Evaluating Machine Learning Models for Prostate Cancer Classification Using Gene Expression Profiles from DNA Microarrays

Sara Haddou Bouazza*, *Jihad* Haddou Bouazza

LAMIGEP, EMSI-Marrakech, Morocco

Abstract. This study evaluates various machine learning models for classifying prostate cancer using gene expression profiles from DNA microarrays. Due to the high dimensionality of these datasets, effective dimensionality reduction through feature selection is essential to identify and remove redundant genes. We applied multiple feature selection methods, including Signal to Noise Ratio (SNR), ReliefF, Correlation Coefficient (CC), Mutual Information (MI), and several others. These methods were combined with classifiers such as K Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Decision Tree Classifier (DTC), Naïve Bayes (NB), and Artificial Neural Network (ANN). Our results demonstrated that the best combination was the Signal to Noise Ratio with Linear Discriminant Analysis, achieving a classification accuracy of 95% using only six genes. This study underscores the importance of effective feature selection and classifier combination for precise and efficient prostate cancer diagnosis, paving the way for improved personalized healthcare strategies. Future work will focus on validating these findings with larger datasets and exploring advanced machine learning techniques to enhance classification performance further.

1 Introduction

In recent years, gene expression analysis has become a vital tool for addressing the complexities of cancer diagnosis and therapeutic research. By examining gene activity, researchers can gain insights into the mechanisms driving cancer onset and progression. Changes in gene expression patterns serve as important indicators for early cancer detection [1] and provide potential targets for drug development, paving the way for more personalized, preventive, and predictive healthcare strategies [2].

Advancements in biotechnology have introduced tools to measure and analyse gene expression, aiding diagnostic and therapeutic decisions for various cancers, including prostate cancer. DNA microarray technology generates high-dimensional datasets, measuring thousands of gene expressions across a limited number of samples [3, 4]. This high dimensionality often leads to overfitting in machine learning models, making dimensionality reduction essential to isolate the most relevant genes for accurate cancer classification [5].

The objective of this paper is to evaluate various feature selection methods and their impact on prostate cancer classification. By comparing the effectiveness of different techniques and classifiers, we aim to identify the optimal combination for accurate and efficient diagnosis. We employed several feature selection methods, including Signal to Noise Ratio (SNR), ReliefF, Correlation Coefficient (CC), Mutual Information (MI), t-Statistics (t-S), Fisher Score, Max-Relevance Min-Redundancy (MRmr), Principal Component Analysis (PCA), Genetic Algorithm (GA), Random Forest (RF), and a hybrid approach combining Support Vector Machines with Recursive Feature Elimination (SVM-RFE). After selecting the informative genes, we trained classifiers—K Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Decision Tree for Classification (DTC), Naïve Bayes (NB), and Artificial Neural Network (ANN)—to categorize new samples as tumor or non-tumor.

This paper is structured as follows: Section 2 provides definitions and details of the feature selection methods and classifiers used. Section 3 presents a comparative analysis of the performance of different feature selection and classification methods for prostate cancer classification. Section 4 discusses the results obtained from the analysis, and Section 5 offers our conclusions and future directions.

2 Materials and methods

2.1 Feature Selection Methods

Feature selection is a crucial preprocessing step for high-dimensional data, such as DNA microarray profiles [6]. It involves identifying relevant genes to reduce noise and enhance classification model

* Corresponding author: sara.hb.sara@gmail.com

performance [7, 8]. Feature selection methods can be categorized into filter, wrapper, and embedded techniques [9].

Signal to Noise Ratio (SNR): SNR evaluates gene importance by comparing mean expression levels between cancerous and non-cancerous samples. It calculates the ratio of the difference in means (signal) to within-class variance (noise), prioritizing genes with higher SNR values for cancer biomarker identification [10].

ReliefF: This instance-based algorithm identifies relevant genes by analyzing their ability to differentiate between samples of the same cancer type versus different types. It updates gene weights based on their contributions to class differentiation and is effective with noisy or incomplete data [11].

Correlation Coefficient (CC): CC measures the linear relationship between gene expression levels and cancer class labels, ranking genes by their Pearson correlation values. High correlations indicate strong relevance to cancer classification, helping identify key genes with significant expression changes [12].

