IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2395~2402
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2395-2402

Feature selection techniques for microarray dataset: a review

Avinash Nagaraja Rao[1], Sitesh Kumar Sinha[1], Shivamurthaiah Mallaiah[2]
[1]Department of Computer Science and Engineering, Rabindranath Tagore University, Bhopal, India
[2]Department of Computer Science, APS College of Engineering, Bengaluru, Bangalore, India

Article Info

Article history:

Received Jun 30, 2023
Revised Oct 12, 2023
Accepted Dec 2, 2023

Keywords:

Correlation
Embedded
Feature selection
Filter
Microarray dataset
Wrapper

ABSTRACT

Automatic speech recognition (ASR) approach is dependent on optimal for many researchers working on feature selection (FS) techniques, finding an appropriate feature from the microarray dataset has turned into a bottleneck. Researchers often create FS approaches and algorithms with the goal of improving accuracy in microarray datasets. The main goal of this study is to present a variety of contemporary FS techniques, such as filter, wrapper, and embedded methods proposed for microarray datasets to work on multi-class classification problems and different approaches to enhance the performance of learning algorithms, to address the imbalance issue in the data set, and to support research efforts on microarray dataset. This study is based on critical review questions (CRQ) constructed using feature election methods described in the review methodology and applied to a microarray dataset. We discussed the analysed findings and future prospects of FS strategies for multi-class classification issues using microarray datasets, as well as prospective ways to speed up computing environment.

Corresponding Author:

Avinash Nagaraja Rao
Department of Computer Science and Engineering, Rabindranath Tagore University
Bhopal, India
Email: avi003@gmail.com

1. INTRODUCTION

Feature is an important element in the classification problem addressed in machine learning (ML) techniques. Selecting the best feature or relevant feature will provide better accuracy in optimization algorithm [1]. Microarray dataset contains minimum number of samples and maximum number of features. Conventionally, most of the microarray datasets include more than 60,000 characteristics or attributes with fewer samples, not surpassing 100 [2]. Due to this scenario important or relevant features are identified using feature selection (FS) techniques viz., filter, wrapper, and embedded methods. This leads to significant reduction in obtaining the optimal results with classification accuracy and computational time [3]. The primary intention of this review is: i) to provide the various FS algorithms to find the best accuracy in microarray dataset analysis, ii) to introduce the various ML techniques involved in the microarray dataset, iii) to examine the publication trends in the domain related to FS using microarray dataset, iv) to consolidate the various research work in different FS techniques by various authors and experts, and v) to provide future scope of research in real time microarray dataset.

Optimum results are obtained in the microarray dataset using appropriate FS techniques. The conventional FS techniques like filter method, wrapper method, embedded method, and hybrid methods are favoured for the subset selection to obtain optimal results. Figure 1 depicts the standard procedure of a FS process. The first step is to use a reliable search strategy to create a subset of the microarray dataset. The second stage involves evaluating the list of subgroups and contrasting the best subset with the one that came before it.

If the newly updated subset is highly suggested than the previous one, then the change is left unmodified. Until the termination condition is met, the procedure will be repeated. Then finally, the best subset score is selected and given to the next classification technique. Selecting a feature subset is achieved using the Algorithm 1 [4].

