Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Microarray Cancer Data Classification using Deep Learning Methods

[1*]Narayan Naik, [2]Sharath Kumar Y H

[1]Assistant Professor, Department of IS&E, Canara Engineering College, Bantwal-574219, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India
[2] Professor, Department of ISE, MIT Mysore-574177, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India

***Abstract*:** A proposal is made in this paper regarding the deep feed-forward neural network for the microarray binary dataset's classification. We have used eight binary class standard datasets of microarray cancer used the purpose of validating the suggested approach, specifically cancers of the brain, colon, prostate, leukemia, ovary, lung-Harvard2, lung-Michigan, and breast. In addition, six multiclass microarray datasets namely 3-class Leukemia, 4-class Leukemia, 4-class SRBCT, 3-class MLL, 5-class Lung cancer and 11-class Tumor are also considered. To come out with curse of dimensionality, the method for reducing dimensionality is PCA in binary class dataset's case. We have crafted architecture of neural network which is fully connected, configuring its parameters with sigmoid initialization for network's input and hidden layer. This includes specifying the number of epochs, batch sizes, and selecting appropriate activation functions. The suggested method's multiclass behavior is made possible by initializing the activation function SoftMax to the output layer. The min-max approach is used for feature scaling. To compute the magnitude of error of the method, binary cross-entropy, and categorical cross-entropy are used on the binary and multi- class datasets and the ADAM optimizer is for optimization. A study is conducted to compare the suggested approach with the most advanced techniques available. According to experimental findings on these common microarray datasets and comparisons with the most advanced technique, the suggested method's performance is quite respectable.

1. Introduction

A class of diseases known as cancer is defined by aberrant cell growth that leads to the development of cancerous cells. It is normal for a healthy body to control the growth of cells, and for those cells to die gradually over time. Damage in the genetic make-up of cells by internal and environ- mental factors result in cells that do not die and continue to grow to form tumors [58]. Some of the main internal factors that cause cancer are incorrect cell division and damage to DNA, while significant external variables include experiencing chemical exposure found in Smoke from tobacco, radiation, and Sunlight's UV rays. [1, 29, 58]. According to molecular biologists, different cancer types have different expression profiles of genes used in the diagnosis of cancer and differentiate cancers. Classification of microarray medical data contributes in identifying such genes that influence a particular biological consequence and predicts outcomes in the event of a new observation. By building a model to predict the category of an object based on the input pattern representing the object, a classification problem can be solved. With the help of the provided test data, a prediction model will be developed to make accurate predictions. [11, 12, 59].

Analyzing micro array gene expression data remains the most challenging research areas in some sectors like statistics, machine learning, bioinformatics, genomics, computational biology, and pattern classification. The tiny sample size and high curse of dimensionality resulting from the existence of irrelevant genes provide the main challenges in microarray cancer investigation [2, 31, 32]. Medical databases frequently contain noise, feature value fluctuations, and class imbalances, which lead to overfitting and decreased classification accuracy [11, 60]. As a means to recognize and understand what factors lead to cancer, microarray data analysis should be investigated, particularly cancer classification. This promotes early-stage cancer identification, which helps doctors create treatment programmes targeted at raising cancer patients' chances of survival. [2, 4, 32].

Labeled microarray cancer data classification is regarded as a problem that involves several key

tasks, including data collection, feature scaling, selection, and classification, and analysis of post-classification. From tens of thousands of highly correlated and informative genes, it is the task of feature selection to select the genes that are most important. These elements of filtered data are then fed into a classifier to improve classification accuracy. [20, 61]. Identifying the ideal and appropriate feature subset will increase computational stability and classification accuracy, feature selection is essential to the classification of cancer data [6, 17, 19, 56].

A deep learning algorithm is employed to classify the data of microarray cancer. To understand the behavior of features, deep learning requires a lot of data. Eight common datasets are— Central Nervous System, Prostate, Breast, Leukemia, Colon, Ovarian, Lung-Michigan, and Lung-Harvard cancer—are used to validate the suggested methodology. The Min-Max approach is utilized to scale feature values in order to mitigate decision bias towards high-value features. The proposed model shows its capability of an accurate prediction when compared withthe latest related works.