Mutual Information (MI): MI quantifies the information each gene's expression provides about class labels (e.g., cancer type). This method captures both linear and non-linear relationships, making it effective for complex gene expression data, with high MI values indicating genes that enhance classification accuracy [13].

t-Statistics (t-S): This method compares gene expression means between cancerous and non-cancerous groups using t-tests. Genes with low p-values indicate significant expression differences and are prioritized for selection, as they are likely important in cancer progression [14].

Fisher Score: The Fisher Score measures the ratio of inter-class variance (between cancer classes) to intra-class variance (within the same class). High scores identify genes that vary significantly between classes, making them effective biomarkers [15].

Max-Relevance Min-Redundancy (MRmr): MRmr selects genes that are relevant to cancer classification while minimizing redundancy. By maximizing each gene's relevance and ensuring diversity among selected genes, MRmr improves classification performance, particularly in gene expression analysis [16].

Principal Component Analysis (PCA): PCA is a dimensionality reduction technique that projects high-dimensional gene expression data into a lower-dimensional space by maximizing variance in the data. It identifies the principal components, which are linear combinations of the original features, representing the directions of maximum variance in the dataset [17]. The PCA Training Process is as follows:

- *Standardize the data* to ensure that each gene has zero mean and unit variance.
- *Compute the covariance matrix* to capture the relationships between genes.
- *Perform eigendecomposition* on the covariance matrix to extract eigenvectors (principal components) and eigenvalues (explained variance).
- *Project the data* onto the principal components to reduce dimensionality.

Genetic Algorithm (GA): GA is an evolutionary algorithm used for feature selection, optimizing gene subsets based on a fitness function such as classification accuracy. It simulates the process of natural selection by evolving a population of gene subsets over generations. [18]. The GA Training Process is as follows:

- *Initialize* a population of random feature subsets (genes).
- Apply selection, *crossover*, and *mutation* operations to generate new feature subsets.
- *Evaluate the fitness* of each subset based on the classification performance.
- *Iterate* through generations, selecting the best-performing subsets and evolving them until convergence.

Random Forest (RF): Random Forest assesses feature importance by measuring how much each gene contributes to reducing impurity (e.g., Gini impurity or entropy) in individual decision trees. The feature importance score is computed by averaging the decrease in impurity across all trees in the forest. [19]. The RF Training Process is as follows:

- *Train a Random Forest model* using the gene expression data.
- *Evaluate feature importance* by calculating how each gene reduces impurity within the trees.
- *Rank* genes based on their contribution to the classification task, with higher-ranking genes being more important for distinguishing cancer types.

Support Vector Machines with Recursive Feature Elimination (SVM-RFE): SVM-RFE is a wrapper-based feature selection method that eliminates the least important genes in a recursive manner using an SVM classifier. It ranks features based on the coefficients of the SVM model and removes those that contribute the least to classification accuracy [20]. The SVM-RFE Training Process:

- *Train an SVM model* on the full set of genes.
- *Rank genes* based on their coefficients in the decision function.
- *Recursively eliminate* the genes with the smallest coefficients, retraining the model at each step until the desired number of genes remains. smallest coefficients, retraining the model at each step until the desired number of genes remains.

2.2 Supervised Classification

Supervised classification aims to categorize data instances into predefined classes based on their features [21]. The classifiers used in this study are:

K Nearest Neighbors (KNN): Instance-based learning where classification is based on the majority vote of the k-nearest neighbors [22]. The KNN Training Process demands no explicit training; classification

happens during inference by calculating distances to neighbors. The KNN Hyperparameters are:

- k: Number of neighbors (in our case K=3)
- distance metric = Euclidean
- weights = 'distance'

Support Vector Machine (SVM): A classifier that finds a hyperplane to separate classes with the maximum margin. [23]. The classifier Training Process is based on finding the optimal hyperplane by solving a convex optimization problem. The SVM Hyperparameters are:

- c = 1
- kernel = Linear
- gamma = 'scale'

Linear Discriminant Analysis (LDA): Projects data onto a lower-dimensional space that maximizes class separation [24]. The LDA Training Process is based on the Compute of class means and within-class scatter and Maximize the between-class scatter.

Decision Tree for Classification (DTC): A tree-based model where each internal node represents a decision based on a feature, and each leaf node represents a class label [25]. It Recursively splits the data based on feature values to maximize information gain (Gini). The DTC Hyperparameters are:

- max_depth = 10
- min_samples_split = 5
- criterion = 'gini'

Naïve Bayes (NB): A probabilistic classifier based on Bayes' theorem with the assumption of feature independence [26]. The NB Training Process involves computing class probabilities and feature likelihoods, followed by classification based on posterior probabilities.

