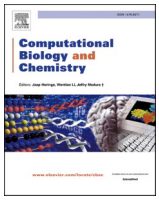
Computational Biology and Chemistry 120 (2026) 108706

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac



MicroarrayCancerNet: Hybrid optimized deep learning with integration of graph CNN with 1D-CNN for cancer classification framework using microarray and seq expression data

B. Shyamala Gowri [a,b,*], S. Anu H Nair [c], K.P. Sanal Kumar [d], S. Kamalakkannan [e]

[a] Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Chidambaram- 608002, Tamil Nadu, India
[b] Assistant Professor, Department of Computer Science and Engineering, Easwari Engineering college, Ramapuram, Chennai-600089, Tamil Nadu, India
[c] Department of Computer Science and Engineering, Annamalai University (Deputed To WPT, Chennai -113), Annamalainagar, Chidambaram-608002, Tamil Nadu, India
[d] Department of Computer Science, RV Government Arts College, Chengalpattu, India
[e] Department of Information Technology, School of Computing Sciences Vels Institute of Science, Technology & Advanced Studies, Pallavaram, Chennai-600117, Tamil Nadu, India

ARTICLE INFO

Keywords:
Cancer Classification
Modified Sandpiper Optimization Algorithm
Optimal Gene Selection
Micro-array Data
Hybrid Deep Learning Framework

ABSTRACT

The key difficulty lies in accurately classifying the relevant genes through analysis and selection. A variety of methods are used to classify the genes. However, in the selection of numerous genes in the huge dimensional microarray data, only a limited amount of success has been achieved. Thus, this study focuses on designing a new cancer classification framework. In the initial stage, the microarray and seq expression information is attained from the standard datasets. Next, the pre-processing is performed using NAN removal and the missing value removal from the samples to convert it into a numeric feature matrix for making the data suitable for further levels of processing. Then, the Modified Sandpiper Optimization Algorithm (MSOA) is suggested for confirming the optimal gene from the pre-processed information. Finally, the chosen optimal gene is fed to the cancer classification stage, where the Hybrid Deep Learning Framework (HDLF) is suggested by incorporating the Graph Convolutional Neural Network (GCNN) with One-Dimensional Convolutional Neural Networks (1D-CNN). The parameters of both Graph CNN and 1D-CNN are tuned via the same MSOA. Finally, the experimental results confirm that the developed model performs well compared to existing machine learning and currently utilized deep learning methods for cancer classification. The precision of the proposed model is 91.78 %.

1. Introduction

Currently, one primary issue of death is cancer (Wang et al., 2007), and microarray data-derived expression of the gene patterns has been discovered as promising cancer diagnostic indicators. In the medical field, cancer research has been going on for hundreds of years. Numerous academic areas are involved in the study of cancer causes. Numerous biological microarray studies have been carried out as a first step in the research of potential treatment that aims only to gather additional information (Muhammad et al., 2023; Jiaji et al., 2023; Xu et al., 2007). Early cancer detection is necessary because treating patients is difficult when it comes final stages of the disease. An accurate cancer prognosis is important for patients to receive appropriate care (Leung and Hung, 2010). Due to the complexity of gene expression levels within the human body, cancer detection is challenging. It is well-recognized that gene expression levels hold significant clues to the fundamental issues surrounding the treatment and prevention of illnesses (Houssein et al., 2021). Detailed and thorough routes and also network-based notes with regulatory linkages should be taken into consideration to unveil the biology of cancer across several scale levels (Chakraborty andMaulik, 2014). To understand the connections between transcription factors and the genes they are targeting, gene regulatory networks have received extensive study. In cancer genomics, modelling the cellular and molecular events that occur during the progression of the tumour by creating networks of gene modification is of utmost importance (Rabia et al., 2023).

* Corresponding author at: Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Chidambaram- 608002, Tamil Nadu, India.
E-mail addresses: shyamalagowribalaraman@gmail.com (B. Shyamala Gowri), anu_jul@yahoo.co.in (S.A. H Nair), sanalprabha@yahoo.co.in (K.P.S. Kumar), kannan.scs@velsuniv.ac.in (S. Kamalakkannan).

https://doi.org/10.1016/j.compbiolchem.2025.108706
Received 26 August 2025; Accepted 30 September 2025
Available online 8 October 2025

Microarrays are utilized to simultaneously evaluate thousands of gene interactions and provide a global picture of cellular activity. The most prevalent and significant feature of functional genomics is the classification of microarray data. Utilizing microarray data entails categorizing patient samples into several classes following their gene expression profiles (Shen and Tan, 2005). To describe the comprehensive aspect of the cell function of a gene by gene methodologies, the microarray method is introduced. Microarray technology is also utilized to detect the activity of every gene throughout the entire genome in a single experiment (Harvey and Ji, 2017). The study of the genetic causes of cancer using microarray studies results in the development of cutting-edge therapeutic designs for the medical sector (Nguyen and Nahavandi, 2016). However, the tiny sample size and huge dimensionality of microarray data, classification is still a challenging and difficult operation (Maji, 2012). Microarray gene expression studies frequently produce a large number of characteristics for a limited number of patients, producing a high-dimensional dataset with a tiny sample size (Almazrua and Alshamlan, 2022). The genes are connected with one another either indirectly or directly, which makes classifying the expression of gene data a highly complicated and complex mission that naturally calls for the usage of an accurate and potent feature selection technique (Rabia, 2022b; 2022a). In recent times, data mining techniques are the methodologies that are further employed to examine enormous quantities of data.

Classification is a vital process in data mining and machine learning that places an instance into the appropriate classification. In (Maulik and Chakraborty, 2014), gene expression data in time series were subjected to Dynamic Bayesian Networks (DBN) and canonical correlation analysis for the involvement of evaluated gene-modifying networks. Similarly to this (Prabhakar and Lee, 2020), a characteristics selection approach based on the Partial Least Squares (PLS) has been used to design gene regulatory networks. By using biological data, particularly time series gene expression measurements, Bayesian techniques were taken into consideration for network analysis (Peng et al., 2021; Samundeeswari and Gunasundari 2023). DBNs have been consistently utilized to simulate changes in gene expression over time (Hsieh and Chou, 2016) among the several approaches for modelling gene regulatory networks (Pham et al., 2006). These methods improve the representation of spatiotemporal input-output interdependence because they have the essential capacity to capture the varying time behaviour of the primary biological network (Wu et al., 2012). It is used to categorize various cancer kinds and can also be used to spot mislabeled data, which aids doctors in making a precise diagnosis. DNA expression data contains a substantial number of genes (Liu et al., 2019). In actuality, few of the characteristics of one sample have significant discriminative information. Recently, various deep learning models have been used to classify cancer efficiently. Further, the diverse research work helps to show the importance of the deep learning model, which is stated below. (Shams et al., 2023) have developed the entropy-controlled deep learning and flower pollination optimization algorithm for detecting breast cancer using mammogram images. For extracting the features, the deep learning technique is adopted. Also, the serial technique was utilized to attain the deep features for a better classification framework (Sethy et al., 2023; Shiyang et al., 2023). Additionally, the accurately classified outcomes were suggested using the neural network. In accordance, the deep learning model can train the inadequate datasets for further improvement. (Mamuna et al., 2023) have stated the computer diagnosis techniques along with the deep learning techniques. Additionally, the Breast cancer (BrC) classification model was adopted using deep learning models to show the reliable performance. To facilitate better outcomes, the data augmentation approach was suggested using the CNN approach. Further, the experimentation was done through the standard CBIS-DDSM datasets. Consequently, the deep learning model was adapted to lower the performance of error rate to enhance performance. This paper introduces a novel idea for cancer classification on microarrays and seq expression data utilizing deep learning, among other newly applied techniques.

The primary contribution of the proposed model is summarised below.

- → ***Hybrid Deep Learning Framework (HDLF) Integration***: The proposed research combines GCNN with 1D-CNN, leveraging both relational gene information and sequential gene expression data for enhanced classification accuracy. The combination of Graph CNN and 1D-CNN within the HDLF allows for more precise identification of cancer types by leveraging the strengths of both models, leading to improved classification performance over existing methods.
- → ***Modified Sandpiper Optimization Algorithm (MSOA):*** Implements a novel optimized approach for simultaneous gene selection and hyperparameter tuning, leading to more accurate and computationally efficient models. MSOA is designed to address the convergence issues faced by traditional SOA, providing more reliable and stable convergence towards the optimal solution, which is crucial for high-dimensional gene selection
- → ***Optimal Gene Selection Strategy:*** Employs MSOA to identify the most relevant genes from high-dimensional microarray and seq expression data, reducing noise and improving model focus on biologically significant features. It searches for an optimal subset of genes that contribute significantly to accurate cancer classification, ensuring that only the genes with the highest predictive power are used.
- → ***Parameter Optimization for Deep Models***: Systematically tunes crucial parameters such as hidden neuron counts and epochs in both GCNN and 1D-CNN, maximizing their individual and combined performance. MSOA explores the parameter space for each model, determining the best values for hidden neurons, epochs, learning rates, and other hyperparameters to maximize classification accuracy.
- → ***Enhanced Feature Extraction:*** Combines structural information from GCNN with sequential pattern recognition from 1D-CNN, providing a richer feature set for classification tasks. This model utilizes GCNN's ability to incorporate gene interaction networks, capturing complex biological relationships that traditional models neglect.

The paper is further provided by the following sections. Section II reviews the traditional approaches to cancer classification. Section III elaborates on the new cancer classification using microarray data with an advanced deep-learning model. Section IV explains the MSOA algorithm based on the optimal gene selection for the microarray cancer classification. Section V explored the proposed HDLF framework with the help of parameter optimization. Section VI explained the discussions and the results of the recommended model. The final Section VII finishes the developed cancer classification model.

2. Literature review

2.1. Systematic works

Fathi et al., (2021) have explained hybrid cancer types and multiple machine learning methodologies were utilized in the hybrid method. To optimise the high depth hyper hyperparameter, Grid Search Cross-Validation (CV) was used. There were seven best microarray cancer datasets were utilized to estimate the methodology. To find out which characteristics were more helpful and related using the existing model, multiple performances were used that contained accuracy for the classification, sensitivity, specificity, F1-score, and AUC. The recommended method highly reduces the amount of genes needed for categorization, chooses the primary informative characteristics, and enhances categorisation correctness based on the results.

Kourou et al., (2019) have found the genes that perform as controllers and mediate the activities of transcription metrics that were found in

Table 1
Benefits and challenges of cancer classification over the traditional models.

Author [citation]	Methodology	Advantages	Limitations
(Fathi et al., 2021)	Decision Tree classifier	• This method has the potential to effectively detect the optimal or near-optimal subsets to provide classification outcomes. • It is used to choose the informative genes to improve the performance of the method.	• However, optimisation forecasting using other algorithms needs to be explored.
(Kourou et al., 2019)	DBN	• This technique can modify the parameters to improve the classification correctness of the model. • It is precise and robust.	• The time consumption as well as the interpretability is higher.
(El Kafrawy et al., 2021)	Ensemble	• This method has provided high classification accuracy and has also successfully resolved the time complexity. • It is regarded as an effective informative gene selection process to detect brain cancer.	• This method has high complexity, limited sample size as well as high dimensionality that degrade the performance.
(Rojas et al., 2020)	MCGA	• It intends to detect the tiny subset of useful genes to attain higher categorisation accuracy.	• This method faces the issue while performing on a large dataset.
(Othman et al., 2020)	Multi-objective cuckoo search algorithms	• It consists of real-world clinical and biological applications, which highly contribute to cancer.	• It has the maximum number of selected genes that increases the time duration.
(Wu and Wang, 2019)	CN	• It has used a standard dataset to attain the ideal classification model.	• This method is expensive regarding time and cost.
(Shah et al., 2020)	LS-CNN	• It is useful in enhancing the treatment strategy as well as in medical discovery.	• The multi-class image dataset to attain better outcomes is limited in this method.
(Haznedar al., 2021)	Ensemble	• It is regarded as a successful model to effectively classify the disease.	• The implementation for training the ANFIS model needs to be explored.

every promoter of our multiple explained gene sets. These characteristics gave the strongest factors for differentiating the tumours from ordinary samples utilizing a Deep Belief Network (DBN)-based classification method. In accordance, the Public functioned differential expression analysis, functional repository, and Gene Expression Omnibus (GEO) of the microarray datasets are gathered. Here, the DBN model is used to select the particular genes and find out the characteristics that could correctly differentiate the samples into the tumors and the control measures.

El Kafrawy et al. (2021) have combined the ensemble mRMRe, in a hybrid method for the selection of the gene considered as (SVM-mRMRe) with embedded SVM coefficients called features ranking. This methodology offers an effective model to combine the ensemble, filter-based, as well as embedded models that were performed. The method was assessed utilizing eight of the highly popular microarray datasets for multiple stages of cancer. Four alternative classifiers like Random Forest (RF), Multilayer Perception (MLP), SVM, and k-Nearest Neighbors (k-NN) were evaluated for the selected subset of features. The computational results have explained that the explained model has improved the distinction of cancer from benign tissues while requiring less time and dimensionality. Additionally, the gene's biological interpretation chosen for the brain cancer dataset accords with the outcomes of pertinent scientific studies and was crucial for predicting the prognosis of patients.

Rojas et al., (2020) have explained a Memetic Cellular Genetic Algorithm (MCGA) to address the characteristics selection issue of cancer microarray datasets. Colon, lymphoma, and leukaemia data from the literature were used for implementation. Other well-known meta--heuristic tactics have been contrasted with MCGA. The outcomes have shown that their approach can offer effective ways.

Othman et al., (2020) have implemented a hybrid multi-objective cuckoo search with the help of evolutionary operators for gene selection. According to this essay, the evolutionary operators' two-time mutation and one-time crossover have been applied. The goal of this study was to enhance the dimensions' values and capacity for exploratory search.

Wu and Wang, (2019) have explained a Complex Network (CN) classifier was allegedly used to carry out the classification task, according to the structure was started using an algorithm, allowing input variables to be chosen across layered various activation functions and connections for various nodes. Then, using the parameters stored in the classifier, an optimal structure was found using a hybrid approach that used particle swarm optimization and genetic programming techniques. We built a basic classifier based on various feature sets, including Spearman's and Pearson's correlation, Cosine coefficient, Fisher ratio, and Euclidean distance, to ensure variety in the ensemble classifiers. According to the experimental findings, a single classifier can be utilized to make cutting-edge results and however, the ensemble made superior outcomes.

Shah et al., (2020) have explained a hybrid deep learning model based on the Laplacian Score-Convolutional Neural Network (LS-CNN) for the categorization of specific cancer data. Haznedar et al., (2021) have suggested a hybrid technique based on the Fuzzy C-Means Clustering (FCM), the Simulated Annealing (SA) algorithm, and the Adaptive Neuro-Fuzzy Inference System (ANFIS). The execution of the recommended model was contrasted to other distinct algorithms and also the other statistical techniques were adapted. The outcomes of the demonstrated FCM-based ANFIS were adapted with the SA algorithm to categorize the cancer datasets.

Shoaib et al. (2025) have proposed a pre-trained CNN model for classifying brain tumours from CT images. The softmax activation of this model was used for extracting the relevant features, and they were given to the principal component analysis for the dimension reduction.

Whig et al. (2025) have presented an unsupervised machine learning model for classifying the different types of bone marrows. This model was well suited for the clinical decision-making process, also enhancing the accuracy of the diagnostic task. The experimental results showed that the proposed model attained compelling accuracy in cancer detection.