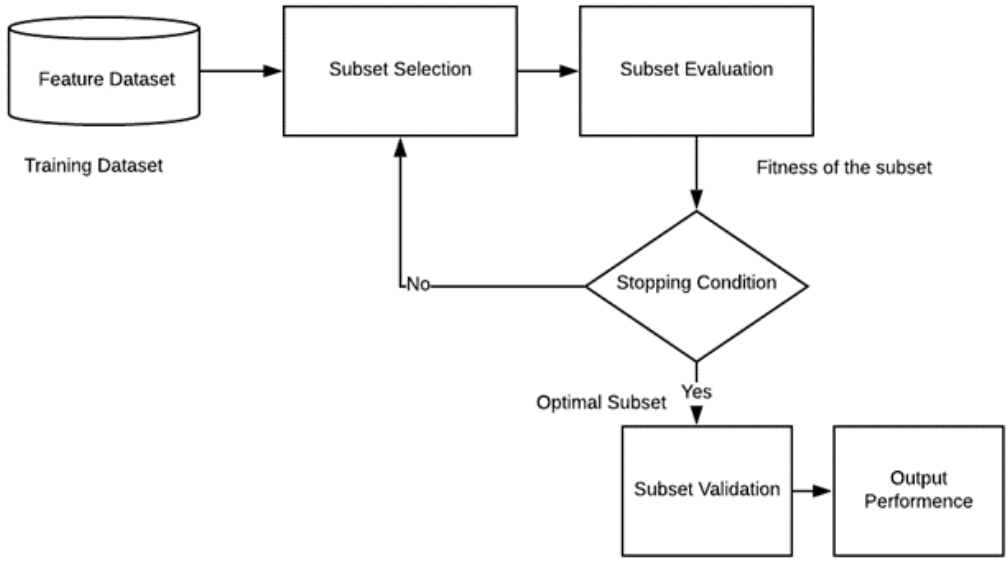

Figure 1. General FS process

Figure 2 shows the complete work flow of filter, wrapper, and embedded model, which finds the best subset in all the features available in the dataset. Both the wrapper and embedded methods have the stopping condition to validate the best subset [5]. Filter method does not involve with particular learning algorithm and validation of best subset whereas wrapper method selects the features using the learning algorithm and validating the optimal feature subset [6]. In embedded methods combines the both filter and wrapper method. As per the study we have represented inferences and future scope of each method in the Table 1.

Algorithm 1: Selecting feature subset
Input: large future microarray dataset
Output: Obtain the best subset
Step 1: Starts with subset generation
Step 2: Subset evaluation
Step 3: Stopping condition
Step 4: Validating the feature subset
if condition satisfies, test n number of times
Select the best feature set
else
Repeat the step2 and step3
Step 5: End process

Table 1. Inference and future scope of FS methods

Method	Inference	Future scope
Filter	As per the review with the microarray dataset, majority of researchers (about 40%) used filter method to obtain the optimal classification performance 2015.	It is observed that the researchers can work on research problems like high computational cost and runtime performance.
Wrapper	As per the review 20% of the researchers have experimented with the wrapper methods using well known population methods to achieve better classification accuracy.	Multi-objective optimization is to be experimented with the various microarray datasets and results should be compared with other algorithms.
Embedded	As per the review 40% of the researchers have used the embedded FS technique in their microarray dataset	