Artificial Neural Network (ANN): consists of multiple layers (input, hidden, output) of interconnected neurons with activation functions [27]. The ANN training process involves training through backpropagation using stochastic gradient descent (SGD) and adjusting weights to minimize a loss function, typically cross-entropy. The ANN Hyperparameters are:

- learning rate = 0.001
- number of hidden layers = 2
- number of neurons per layer = 100
- activation function = 'ReLU'
- epochs = 50
- batch size = 16

These classifiers, with their respective training processes and Hyperparameters, were selected to provide a comprehensive evaluation of the dataset's predictive performance.

2.3 Performance Evaluation

The classifiers' performance was evaluated using classification accuracy [28], which measures the proportion of correctly classified instances:

$$Acc = 100* (TP + TN) / (TN + TP+ FN+ FP) \quad (1)$$

Where: TP: true positive; FP: false positive; TN: true negative; FN: false negative.

3 Stat of art

Machine learning is vital in cancer diagnosis, particularly through effective feature selection that identifies relevant biomarkers from high-dimensional datasets. By reducing dimensionality, these algorithms focus on the most informative genes, crucial for managing large genomic data [29]. This strategy aids in discovering potential biomarkers for early diagnosis and targeted therapy [8] while improving model interpretability by highlighting key genes in cancer progression [30]. Additionally, reducing features helps mitigate overfitting, enhancing the model's generalization on new data and increasing its clinical utility [31].

Recent research has introduced innovative classification methodologies with impressive accuracy rates across various domains. For example, [32] uses SVM with a radial basis function kernel, employing Discrete Cosine Transform (DCT) and Enhanced Harmony Search (EHO) for feature selection, achieving 94.8% accuracy. Similarly, [33] presents a hybrid model with multi-objective particle swarm optimization (MOPSO) at 94%. In contrast, [34] reports a lower accuracy of 77.79% using MS-SVM-RFE (MSR). Additionally, [35] introduces a Multi-Objective Binary Cuckoo Search Algorithm (MOBCSA) with KNN, reaching 94.89% accuracy. These results highlight the evolving landscape of classification techniques and the effectiveness of tailored methodologies in achieving high precision in classification tasks.

4 Results

The study utilized Matlab for simulations and assessed various methods on a prostate cancer dataset containing 12,600 genes and 102 samples, categorized into tumor (52 samples) and non-tumor classes (50 samples). The data is available for download at broadinstitute.org/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=75. The dataset was divided into training and testing sets, with the training samples used for feature selection and model development, while the test samples were set aside for performance evaluation. Following a standard machine learning pipeline, the study implemented feature selection to reduce dataset dimensionality, followed by classification algorithms for predictive modeling.

Given the dataset's high dimensionality, feature selection was essential for narrowing down the number of genes while retaining the most informative ones. The feature selection methods employed included SNR, ReliefF, CC, MI, T-S, Fisher, MRmr, PCA, GA, RF, and SVM-RFE. Once the relevant features were identified, several classification algorithms—such as KNN, SVM, LDA, DTC, NB, and ANN—were tested to build predictive models. The training and testing process involved applying a feature selection algorithm to the training samples, which helped reduce dimensionality and retain key features for model training. Cross-validation was used to fine-tune hyperparameters, preventing overfitting and enhancing the models' generalization. After training, the models were

evaluated on the test set to measure accuracy. By splitting the dataset into training and testing subsets, the study validated unseen data, providing a reliable measure of predictive performance. The integration of advanced feature selection methods with various classification algorithms enabled a comprehensive analysis of the prostate cancer dataset, yielding robust and interpretable results.

4.1 Feature Selection and Classification Results

Table 1 summarizes the performance of various classifiers using different feature selection methods, presenting the maximum accuracy (Acc) achieved for subsets of genes (Nbr genes). The lowest recorded accuracy, at 58.8%, was associated with the feature selection methods MI, Fisher, and MRmr, as well as the classifiers DTC, NB, and ANN. In contrast, the highest accuracy of 95% was achieved with SNR, CC, and SVM-RFE in combination with LDA.