Badashah et al. (2025) have developed an image processing-based machine learning model for detecting bone cancer. Here, the Gaussian elimination was done to enhance the quality of the images. Finally, sufficient data was used for training and testing the proposed model to get accurate results.

Nurtay et al. (2025) have presented DCNN for the effective assessment of brain tumours in humans. The customised CNN with the specific

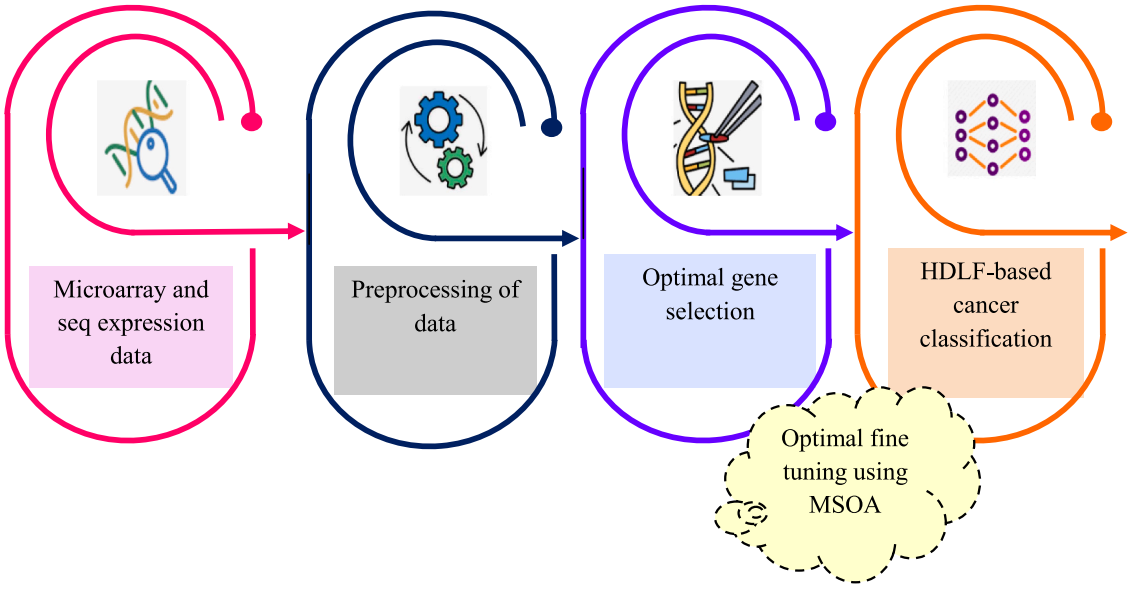


Fig. 1. Architectural illustration of the recommended cancer classification using microarray and seq expression data.

layer architecture was trained to accurately detect the brain tumour in humans. The robust performance of the model in the cancer classification is confirmed by the validation.

2.2. Problem statement

The microarray, depending on the gene expression technique, is regarded as the essential method for cancer prognosis, treatment as well as in diagnosis. It is significantly utilized to measure the gene expression level. The microarrays help measure the interaction among thousands of genes randomly as well as design the global picture for cellular function. The benefits and drawbacks are listed in Table 1.

→ Conventional models struggle with the curse of dimensionality due to the large number of gene features relative to limited data samples, leading to overfitting and reduced accuracy. The model employs the MSOA to select optimal genes, reducing dimensionality, eliminating irrelevant features, and mitigating noise, thereby improving classification accuracy.
→ The presence of noisy, irrelevant, or redundant genes hampers the reliability and interpretability of classification results. In the proposed model, Missing values and noise are handled through data preprocessing steps such as NAN removal and missing value handling, ensuring cleaner input data for the model.
→ Identifying the most informative genes from high-dimensional data is complex but crucial for improving model performance. Combining GCNN and 1D-CNN allows the model to capture complex patterns in high-dimensional data while being robust against overfitting.
→ Complex models tend to overfit training data, reducing their ability to accurately classify unseen data. Combining GCNN and 1D-CNN allows the model to capture complex patterns in high-dimensional data while being robust against overfitting.
→ Tuning hyperparameters such as the number of hidden neurons and epochs in deep learning models is challenging and time-consuming. MSOA optimizes hyperparameters like hidden neurons and epochs, enabling efficient and effective training without extensive manual tuning.

3. A new cancer classification model using microarray and seq data with an advanced deep-learning model

3.1. Microarray cancer classification framework

A group of disorders known as cancer is characterised by unnatural cell proliferation. Also, cell proliferation is normal in a healthy body, which shows the growth of the cells (Rabia et al., 2019). Based on external and internal factors, the genetic cells are harmed, which causes them to become tumours. While exposure to substances like radiation, UV light from the sun, and chemicals in cigarette smoke are important external variables that contribute to cancer, incorrect cell damage and division of DNA are the main internal contributors. Due to the intrinsic complexity in the data's nature, including smaller sample size, improved dimensionality, an unbalanced number of classes, higher variation of feature values, and noisy data structure, analyzing microarray data is a very hard job (Jose et al., 2017). Less accurate classification and an over-fitting issue have resulted from this. For categorization, the microarray cancer data is split into a set of classes, where the authors set out to design a machine-learning algorithm (Kiran et al., 2022; 2023). It is lacking in strategy for expanding the recommended method to multi-class microarray cancer datasets. Additionally, the correctness of the categorization on those binary datasets with lower classification accuracy scores does not improve. Therefore, a new hybrid-optimized deep learning method utilizing microarray data has been implemented. Fig. 1 displays the fundamental architecture of the recommended model.

The main scope of this study is to implement a novel framework for classifying cancer. The microarray and seq expression data is obtained from established datasets. The collected data undergoes pre-processing through NAN removal and the missing value removal for further processing. NAN removal and missing value removal are essential preprocessing steps aimed at improving data quality and reliability for subsequent analysis. Specifically, they serve to eliminate incomplete or invalid data points that could negatively impact the performance of the classification model. NAN removal involves identifying and removing data entries where the value is NaN, which indicates missing or corrupt data. This step ensures that only valid numerical data is retained for analysis, preventing errors during feature extraction and model training. Missing value removal refers to further eliminate samples or features that contain missing data, ensuring the dataset consists solely of complete information. This process helps to reduce noise and biases caused

Table 2
Number of features and instances of different types of cancer in Datasets1 and 2.

Dataset	Types of Cancer / Sample Labels	Number of Features	Number of Instances
Dataset 1 (Microarray Gene Expression Cancer Data)	BRCA, KIRC, COAD, LUAD, PRAD, Brain_CG_1 to Brain_NG_14 (multiple brain cancer subtypes)	12000 attributes	1627 instances
Dataset 2 (Gene Expression Cancer RNA-Seq Data)	Endometrial Cancer, Lung Cancer, Prostate Cancer, Central Nervous System Cancer, Brain Cancer	20,531 attributes (genes)	801 instances

by incomplete data, thereby enhancing the accuracy and robustness of gene selection and classification. As a result, the performance of the data is enhanced. Then, for choosing the ideal gene from the pre-processed data, the MSOA is recommended. The optimal gene selection is described as a crucial step to improve the accuracy and efficiency of cancer classification. It involves selecting the most relevant genes from pre-processed data to reduce dimensionality and eliminate redundant or irrelevant features. Specifically, the MSOA algorithm is utilized for this purpose, helping to identify the best subset of genes that contribute significantly to differentiate various cancer types. By selecting these optimal genes that focus on the most informative features, which enhances classification performance. This process ultimately leads to a more accurate and efficient classification framework, as it reduces noise and overfitting associated with high-dimensional data and irrelevant features. The optimized gene set serves as a refined input for the deep learning classifiers, improving their ability to correctly categorize cancer types. The selected best gene is then fed into the stage for classifying cancer, where the HDLF is suggested by combining the GCNN and 1D-CNN.The GCNN is adept at handling the complex relationships and topological structures inherent in microarray data, capturing spatial and relational features effectively. Meanwhile, the 1D-CNN focuses on extracting feature patterns from sequential data, enhancing the model's ability to recognize distinctive gene expression signatures. By integrating these two architectures, the HDLF can exploit both the relational structure and feature patterns of the data, leading to more precise classification outcomes. Additionally, the parameters of both networks, such as hidden neuron counts and epochs are optimized using the MSOA algorithm, ensuring the model is finely tuned for maximum performance. This hybrid approach significantly improves the classifier's accuracy, robustness, and ability to assist clinicians in early and reliable cancer diagnosis. Here, the same MSOA is used to control the parameters of both the GCNN and the 1D-CNN. By utilizing these two methods optimal gene is acquired. Then, the developed HDLF model is used to classify the cancer effectively. Here, the clinician shows better treatment based on the final classified outcomes. Based on the outcomes, the multi-disease cancer classification is performed while classifying cancers like Central Nervous System Cancer, Lung Cancer, Endometrial Cancer, Brain Cancer, Prostate Cancer, Gene Expression, and Microarray. Finally, the numerical findings show that the developed model employs deep learning approaches for cancer classification.

3.2. Description of microarray cancer data

Initially, the data is collected from the benchmark datasets, which are depicted as follows.

Dataset 1 (Haznedar et al., 2017; Statnikov et al., 2005):It is gathered from https://data.mendeley.com/datasets/ynp2tst2hh/4: Access Date: 2023–02–06. The dataset is Microarray Gene Expression Cancer Data. It is mostly utilized for the categorization of microarray cancer data and is collected from Rutgers University. It is utilized to record the expression stage of thousands of genes simultaneously. It contains only a small set of genes that are appropriate for cancer recognition. The total number of samples available in this dataset is 1627.

Dataset 2 (Fiorini, 2016; García-Díaz et al., 2020): It collects from the given link https://archive.ics.uci.edu/dataset/401/gene+expres: Access date: 2023–02–06. The dataset is Gene Expression Cancer RNA-Seq Dataset. RNA-Seq (HiSeq) PANCAN data set is part of the dataset collection, which includes the gene expressions of the patients having multiple types of tumors called PRAD, BRCA, COAD, LUAD, and KIRC. It has 801 amounts of instances and 20531 attributes. The tasks which are associated with this dataset are clustering and classification. Each sample's RNA-Seq level of the gene expression, as determined by the Illumina HiSeq platform, is its variable (attribute). Table 2 shows the number of features and instances in Datasets 1 and 2.

The needed data is gathered from the above-mentioned datasets that are expressed$I_x$here$x = 1,2,\ldots,X$; in turn, the variable Xdefines the total number of collected data.

Adaptability of the model in handling both types of data: The model is adaptable for handling both types of data because it processes standardised gene expression features, which are numerical values representing gene activity levels. The underlying input to the network is process the gene expression vectors, which are comparable regardless of whether they originate from DNA microarray data or RNA-seq data, and further they are pre-processed into a consistent format.

Input to the Network for Both Data Types: For DNA microarray data, the input consists of gene expression levels obtained from microarray experiments, formatted as a feature vector where each feature corresponds to a specific gene's expression level. For RNA-seq data, after initial processing (such as normalization and feature selection), the input similarly becomes a numerical feature vector of gene expression levels. In essence, the model's input layer accepts numerical vectors of gene expression data. Since both datasets are converted into this common format, the model operates on the features directly, making it flexible and adaptable to handle both data types effectively. The key is the feature extraction and pre-processing steps that standardize the input data, rather than relying on the original data acquisition architecture.

3.3. Pre-processing

The raw data$I_x$is given as input to the pre-processing step. It is a step where the data analysis process and data mining convert the raw data into a format that machine learning algorithms and computers can evaluate and understand. It has two steps. One is NAN removal, and the other is missing value removal.

NAN removal: The first step of the data pre-processing is called NAN removal. Here, the original data $I_x$is given as input. NaN, or Not a Number, is a special value used in data frames and numpy arrays to indicate a cell's missing value. Due to this, it affects the data quality and does not have the potential to determine the essential features. To overcome this, these values are removed and acquiredI_x^{nan}.

Missing Value Removal: This is the second phase of the pre-processing data. Because of the missing value, the data is not being collected properly. So, it creates collection and management errors. The elimination of missing values from the dataset may be one way to handle the problem. Here, the input is taken asI_x^{nan}, where the missing value is removed, and the final pre-processed data is indicated byI_x^{pre}.

4. Modified Sandpiper optimization algorithm-based optimal gene selection for microarray cancer classification

4.1. Proposed MSOA

Motivation for using the MSOA algorithm: The motivation for developing the MSOA arises from the need to overcome the inherent limitations of the original Sandpiper Optimization Algorithm (SOA), which was initially designed to address complex optimization problems. While SOA demonstrated superior performance in evaluating standard

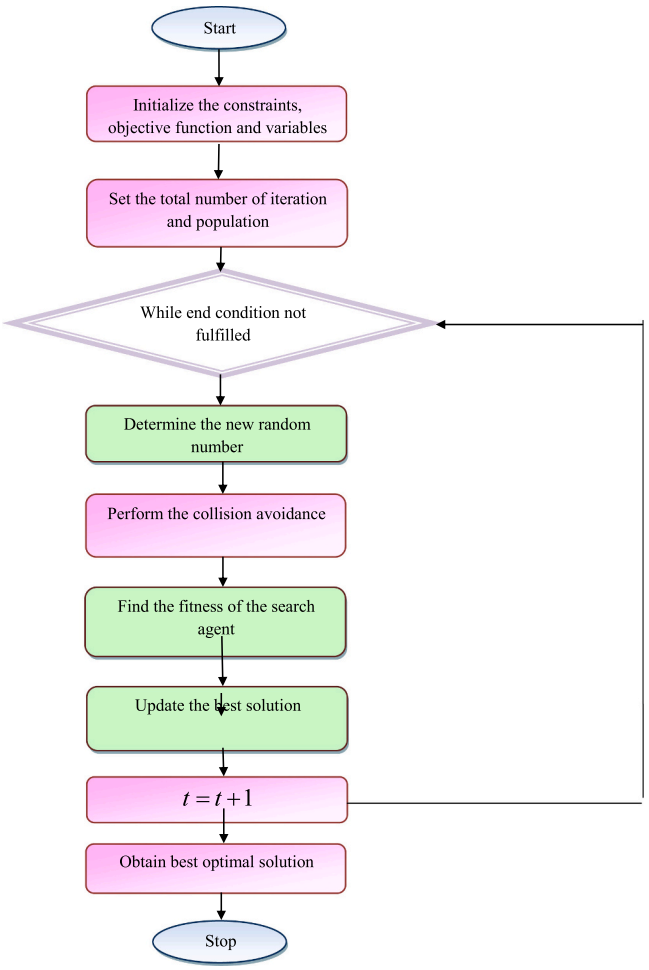


Fig. 2. Flowchart of developed MSOA.

test functions compared to other algorithms, it exhibited several drawbacks that hinder its effectiveness in real-world applications, particularly in high-dimensional and intricate problems such as gene selection for cancer classification. These limitations include a tendency to converge prematurely, getting trapped in local minima, and lacking the capability to handle multi-objective and binary optimization problems directly, which are crucial aspects when dealing with gene expression data, where selecting optimal gene subsets involves multi-faceted criteria. Additionally, the original SOA suffers from relatively low response times, limiting its efficiency for large datasets and time-sensitive scenarios. Recognizing these challenges, researchers proposed MSOA to enhance the exploration and exploitation balance, improve convergence speed, and incorporate multi-objective and binary functionalities effectively. These improvements ensure that the algorithm can better navigate complex search spaces, avoid suboptimal solutions, and provide more accurate and stable outcomes. Consequently, MSOA offers a more robust and versatile optimization tool suitable for critical tasks such as optimal gene selection and parameter tuning in cancer diagnosis frameworks, ultimately contributing to higher classification accuracy and more reliable results in computational bioinformatics applications. Based on the cancer classification, various researchers are explored to show their significant performance. However, the existing research often fails to provide better convergence, and it also easily traps into the local minima problems. Based on these, an SOA algorithm is adopted to achieve its superior outcomes. Here, the resolution of optimization issues in the complex structure becomes challenging. To alleviate these problems, an MSOA algorithm is enhanced to provide the optimal performance. The developed MSOA algorithm is implemented for estimating the optimal solutions. The previous SOA algorithm evaluated forty-four standard test functions. The results of the SOA show that it performs better than the other competing optimized algorithms. However, it lacks multi-objective and binary versions, and it has a low response time. To overcome that, a new proposed MSOA algorithm has been implemented. The MSOA algorithm helped to evaluate the optimal gene and parameter tuning on the proposed HDLF model.