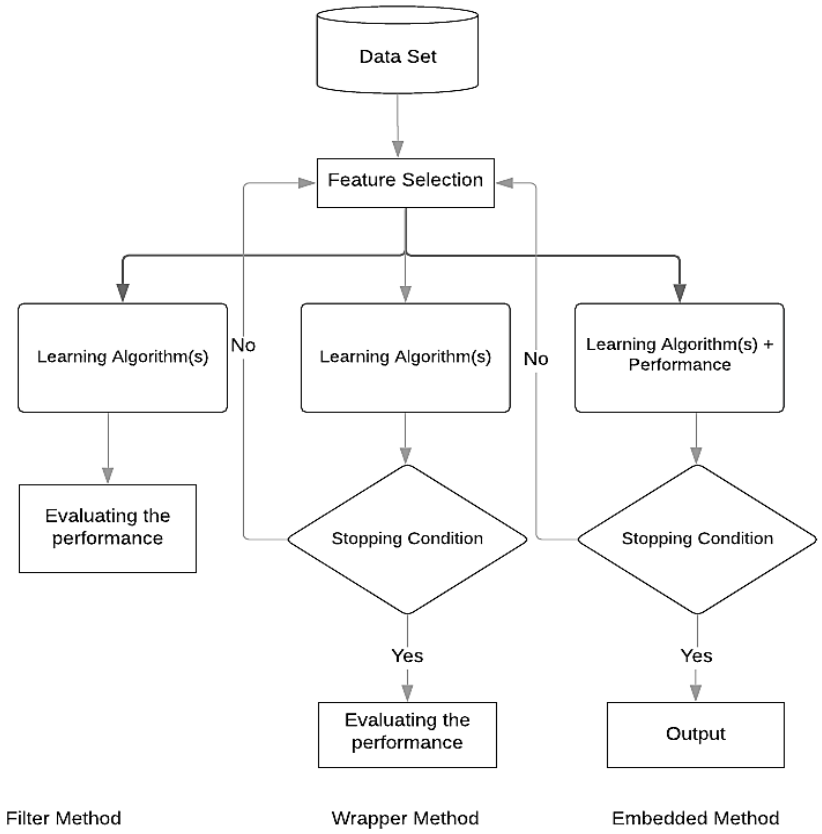

Figure 2. Basic flow of filter, wrapper, embedded methods

2. LITERATURE REVIEW

This section dives into two key aspects of analyzing microarray data: FS techniques and the overall review technology roadmap. We'll explore the fundamental methods for selecting the most informative genes from these VAST datasets, followed by a roadmap outlining the steps involved in effectively reviewing and analyzing this technology.

2.1. Filter method

The filter method checks whether the particular feature is relevant or not in the microarray dataset apart from calculating the error rate. Initially, feature/relevance score is calculated and then it is compared with the remaining columns in the dataset, then it finalizes the low feature/relevance scoring subset to be ignored form the microarray dataset. The optimal subset features are fed into the classification technique for the next process. The filter methods can be classified into two types, namely univariate and multivariate. The univariate evaluates the features separately and the multivariate filters evaluate features concerning the relationship between the two features [7]. Feature scoring is calculated using the following methods like pearson correlation, chi-squared test, entropy, Fisher's score, Spearman correlation, Kendall correlation, analysis of variance (ANOVA), count based, linear discriminant analysis (LDA), relief, information gains fast correlation-based filter and maximal relevance and minimal redundancy (mRMR) [8].

2.2. Wrapper method

In the wrapper method, subset features in the dataset are evaluated independently. The evaluation criteria of the specific subset are achieved by using the learning algorithm(s). Due to the exponential growth of the subset, the number of features will increase in this scenario. The learning algorithms are used to find the best fitness of the subset [9]. The wrapper model algorithm gives more understanding of the selection of the best subset in the features. In the wrapper model, the search methods are classified into two major categories, namely sequential and randomized. The sequential search methods consist following algorithms like sequential forward selection (SFS), sequential backward selection (SBS), plus q take away r and beam search. The second method is a population-based algorithm or sequential search [10] that finds better optimum fitness of the subset, using techniques such as genetic algorithm, particle swarm algorithm, advanced binary ant colony

optimization, harmony [11], differential evolution, whale optimization, artificial bee colony, and bacterial colony optimization [12].

2.3. Embedded method

The embedded FS technique is a classifier dependent FS method [13]. The learning algorithm is playing a vital role in the embedded method. The researchers prefer the embedded method commonly due to the low computational cost. Irrelevant features are removed using widely used techniques like weight vector of support vector machine (SVM), decision tree, weighted Naive Bayes (NB). The following embedded methods, namely first-order inductive learner (FOIL) feature subset selection algorithm, probably approximately correct (PAC) Bayes, kernel-penalized support vector machine (KP-SVM), least absolute shrinkage and selection operator (LASSO).

The committed writing uncovers an assortment of approaches managing with critical resource questions (CRQ) issues within this section details about review methodology employed throughout this paper. As per the intension framed for this study along with CRQ is framed and the road map of the review methodology is represented in Table 2. In the initial stage we have framed the CRQ based FS techniques used in microarray dataset. While in the second stage we have selected 41 manuscripts related to the topic in recent years. In the next stage FS techniques used in these articles are analyzed. Then we have focused on the classification accuracy using various ML techniques. The final stage of the review methodology is to provide the future directions to the researchers, from the insights. The review is based on the key questions which help to form a study of this paper is represented in Table 2.

Table 2. Research questions

ID	Question	Motivation
CRQ1	What are the recent publications that interpret the outcome and feature scopes to the benefit of researchers?	To examine the publication trends in the domain related to FS using microarray dataset.
CRQ2	What are the different FS techniques used in microarray dataset?	To provide the various FS algorithms to find the best accuracy in microarray dataset analysis
CRQ3	What are the common microarray datasets used in recent years?	To provide the frequently used microarray dataset in recent years.
CRQ4	What tools are used to implement the FS techniques when we select high dimensional dataset?	To explore the tools used by the researchers in high dimensional dataset.
CRQ5	What is the future scope of research in real time microarray dataset?	To provide the future scope to the researchers based on this study.