Following the accuracy data in Table 1, we created Figure 1 to illustrate the accuracies attained by each feature selection method, enabling easy comparison to identify the best-performing techniques. Additionally, Figure 2 compares the classifiers based on their accuracies as reported in Table 1.

Table 1. Evaluating the effectiveness of proposed classifiers and selection methods for prostate classification

	KNN		SVM		LDA		DTC		NB		ANN	
	Acc (%)	Nbr genes	Acc (%)	Nbr Genes	Acc (%)	Nbr genes	Acc (%)	Nbr genes	Acc (%)	Nbr genes	Acc (%)	Nbr Genes
SNR	91	22	92	8	95	4	92	18	92	22	92	12
ReliefF	91	32	94	34	94	75	90	42	90	27	92	28
CC	91	6	94	44	95	6	91	31	91	37	92	18
MI	65	1	65	56	72	10	58,8	12	58,8	21	65	16
T-S	62	12	78,4	31	78,4	15	60	31	60	22	65	22
Fisher	60	22	78,4	12	78,4	22	58,8	34	58,8	12	58,8	13
MRmr	60	7	65	60	65	49	54,9	3	58,8	23	60	19
PCA	90	25	92	18	92	27	85	22	85	15	90	32
GA	90	12	90	8	90	12	83	13	85	23	90	14
Random Forest	90	24	93	18	94	6	88	18	90	24	92	14
SVM-RFE	90	16	92	14	95	11	90	22	90	14	90	7

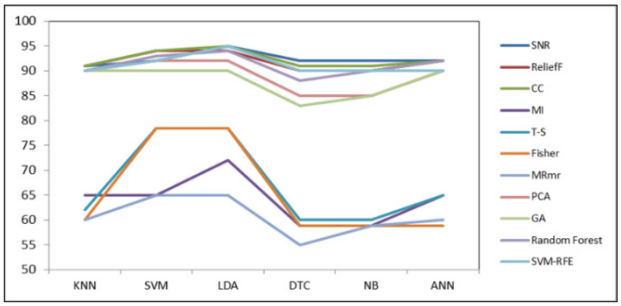

Fig. 1: Comparative analysis of feature selection methods for prostate cancer classification.

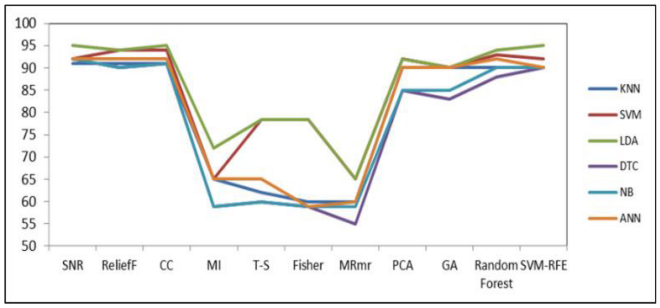

Fig. 2: Comparative evaluation of classifiers for prostate cancer classification

4.2 Detailed Analysis

The feature selection methods exhibited significant variability in their performance, with SNR, CC, and SVM-RFE consistently achieving high classification accuracy. Notably, SNR attained 95% accuracy with LDA by selecting only six genes, highlighting its effectiveness in reducing dimensionality while maintaining relevant information. The genes selected by each method include: for SNR-LDA: x.V4365, x.V5757, x.V6185, and x.V9172; for CC-LDA: x.V7247, x.V10494, x.V6866, x.V6462, x.V9850, and x.V8850; and for SVM-RFE-LDA: x.V10494, x.V7820, x.V9172, x.V10956, x.V3794, x.V6185, x.V4365, x.V11818, x.V9850, x.V6462, and x.V6620.

Among the classifiers tested, LDA emerged as the most effective when combined with various feature selection methods, particularly SNR, achieving the highest accuracy of 95%. Other classifiers, such as SVM and ANN, also demonstrated commendable performance, especially when paired with the ReliefF and Random Forest feature selection techniques. This comprehensive approach to feature selection and classification enabled robust predictive modeling in the study.

4.3 Discussion

Our study demonstrates the transformative potential of combining advanced feature selection methods with robust machine learning classifiers for classifying

prostate cancer using gene expression data from DNA microarrays. By narrowing down the extensive gene list to the most relevant ones, we significantly enhance classification accuracy and reduce computational complexity.