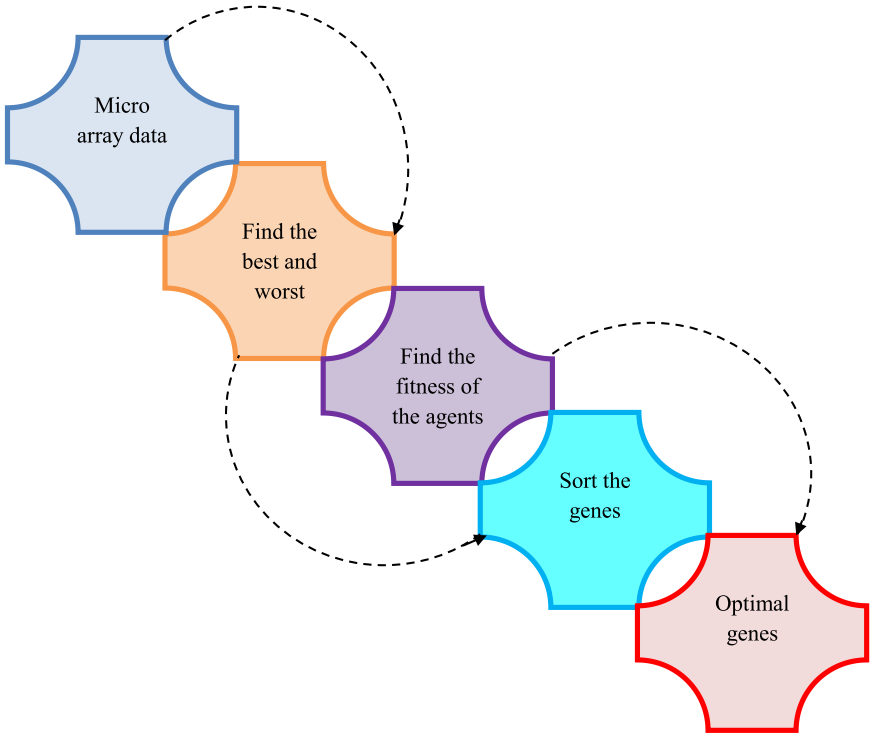


Fig. 3. Diagrammatic representation of the step-by-step process of gene selection using MSOA.

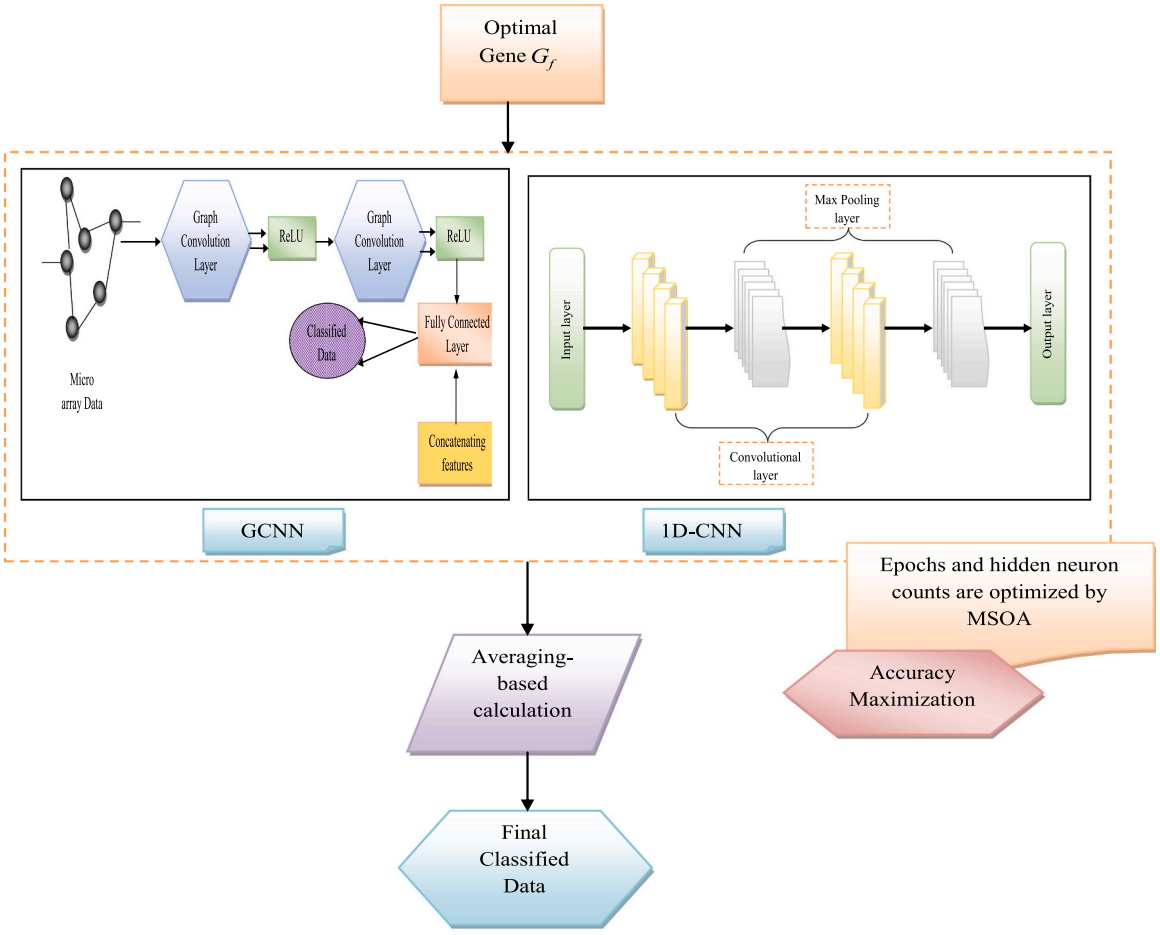


Fig. 4. Depicts HDLF for the proposed MSOA methodology.

SOA (Kaur et al., 2020): The Sandpiper algorithm is used to solve real-life problems. This algorithm is inspired by the attacking and migration characteristics of the sandpipers. The mathematical model of MSAO is explored below.

i) Collision avoidance: Here, an extra variable calledD_A is used for the estimation of the location for the new search agent to prevent collision among the neighbouring sandpipers is represented in Eq. (1)

$$\vec{D}_{PS} = D_A \times \vec{Q}_{PS}(w) \tag{1}$$

Here, the term $\vec{D}_{PS}$denotes the place of the search agent that didn't mix with the other search agent, $\vec{Q}_{PS}$ indicates the existing position of the search agent, w denotes the existing looping, and D_A refers to the motion of the search agent in the place of searching. In the SOA algorithm, there is a variable called $D_A$but it takes a random number. So it reduces the accuracy of the result. So, this proposed method uses a new $D_A$estimation that is expressed using Eq. (2).

$$D_A = \frac{CF - BF}{WF - BF} \times 2 \tag{2}$$

Here CFdenotes the current fitness, BFwhich is expressed as best fitness, and also WFrefers to the worst fitness.

The flowchart of the developed MSOA is shown in Fig. 2.

4.2. Optimal gene selection

The pre-processed data is represented asI_x. To effectively decrease the data, the feature selection method is applied in data pre-processing. This helps to locate the precise data models. For a model to predict the target variable, feature selection models aim to minimize input variables to those that are thought to be most helpful. It eliminates the duplicate or unused predictors from the model. Feature selection is a crucial and popular dimensionality reduction strategy for data mining that involves selecting the appropriate features based on specific criteria. In most situations, an exhaustive search for the ideal feature subset is not practicable. The most fundamental and difficult problems in feature selection are stable feature selection, optimal redundancy removal, and the use of auxiliary data and prior knowledge. To overcome these problems, a new feature called optimal gene selection usingthe MSOA algorithm. The utilization ofthe MSOA algorithm removes the unwanted genes to provide the optimal genes. The output of the optimal gene selection is referred to asG_f. There are 100 features that have been chosen before feature selection, and 50 features have been chosen after feature selection.

The following Fig. 3 explains the process of gene selection.

5. Advanced microarray cancer classification using a hybrid deep learning framework

5.1. Proposed hybrid deep learning framework with parameter optimization

The incorporation of GCNN with 1D-CNN in the proposed HDLF is driven by the need to effectively leverage the complementary strengths of both models to enhance cancer classification accuracy using microarray data. GCNNs are specially designed to handle graph-structured data, capturing complex relationships and interactions between genes that are not easily modelled by traditional Euclidean-based neural networks. By exploiting local connectivity, shift invariance, and the ability

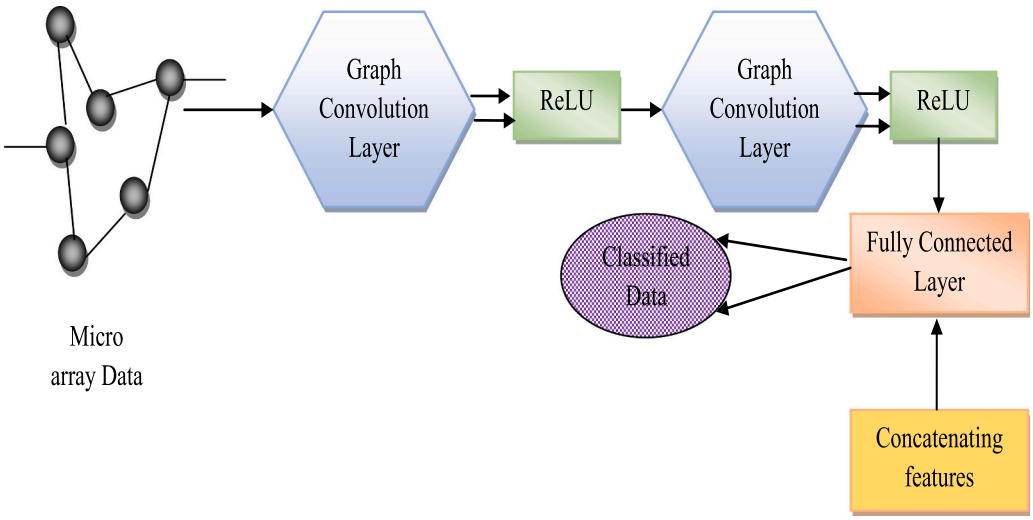


Fig. 5. The diagrammatic representation of GCNN.

to incorporate topological information, GCNNs can learn rich, context-aware feature representations from gene interaction networks, which are critical in understanding the intricate biological pathways involved in cancer. Meanwhile, 1D-CNNs excel in processing sequential or one-dimensional data, such as gene expression profiles, by automatically extracting salient spectral features and capturing local patterns within the gene expression sequences. In the integrated HDLF, the GCNN component first captures the relational and topological characteristics among genes, enhancing the feature space with biologically meaningful information. The 1D-CNN then processes the gene expression data to extract features based on local sequential patterns. These two modules are interconnected and trained jointly, including the parameters such as hidden neurons and epochs optimized via the MSOA. This synergy allows the framework to effectively model both the gene interaction networks and the sequential gene expression patterns, leading to more discriminative features and, consequently, improved cancer classification performance. The combined use of GCNN and 1D-CNN thus provides a comprehensive, multi-faceted approach for analyzing microarray data, capturing both structural and sequential information inherent in biological data, which significantly enhances the robustness and accuracy of the classification system. In this proposed model, GCNN and 1D-CNN have been combined to acquire the final classified data. CNN data representation and classification abilities are aimed to be enhanced by GCNNs. Because of the graph's arbitrary size and complex topology, there is no spatial locality, and CNN on graphs is particularly challenging to perform. Node ordering is also not fixed. CNNs leverage two key aspects that account for their effectiveness: local connection and shift in variance. The number of unknown parameters or weights that must be calculated is drastically reduced, and the computing cost is significantly reduced by parameter sharing. The 1D-CNN, a specific type of deep learning neural network, is developed specifically to handle one-dimensional input, such as time series data. Despite sharing many characteristics with regular CNNs, 1D-CNNs have a few key distinctions that make them the best choice for processing one-dimensional data. A CNN is good at spotting simple patterns in data that are then used to generate more complex patterns in higher layers. The primary benefits of the 1D-CNN-based method is suitable for real-time fault detection and monitoring. However, the CNN cannot effectively encode the position and orientation of objects without a large amount of trained data. The position and orientation of objects are not encoded. They struggle to categorize photos with various positions. The proposed hybrid deep learning framework is used with parameter optimization via MSOA. The optimization parameters like epochs in GCNN, hidden units in GCNN, Epochs in 1D-CNN, and hidden units in 1D-CNN. The Objective Function *OF* is presented in Eq. (3).

$$OF = \underset{\{ge, ep^{gcnn}, hn^{gcnn}, ep^{1dcnn}, hn^{1dcnn}\}}{\arg\max} \left[Acy \right] \tag{3}$$

The gene selection is represented by*ge*, an epoch value of GCNN denoted byep^{gcnn}, also the hidden units of GCNN are represented byhn^{gcnn}. Similarly, an epoch value of 1DCNN is denoted byep^{1dcnn}; also, the hidden units of 1DCNN are represented byhn^{1dcnn}. The gene selection has 10 values, which range from 0 to the data length. The epoch for GCNN limits from 50 to 100. The hidden units of GCNN range from 5 to 255. The epoch for GCNN limits from 50 to 100. Also, the hidden units of 1D-CNN range from 5 to 255. Here, the parameters like epochs and hidden neuron count are in the ranges of 52 and 145. Here, the ReLu activation function is performed. Moreover, the batch size lies between 32, and the learning rate of 0.01 is considered for the training process. Further, the term *Acy* defines the accuracy measure as how close a given set of calculations is to their true value, and it is computed by Eq. (4).

$$Acy = \frac{XY + NM}{XY + NM + KL + AS} \tag{4}$$

The terms*AS*and *KL*are denoted as false negative and false positive values. Also, the values of true negative and true positive represent*NM*and*XY*. The upcoming Fig. 4 shows the framework of HDLF.

5.2. Graph convolutional neural network

CNN is very useful for signals explained on regular Euclidean domains. GCNNs attempt to improve the data indication and categorizing abilities of GCNN (Zhang et al., 2021) of individuals on the spectrum using the GCNN classifier into four groups: late mild cognitive impairment, early mild cognitive impairment, AD, and cognitively normal. By margins that depend on the illness category, the GCNN classifier surpasses an SVM classifier.

