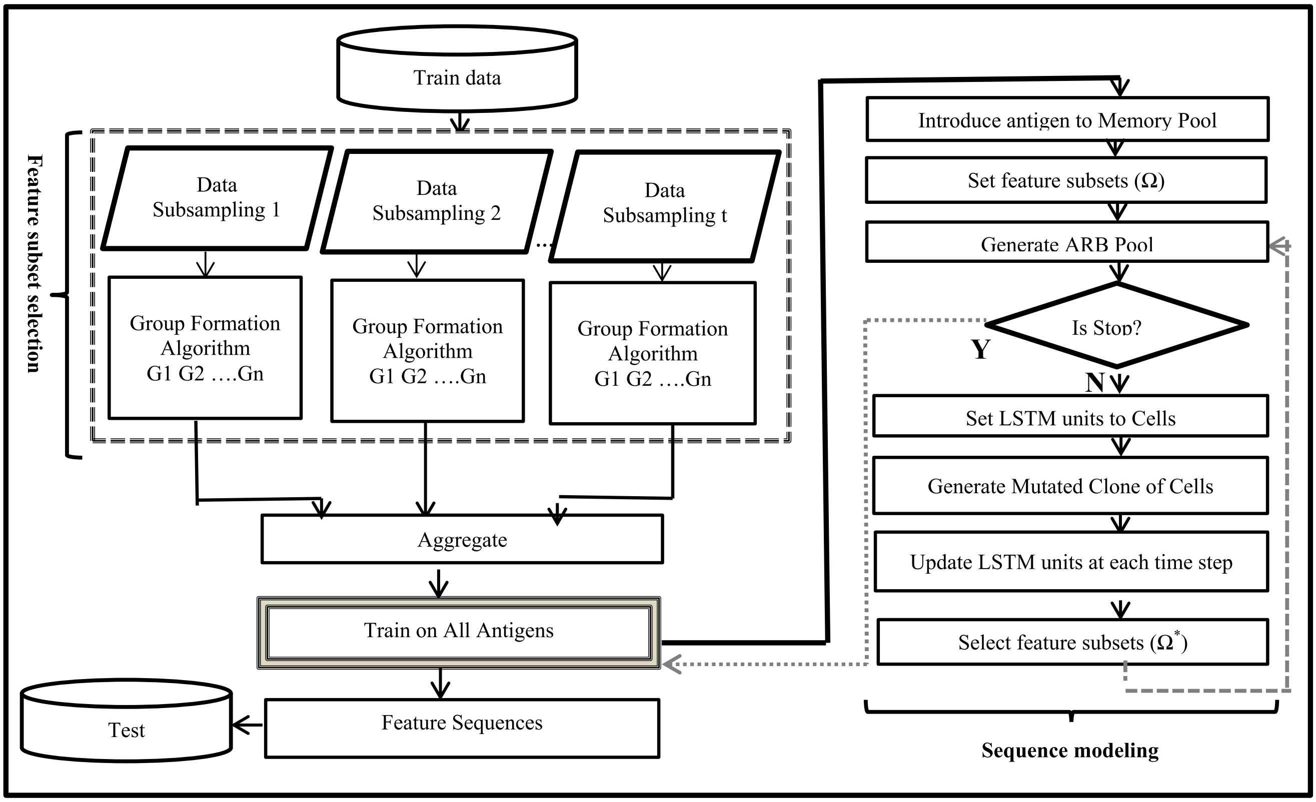

FIGURE 9. Flowchart of the suggested framework.

15% validation, and %15 test set. A number of bootstrap data sets were obtained from the training data set in order to ensure the resistance of training samples against variations. Then, *n* is the number of feature groups selected by separately running the group formation algorithms which are DGF, CFG, and IGFG based number bootstrap data sets. We set the values of the parameters of *t* and *n* to 10 and 10. The number of features determined from the group formation algorithms vary for each data set. Within the scope of this study, for Colon, Lung, Prostate, SRBCT, Lymphoma and Leukemia data sets the number of attributes obtained based on the DGF algorithm is 179, 130, 113, 153,132 and 162, and for the CFG algorithm is 100,123,124,136,192 and 248 and the IGFG algorithm is 20, 250, 96, 26, 30 and 306. The presence of the feature in a feature subset was encoded with 1 while the absence of it was encoded with 0. Each group of feature sets treated based on the suggested LSTM-AIRS1, LSTM-PAIRS1, LSTM-AIRS2, LSTM-PAIRS2, GA, and ANN+GA algorithms like a single cell. Then these algorithms compared with FCBF, CGS, SFS and BES algorithms. While stability results were evaluated by the Jaccard test, their classifying accuracy was evaluated using SVM, Naïve Bayes, and Random Forest classifiers. Weka was used to obtain classification accuracies. For all algorithms, the classification accuracy of the most optimal solution candidate obtained at the end of each run was evaluated using only test data set with 10 cross fold validations. The performance values added to the results were calculated by taking the average of the number of tests. In this paper, the validation set is used to tuning parameters of Genetic Algorithms and Artificial Neural Network with Genetic Algorithms. For standard AIRS the *affinity threshold value*, *clonal ratio*, *mutation ratio*, *np*, *total resource*, *stimulating value*, *hypermutation ratio, number of tests* parameters set to the values of 0.2, 10, 0.1, 2, 150, 0.9, 2.0, *10* respectively and the k-NN used as the classifier. In suggested LSTM based AIRS the *affinity threshold value*, *clonal ratio*, *mutation ratio*, *np*, *total resource*, *stimulating value*, *hypermutation ratio*, *number of tests* and *iteration number* parameters set to the values of *0.2*, *10*, *0.1*, *2*, *250*, *0.91*, *2.0*, *20* and *30*, respectively. The Wekaclassalgos project source code used.