The results underscore the vital role of feature selection methods in improving prostate cancer classification accuracy. Our analysis shows that methods like Signal to Noise Ratio (SNR), Correlation Coefficient (CC), and Support Vector Machines with Recursive Feature Elimination (SVM-RFE) consistently yield high accuracies. Notably, SNR achieved an impressive 95% accuracy with a minimal gene subset, effectively distinguishing between tumor and non-tumor samples. Similarly, CC's quantification of linear relationships between gene expression and class labels proved effective, identifying key genes that could serve as important biomarkers. SVM-RFE's recursive elimination of the least significant genes contributed to its success with LDA, enhancing both classification performance and model interpretability. In contrast, methods like Mutual Information (MI) and Fisher Score showed lower accuracies, highlighting the need for tailored feature selection techniques in high-dimensional analyses. Overall, the interplay between feature selection methods and classification algorithms is crucial for developing reliable predictive models in cancer diagnosis, advocating for the integration of advanced techniques to enhance interpretability and accuracy.

5 Conclusion

This study emphasizes the significant impact of feature selection methods on prostate cancer classification accuracy. By comparing multiple feature selection techniques, including SNR, CC, and SVM-RFE, alongside various classifiers, we identified optimal strategies for accurate tumor identification. The high accuracy achieved with the selected feature subsets not only demonstrates the effectiveness of these methods but also paves the way for personalized and precise cancer diagnostics.

The selected genes highlighted in our analysis, particularly from the SNR-LDA and CC-LDA combinations, provide promising avenues for further research into prostate cancer biomarkers. Their biological relevance and potential roles in tumorigenesis warrant deeper exploration to enhance understanding and treatment approaches for prostate cancer.

Future work should focus on validating these findings through larger, more diverse datasets, potentially integrating additional biological and clinical data to strengthen predictive models. Moreover, exploring the combination of feature selection and classification methods within ensemble learning frameworks may further improve classification performance, offering a comprehensive toolkit for prostate cancer diagnosis and management.

Ultimately, this study contributes to the growing body of research aimed at harnessing machine learning and gene expression analysis in the fight against cancer, providing valuable insights for both clinical practice and ongoing research endeavors.

References

1. Naeem, A., Farooq, M. S., Khelifi, A., & Abid, A. (2020). Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. IEEE access, 8, 110575-110597.
2. Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Classification of breast cancer based on histology images using convolutional neural networks. Ieee Access, 6, 24680-24693.
3. Gupta, S., Gupta, M. K., Shabaz, M., & Sharma, A. (2022). Deep learning techniques for cancer classification using microarray gene expression data. Frontiers in Physiology, 13, 952709.
4. Alhassan, A. M., & Zainon, W. M. N. W. (2021). Review of feature selection, dimensionality reduction and classification for chronic disease diagnosis. IEEE Access, 9, 87310-87317.
5. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access, 9, 71194-71209.
6. El Kafrawy, P., Fathi, H., Qaraad, M., Kelany, A. K., & Chen, X. (2021). An efficient SVM-based feature selection model for cancer classification using high-dimensional microarray data. IEEE Access, 9, 155353-155369.
7. Sara, H. B., & Jihad, H. B. (2024, April). Artificial Intelligence Application for the Classification of Central Nervous System Tumors Based on Blood Biomarkers. In 2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST) (pp. 1-5). IEEE.
8. Singh, N., & Singh, P. (2021). A hybrid ensemble-filter wrapper feature selection approach for medical data classification. Chemometrics and Intelligent Laboratory Systems, 217, 104396.
9. Raj, R. J. S., Shobana, S. J., Pustokhina, I. V., Pustokhin, D. A., Gupta, D., & Shankar, K. J. I. A. (2020). Optimal feature selection-based medical image classification using deep learning model in internet of medical things. IEEE Access, 8, 58006-58017.
10. Mishra, D., & Sahu, B. (2011). Feature selection for cancer classification: a signal-to-noise ratio approach. International Journal of Scientific & Engineering Research, 2(4), 1-7.
11. Haq, A. U., Li, J. P., Saboor, A., Khan, J., Wali, S., Ahmad, S., ... & Zhou, W. (2021). Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques. IEEE Access, 9, 22090-22105.
12. Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction methods with deep learning to predict clinical

outcome of breast cancer. Ieee Access, 6, 28936-28944.