GCNN is a cutting-edge technology for topic staging and AD spectrum categorization. Local connection and shift in variance are two essential characteristics that CNNs make use of and which explain their effectiveness. Based on the receptive places that work in nearby neighborhoods, CNN feature extraction. To take advantage of translation invariance, this results in global parameter sharing across spatial regions. In neural networks, parameter sharing greatly minimizes the number of unknown parameters or weights that must be computed during the training stage and dramatically lowers the computational cost. Graph convolution, a linear layer, and a nonlinear activation function are the three crucial parts of a GCN. The GCN is a method for

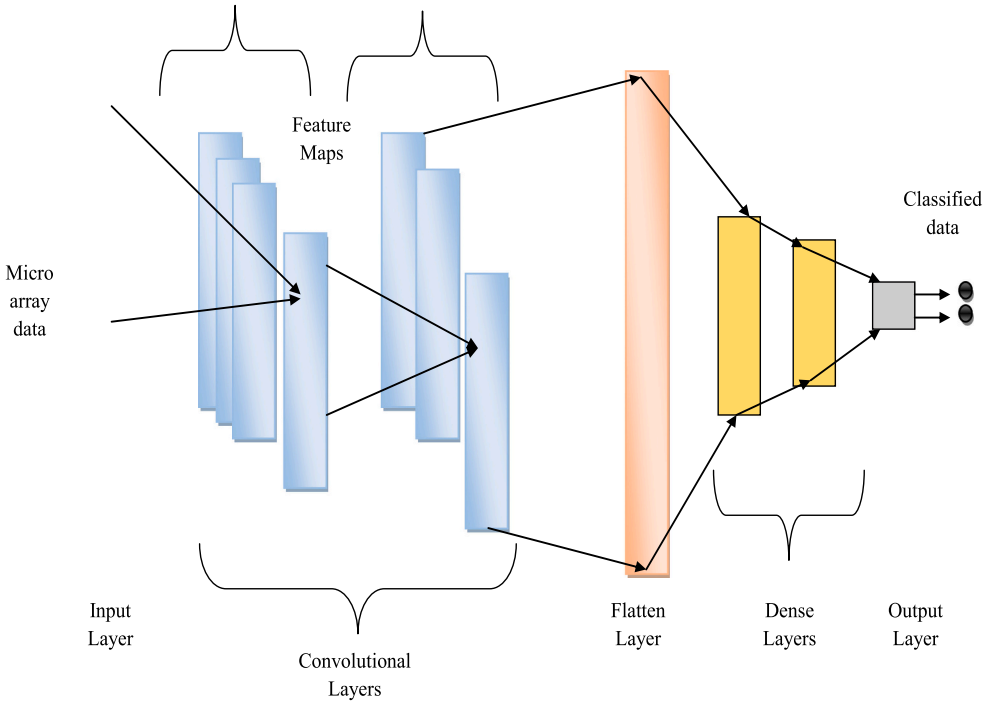


Fig. 6. The architecture of 1D-CNN is used in the proposed method.

Table 3
Simulation parameters of the designed method for the cancer classification framework.

Algorithm	Parameters	Values
SOA	Iteration	25
	Number of population	10
	Sandpiper control frequency	2–0
Proposed MSOA	Total number of population	10
	Maximum iteration	25

learning graph-structured data while under semi-supervised supervision. The diagram of the GCNN architecture is shown in Fig. 5.

5.3. 1D-CNN

One particular kind of deep learning neural network, known as a 1D-CNN (Mostavi et al., 2020), is created expressly to handle one-dimensional input, including audio signals or time series data. Although 1D-CNNs are similar to standard CNNs, they differ in a few significant ways that make them ideal for handling one-dimensional data. A CNN is useful for finding basic patterns in information that are used to create more significant patterns in higher layers. The outcomes show that when used on spectroscopic datasets, the 1D-CNN had an average performance that was better than the other methods examined. The main benefit of CNN is that it provides an analytical way to directly extract characteristics from the input data's raw form. Its classification and learning skills surpass those of conventional neural networks. As a result, a 1D-CNN method for the extraction of spectral character is introduced. The CNN contains different layers like pooling, convolution, and fully connected. Compared to the other AI methods, CNN gives the results. After the augmentation of the data, the specificity and sensitivity of the 1D-CNN methods have improved. Hence, the consideration of this flattened layer in the 1D-CNN model effectively minimizes the data dimensionality issues and could significantly reduce the number of parameters in the fully connected layers. Finally, the accurate classification performance is achieved using the learned features, which shows accurate performance. To acquire dimensionality-reduced feature data, the convolutional layer is located behind the input layer, where the local feature extraction is carried out. The convolution layer $d(i)$is expressed below in Eq. (5).

$$d(i) = Y \otimes G_i, \forall i \in [1, \ldots, I] \tag{5}$$

The convolution layer is followed by the pooling layer and has the ability to further reduce the feature vector's dimensionality, improve the network's robustness, and retrieve lower-resolution feature data. Thus, it prevents overfitting during training and boosts the network's common efficiency. Fig. 6 shows the diagrammatic representation of a 1D-CNN.

Reshaping process in the flatten layer: The flatten layer in the 1D-CNN model plays a crucial role in reshaping the extracted feature maps into a one-dimensional vector, making them compatible for subsequent fully connected layers. In our implementation, after the convolutional and pooling layers, the feature maps are flattened using the Flatten () function in Python's Keras library. Specifically, the parameters involved include the dimensions of the feature maps prior to flattening, which depend on the input data size, the kernel size, the stride, and the pooling parameters are set during the convolution and pooling operations. For example, assuming the input data has a length of 10,000 features, and the convolutional layer utilizes a kernel size of 3 with stride 1 and 'valid' padding, the output size after convolution would be (10,000–3 +1)= 9998 features per filter. If 64 filters are applied, the resulting feature map would have dimensions (9998, 64). When passing through a pooling layer (e.g., max pooling with pool size 2), the dimensions reduce further by half, resulting in an output shape of approximately (4999, 64). The flattened layer then reshapes this multi-dimensional tensor into a 1D vector of size $4999 \times 64 = 319{,}936$ features, which serve as input to the dense layers. This reshaping process minimizes data dimensionality issues by transforming complex feature maps into a manageable vector form, significantly reducing the number of parameters in subsequent layers.

6. Simulation findings

6.1. Implementation platform

This designed cancer classification method using microarray and seq

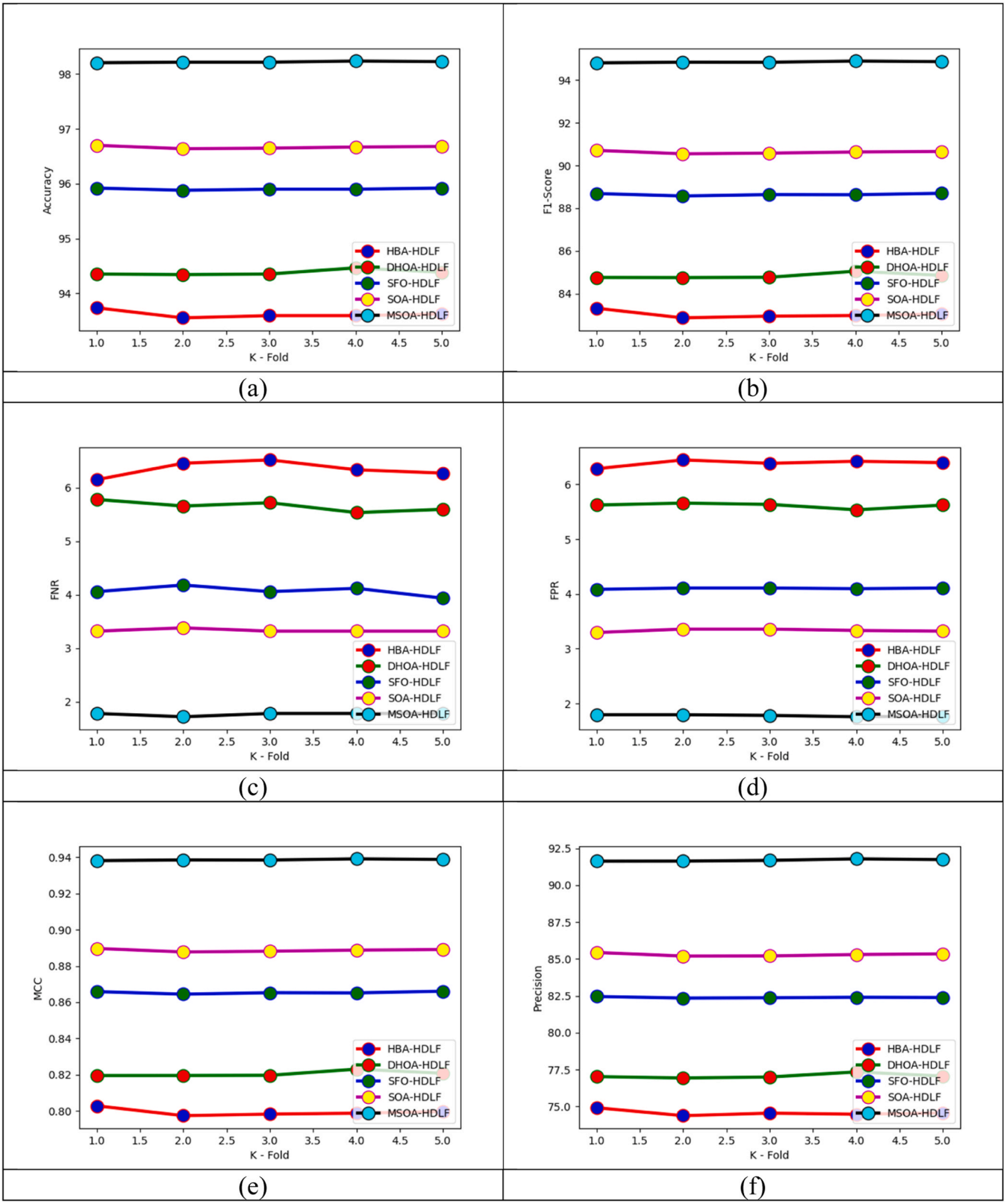


Fig. 7. K-fold performance evaluation of developed cancer classification method using microarray and seq data compared with dataset 1 concerning (a) Accuracy, (b) F1-Score, (c) FNR,(d) FPR,(e) MCC and (f) Precision.

data was implemented in the Python platform. Based on the maximum iteration and population size, the offered model is validated, which shows the value of 25 and 10. Thus, the comparison algorithms like Honey Badger Algorithm (HBA)-HDLF (Vaiyapuri et al., 2022), Deer Hunting Optimisation Algorithm (DHOA)-HDLF (Brammya et al., 2019), Sunflower Optimization (SFO)-HDLF (Gomes et al., 2019), and SOA-HDLF (Kaur et al., 2020) were taken. Consequently, conventional classifier models were considered as DT (Fathi et al., 2021), DBN

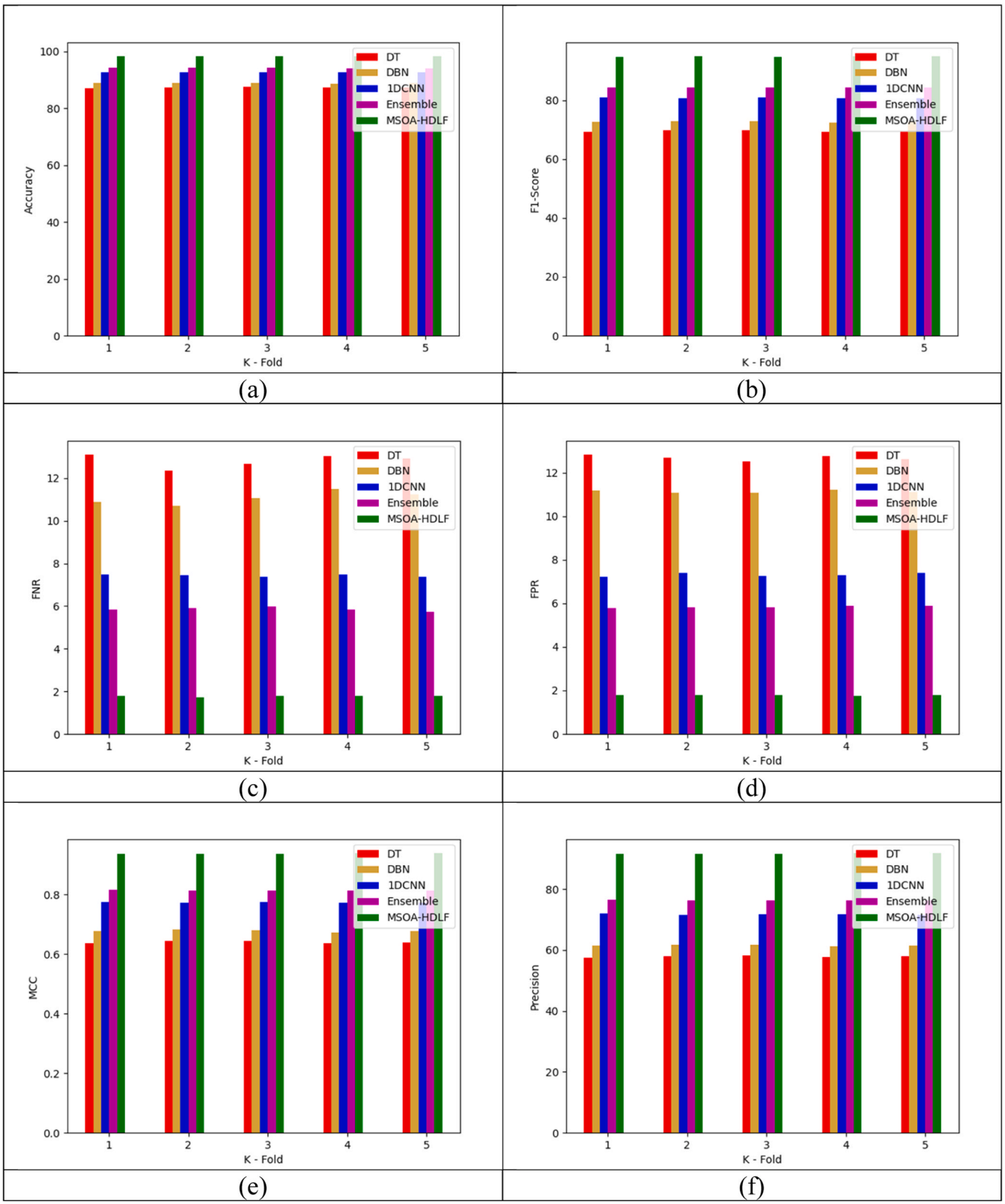


Fig. 8. K-fold performance evaluation of the proposed method using microarray and seq data compared with a traditional classifier model for dataset 1 in terms of(a) Accuracy, (b) F1-Score, (c) FNR,(d) FPR,(e) MCC and (f) Precision.

(Kourou et al., 2019), 1D-CNN (Mostavi et al., 2020), and Ensemble (Haznedar et al., 2021), respectively. The data is split into two phases: training and testing. Further, 75 % of the data is validated in the training phase and the remaining 25 % of the data is given in the testing phase. The simulation parameter for the developed model is listed in Table 3.