In the GA, *population size* takes the value *100* and *chromosome length* varies according to the feature size of the DGF, CFG, and IGFG algorithms. The uniform crossover technique and tournament selection method used in this study. The *mutation ratio*, *number of the tournament*, *number* of tests and *iteration number* set to *0.8*, *2, 20 and 30*. The fitness function of AIRS and GA was calculated according to the accuracy of the KNN classifier.

In ANN+GA, *learning rate*, *momentum*, *training time* parameters set to *0.3*, *0.2* and *1000*. ANN training is done by gradient descent trainer. GA selected HN size for DGF, CFG, and IGFG algorithms are respectively *5.8*, *5.4* and *4.5*. The fitness function of the ANN+GA was calculated according to mean square error (MSE) of the KNN classifier.

All parameters of the GA and ANN+GA tuned using the validation set.

In the FCBF algorithm, the Symmetrical Uncertainty (SU) is computed with entropy in (15). The conditional entropy computes the dependencies of features with the (16). If X is a random variable and P(x) is the probability of x, the information gain is computed with the (17), SU value computed with using the equation of (18).

$$\mathrm{H}(\mathrm{X}) = -\sum\nolimits_{\mathrm{i}} \mathrm{P}(\mathrm{xi}) \log 2\mathrm{P}(\mathrm{Xi}) \tag{15}$$

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = \sum\nolimits_{\mathrm{j}} \mathrm{P}(\mathrm{yj}) \log 2 \times \sum\nolimits_{\mathrm{i}} \mathrm{P}(\mathrm{xi} \mid \mathrm{yj}) \log 2(\mathrm{P}(\mathrm{xi} \mid \mathrm{yj})) \tag{16}$$

$$\mathrm{IG}(\mathrm{X}|\mathrm{Y}) = \mathrm{H}(\mathrm{X}) - \mathrm{H}(\mathrm{X}|\mathrm{Y}) \tag{17}$$

$$\mathrm{SU}(\mathrm{X}, \mathrm{Y}) = 2\left[\mathrm{x} = \frac{\mathrm{IG}(\mathrm{X} \mid \mathrm{Y})}{\mathrm{H}(\mathrm{X}) + \mathrm{H}(\mathrm{Y})})\right] \tag{18}$$

In CGS algorithm, 10 fold cross-validation procedure used. The CV accuracies calculated using the KNN classifier. In SFS and BES algorithms the number of tests and iteration number parameters took the values of 20 and 30, respectively. KNN classifier used and the k parameter set to the value 5.

VII. RESULTS AND DISCUSSION

The stability of feature selection approaches is obtained by measuring the similarity between feature sets. In this study, the Jaccard index was calculated using the formula given in (19). Parameter m was used to specify the number of feature sets while expressing two feature sets used for Si and Sj similarity measurement.

$$\mathrm{Jaccard}_{\mathrm{Index}} = \mathrm{I_j}\left(\mathrm{S_i}, \mathrm{S_j}\right) = \left|\mathrm{S_i} \cap \mathrm{S_j}\right| / \left|\mathrm{S_i} \cup \mathrm{S}_j\right| \tag{19}$$

The stability estimation was calculated using the Jaccard test formula specified in (20).

$$\sum(\mathrm{S}) = \frac{2}{\mathrm{m}(\mathrm{m}-1)} \sum_{\mathrm{i}=1}^{\mathrm{m}-1} \sum_{\mathrm{j}=\mathrm{i}+1}^{\mathrm{m}} \mathrm{I_j}\left(\mathrm{S_i}, \mathrm{S_j}\right) \tag{20}$$

In this paper, it was focused on the problem of instability encountered in feature selection algorithms. As a solution to this problem, stable feature groups were obtained by combining group-level learning and meta-dynamics of suggested approaches. For all algorithms, the fitness function was calculated with the accuracy of the k-NN classifier. The fitness function is computed using the formula given in (21).

$$\text{Fitness function} = \text{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}} \tag{21}$$

We applied the 10 fold cross-validation procedures. The average values of the 10 × 10 CV were obtained for each microarray data set. The best accuracy results for tested methods are written in bold and the worst accuracy results represented by using the asterisk symbol after the value in all tables. Table 2 shows the accuracy performances of the AIRS classifiers based on six common microarray data sets. The results in Table 2 also show the highest classification accuracy which was reached by the PAIRS2 algorithm based on Lymphoma data set with 99.3% and the obtained highest classification accuracy was 98.8% for the AIRS2 algorithm based on Colon data set. The AIRS1 and PAIRS1 algorithms reached the best accuracy performances with 91.8% and 89.2% based on Colon and Prostate data sets respectively. The comparison of accuracy performance in Table 2 shows the lowest classification accuracy that was obtained by the AIRS1 algorithm based on SRBCT dataset with the accuracy of 70.1%.

TABLE 2. Accuracy of airs based classifier methods.