2.4. Literature resources

This section is to identify the candidate papers which can helps to answer the CRQs represented in the Table 2. Research articles are collected from different online sources like Scopus, association for computing machinery (ACM) digital library, and PubMed. The selection of the research paper is done using search selection using keywords, tittle, and abstract. Another criterion for selection is we have year constrains starts from 2014 to 2022.

2.5. Search process and study selection

The automatic search process is done with different online sources with the two criteria explained. The potential research papers are identified and collected for the study. The frequent checks are done with online sources and confirmed that there are relevant contents are collected for our study. Based on those content we have formed two criteria named inclusion criteria and exclusion criteria for our study.

2.6. Inclusion criteria

Inclusion criteria are characteristics that are to be included in the study. In our study we have included following characterstics. i) research articles which address the microarray dataset (mandatory), ii) research articles using FS techniques, iii) research articles performing comparison of various FS techniques on different microarray dataset, and iv) research articles which are published recent years.

2.7. Exclusion criteria

Exclusion criteria comprise characteristics used to identify potential research participants who should not be included in a study. In our study following are comes under exclusion criteria: i) research articles which are not written in English language and ii) research articles which contain less FS techniques. The mandatory criterion helps us to select the papers and if any one of the exclusion criteria is satisfied means we have rejected the papers for our study. Initially the tittle, keywords, and abstracts are examined and if it is required further study is done on the full article.

2.8. Data extraction

After identifying the relevant research articles which satisfied our CRQs information's are extracted from the selected papers. We have created a checklist depends on our CRQs represented in Table 3. The aim of the check list is to extract sufficient information on the CRQs which is maintained in the separate excel sheet. We have extracted major contents to our study in this section. The second phase is cross validation between the authors. We are acknowledged that all the selected papers are not met the five CRQs.

2.9. Recent micro arraydatasets

Datasets plays a vital role in any ML applications [14]. The curse of dimensionality is the primary research issue, faced by researchers while dealing with a microarray dataset [15]. Due to the high number of samples, it is hard to find the optimal results with a high dimensional dataset. Generally, microarray datasets are grouped into two different categories named binary classification and multiclass classification. Many researchers have experimented in binary class classification [16].

2.10 Challenges

The following obersrvation are the challenges faced during the exection features and dataset: i) the computational time is high due to a greater number of features in the gene expression with the noise, and ii) the unbalanced dataset will affect the training and test dataset so that the accuracy will be another issue. Table 4 (see an Appendix) explains review methodology, for this different journals were refered from each different methods were proposed for FS and accuracy is measured from each method and feature enhanchements that can be done from each were dissussed.

Table 3. Check list form

CRQ	Validation
1	What are the recent publications that interpret the outcome and feature scopes to the benefit of researchers? – Year – Source name – Feature scope
2	What are the different FS techniques used in microarray dataset? – Filter – Wrapper – Embedded – Hybrid
3	What are the common microarray datasets used in recent years? – Dataset name – Binary classification 1 0 – Multi class classification
4	What tools are used to implement the FS techniques when we select high dimensional dataset? – Name of the tools used to perform FS
5	What is the future scope of research in real time microarray dataset? – Future scope to implement in real time dataset

3. RESEARCH PERSPECTIVE AFTER LITREATURE REVIEW

3.1. Stability

When developing FS approach for high-dimensional datasets, stability is of paramount importance. If the FS method yields the same result independent of changes to the input data, it is considered resilient [26]. If the stability issue of the FS method is ignored, erroneous inference and inaccurate results may result. One of the most glaring causes of unpredictability is the neglect of features that were selected because they were aligned with the dependent variables. Stability facilitates the trade-off between bias and volatility in the classification error rate [27]. The dimensionality of the dataset, the number of features selected, the size of the sample, and the unpredictability of the data all have a role in the FS algorithm's stability [28].

3.2. Choosing appropriate measures of fitness

Wrapper FS techniques seek to optimise a feature subset by maximising an objective function. The classification problem influences the development of the FS goal function. At first, an objective function was developed to maximise classification success while minimising the number of features. Feature counts and classification accuracy are merged into a single fitness function in a single-objective technique [29].