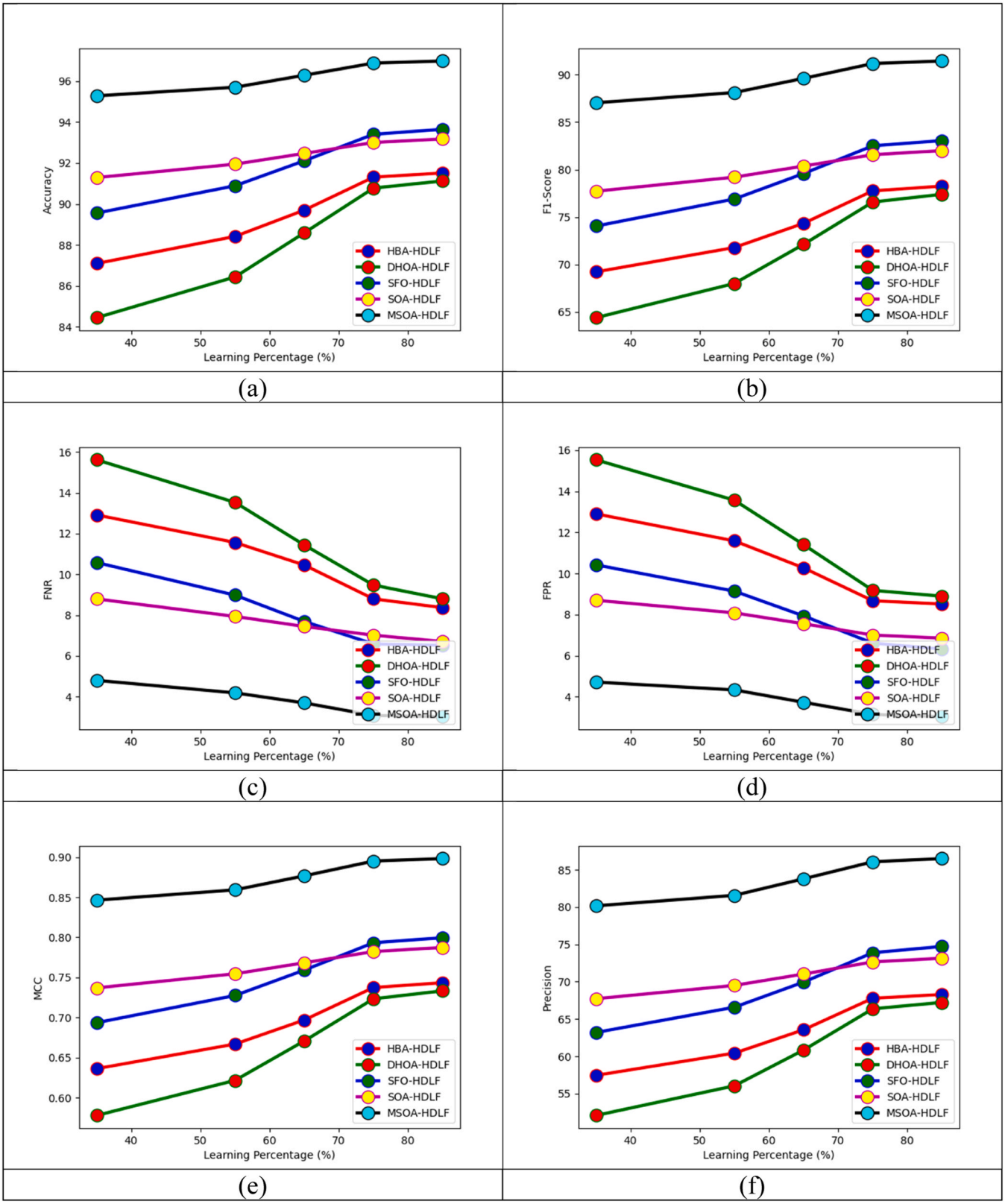


Fig. 9. Performance analysis of the developed method using microarray and seq data for dataset 1 concerning (a) Accuracy, (b) F1-Score, (c) FNR,(d) FPR,(e) MCC, and (f)Precision.

6.2. *Performance measures*

Precision: The localised results and the related value of the anomaly detected are known as precision.

$$A_s = \frac{XY}{XY + AS} \tag{6}$$

Specificity: Specificity is estimated by the probability of a negative

Fig. 10. Performance evaluation of proposed cancer classification method using microarray and seq data compared with traditional classifier model for dataset 1 concerning (a) Accuracy, (b) F1-Score, (c) FNR,(d) FPR,(e) MCC, and (f) Precision.

rate.

$$spec = \frac{NM}{NM + AS} \tag{7}$$

FPR and FNR: The False Positive Rate evaluates the value that is identified by mistake. On the other side, the False Negative Rate estimates the abnormalities not correctly, even if it has the images.

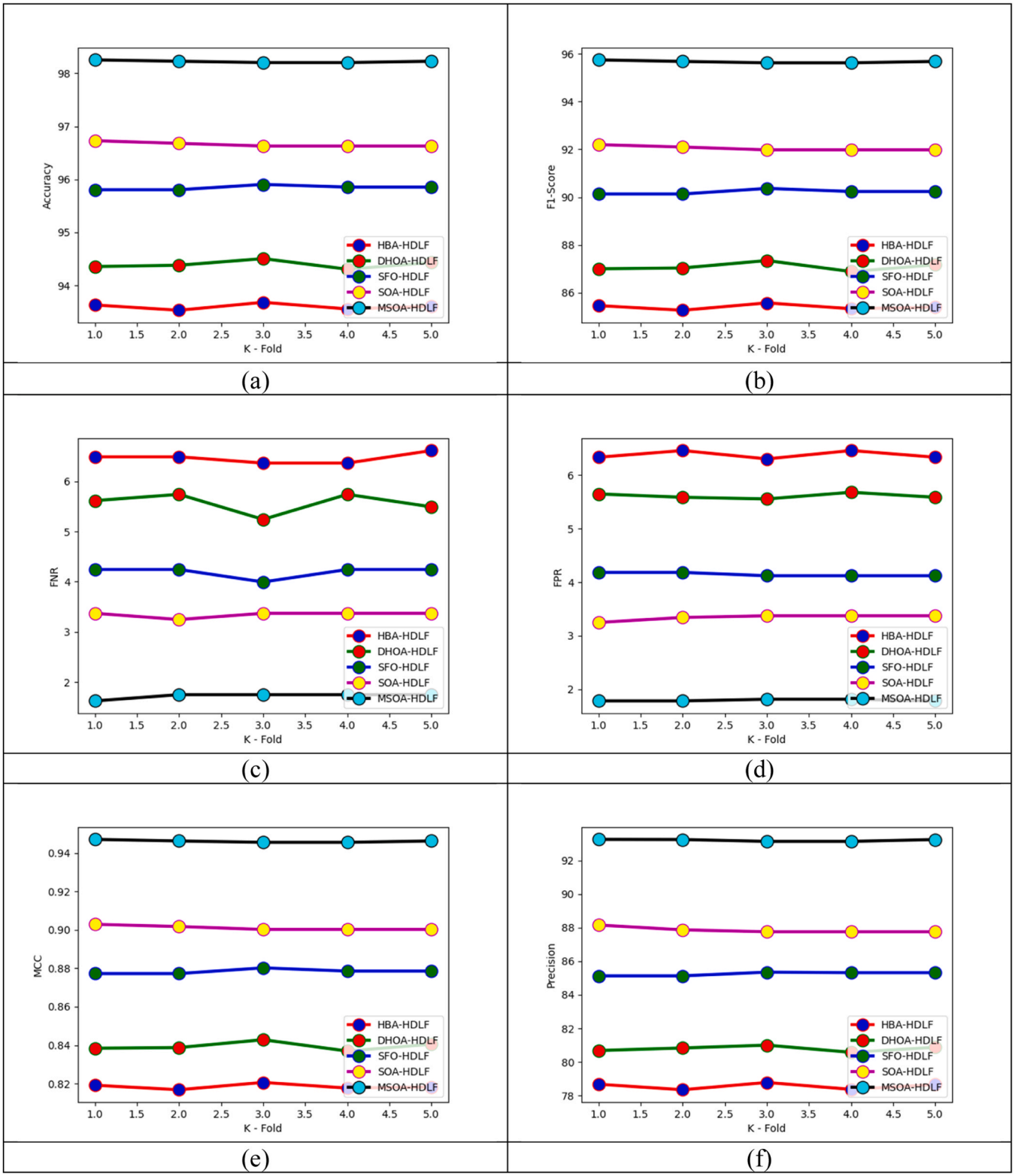


Fig. 11. K-fold performance evaluation of developed cancer classification model for dataset 2 regarding(a) Accuracy, (b) F1-Score, (c)FNR,(d)FPR,(e)MCC, (f)Precision.

$$FPR = \frac{KL}{KL + XY} \tag{8}$$

$$FNP = \frac{NM}{NM + XY} \tag{9}$$

Recall: It is the Metric that calculates the number of correct positive values out of all positive rates.

$$\mathrm{Re} = \frac{XY}{XY + AS} \tag{13}$$

F1-Score: The ratio between the harmonic value of recall as well as precision.

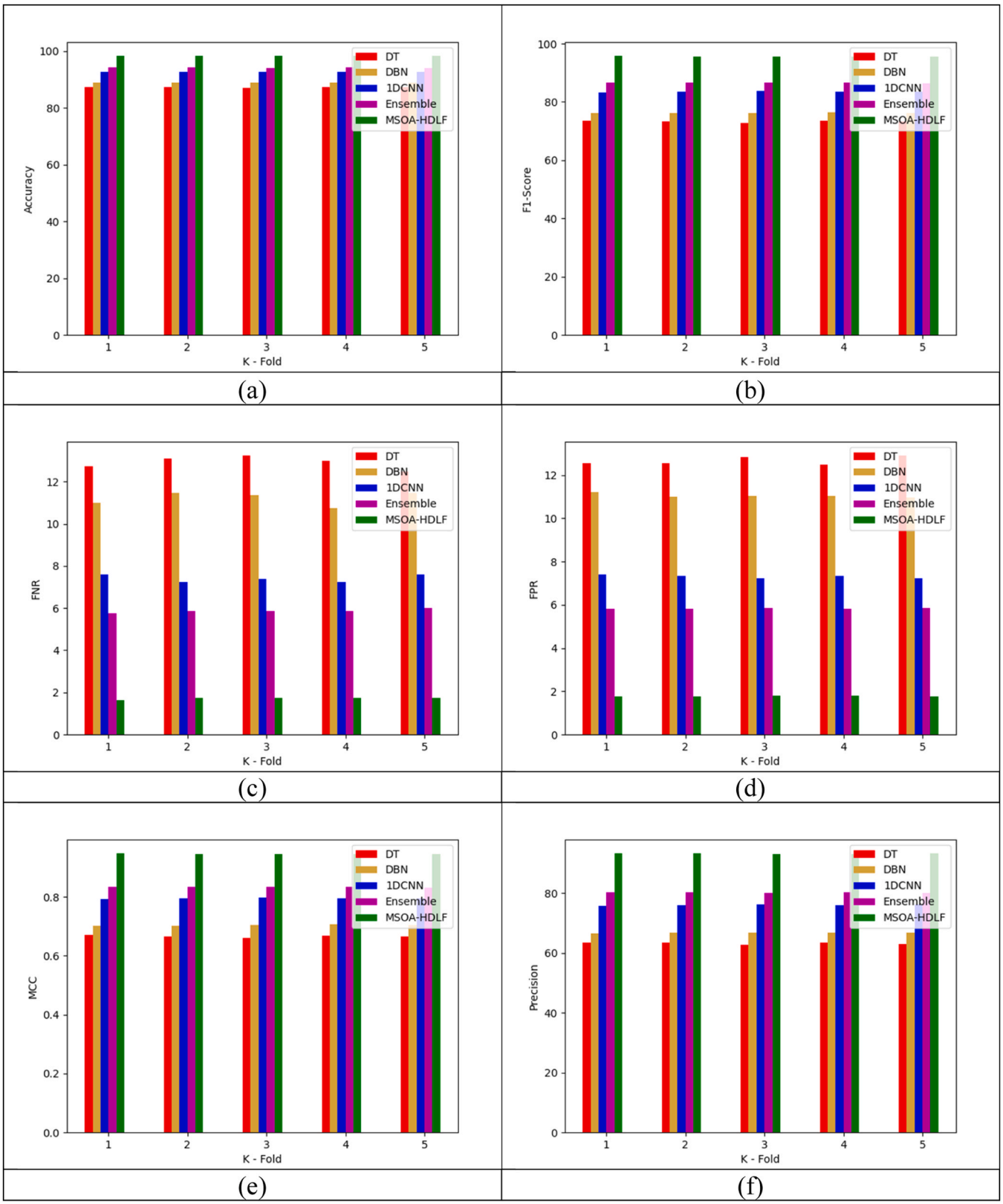


Fig. 12. K-fold performance evaluation of proposed cancer classification model using microarray and seq data for dataset 2 concerning (a) Accuracy, (b) F1-Score, (c) FNR, (d) FPR,(e) MCC and (f) Precision.

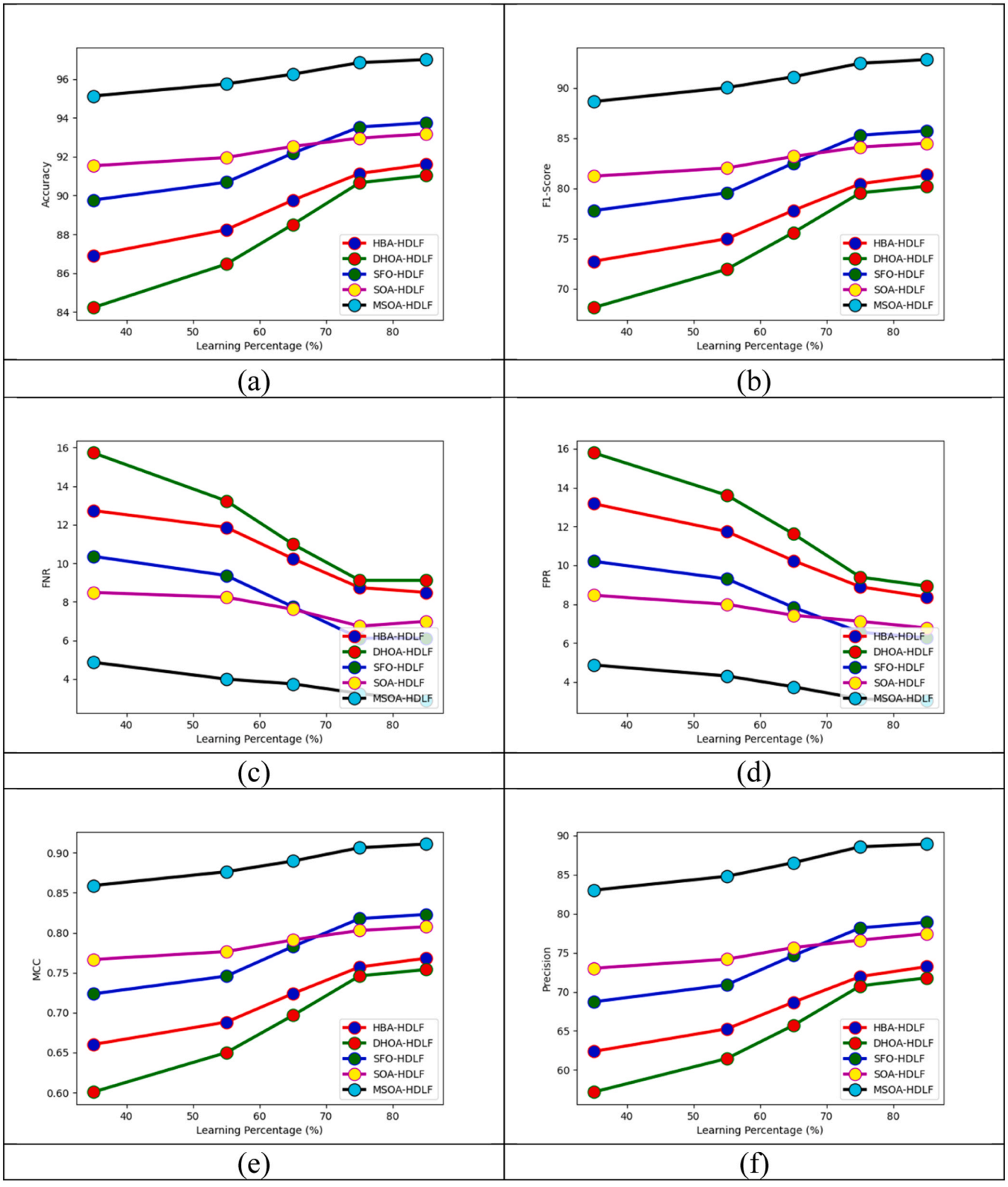


Fig. 13. Performance evaluation of proposed cancer classification model using microarray and seq data for dataset 2 in terms of(a) Accuracy, (b) F1-Score, (c) FNR, (d) FPR,(e) MCC, and (f)Precision.