Dataset	AIRS1 %	PAIRS1 %	AIRS2 %	PAIRS2 %
Colon	91.8	84.2	98.8	96.9
Lung	89.3	88.2	91.2	98.7
Prostate	86.8	89.2	97.5	95.5
SRBCT	70.1*	77.9	75.7	70.9
Lymphoma	90	86.6	97.5	**99.3**
Leukemia	86.9	85.1	96.5	98
Average	85.8	85.2*	92.8	**93.2**

TABLE 3. Accuracy of standard classifier methods.

Dataset	k-NN %	SVM %	NB %	RF %
Colon	77.4	85.4	53.2*	82.2
Lung	98.8	99.4	98.3	99.4
Prostate	85.2	88.2	62.7	90.1
SRBCT	87.3	98.4	95.2	93.6
Lymphoma	98.3	**100**	93.5	82.2
Leukemia	84.7	98.6	98.6	90.2
Average	88.6	**95**	83.5*	89.6

The average classification accuracy of AIRS2 and PAIRS2 algorithms are better than the other algorithms. The average classification accuracy of the AIRS2 algorithm is 92.8% and the average accuracy performance of the PAIRS2 algorithm is 93.2%.

The classification accuracy of the k-NN, SVM, NB, and RF classifier methods is given in Table 3. Classification results in Table 3 are obtained from WEKA by using 10 fold cross validations for classifiers. The SMO was used for the SVM classifier method. For RF classifier method the maxDepth, numFeatures, numTrees parameters are set to 0, 0, 100 respectively. It can be concluded from the results in Table 3 that SVM classifier achieved the highest average classification accuracy by 95% when compared with other classifiers. The SVM classifier achieved the highest classification accuracy based on Lymphoma data set with 100%. SVM and RF classifiers give the same results for the Lung data set by the accuracy of 99.4%. SVM and NB classifiers achieved the highest classifier results for Leukemia data set with the accuracy of 98.6%. k-NN classifier achieved the highest classifier results for Lung data set with the accuracy of 98.8%. Consequently, looking at the averaged classification results generally in all cases of data sets SVM classifier has better accuracy result than other classifiers.

TABLE 4. Classification accuracy of feature selection algorithms.

Dataset	FCBF SVM %	FCBF NB %	FCBF RF %	CGS SVM %	CGS NB %	CGS RF %	SFS SVM %	SFS NB %	SFS RF %	BES SVM %	BES NB %	BES RF %
Colon	69.2	54.8	74.1	82.2	84.6	79.8	70.7	59.2	76.9	78.8	76.9	71.1
Lung	86.1	93.3	90.6	**98.6**	97.2	96.7	93.3	95.8	95	94	94.5	95
Prostate	79.4	63.7	79.4	89.1	84.3	86	75.2	75.7	64.7	78.2	76.1	74.5
SRBCT	36.5*	53.9	52.3	98.5	96.1	97.5	60	61.5	57.6	82	41	46.1
Lymphoma	77.4	88.7	79	97.8	94.6	92.3	76.9	76.9	81.5	80.7	76.9	76.9
Leukemia	65.2	58.3	58.3	97.2	93.3	90	86.6	80.6	80.6	82.8	80.6	83.1
Average	67.2*	68.7	72.2	**93.3**	91.7	90.3	77.1	75	76	82.7	74.3	74

TABLE 5. Classification accuracy of algorithms based on DGF.

Dataset	LSTM-AIRS1 SVM %	LSTM-AIRS1 NB %	LSTM-AIRS1 RF %	LSTM-PAIRS1 SVM %	LSTM-PAIRS1 NB %	LSTM-PAIRS1 RF %	LSTM-AIRS2 SVM %	LSTM-AIRS2 NB %	LSTM-AIRS2 RF %	LSTM- PAIRS2 SVM %	LSTM- PAIRS2 NB %	LSTM- PAIRS2 RF %	GA SVM %	GA NB %	GA RF %	ANN+GA SVM %	ANN+GA NB %	ANN+GA RF %
Colon	83.2	82	79.4	79.4	76.9	84.6	84.6	96.1	92.3	88.4	84.6	76.9	82.1	75.5	73.2	87.1	82.3	85.6
Lung	92.1	96.9	92.9	**99.6**	95.6	93.7	91.6	97.2	88.8	92.8	99.1	91.6	88.8	91.6	91	91.5	89.1	91.6
Prostate	80.9	80.9	71.4	71.1	73.3	70.7	76.1	77.7	66.9	89	75	71.7	71.4	67.1	66.6	73.6	67.8	68.6
SRBCT	84.5	61.7	75	76.9	65.1	74.6	88.4	69.1	73	92.3	90.2	91	76.9	53.8*	61.5	80.1	72.3	64.7
Lymphoma	84.6	78.7	88.4	93.3	93.8	88.2	92.3	76.9	87.1	94.7	76.9	88.4	83.7	76.9	84.6	90	88.5	80
Leukemia	84.9	80	83.5	83.3	83.3	89.1	86.6	83.3	83.3	89.7	86.6	89.1	83.3	80	73.3	86	80	78.2
Average	85	80	81.7	83.9	81.3	83	86.6	83.3	81.8	**91.1**	85.4	84.7	81	74.1*	75.1	84.7	80	78.1