3.3. Selection of classifiers

Various classifiers have been used to address FS issues, including k-nearest neighbour (KNN), NB, SVM, random forest (RF), and artificial neural networks (ANN). The most crucial step in optimising results from a high-dimensional dataset is selecting a classifier to use. The KNN was found to be the most popular classifier in the reviewed literature, and in the classification process, SVM is essential [30].

3.4. Reduced size of sample

Numerous studies have sought to address the widespread problem of limited data from individual microarray experiments. One primary worry is that learning methods suffer significantly when working with little data. The right validation strategy for calculating the misclassification rate is essential for fixing this problem [31].

3.5. Disparity in class

An unfavourable learning environment is created when one class dominates the dataset at the expense of the others. Unbalanced microarray datasets [32] are common, and multiclass microarray datasets are a prime example. When the test set's imbalance is more pronounced than in the training sets, solving this issue becomes difficult. To get around this issue, several researchers use preprocessing techniques like under sampling and oversampling. Ensemble classifiers have recently been proposed as a way to address class imbalance [33].

3.6. Outliers

One of the most important yet under-discussed aspects of microarray data is the identification of outliers. Polluted database samples are known as outliers, and they occur when instruments or humans make mistakes during data collection or analysis [34]. The learning process is hindered by outliers because they prevent the useful genes from being chosen.

4. DISCUSSION AND FUTURE DIRECTIONS

The examination of microarray data provides important clues that facilitate the identification of new diseases. High complexity of gene expression data and small sample size make categorization a difficult process. Therefore, FS strategy is the most workable answer to these issues. This research meticulously synthesises the approaches, techniques, datasets, and future possibilities of the microarray dataset during the last three years. Reviewing the literature critically, this article examines a wide range of topics, including multiclass classification, methods for enhancing learning algorithm performance, resolving the issue of data set imbalance, and verifying the results of previous work using microarray datasets. This work provides a comprehensive examination of FS procedures in the context of search strategies, offering insights into the current techniques used for FS on microarray datasets. This paper presents a complete literature evaluation of five distinct kinds of FS strategies, including filter, wrapper, and embedding methods. This paper examines the challenges and research considerations that arise during the development of a FS method.

APPENDIX

Table 4. Review methodology

S.no	Ref	Proposed method	Evaluation measures	Findings or future enhancements
1	[17]	Improved version of salp swarm algorithm (SSA) with transfer function and cross over scheme to overcome the issues of FS.	Accuracy and fitness value	The converging ability is not investigated. In addition, S-Shaped and V-shaped transfer functions are to be explored in future enhancement.
2	[18]	Hybrid variant of Harris Hawk optimization (HHO) based on bitwise and simulated annealing (SA) to solve the FS problem.	Accuracy and fitness value	High dimensional datasets are not investigated.
3	[19]	Improved SSA with inertia weight parameter to enhance both exploration and exploitation capabilities.	Classification accuracy and feature reduction.	The converging ability of the model is not tested.
4	[20]	Binary equilibrium optimization combined with SA for FS problems.	Accuracy and fitness value	High dimensional datasets are not investigated.
5	[21]	A novel single objective optimization based on chaotic theory for FS problems.	Accuracy and fitness value	The medical data is not experimented in the proposed system.
6	[21]	Opposition-based self-adaptive cohort intelligence (OSCI), optimal features have been selected.	Accuracy and fitness value	High dimensional datasets are not investigated.
7	[22]	Dynamic version of SSA to solve the local optima problem to balance both exploration and exploitation.	Accuracy and fitness value	The updating rules can be modified based on the problem statement.
8	[23]	Hybrid version of SSA and simulated annealing to improve the exploitation capability.	Accuracy and number of features are tested	The model has been implemented for multi objective optimization.
9	[24]	Binary version of SSA with three updating rules to solve the FS problems.	Accuracy, fitness value, and converging ability.	The exploration search space is improved. Furthermore, updating rules can be modified based on the problem statement.
10	[25]	Binary version of SSA with random weight network (RWN) to solve the FS problem.	Classification accuracy	The investigation is performed on binary classification problems. The exploration and exploitation are not considered for evaluating the model.

REFERENCES

[1] M. A. Hambali, T. O. Oladele, and K. S. Adewole, "Microarray cancer feature selection: review, challenges and research directions," *International Journal of Cognitive Computing in Engineering*, vol. 1, pp. 78–97, 2020, doi: 10.1016/j.ijcce.2020.11.001.