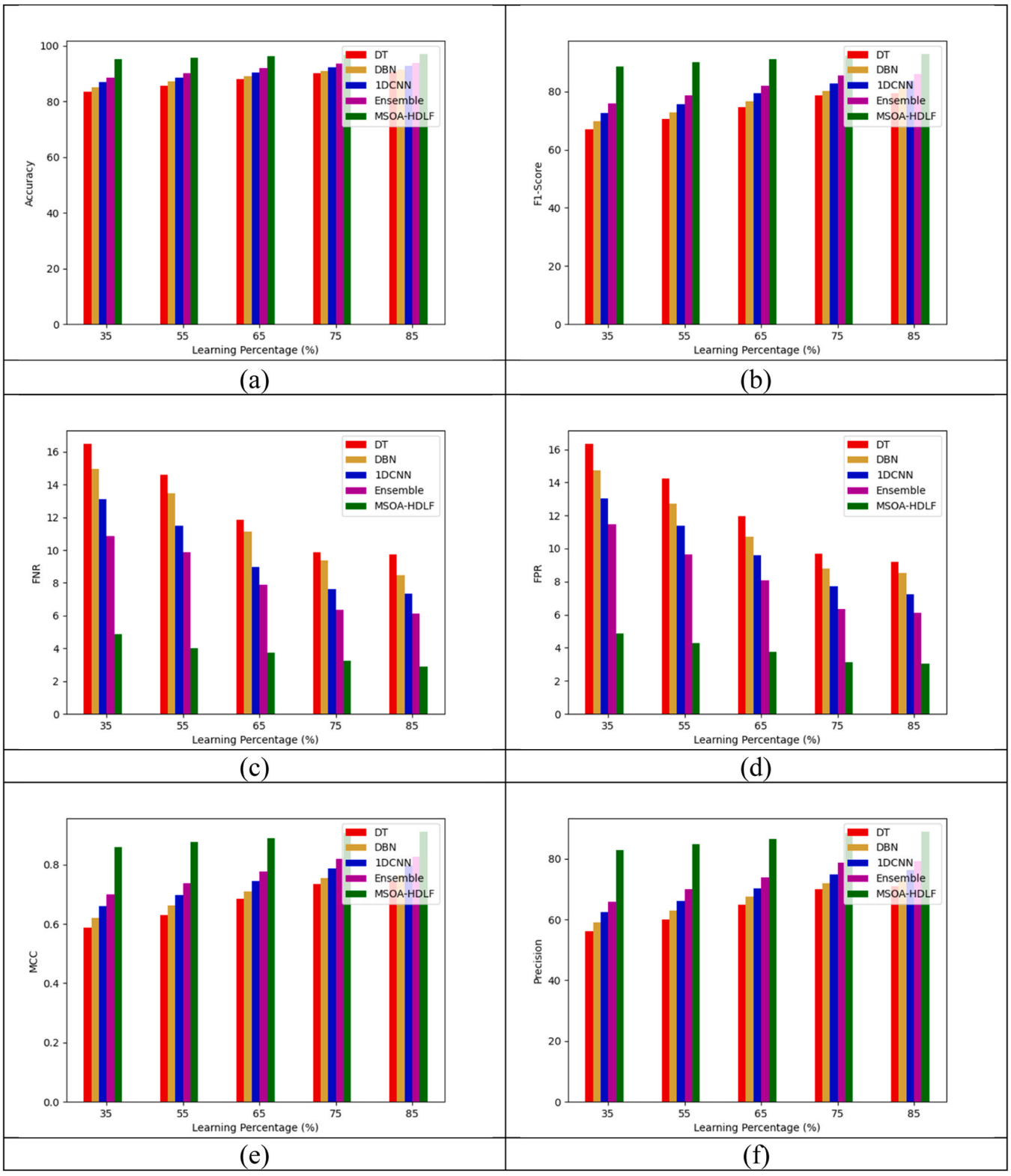


Fig. 14. Performance analysis of proposed model using microarray and seq data for dataset 2 regarding (a) Accuracy, (b) F1-Score, (c) FNR,(d) FPR,(e) MCC, and (f) Precision.

Table 4
Performance validation of the suggested cancer classification model using microarray and seq data over algorithms.

Measures	HBA-HDLF (Vaiyapuri al., 2022)	DHOA-HDLF (Brammya al., 2019)	SFO-HDLF (Gomes et al., 2019)	SOA-HDLF (Kaur et al., 2020)	MSOA-HDLF
Dataset 1					
Accuracy	91.51292	91.12341	93.64494	93.17343	96.97622
FDR	31.71402	32.77425	25.2704	26.85632	13.49424
Sensitivity	91.63592	91.20541	93.48093	93.29643	96.98647
FNR	8.364084	8.794588	6.519065	6.703567	3.01353
MCC	0.743435	0.733322	0.799606	0.787418	0.898337
FPR	8.511685	8.892989	6.322263	6.851169	3.02583
Specificity	91.48831	91.10701	93.67774	93.14883	96.97417
Precision	68.28598	67.22575	74.7296	73.14368	86.50576
NPV	91.48831	91.10701	93.67774	93.14883	96.97417
F1-Score	78.2563	77.40084	83.06011	82	91.4468
Dataset 2					
Accuracy	91.61049	91.0362	93.7578	93.18352	97.00375
Sensitivity	91.51061	90.88639	93.88265	93.00874	97.12859
Specificity	91.63546	91.07366	93.72659	93.22722	96.97253
Precision	73.22677	71.79487	78.90871	77.44283	88.91429
FPR	8.364544	8.926342	6.273408	6.772784	3.027466
FNR	8.489388	9.113608	6.117353	6.991261	2.871411
NPV	91.63546	91.07366	93.72659	93.22722	96.97253
FDR	26.77323	28.20513	21.09129	22.55717	11.08571
F1-Score	81.35405	80.22039	85.74686	84.51503	92.8401
MCC	0.768134	0.753941	0.82295	0.807444	0.910922

Table 5
Comparative analysis of the suggested cancer classification model using microarray and seq data over techniques.

Measures	DT (Fathi al., 2021)	DBN (Kourou et al., 2019)	1D-CNN (Mostavi al., 2020)	Ensemble (Haznedar al., 2021)	MSOA-HDLF
Dataset 1					
Sensitivity	90.77491	91.57442	92.68143	94.09594	96.98647
Specificity	90.72571	91.53752	92.65683	93.97294	96.97417
Accuracy	90.73391	91.54367	92.66093	93.99344	96.97622
NPV	90.72571	91.53752	92.65683	93.97294	96.97417
MCC	0.723322	0.743933	0.773284	0.810017	0.898337
F1-Score	76.55602	78.3066	80.80429	83.92759	91.4468
Precision	66.18834	68.39688	71.62548	75.74257	86.50576
FPR	9.274293	8.462485	7.343173	6.02706	3.02583
FNR	9.225092	8.425584	7.318573	5.904059	3.01353
FDR	33.81166	31.60312	28.37452	24.25743	13.49424
Dataset 2					
Sensitivity	90.26217	91.51061	92.63421	93.88265	97.12859
FPR	9.17603	8.520599	7.209738	6.117353	3.027466
FNR	9.737828	8.489388	7.365793	6.117353	2.871411
Accuracy	90.71161	91.48564	92.75905	93.88265	97.00375
Specificity	90.82397	91.4794	92.79026	93.88265	96.97253
F1-Score	79.53795	81.12894	83.65276	85.992	92.8401
Precision	71.09145	72.86282	76.25899	79.32489	88.91429
NPV	90.82397	91.4794	92.79026	93.88265	96.97253
FDR	28.90855	27.13718	23.74101	20.67511	11.08571
MCC	0.745175	0.765422	0.796753	0.825911	0.910922

$$F1Score = 2 \times \frac{XY \times KL}{XY + KL} \tag{10}$$

FDR: False Discovery Rate is estimated by defining the ratio of false positive and true negative and false positive.

$$FDR = \frac{KL}{XY + KL} \tag{11}$$

NPV: Negative Predictive Value is estimated by the ratio between true negative and true negative and false negative.

$$NPV = \frac{NM}{NM + AS} \tag{12}$$

MCC: It evaluates the difference between the detected image output and actual image.

$$MCC = \frac{XY \times NM - KL \times AS}{\sqrt{(XY + KL)(XY + AS)(NM + KL)(NM + AS)}} \tag{13}$$

6.3. K-fold analysis of the designed cancer classification model using Dataset 1

Fig. 7 and Fig. 8 visualise the K-fold validation of diverse algorithms and classifiers for dataset 1 is validated using the developed model. The validation is carried out with standard measures to provide efficiency in the developed model. While evaluating with the K-fold analysis, it helps to build the recommended framework into a more generalized one. Moreover, it avoids the problem of overfitting. For the evaluation of k-fold analysis, we have divided the whole dataset into 5 sets. For example, the K-fold is performed based on a total of 100 data. Here, the 1-fold takes 1–20, followed by the 2-fold considered 21–40, followed by the 3-fold shows 41–60, followed by the 4-fold takes 61–80, and finally the 4-fold contains 81–100. If the 1-fold analysis is considered, the testing is performed on 1 set, and the rest of the data are in the process of training. This execution helps to maximize the performance. This process is repeated until better solutions are attained. Considering the algorithm-based analysis, the F1-score of the proposed model is enhanced by 13.2 % of HBA-HDLF, 11.2 % of DHOA-HDLF, 7.3 %, and 4.9 % of SFO-HDLF and SOA-HDLF. However, these evaluations are utilized to provide accurate outcomes. Here, the existing HBA-HDLF shows a higher error rate. Certainly, it leads to misclassification errors that affect the system's performance. The classifier-based analysis shows that the performance of the offered model achieves 65 %, 60 %, 37 %, and 23 % than DT, DBN, 1D-CNN, and Ensemble regarding precision. The entire validation shows that the recommended method attained enhanced performance.

6.4. Performance evaluation of the developed model using Dataset 1

The overall validation for the cancer classification model is suggested to show better outcomes of the designed MSOA-HDLF model. The effective analysis is provided based on diverse methods for dataset 1 is validated and it is represented in Figs. 9 and 10. Here, the learning percentage-based analysis is provided for both existing approaches. While considering the accuracy analysis, the learning percentage varies based on the different variations like 40, 50, 60, 70, and 80. Here, the FNR value of the proposed method is decreased by 18.42 % of HBA-HDLF, 22.10 % of DHOA-HDLF, 12.89 %, and 12.36 % of SFO-HDLF and SOA-HDLF. The existing SOA-HDLF shows the second greatest outcomes while considering the accuracy. Certainly, the DHOA and DT approaches cannot provide accurate outcomes, and they cannot handle larger datasets. The accuracy analysis recommends that the implemented MSOA-HDLF model offers 17 %, 15 %, 12 %, and 9 % better than DT, DBN, 1D-CNN, and Ensemble. The findings of the offered method proved the effective outcomes.

6.5. Validation of 5-fold of the offered cancer classification model for dataset 2

For dataset 2, the K-fold analysis is experimented with various approaches for the cancer classification model is visualized in Figs. 11 and 12. Also, the five sets of K-fold validation are considered, which helps to show the different variations of each set to prove its efficiency. Here, the 5-fold validation is a technique that is utilized for cross-validation. In 5-fold validation, 25 % of the data is utilized for testing. Cross-validation involves dividing a dataset into test and training data. In addition, the dataset is split to employ a cross-validation test. It is utilized to evaluate the expert systems and also to detect overfitting issues. It is mainly utilized for understanding the performance of the algorithm. Further, the performance of the developed method is improved by 4.9 % of HBA-

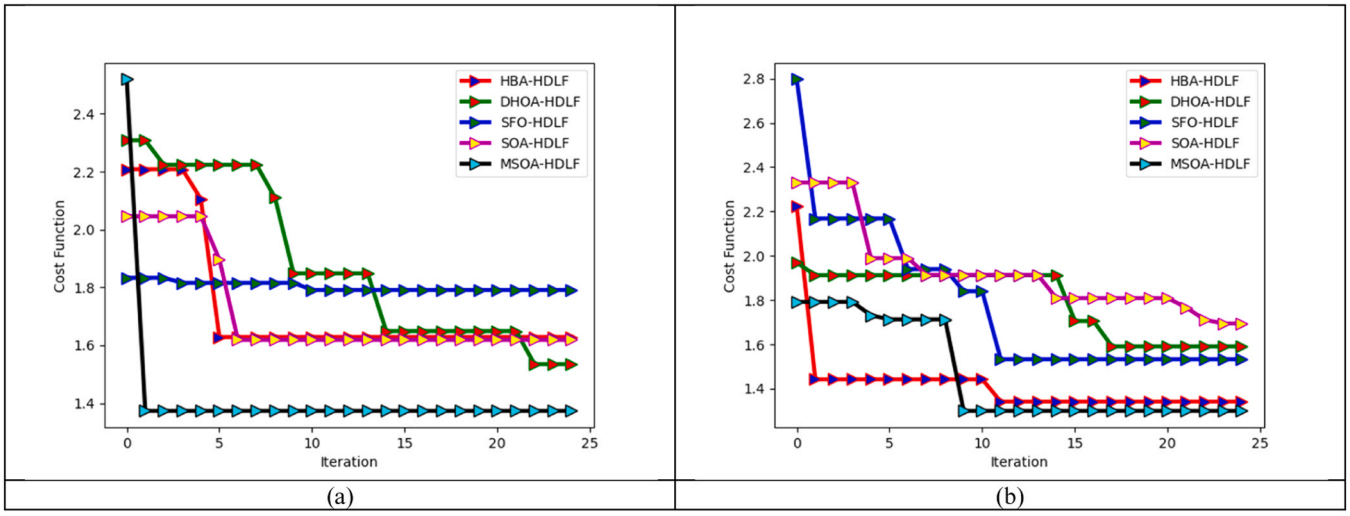


Fig. 15. Performance estimation based on convergence regarding (a) Dataset 1 and (a) Dataset 2.