Table 4 summarizes the classification accuracy performance of the FCBF, CGS, SFS and BES algorithms by using microarray datasets. Each of the selected feature sets evaluated based on the SVM, NB and RF classifiers. There is a remarkable average accuracy performance difference between the CGS and FCBF algorithms. The highest classification accuracy was achieved by CGS algorithm based on Lung data set with an accuracy of 98.6% by SVM classifier. The lowest classification accuracy was obtained by FCBF algorithm based on SRBCT data set with the accuracy of 36.5% by SVM classifier. According to the average classification accuracies of feature selection algorithms in Table 4, the best average accuracy performance was achieved by CGS algorithm with the accuracy of 93.3% and the lowest accuracy performance was obtained through FCBF algorithm with the accuracy of 67.2% by based on the SVM classifier. SFS and BES algorithms achieved the highest classification accuracy based on Lung and Leukemia datasets respectively. The SFS and BES algorithms respectively reached the best average classification performances with the accuracy of 77.1% and 82.7% by SVM classifier.

Table 5 lists the classification accuracy of the algorithms based on the DGF group formation algorithm. The LSTMPAIRS1 algorithm achieved the highest classification performance by SVM classifier with an accuracy of 99.6%. The lowest classification accuracy was obtained through the GA algorithm based on SRBCT data set with the accuracy of 53.8% by NB classifier. According to the average classification accuracies of the comparative algorithms in Table 5, the best average accuracy performance was achieved through the LSTM-PAIRS2 algorithm with the accuracy of 91.1% by SVM classifier and the lowest accuracy performance was obtained through GA algorithm by based on the NB classifier with the accuracy of 74.1%.

The classification accuracy of the algorithms based on the CFG group formation algorithm was evaluated in Table 6. While the LSTM-AIRS2 algorithm given in Table 6 has the highest classification with the accuracy of 98.6% based on NB classifier, the lowest classification accuracy was obtained through GA with the accuracy of 60.2% by NB classifier. Average classifier performance shows that the highest accuracy performance was achieved through the LSTM-PAIRS2 algorithm with the accuracy of 83.3% based on SVM classifier and the lowest accuracy performance was obtained through GA with the accuracy of 72.2% based on NB classifier.

Table 7 shows the classification accuracy of the algorithms based on the IGFG group formation algorithm. ANN+GA algorithm achieved the highest performance with the accuracy of 95.8% by using SVM classifier in Lung data set. The lowest classification performance was obtained through GA by NB classifier with the accuracy of 57.8% in Prostate data set. According to the average classification accuracies of feature selection algorithms in Table 7, the best average accuracy performance was achieved through the LSTM-PAIRS2 algorithm by based on SVM classifier with the accuracy of 78.9%

TABLE 6. Classification accuracy of algorithms based on CFG.

Dataset	LSTM- AIRS1 SVM %	LSTM- AIRS1 NB %	LSTM- AIRS1 RF %	LSTM-PAIRS1 SVM %	LSTM-PAIRS1 NB %	LSTM-PAIRS1 RF %	LSTM-AIRS2 SVM %	LSTM-AIRS2 NB %	LSTM-AIRS2 RF %	LSTM- PAIRS2 SVM %	LSTM- PAIRS2 NB %	LSTM- PAIRS2 RF %	GA SVM %	GA NB %	GA RF %	ANN+GA SVM %	ANN+GA NB %	ANN+GA RF %
Colon	77.6	74.6	83.1	80.7	75.5	73.4	76.9	75.8	70.9	80.7	87.2	73	74.3	76.9	69.2	80	81.3	84.6
Lung	86.1	88.8	87.9	93.5	96.6	90	95.6	**98.6**	94.4	91.6	90.2	95.7	86.1	90	83.3	89.1	92.5	89.4
Prostate	70.7	72.1	71.4	70	72.8	75	71.4	65.9	76.1	73.5	78.9	77	61.9	60.2*	61.9	72.3	65.8	64.1
SRBCT	73.1	70	67.9	73.8	66.3	72.4	77.5	69.2	71.2	78.3	67.2	80	73.2	60.7	61.2	79.9	69.3	64
Lymphoma	78.5	72.7	84.6	78.3	76.9	86.6	86.2	78.2	80.7	87.9	77.8	82.6	76.9	72.1	80	85.1	78.1	79.8
Leukemia	83.1	84.3	85.2	82.5	87.6	88	86.6	86.6	82.2	88.3	84.8	86	83.6	73.3	80	84.4	79.8	75.4
Average	78.1	77.1	80	79.8	79.3	81	82.3	79	79.3	**83.3**	69.4	82.3	76	72.2*	72.6	81.1	77.8	76.2

TABLE 7. Classification accuracy of algorithms based on IGFG.