[2] V. Bolón-Canedo and B. Remeseiro, "Feature selection in image analysis: a survey," *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2905–2931, 2020, doi: 10.1007/s10462-019-09750-3.

[3] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, pp. 1–26, 2020, doi: 10.1186/s40537-020-00327-4.

[4] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, pp. 1–9, 2019, doi: 10.1016/j.compbiomed.2019.103375.

[5] S. Shadravan, H. R. Naji, and V. K. Bardsiri, "The sailfish optimizer: a novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 20–34, 2019, doi: 10.1016/j.engappai.2019.01.001.

[6] K. Tadist, S. Najah, N. S. Nikolov, F. Mrabti, and A. Zahi, "Feature selection methods and genomic big data: a systematic review," *Journal of Big Data*, vol. 6, no. 1, pp. 1–24, 2019, doi: 10.1186/s40537-019-0241-0.

[7] T. Saw and P. Hnin, "Swarm intelligence based feature selection for high dimensional classification: a literature survey," *International Journal of Computer (IJC)*, vol. 33, no. 1, pp. 69–83, 2019.

[8] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007, doi: 10.1093/bioinformatics/btm344.

[9] A. Mangal and E. A. Holm, "A comparative study of feature selection methods for stress hotspot classification in materials," *Integrating Materials and Manufacturing Innovation*, vol. 7, no. 3, pp. 87–95, 2018, doi: 10.1007/s40192-018-0109-8.

[10] R. R. Rani and D. Ramyachitra, "Microarray cancer gene feature selection using spider monkey optimization algorithm and cancer classification using SVM," *Procedia Computer Science*, vol. 143, pp. 108–116, 2018, doi: 10.1016/j.procs.2018.10.358.

[11] Y. Prasad, K. K. Biswas, and M. Hanmandlu, "A recursive PSO scheme for gene selection in microarray data," *Applied Soft Computing Journal*, vol. 71, pp. 213–225, 2018, doi: 10.1016/j.asoc.2018.06.019.

[12] B. Sahu, S. Dehuri, and A. K. Jagadev, "Feature selection model based on clustering and ranking in pipeline for microarray data," *Informatics in Medicine Unlocked*, vol. 9, pp. 107–122, 2017, doi: 10.1016/j.imu.2017.07.004.

[13] M. K. Ebrahimpour, H. Nezamabadi-pour, and M. Eftekhari, "CCFS: a cooperating coevolution technique for large scale feature selection on microarray datasets," *Computational Biology and Chemistry*, vol. 73, pp. 171–178, 2018, doi: 10.1016/j.compbiolchem.2018.02.006.

[14] C. Arunkumar and S. Ramakrishnan, "Attribute selection using fuzzy rough set based customized similarity measure for lung cancer microarray gene expression data," *Future Computing and Informatics Journal*, no. 3, pp. 131–142, 2018.

[15] H. Dong, T. Li, R. Ding, and J. Sun, "A novel hybrid genetic algorithm with granular information for feature selection and optimization," *Applied Soft Computing Journal*, vol. 65, pp. 33–46, 2018, doi: 10.1016/j.asoc.2017.12.048.

[16] S. Maldonado, R. Weber, and F. Famili, "Feature selection for high-dimensional class-imbalanced data sets using support vector machines," *Information Sciences*, vol. 286, pp. 228–246, 2014, doi: 10.1016/j.ins.2014.07.015.

[17] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, 2018, doi: 10.1016/j.jbi.2018.07.014.

[18] E. H. Houssein, M. E. Hosney, M. Elhoseny, D. Oliva, W. M. Mohamed, and M. Hassaballah, "Hybrid Harris hawks optimization with cuckoo search for drug design and discovery in chemoinformatics," *Scientific Reports*, vol. 10, no. 1, pp. 1–22, 2020, doi: 10.1038/s41598-020-71502-z.

[19] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 3, pp. 335–344, 2020, doi: 10.1016/j.jksuci.2018.06.003.

[20] Y. Gao, Y. Zhou, and Q. Luo, "An efficient binary equilibrium optimizer algorithm for feature selection," *IEEE Access*, vol. 8, pp. 140936–140963, 2020, doi: 10.1109/ACCESS.2020.3013617.