13. Ramaswamy, R., Kandhasamy, P., & Palaniswamy, S. (2023). Feature selection for Alzheimer's gene expression data using modified binary particle swarm optimization. IETE Journal of Research, 69(1), 9-20.
14. Mohammed, B., Hamdan, M., Bassi, J. S., Jamil, H. A., Khan, S., Elhigazi, A., ... & Marsono, M. N. (2020). Edge computing intelligence using robust feature selection for network traffic classification in internet-of-things. IEEE Access, 8, 224059-224070.
15. Bugata, P., & Drotar, P. (2020). On some aspects of minimum redundancy maximum relevance feature selection. Science China Information Sciences, 63(1), 112103.
16. Huang, M., Sun, L., Xu, J., & Zhang, S. (2020). Multilabel feature selection using relief and minimum redundancy maximum relevance based on neighborhood rough sets. IEEE Access, 8, 62011-62031.
17. Wu, W., & Zhou, H. (2017). Data-driven diagnosis of cervical cancer with support vector machine-based approaches. IEEE Access, 5, 25189-25195.
18. Houssein, E. H., Abdelminaam, D. S., Hassan, H. N., Al-Sayed, M. M., & Nabil, E. (2021). A hybrid barnacles mating optimizer algorithm with support vector machines for gene selection of microarray cancer classification. IEEE Access, 9, 64895-64905.
19. Richhariya, B., Tanveer, M., Rashid, A. H., & Alzheimer's Disease Neuroimaging Initiative. (2020). Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomedical Signal Processing and Control, 59, 101903.
20. Salman, O., Elhajj, I. H., Kayssi, A., & Chehab, A. (2020). A review on machine learning–based approaches for Internet traffic classification. Annals of Telecommunications, 75(11), 673-710.
21. Chaudhari, P., Agarwal, H., & Bhateja, V. (2021). Data augmentation for cancer classification in oncogenomics: an improved KNN based approach. Evolutionary Intelligence, 14, 489-498.
22. Wazery, Y. M., Saber, E., Houssein, E. H., Ali, A. A., & Amer, E. (2021). An efficient slime mould algorithm combined with k-nearest neighbor for medical classification tasks. IEEE Access, 9, 113666-113682.
23. Varan, M., Azimjonov, J., & MaÇal, B. (2023). Enhancing Prostate Cancer Classification by Leveraging Key Radiomics Features and Using the Fine-Tuned Linear SVM Algorithm. IEEE Access.
24. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access, 9, 71194-71209.
25. Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., ... & Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. IEEE Access, 9, 24273-24287.
26. Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. IEEE Access, 6, 29637-29647.
27. Lee, S. A., Cho, H. C., & Cho, H. C. (2021). A novel approach for increased convolutional neural network performance in gastric-cancer classification using endoscopic images. IEEE Access, 9, 51847-51854.
28. Kabir, M. F., Chen, T., & Ludwig, S. A. (2023). A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. Healthcare Analytics, 3, 100125.
29. Ramírez-Mena, A., Andrés-León, E., Alvarez-Cubero, M. J., Anguita-Ruiz, A., Martinez-Gonzalez, L. J., & Alcala-Fdez, J. (2023). Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. Computer Methods and Programs in Biomedicine, 240, 107719.
30. Zarbakhsh, P. (2023). Spatial attention mechanism and cascade feature extraction in a U-Net model for enhancing breast tumor segmentation. Applied Sciences, 13(15), 8758.
31. Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., ... & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. Physica medica, 83, 9-24.
32. Mani, K., & Rajaguru, H. (2024). A framework for performance enhancement of classifiers in detection of prostate cancer from microarray gene. Heliyon, 10(9).
33. Rahimi, M. R., Makarem, D., Sarspy, S., Mahdavi, S. A., Albaghdadi, M. F., & Armaghan, S. M. (2023). Classification of cancer cells and gene selection based on microarray data using MOPSO algorithm. Journal of Cancer Research and Clinical Oncology, 149(16), 15171-15184.
34. Ding, X., Yang, F., & Ma, F. (2022). An efficient model selection for linear discriminant function-based recursive feature elimination. Journal of Biomedical Informatics, 129, 104070.
35. Abdulwahab, H. M., Ajitha, S., Saif, M. A. N., Murshed, B. A. H., & Ghanem, F. A. (2024). MOBCSA: Multi-Objective Binary Cuckoo Search Algorithm for Features Selection in Bioinformatics. IEEE Access.