Dataset	LSTM - AIRS1 SVM %	LSTM - AIRS1 NB %	LSTM - AIRS1 RF %	LSTM-PAIRS1 SVM %	LSTM-PAIRS1 NB %	LSTM-PAIRS1 RF %	LSTM-AIRS2 SVM %	LSTM-AIRS2 NB %	LSTM-AIRS2 RF %	LSTM- PAIRS2 SVM %	LSTM- PAIRS2 NB %	LSTM- PAIRS2 RF %	GA SVM %	GA NB %	GA RF %	ANN+GA SVM %	ANN+GA NB %	ANN+GA RF %
Colon	76.9	69.2	65.3	69.2	71.1	63.2	69.2	61.5	69.2	68.6	71.2	67.9	63.5	61.5	60	70.1	65.7	66.8
Lung	88.8	94.5	91.6	90.7	90.8	89.8	86.1	88.8	88.8	94.6	93.3	93.7	84.3	87.3	81.3	**95.8**	94.7	90.1
Prostate	69.6	65.8	70	71.9	61.9	77	70.3	62.7	72.8	70.5	67.9	76.1	60	57.8*	59.7	70.5	63.2	63
SRBCT	60.1	59.8	62	68.4	62.1	69	63.5	60.5	73	67.8	61.2	63	83.7	59.2	60	65.1	63	61.1
Lymphoma	70	71.3	74.3	80	75.4	78.4	84.6	76.9	61.3	85.4	79.3	80.2	83.3	70	73.3	81.4	76.2	77
Leukemia	85	79	87.1	86.6	84.5	87.1	86.6	82.6	84	86.6	86.6	81.3	81.1	74.5	79.2	82.6	76	76.3
Average	75	73.3	75	77.8	74.3	77	76.7	72.1	74.8	**78.9**	76.6	77	75.9	68.3*	68.9	77.5	73.1	72.1

and the lowest average accuracy performance was obtained through GA algorithm by based on the NB classifier with the accuracy of 68.3%. These results clearly indicate that the in general for all algorithms highest average classification accuracies are achieved respectively by the DGF, CFG and IGFG group formation algorithms. According to the average classification accuracies the highest classification accuracy was achieved by the LSTM-PAIRS2 algorithm and the lowest classification accuracy was obtained via GA in terms of DGF, CFG and IGFG group formation algorithms. Generally, accuracy performances of the ANN+GA tend to be better than the GA accuracy results.

Figure 10-12 shows the stability performances of the comparative algorithms in terms of the group formation algorithms based on the six common datasets. The highest stability performance was achieved via LSTM-PAIRS1, LSTM-AIRS2, LSTM-PAIRS2 and ANN+GA algorithms based on the DGF feature groups in terms of Lung dataset. The lowest stability performance was obtained by GA based on the DGF feature groups in terms of the SRBCT datasets.

According to the stability performances of the comparative algorithms based on CFG feature group in figure 11, the highest stability performance was achieved by the LSTM-PAIRS2algorithm in terms of the Lung dataset. The lowest stability performance was obtained by GA in terms of the Prostate datasets. ANN+GA and GA algorithms have the same stability results in terms of the SRBCT dataset. LSTM-AIRS1and LSTM-PAIRS1 algorithms have higher stability performance than LSTM-AIRS2 and LSTM-PAIRS2 algorithms in terms of Lymphoma and Leukemia datasets.

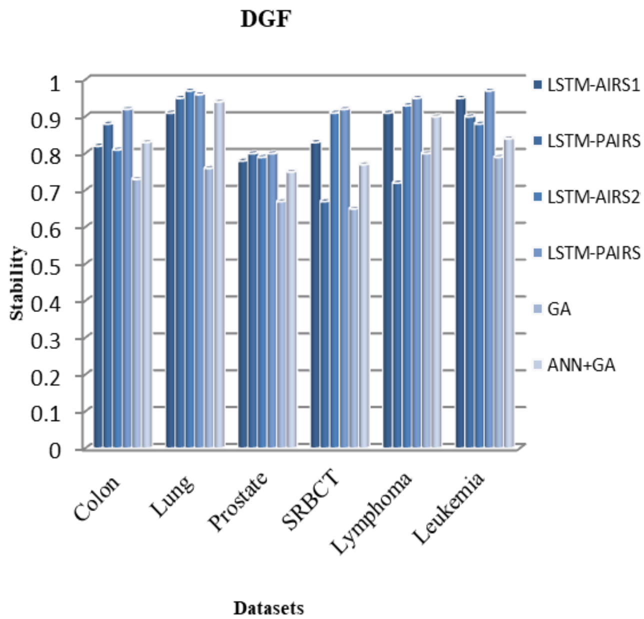

FIGURE 10. Stability performance of the comparative algorithms based on DFG.

The stability performance of the comparative algorithms based on IGFG presents the LSTM-PAIRS1 and LSTM-AIRS2 have highest stability performance in terms of the Lung datasets.

ANN+GA algorithm achieves the highest stability results based on IGFG in terms of the Lung and Lymphoma datasets.

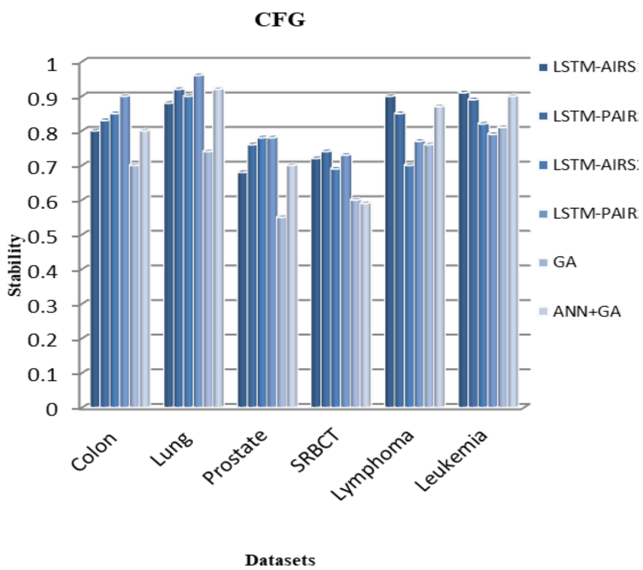

FIGURE 11. Stability performance of the comparative algorithms based on CFG.