[21] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Feature selection using chaotic salp swarm algorithm for data classification," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3801–3816, 2019, doi: 10.1007/s13369-018-3680-6.

[22] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, no. 11, pp. 341–359, 1997.

[23] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[24] W. Gao, S. Liu, and L. Huang, "A global best artificial bee colony algorithm for global optimization," *Journal of Computational and Applied Mathematics*, vol. 236, no. 11, pp. 2741–2753, 2012, doi: 10.1016/j.cam.2012.01.013.

[25] S. H. Bouazza, K. Auhmani, A. Zeroual, and N. Hamdi, "Selecting significant marker genes from microarray data by filter approach for cancer diagnosis," *Procedia Computer Science*, vol. 127, pp. 300–309, 2018, doi: 10.1016/j.procs.2018.01.126.

[26] A. K. Shukla, P. Singh, and M. Vardhan, "A hybrid gene selection method for microarray recognition," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 975–991, 2018, doi: 10.1016/j.bbe.2018.08.004.

[27] M. J. Rani and D. Devaraj, "Two-stage hybrid gene selection using mutual information and genetic algorithm for cancer data classification," *Journal of Medical Systems*, vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1372-8.

[28] S. Maldonado and J. López, "Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification," *Applied Soft Computing Journal*, vol. 67, pp. 94–105, 2018, doi: 10.1016/j.asoc.2018.02.051.

[29] J. R. Anaraki and H. Usefi, "A comparative study of feature selection methods on genomic datasets," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2019, pp. 471–476, doi: 10.1109/CBMS.2019.00097.

[30] R. Aziz, C. K. Verma, and N. Srivastava, "Dimension reduction methods for microarray data: a review," *AIMS Bioengineering*, vol. 4, no. 1, pp. 179–197, 2017, doi: 10.3934/bioeng.2017.1.179.

[31] K. Balakrishnan, R. Dhanalakshmi, and U. M. Khaire, "Improved salp swarm algorithm based on the levy flight for feature selection," *Journal of Supercomputing*, vol. 77, no. 11, pp. 12399–12419, 2021, doi: 10.1007/s11227-021-03773-w.

[32] M. Liu, X. Yao, and Y. Li, "Hybrid whale optimization algorithm enhanced with Lévy flight and differential evolution for job shop scheduling problems," *Applied Soft Computing Journal*, vol. 87, 2020, doi: 10.1016/j.asoc.2019.105954.

[33] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: homogeneous and heterogeneous approaches," *Knowledge-Based Systems*, vol. 118, pp. 124–139, 2017, doi: 10.1016/j.knosys.2016.11.017.

[34] F. Yang and K. Z. Mao, "Robust feature selection for microarray data based on multicriterion fusion," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 4, pp. 1080–1092, 2011, doi: 10.1109/TCBB.2010.103.

BIIOGRAPHIES OF AUTHORS

Mr. Avinash Nagaraja Rao is a research scholar in the Department of Computer Science and Engineering, RNTU, Bhopal. He has obtained his M.Tech. degree in Computer Science and Engineering from VTU. He has 13 years of teaching experience and published more than 10 research papers in reputed international journals. His main research work focuses on machine learning, deep learning, network security, image processing, and data mining. He can be contacted at email: avi003@gmail.com.

Dr. Sitesh Kumar Sinha completed his Ph.D. from degree from BRAB MIT, Muzaffarpur, Bihar. Currently, he working as registrar (administration) Dr C. V. Raman University Vaishali Bihar. He is completed government funding project during his Ph.D. work related on computer network. He is published more than 40 research paper in various international and national journals. His main research work focuses on network security, image processing, and software engineering. He can be contacted at email: siteshkumarsinha@gmail.com.

Dr. Shivamurthaiah M completed his Ph.D. from RNTU and currently working as professor in Department of CSE at Dr. S. M. College of Engineering, Bengaluru. He has 17 years of teaching and 9 years of research experience and published more than 30 research papers in reputed international journals. He has published 9 patents in that 2 patent are granted. His main research work focuses on cryptography algorithms, network security, image processing and software engineering data mining, IoT, and computational intelligence-based education. He can be contacted at email: shivamurthaiah@gmail.com.