Table 6
Overall performance analysis of the developed model.

Measures	DSCCNet (Maryam et al., 2023)	Inception ResNetV2 (Mamoona et al., 2023)	DCNNBT (Mohd et al., 2023)	MSOA-HDLF
Dataset 1				
Accuracy	93.34	93.66	93.97	96.97
Sensitivity	93.17	93.54	93.84	96.98
Specificity	93.38	93.69	93.99	96.97
Precision	73.79	74.77	75.76	86.50
FPR	6.617	6.309	6.002	3.025
FNR	6.826	6.457	6.150	3.013
NPV	93.38	93.69	93.99	96.97
FDR	26.205	25.22	24.23	13.49
F1-Score	82.35	83.11	83.84	91.44
MCC	79.14	80.03	80.89	89.83
Dataset 2				
Accuracy	93.56	93.96	94.28	97.00
Sensitivity	93.51	93.88	94.13	97.13
Specificity	93.57	93.98	94.32	96.97
Precision	78.43	79.58	80.56	88.91
FPR	6.43	6.02	5.68	3.03
FNR	6.49	6.12	5.87	2.87
NPV	93.57	93.98	94.32	96.97
FDR	21.57	20.42	19.44	11.09
F1-Score	85.31	86.14	86.82	92.84
MCC	81.74	82.77	83.61	91.09

Table 7
Ablation study for the recommendedMSOA-HDLF model.

Measures	1DCNN	GCNN	GCNN+ 1DCNN	MSOA-HDLF
Dataset 1				
Accuracy	92.06	93.15	94.04	96.98
Sensitivity	92.00	93.05	93.97	96.99
Specificity	92.07	93.17	94.06	96.97
Dataset 2				
Accuracy	92.48	93.31	94.26	97.00
Sensitivity	92.76	93.26	94.13	97.13
Specificity	92.42	93.32	94.29	96.97

HDLF, 3.8 % of DHOA-HDLF2.4 %, and 1.5 % of SFO-HDLF and SOA-HDLF regarding accuracy. While considering the error analysis in classifiers, the precision of the model shows 57 %, 52 %, 26 %, and 21 % improvement over DT, DBN, 1D-CNN, and Ensemble. This validation shows that the offered MSOA-HDLF-based cancer classification model provides better performance.

6.6. Performance estimation of the developed model for Dataset 2

Fig. 13 shows a superior analysis of the recommended approach in contrast with various algorithms. Here, the proposed model is improved over 24.0 % of HBA-HDLF, 32.5 % of DHOA-HDLF 18.0 %, and 12.0 % of SFO-HDLF and SOA-HDLF based on precision. Fig. 14 shows the performance evaluation of the recommended model in contrast with the existing classifier model for dataset 2. Considering 35, the precision of the offered method is enhanced by 29.2 % of DT, 28.0 % of DBN, 25.6 % 1D-CNN, and 23.1 % of Ensemble accordingly. Thus, the proposed model is used for the effective and early detection of cancer to save the lifespan of humans. The proposed model can determine the level of a thousand genes in a single experiment to accurately determine the presence of cancer in humans, which assists clinicians in making decisions for cancer-affected patients.

6.7. Numerical analysis

The comparison of the designed cancer classification model on diverse metrics is given in Tables 4 and 5. The MSOA-HDLF shows better results than the other traditional methods. The proposed MSOA-HDLF methodology has 96.9 % accuracy. In addition, the FNR of the proposed cancer classification model is 3.01 % while the existing models, such as HBA-HDLF, DHOA-HDLF, SFO-HDLF, and SOA-HDLF, attained the FNR of 8.364 %, 8.79 %, 6.519 %, and 6.703 % in the cancer classification. The lower FNR of the suggested model indicates that it can effectively classify the cancer-affected patient to extend their overall life span.

6.8. Analysis based on Convergence

The developed model is validated based on the Convergence analysis in Fig. 15. As the iteration in the analysis increased, the cost function of the model would decrease automatically. The rapid convergence rate of the proposed algorithm is verified by Fig. 15, and it is attained because of the improved random parameter of the proposed MSOA. This improved random number helps reach the global optimal solution with minimum iteration value, and it does not fall into the condition of the local optimal while searching for the solution in the search space.

Table 8
Statistical analysis of the recommended framework.

Algorithms	HBA-HDLF (Vaiyapuri al., 2022)	DHOA-HDLF (Brammya al., 2019)	SFO-HDLF (Gomes et al., 2019)	SOA-HDLF (Kaur et al., 2020)	MSOA-HDLF
Dataset 1					
Mean	1.740188	1.884313	1.803032	1.715145	1.419061
Worst	2.208421	2.308278	1.833202	2.045684	2.519547
Standard Deviation	0.224602	0.276997	0.01555	0.173944	0.224636
Best	1.628263	1.534511	1.791018	1.618597	1.373207
Median	1.628263	1.848495	1.791018	1.618597	1.373207
Dataset 2					
Mean	1.417262	1.794925	1.783624	1.92823	1.461948
Best	1.341805	1.590551	1.532606	1.694477	1.299923
Standard Deviation	0.171732	0.151358	0.330178	0.194216	0.217222
Worst	2.224335	1.969047	2.796438	2.330428	1.791842
Median	1.341805	1.912086	1.532606	1.913051	1.299923

Table 9
Computational time analysis of the recommended method among algorithms.

TERMS	HBA-HDLF (Vaiyapuri al., 2022)	DHOA-HDLF (Brammya al., 2019)	SFO-HDLF (Gomes et al., 2019)	SOA-HDLF (Kaur et al., 2020)	MSOA-HDLF
Dataset 1					
Time (sec)	41.4577	49.4671	42.5783	45.8434	37.2147
Dataset 2					
Time (sec)	31.3557	36.5472	30.6898	31.8455	37.2147

Table 10
Analysis based on computational time using diverse classifiers.

TERMS	DT (Fathi al., 2021)	DBN (Kourou et al., 2019)	1D-CNN (Mostavi al., 2020)	Ensemble (Haznedar al., 2021)	MSOA-HDLF
Dataset 1					
Time (sec)	55.6843	58.4682	52.2356	47.3689	37.2147
Dataset 2					
Time (sec)	48.5478	50.7695	46.9432	41.4578	37.2147

6.9. Comparative validation using recent methods

The comparative analysis is performed to validate the recent methods for the cancer classification model for datasets 1 and 2 are tabulated in Table 6. The performance of the designed method achieves 13.5 %, 12.2 %, and 11.0 % better performance than DSCC_Net, InceptionResNetV2, and DCNNBT regarding MCC. Throughout the empirical analysis, the developed model shows effective outcomes. The results from the model is helpful for determining the most effective treatment for improving the life span of the patient affected with the cancer. The proposed model uses gene expression, which helps to offer effective therapies to the patient by making accurate predictions.

6.10. Ablation experiment for the developed cancer classification model

The ablation study conducted for the offered cancer classification framework is validated in Table 7. The ablation study showed that the recommended model attained a higher accuracy of 96.98 % when combining all models such as1DCNN, GCNN, and MSOA. The results from the ablation study showed that the hybrid model offers a precise level of accuracy in the cancer classification and also reduces the mortality rate of humans by suggesting effective treatment therapies.

Table 11
Comparision over prior algorithms.

TERMS	HBA-HDLF (Vaiyapuri al., 2022)	DHOA-HDLF (Brammya al., 2019)	SFO-HDLF (Gomes et al., 2019)	SOA-HDLF (Kaur et al., 2020)	MSOA-HDLF
Dataset 1					
Accuracy	91.32	90.08	95.04	93.12	97
Sensitivity	91.28	90.05	95.07	93.17	96.96
Specificity	91.36	90.11	95.02	93.07	97.04
Precision	90.91	89.61	94.75	92.72	96.88
FPR	8.64	9.89	4.98	6.93	2.96
FNR	8.72	9.95	4.93	6.83	3.04
NPV	91.36	90.11	95.02	93.07	97.04
FDR	9.09	10.39	5.25	7.28	3.12
F1-Score	91.10	89.83	94.91	92.95	96.92
MCC	82.63	80.15	90.07	86.23	94.00
Dataset 2					
Accuracy	93.58	91.58	89.86	95.4	97.3
Sensitivity	93.55	91.62	89.94	95.39	97.31
Specificity	93.61	91.54	89.78	95.41	97.29
Precision	93.58	91.51	89.76	95.39	97.28
FPR	6.39	8.46	10.22	4.59	2.71
FNR	6.45	8.38	10.06	4.61	2.69
NPV	93.61	91.54	89.78	95.41	97.29
FDR	6.42	8.49	10.24	4.61	2.72
F1-Score	93.57	91.57	89.85	95.39	97.30
MCC	87.16	83.16	79.72	90.80	94.60

6.11. Statistical performance of the designed framework using diverse algorithms

The statistical outcome of the recommended cancer classification model using existing algorithms is listed in Table 8. The different measures like best, worst, mean, median, and standard deviation. While considering the median-based analysis, the offered MSOA-HDLF model shows 3.1 %, 32.0 %, 15.1 %, and 32.% enhancements over HBA-HDLF, DHOA-HDLF, SFO-HDLF, and SOA-HDLF. These empirical findings suggest that the recommended model shows superior performance over the existing approaches. Here, the proposed model uses the microarray and seq gene expression data that is more efficiently processed by the deep learning model to get an efficient and effective classification outcome compared to the other models. In addition, the proposed model uses the MSOA for the parameter optimization that greatly reduces the chance of misclassification and reduces the error in the cancer classification process.

Table 12
Comparison over prior techniques.

TERMS	DT (Fathi al., 2021)	DBN (Kourou et al., 2019)	1D-CNN (Mostavi al., 2020)	Ensemble (Haznedar al., 2021)	MSOA-HDLF
Dataset 1					
Accuracy	90.12	94.16	92.28	95.92	97
Sensitivity	90.13	94.33	92.35	95.89	96.96
Specificity	90.11	94.00	92.21	95.95	97.04
Precision	89.62	93.71	91.82	95.73	96.88
FPR	9.89	6.00	7.79	4.05	2.96
FNR	9.87	5.67	7.65	4.11	3.04
NPV	90.11	94.00	92.21	95.95	97.04
FDR	10.38	6.29	8.18	4.27	3.12
F1-Score	89.87	94.02	92.09	95.81	96.92
MCC	80.23	88.32	84.55	91.83	94.00
Dataset 2					
Accuracy	92.58	90.66	94.36	96.34	97.3
Sensitivity	92.55	90.74	94.35	96.35	97.31
Specificity	92.61	90.58	94.37	96.33	97.29
Precision	92.58	90.56	94.35	96.31	97.28
FPR	7.39	9.42	5.63	3.67	2.71
FNR	7.45	9.26	5.65	3.65	2.69
NPV	92.61	90.58	94.37	96.33	97.29
FDR	7.42	9.44	5.65	3.69	2.72
F1-Score	92.56	90.65	94.35	96.33	97.30
MCC	85.16	81.32	88.72	92.68	94.60

Table 13
Comparison of data and parameters before and after optimization.

Model	Accuracy (Before Optimization)	Accuracy (After Optimization - MSOA)	Improvement (%)
1D-CNN	88.50 %	91.20 %	2.70 %
GCNN	87.30 %	90.80 %	3.50 %
HDLF	89.60 %	93.70 %	4.10 %

6.12. Computational time analysis

The analysis based on computational time is evaluated using the diverse methods for datasets 1 and 2 are provided in Tables 9 and 10. The proposed MSOA-HDLF model consumes the time of 37.21 sec in the cancer classification process. The lower computational time of the proposed model is mainly due to the incorporation of the hybrid model, which incorporates the advantages of both models, reducing the overall time involved in the cancer classification process.

6.13. Overall analysis of the proposed model

The significant analysis is made using the developed model, where the K-fold is performed to provide a reliable performance, which is depicted in Tables 11 and 12. Here, the classifiers and algorithms are evaluated using the offered MSOA-HDLF model for the cancer classification model. The performance of the developed model achieves 5.12 %, 7.3 %, 3.1 %, and 1 % to HBA-HDLF, DHOA-HDLF, SFO-HDLF, and SOA-HDLF. The findings show better performance in the developed model.

6.14. Comparison of data and parameters before and after optimization

The comparison of the proposed HDLF before and after the parameter optimization is given in Table 13. It can be seen from Table 13 that the proposed HDLF model provides an accuracy of 93.7 % and the after the optimization, thr accuracy of the HDLF model is 4.10 % lower if the optimization is not applied to the recommended model. So, the results indicate that the MSOA shows a great impact on the proposed HDLF, so it attained excellent accuracy in the cancer classification.

6.15. Comparison with recent models

Table 14 indicates the comparison of the proposed model with recent approaches. Here, the accuracy of the proposed model is 88.30 %, and precision of 90.32 %. The higher value of the proposed model indicates its efficiency in the cancer classification compared to the pre-trained CNN, unsupervised machine learning model, CNN, and Ensemble models.

6.16. Discussion on better results of the proposed model

The proposed method is capable of producing excellent results primarily due to its comprehensive and integrated approach that combines advanced feature selection, parameter optimization, and sophisticated deep learning architectures. The utilization of the MSOA for optimal gene selection ensures that only the most informative genes are chosen from high-dimensional microarray and seq data, reducing redundancy and noise that can adversely affect model accuracy. This targeted feature selection not only streamlines the data but also mitigates overfitting, leading to more reliable and generalizable models. Additionally, the systematic tuning of hyperparameters such as hidden neuron counts and epochs in both the GCNN and 1D-CNN via MSOA ensures that the models are configured at their optimal settings, further boosting classification performance. The hybrid deep learning framework leverages

Table 14
Comparison of the proposed model with recent models.

TERMS	Pretrained CNN (Shoaib et al. 2025)	Unsupervised machine learning model (Whig, et al. 2025)	CNN Nurtay, et al. (2025)	Ensemble (Haznedar al., 2021)	MSOA-HDLF
Dataset 1					
Accuracy	84.85378	87.82191	80.44522	82.71497	88.30205
Sensitivity	82.65993	85.28428	77.68663	79.722	86.38298
Specificity	87.21668	90.59361	83.58209	86.14232	90.32258
Precision	87.44435	90.82814	84.32769	86.82102	90.3829
FPR	12.78332	9.406393	16.41791	13.85768	9.677419
FNR	17.34007	14.71572	22.31337	20.278	13.61702
NPV	82.36301	84.93151	76.71233	78.76712	86.30137
FDR	12.55565	9.171861	15.67231	13.17898	9.617097
Dataset 2					
Accuracy	84.88132	87.7709	80.03096	82.97214	89.06089
Sensitivity	81.59923	85.24752	77.17602	80.0194	86.66667
Specificity	88.66667	90.51724	83.29646	86.32856	91.63987
Precision	89.25184	90.72708	84.08851	86.93361	91.78082
FPR	11.33333	9.482759	16.70354	13.67144	8.360129
FNR	18.40077	14.75248	22.82398	19.9806	13.33333
NPV	80.68756	84.93428	76.13751	79.17088	86.45096
FDR	10.74816	9.272919	15.91149	13.06639	8.219178

the strengths of both GCNN's ability to handle complex graph-structured data and 1D-CNN's proficiency with sequential data, providing a robust architecture capable of capturing intricate patterns within gene expression data. Moreover, the model's design incorporates rigorous validation techniques like K-fold cross-validation, which enhances its stability and reliability. The combination of these elements provides precise feature selection, automated hyperparameter tuning, and a hybrid architecture that contributes to the model's high accuracy, robustness, and adaptability, making it well-suited for complex tasks such as cancer classification from microarray and seq data. This comprehensive methodology not only enhances the deep learning models' effectiveness but also ensures that the system can generalize well to unseen data, resulting in consistently good performance across diverse datasets. It shows a brief discussion of the sustainability of the developed model in terms of scalability, interpretability, and computational effectiveness, which is described below.