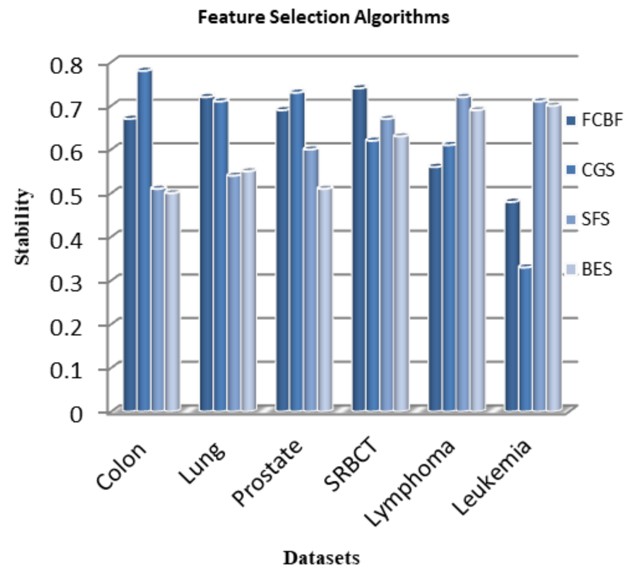

FIGURE 13. Stability performance of the feature selection algorithms.

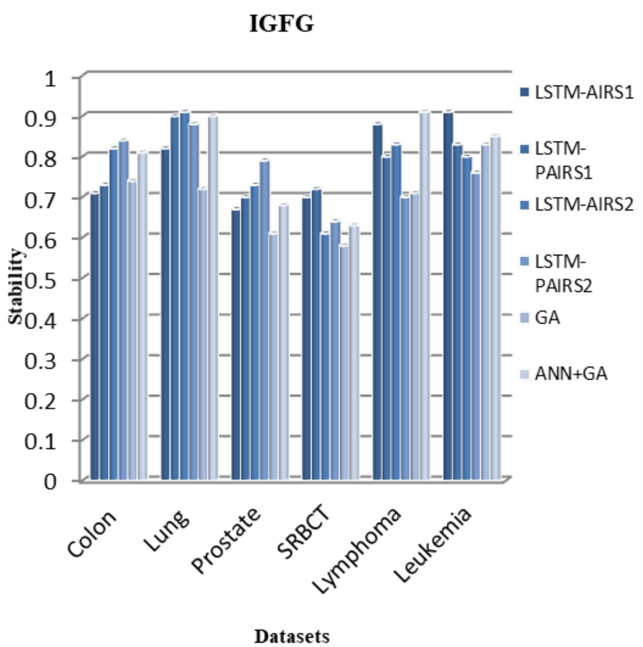

FIGURE 12. Stability performance of the comparative algorithms based on IGFG.

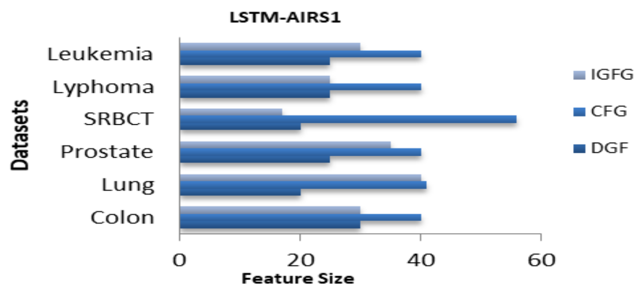

FIGURE 14. Feature size of the LSTM-AIRS1 algorithm.

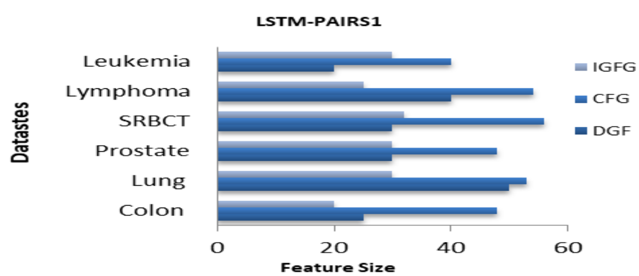

FIGURE 15. Feature size of the LSTM-PAIRS1 algorithm.

Suggested algorithms achieve the highest stability performance based on Lung dataset generally. The stability performances of ANN+GA are more stable than the GA algorithm for DGF, CFG and IGFG feature groups. It can be inferred that the initially selected genes based on the group formation algorithm affect the stability results.

The stability performances of the feature selection algorithms in figure 13 show that the highest stability results were achieved through CGS and FCBF algorithms based on Colon, Lung, and Prostate data sets. The SFS and BES algorithms achieved the highest stability results in terms of Lymphoma, Leukemia and SRBCT datasets respectively.

Figure 14-20 shows the feature size of the comparative algorithms. The LSTM-AIRS1 algorithm has results close to each other based on the DGF and IGFG feature groups in terms of Colon, SRBCT and Lymphoma datasets and for Lung and Prostate data sets based on CFG and IGFG feature groups.

In LSTM-PAIRS1 algorithm, the highest feature size was obtained based on the CFG and lowest feature size was achieved based on IGFG in general. In the LSTM-AIRS2 algorithm, the highest feature size was obtained by based on the CFG in terms of Colon, Lymphoma and Leukemia datasets, and the lowest feature size was achieved

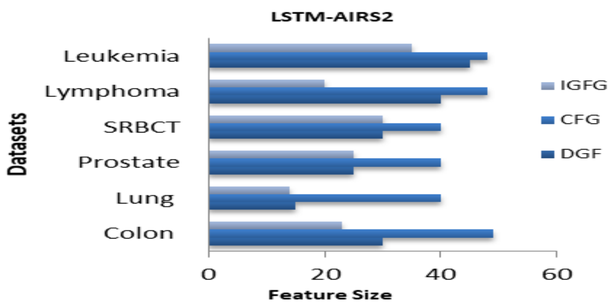

FIGURE 16. Feature size of the LSTM-AIRS2 algorithm.

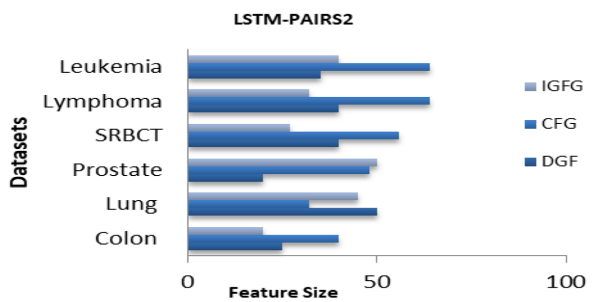

FIGURE 17. Feature size of the LSTM-PAIRS2 algorithm.

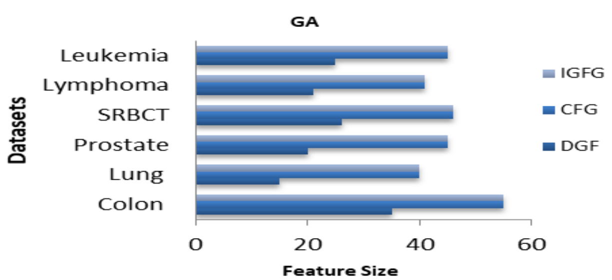

FIGURE 18. Feature size of the GA algorithm.

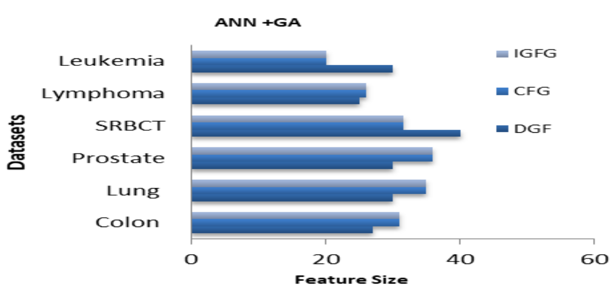

FIGURE 19. Feature size of the ANN+GA algorithm.

based on DGF and IGFG in terms of Lung data set. The Lymphoma and Leukemia data sets had the highest feature size in the LSTM-PAIRS2 algorithm based on CFG feature group. The lowest feature size was achieved based on the IGFG feature group in terms of the Colon data set. The feature size of the ANN+GA algorithm was smaller than the GA except Lung and Prostate datasets based on DGF. GA generally has the highest feature size by based on IGFG.

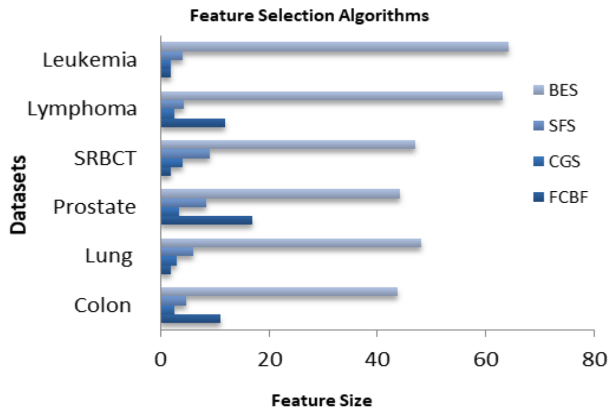

FIGURE 20. Feature size of the feature selection algorithms.

ANN+GA have the highest feature size based on CFG like the suggested algorithms. Figure 20 represents the feature size of the feature selection algorithms. The average feature size of the CGS, FCBF, SFS and BES algorithms are 2.9, 7.6, 6.1 and 51.6, respectively. CGS selects the minimum average feature size that is shown in figure 20.

VIII. CONCLUSIONS

The main motivation of this work is to create a framework in order to model a stability mechanism for robust feature selection. The development of stable associative immune memory is modeled based on the LSTM memory blocks. LSTM recurrent neural networks are trained with the AIRS in order to be effective for sequential learning problems. The aim of this study was to better understand the immunological principles and mechanisms of the biological processes of AIRS. This paper presented a novel sequence model with an LSTM embedded into AIRS to solve the problem of sequence learning. A general framework was used to combine the gene selection framework with the evolution of immune memory using LSTM recurrent neural networks. This study also has an analysis of using a different type of feature groups based on the different feature selection algorithms. The suggested algorithms achieved remarkable increases in their robustness and classification accuracy. In conclusion, LSTM based AIRS is a robust and effective method for microarray analysis.

Future studies may explore the issue of building model for the associative memory based on different type of Recurrent Neural Networks (RNNs). For example, Gated Recurrent Unit (GRU), could be used which may further improve the solutions generated by LSTM. Since the GRU has a less complex structure than the LSTM, it is computationally more efficient and has less execution time.