Scalability: Accurate data is needed to provide strong performance regarding accuracy to select the best features. If the size of the data increases, then it provides significant performance in the larger expression of gene data. Thus, the recommended model provides better scalable performance in the cancer classification model.

Interpretability: The observation that shows the cause and effect of the system. In some instances, the deep learning model shows more hidden layers, which makes the interpretability of the model more difficult. In the future, the interpretability will be considered using the developed model.

Computational analysis*:* Computational analysis is the amount of memory or time that is taken for each step in the calculation. Moreover, the computational analysis of the developed model shows better outcomes.

Limitations of the database*:* In larger datasets, it may not have the ability tofit into the RAM of desktop computers. Further, it may not be reliable and may lose the data. It tends to affect the accuracy of the model. In addition to this, a novel method is developed where the pre-processing is performed using the NAN removal and missing values. Although the larger dataset has been considered for the validation process using the developed model. Thus, it significantly chooses the optimal features based on the gene selection process. Owing to these, larger data are suitable in this proposed methodology to provide accurate outcomes. While training the data, an accurate outcome is provided to enhance the system's performance. Hence, the accurate classification of cancer is suggested for attaining superior performance.

7. Conclusion

This research work has explored the implementation of a new cancer classification framework using microarray and seq data. The microarray and seq data were obtained from the different standard datasets. The collected data were passed through the phase of pre-processing. Further, the MSOA algorithm is suggested to select the optimal genes from the pre-processed data. At last, the HDLF model has recommended that the optimal gene fed into the cancer classification for attaining better-classified outcomes. In accordance, the execution of the method was evaluated using different metrics and compared with other existing methodologies. By using dataset 1, the k-fold is 2, and the accuracy of the proposed model was improved by 5.4 % of HBA-HDLF, 4.8 % of DHOA-HDLF, 3.2 % and 2.3 % of SFO-HDLF and SOA-HDLF, respectively. The results confirmed the effectiveness of the proposed model in the cancer classification process.

Advantages and limitations of the developed model

The implemented model can classify the cancer in the appropriate location in an effective way. Thus, it would be more beneficial for the clinicians to provide better treatment. Here, the recommended MSOA algorithm is performed to optimize the complex parameters to provide an optimal solution. Therefore, it enhances the feature propagation, which helps to minimize the false errors, and thus it boosts the training speed of the model. In accordance, the K-fold analysis is evaluated to prove that the designed HDLF model tends to solve issues like over-fitting. Thus, it enhances the performance of the system. To the best of our knowledge, the recommended approach shows superiority over the existing algorithms, but it also has certain limitations. It does not have the facility to be applicable in real-time environments.

Future scope of the developed model

In the future, the developed model will be utilized to evaluate the model using real-time data. Consequently, the artificial intelligence will be adapted to provide precise performance in the cancer classification model. Also, the ensemble models will be used to show the enhanced performance of the cancer classification model. The future work will focus on integrating larger datasets with more number of instances and SMOTE analysis to validate the model's generalizability.

Declaration of Competing Interest

The authors declare no conflict of interest

Data availability

No data was used for the research described in the article.

References

Almazrua, H., Alshamlan, H., 2022. A comprehensive survey of recent hybrid feature selection methods in cancer microarray gene expression data. IEEE Access.

Badashah, S.J., Alam, A., Jawarneh, M., Moharekar, T.T., Hariram, V., Poornima, G., Jain, A., 2025. Cancer classification and detection using machine learning techniques. Nat. Lang. Process. Softw. Eng. 95–111.

Brammya, G., Praveena, S., Ninu Preetha, N.S., Ramya, R., Rajakumar, B.R., Binu, D., 2019. Deer hunting optimization algorithm: a new nature-inspired meta-heuristic paradigm. Comput. J.

Chakraborty, D., Maulik, U., 2014. Identifying cancer biomarkers from microarray data using feature selection and semisupervised learning. IEEE J. Transl. Eng. Health Med. 2, 1–11.

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.

Fathi, H., AlSalman, H., Gumaei, A., Manhrawy, I.I., Hussien, A.G., El-Kafrawy, P., 2021. An efficient cancer classification model using microarray and high-dimensional data. Comput. Intell. Neurosci. 2021.

Fiorini, S., 2016. Gene expression cancer RNA-Seq [Dataset. UCI Mach. Learn. Repos. https://doi.org/10.24432/C5R88H.

García-Díaz, P., Sánchez-Berriel, I., Martínez-Rojas, J.A., Diez-Pascual, A.M., 2020. Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. Genomics 112 (2), 1916–1925.

Gomes, G.F., da Cunha, S.S., Ancelotti, A.C., 2019. A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates. Eng. Comput. 35, 619–626.

Harvey, B.S., Ji, S.Y., 2017. Cloud-scale genomic signals processing for robust large-scale cancer genomic microarray data analysis. IEEE J. Biomed. Health Inform. 21 (1), 238–245.

Haznedar, B., Arslan, M.T., Kalinli, A., 2017. Microarray gene expression cancer data. Mendeley data 2017. https://doi.org/10.17632/ynp2tst2hh.4.

Haznedar, B., Arslan, M.T., Kalinli, A., 2021. Optimizing ANFIS using simulated annealing algorithm for classification of microarray gene expression cancer data. Med. Biol. Eng. Comput. 59, 497–509.

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.

Hsieh, S.Y., Chou, Y.C., 2016. A faster cDNA microarray gene expression data classifier for diagnosing diseases. IEEE/ACM Trans. Comput. Biol. Bioinforma. 13 (1), 43–54.

Jiaji, W., Muhammad, A.K., Shuihua, W., Yudong, Z., 2023. SNSVM: SqueezeNet-Guided SVM for breast cancer diagnosis. Comput. Mater. Contin. 76 (20).

Jose, D., Chithara, A.N., Kumar, P.N., Kareemulla, H., 2017. Automatic detection of lung cancer nodules in computerized tomography images. Natl. Acad. Sci. Lett. 40, 161–166.

Kaur, A., Jain, S., Goel, S., 2020. Sandpiper optimization algorithm: a novel approach for solving real-life engineering problems. Appl. Intell. 50, 582–619.

Kiran, J., Muhammad, A.K., Majed, A., Usman, T., Yu, D.Z., Ameer, H., Artūras, M., Robertas, D., 2022. Breast cancer classification from ultrasound images using Probability-Based optimal deep learning feature fusion. Sensors 22 (3), 807.

Kiran, J., Muhammad, A.K., Jamel, B., Majed, A., Nouf, A.A., Huda, A., Usman, T., Jae-Hyuk, C., 2023. BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and Equilibrium-Jaya controlled regula Falsi-Based features selection. Diagnostics 13 (7), 1238.
Kourou, K., Rigas, G., Papaloukas, C., Mitsis, M., Fotiadis, D.I., 2019. Cancer classification from time series microarray data through regulatory dynamic Bayesian networks. Comput. Biol. Med., 103577
Leung, Y., Hung, Y., 2010. A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification. IEEE/ACM Trans. Comput. Biol. Bioinforma. 7 (1), 108–117.
Liu, K.H., Ng, V.T.Y., Liong, S.T., Hong, Q., 2019. Microarray data classification based on computational verb. IEEE Access 7, 103310–103324.
Maji, P., 2012. Mutual information-based supervised attribute clustering for microarray sample classification. IEEE Trans. Knowl. Data Eng. 24 (1), 127–140.
Mamoona, H., Muhammad, I.K., Saleh, N.A., Jhanjhi, N.Z., 2023. Framework for detecting breast cancer risk presence using deep learning. Electronics 12 (2), 403.
Mamuna, F., Muhammad, A.K., Saima, S., Nouf, A.A., Shui-Hua, W., 2023. B2C3NetF2: breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled newton raphson feature selection. CAAI Trans. Intell. Technol. 8 (3).
Maryam, T., Ahmad, N., Hassaan, M., Jawad, T., Rizwan, A.N., Seung, W.L., 2023. DSCC_Net: Multi-Classification deep learning models for diagnosing of skin cancer using dermoscopic images. Cancers 15 (7), 2179.
Maulik, U., Chakraborty, D., 2014. Fuzzy preference based feature selection and semisupervised SVM for cancer classification. IEEE Trans. nanobioscience 13 (2), 152–160.
Mohd, A.H., Ilyas, K., Ahsan, A., Sayed, M.E., Ali, A., Ghamry, N.A., 2023. DCNNBT: a novel deep convolution neural Network-Based brain tumor classification model. Fractals 31 (6), 2340102.
Mostavi, M., Chiu, Y.C., Huang, Y., Chen, Y., 2020. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med. Genom. 13, 1–13.
Muhammad, I., Ashraf, I., Usman, Z., Awais, Y., Siddique, A., Muhammad, A.K., Salman, A.A., Yu, D.Z., 2023. DS2LC3Net: a decision support system for lung colon cancer classification using fusion of deep neural networks and normal distribution based gray wolf. ACM Trans. Asian Low. Resour. Lang. Inf. Process.
Nguyen, T., Nahavandi, S., 2016. Modified AHP for gene selection and cancer classification using type-2 fuzzy logic. IEEE Trans. Fuzzy Syst. 24 (2), 273–287.
Othman, M.S., Kumaran, S.R., Yusuf, L.M., 2020. Gene selection using hybrid multi-objective cuckoo search algorithm with evolutionary operators for cancer microarray data. IEEE Access 8, 186348–186361.
Peng, C., Wu, X., Yuan, W., Zhang, X., Zhang, Y., Li, Y., 2021. MGRFE: multilayer recursive feature elimination based on an embedded genetic algorithm for cancer classification. IEEE/ACM Trans. Comput. Biol. Bioinforma. 18 (2), 621–632.
Pham, T.D., Beck, D., Yan, H., 2006. Spectral pattern comparison methods for cancer classification based on microarray gene expression data. IEEE Trans. Circuits Syst. I Regul. Pap. 53 (11), 2425–2430.
Prabhakar, S.K., Lee, S.W., 2020. Transformation based tri-level feature selection approach using wavelets and swarm computing for prostate cancer classification. IEEE Access 8, 127462–127476.
Rabia, A.M., Verma, C.K., Srivastava, N., 2019. Novel machine learning approach for classification of high-dimensional microarray data. Soft Comput. 23, 13409–13421.
Rabia, M.A., 2022b. *application of* nature inspired soft computing techniques for gene selection: a novel framework for classification of cancer. *application of*. soft Comput. 26, 12179–12196.
Rabia, M.A., 2022a. Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data. Med. Biol. Eng. Comput. 60, 1627–1646.
Rabia, M.A., Joshi, A.A., Kumar, K., Gaani, A.H., 2023. Hybrid feature selection techniques utilizing soft computing methods for cancer data. Comput. Anal. Methods Biol. Sci. 17.
Rojas, M.G., Olivera, A.C., Carballido, J.A., Vidal, P.J., 2020. A memetic cellular genetic algorithm for cancer data microarray feature selection. IEEE Lat. Am. Trans. 18 (11), 1874–1883.
Samundeeswari, P., Gunasundari, R., 2023. An efficient fully automated lung cancer classification model using GoogLeNet classifier. J. Circuits Syst. Comput. 32 (14), 2350246.
Sethy, P.K., Geetha Devi, A., Padhan, B., Behera, S.K., Sreedhar, S., Das, K., 2023. Lung cancer histopathological image classification using wavelets and AlexNet. J. XRay Sci. Technol. 31 (1), 211–221.
Shah, S.H., Iqbal, M.J., Ahmad, I., Khan, S., Rodrigues, J.J., 2020. Optimized gene selection and classification of cancer from microarray gene expression data using deep learning. Neural Comput. Appl. 1–12.
Shams, R., Muhamamd, A.K., Anum, M., Nouf, A.A., Jamel, B., Majed, A., Usman, T., Yu, D.Z., 2023. Regula Falsi-Based feature selection framework for breast cancer recognition in mammography images. Diagnostics 13 (9), 1618.
Shen, L., Tan, E.C., 2005. Dimension reduction-based penalized logistic regression for cancer classification using microarray data. IEEE/ACM Trans. Comput. Biol. Bioinforma. 2 (2), 166–175.
Shiyang, L., Chengquan, L., Qicai, L., Yilin, P., Liyang, W., Zhu, S., 2023. An actinic keratosis auxiliary diagnosis method based on an enhanced MobileNet model. Bioengineering 10 (6), 732.
Shoaib, M.R., Zhao, J., Emara, H.M., Mubarak, A.S., Omer, O.A., Abd El-Samie, F.E., Esmaiel, H., 2025. Improving brain tumor classification: an approach integrating pre-trained CNN models and machine learning algorithms. Heliyon 11, 10.
Statnikov, A., Tsamardinos, I., Dosbayev, Y., Aliferis, C.F., 2005. GEMS: a system for automated cancer diagnosis and biomarker discovery from microarray gene expression data. Elseiver 74, 491–503.
Vaiyapuri, T., Jothi, A., Narayanasamy, K., Kamatchi, K., Kadry, S., Kim, J., 2022. Design of a honey badger optimization algorithm with a deep transfer Learning-Based osteosarcoma classification model. Cancers 14 (24), 6066.
Wang, L., Chu, F., Xie, W., 2007. Accurate cancer classification using expressions of very few genes. IEEE/ACM Trans. Comput. Biol. Bioinforma. 4 (1), 40–53.
Whig, P., Kasula, B.Y., Yathiraju, N., Jain, A., Sharma, S., 2025. Bone cancer classification and detection using machine learning technique. In Diagnosing Musculoskeletal Conditions using Artifical Intelligence and Machine Learning to Aid Interpretation of Clinical Imaging. Academic Press, pp. 65–80.
Wu, M.Y., Dai, D.Q., Shi, Y., Yan, H., Zhang, X.F., 2012. Biomarker identification and cancer classification based on microarray data using laplace naive Bayes model with mean shrinkage. IEEE/ACM Trans. Comput. Biol. Bioinforma. 9 (6), 1649–1662.
Wu, P., Wang, D., 2019. Classification of a DNA microarray for diagnosing cancer using a complex network based method. IEEE/ACM Trans. Comput. Biol. Bioinforma. 16 (3), 801–808.
Xu, R., Anagnostopoulos, G.C., Wunsch, D.C., 2007. Multiclass cancer classification using semisupervised ellipsoid ARTMAP and particle swarm optimization with gene expression data. IEEE/ACM Trans. Comput. Biol. Bioinforma. 4 (1), 65–77.
Zhang, Y.D., Satapathy, S.C., Guttery, D.S., Górriz, J.M., Wang, S.H., 2021. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. Inf. Process. Manag. 58 (2), 102439.