REFERENCES

[1] A. Ridok, W. F. Mahmudy, and M. Rifai, "An improved artificial immune recognition system with fast correlation based filter (FCBF) for feature selection," in *Proc. 4th Int. Conf. Image Inf. Process.*, Dec. 2017, pp. 1–6.
[2] D. Dasgupta and F. Nino, *Immunological Computation: Theory and Applications*. London, U.K.: Taylor & Francis, 2009.

[3] D. Grzegorz, "An artificial immune system for classification with local feature selection," *IEEE Trans. Evol. Comput.*, vol. 16, no. 6, pp. 847–860, Dec. 2012.
[4] F. Ahmad, N. A. Mat-Isa, Z. Hussain, R. Boudville, and M. K. Osman, "Genetic algorithm-artificial neural network (GA-ANN) hybrid intelligence for cancer diagnosis," in *Proc. 2nd Int. Conf. Comput. Intell., Commun. Syst. Netw.*, Jul. 2010, pp. 78–83. doi: 10.1109/CICSyN.2010.46.
[5] J. Brownlee, "Artificial immune recognition system (AIRS) a review and analysis," Tech. Rep., 2005. doi: 10.1.1.67.9753.
[6] J. Brownlee, "Clonal selection theory & clonalg the clonal selection classification algorithm (CSCA)," Tech. Rep., 2005.
[7] K.-J. Wang, K.-H. Chen, and M.-A. Angelia, "An improved artificial immune recognition system with the opposite sign test for feature selection," *Knowl.-Based Syst.*, vol. 71, pp. 126–145, Nov. 2014.
[8] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, Washington, DC, USA, vol. 2, 2003, pp. 856–863.
[9] M. Farhan, M. Mohsin, A. R. Hamdan, and A. A. Bakar, "An evaluation of feature selection technique for dendrite cell algorithm," in *Proc. Int. Conf. IT Converg. Secur. (ICITCS)*, Oct. 2014, pp. 1–5.
[10] M. García-Torres, F. Gómez-Vela, B. Melián-Batista, and J. M. Moreno-Vega, "High-dimensional feature selection via feature grouping: A variable neighborhood search approach," *Inf. Sci.*, vol. 326, pp. 102–118, Jan. 2016. doi: 10.1016/j.ins.2015.07.041.
[11] M. J. Gangeh, H. Zarkoob, and A. Ghodsi, "Fast and scalable feature selection for gene expression data using Hilbert-Schmidt independence criterion," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 167–181, Jan./Feb. 2017.
[12] M. Rußvurm and M. Körner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1496–1504.
[13] M. Ghosh *et al.*, "Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 159–176, 2019. doi: 10.1007/s115177-018-1874-4.
[14] N. C. de Leandro and T. Jonathan, *Artificial Immune Systems: A New Computational Intelligence Approach*. London, U.K.: Springer, 2002.
[15] R. Feldman and E. S. Kim, "Prognostic and predictive biomarkers post curative intent therapy," *Ann. Transl. Med.*, vol. 5, no. 18, p. 374, Sep. 2017.
[16] L. Yu, C. Ding, and S. Loscalzo, "Stable feature selection via dense feature groups," in *Proc. 14th ACM Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2008, pp. 803–811. doi: 10.1145/1401890.1401986.
[17] S. Loscalzo, L. Yu, and C. Ding, "Consensus group stable feature selection," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Paris, France, Jun./Jul. 2009, pp. 567–576.
[18] S. Vijendra and S. Laxman, "Subspace clustering of high-dimensional data: An evolutionary approach," *Appl. Comput. Intell. Soft Comput.*, vol. 2013, Nov. 2013, Art. no. 863146. doi: 10.1155/2013/863146.
[19] W. Zhong, X. Lu, and J. Wu, "Feature selection for cancer classification using microarray gene expression data," Tech. Rep., 2017. doi: 10.19080/BB0AJ.2017.01.555557.
[20] Y. Jiang and X. Zheng, "LSTM-scheduler project on github," Tech. Rep., 2017.
[21] Z. He and W. Yu, "Stable feature selection for biomarker discovery," *Comput. Biol. Chem.*, vol. 34, pp. 215–225, Aug. 2010. doi: 10.1016/j.compbiolchem.2010.07.002.
[22] Z. Zareizadeh, M. S. Helfrous, A. Rahideh, and K. Kazemi, "A robust gene clustering algorithm based on clonal selection in multiobjective optimization framework," *Expert Syst. Appl.*, vol. 113, pp. 301–314, Dec. 2018. doi: 10.1016/j.eswa.2018.06.047.
[23] Z. Ahmed and S. Zeeshan, "Applying WEKA towards machine learning with genetic algorithm and back-propagation neural networks," *J. Data Mining Genomics Proteomics*, vol. 5, no. 2, p. 1, 2014. doi: 10.4172/2153-0602.1000157.

CANAN BATUR ŞAHİN received the Diploma and master's degrees in computer engineering from Istanbul Technical University. She is currently a Research Assistant with Siirt University. Her research interests include optimization, artificial intelligence, and machine learning.

BANU DİRİ is a Professor with Yildiz Technical University and she works on Natural Language Processing. She authored more than 150 publications in this field. Her research interests include speech recognition, natural language processing, and machine learning.