2. Related Works

Researchers in a variety of fields, including computational biology, bioinformatics, genomic studies, pattern recognition, machine learning and statistics are becoming interested in microarray data analysis. The following discussion includes the most recent microarray data analyses in certain fields like pattern recognition, artificial intelligence and machine learning. Applying deep learning methods in analyzing bioinformatics data such as microarray cancer datasets is a challenging task as such data has limited size of sample and unequal classes [62]. Zeebaree et al. [63] utilized a convolutional neural network for the selection of gene and the categorization of microarray cancer data. This work's creators don't disclose how they have got these dimensions of the features. With the noteworthy exception that the weights were chosen at random, Mohapatra *et al*. [2] recommended utilizing one Hidden Layer Feed-Forward Network (SLFN) in combination with Ridge Regression (RR). The breast, leukaemia, prostate, and colon tumour binary microarray datasets were utilized to validate the methodology. Our observation is that they should have been applied the training and test samples similar to the standard train/test as in the dataset of Breast cancer case. A Genetic Programming oriented classifier combined with Information Gain for feature selection was suggested by Salem et al. [1]. In the paper Lin *et al*. [59] reported [59] on feature selection and classification based on silhouette statistics, a genetic algorithm with silhouette statistics has been designed. They did note, however, that the feature selection strategy was not ideal because it produced a high amount of features, which would cause the model to over fit. The K-Nearest Neighbours (KNN) classifier was used by Kumar *et al*. [17] in combination with a choice of features and the algorithm categorization centered upon the concept of MapReduce. The DLBCL, colon datasets, leukaemia, and prostate are four common datasets that Nguyen *et al*. [38] assessed their model and recommended a combined gene selection approach for microarray data classification. Although the proposed is stable among classifiers, they were unable to verify the stability claim for over 5 classifiers. The existing 5 classifiers which includes MLP (Multilayer Perceptron), Linear Discriminant Analysis, Support Vector Machine, K-Nearest Neighbour, and PNN (Probabilistic Neural Network) were used to validate the technique. Brain Emotional Learning and principal Component Analysis and were combined by Lofti and Azita [64] to classify microarray cancer data. Three datasets were used to evaluate their approach, however this was insufficient to verify the method's generalizability. Sharbaf *et al*. [6] put forward an approach of hybrid for the selection of gene and the microarray datasets' classifications, employing ant colony optimization techniques as well as cellular learning automata. They explored the influence of different methods of feature selection that utilize ranking techniques to identify the most effective features. To validate their approach, they employed 3 classifiers: SVM, NB and KNN. Ravi *et al*. [57] conducted a comprehensive review aimed at unveiling the models of deep learning potential in medical data sectors. Various deep learning designs, including recurrent networks, convolutional networks, and deep feed-forward have been demonstrated and are useful for addressing various problem domains. Kar *et al*. [25] presented a technique of feature selection for classifying microarray cancer data, as per Particle Swarm Optimization (PSO). They validated their method using the Leukemia ALL-AML as well as SRBCT datasets. For every dataset, they execute the experiment ten times, and these ten runs' average is given as

the final outcome. Microarray classification was put forward in two stages by Garcia and Salvador [51]. A relief ranking algorithm was then used to identify features, and a classifier was trained using the feature space of lower-dimensional to categorize every sample. The Relief algorithm has been employed for feature selection, and the outcomes of the experiments are reported on 8 distinct cancer datasets. They utilized 3 models of linear classification models: the Fisher linear discriminant, Multilayer Perceptron neural network, and Support vector machine classifiers. Chen *et al*. [14] presented a particle swarm optimization-based feature selection technique, and the C4.5 decision tree is utilised for classification. Have reported the outcomes of an experimental study using an approach called 5-fold cross-validation on several datasets related to tumor cancer. For large biological datasets, Farid *et al*. [52] considered the K-nearest neighbours and decision trees and suggested an adaptive rule classifier. Lyu *et al*. [65] suggested a feature selection technique that uses filters and relies on the Gram-Schmidt orthogonalization approach and the maximum information coefficient. Li *et al*. [66] created data-driven weights for lung cancer classification according to information theory and an overlapping grouping technique. An ensemble strategy based on feature subsets and learning from several projections of the original feature space was presented by Piao et al. [22] to identify multi-class microarray cancer data. To address the increased computational complexity stemming from redundant features in feature selection, Wang et al. [67] ingeniously combined Markov blanket techniques with Wrapper-based feature selection methods. As a result, there are a lot of studies being conducted in feature selection sectors, dimensionality reduction, and microarray data classification. We attempted to investigate the Principal Component Analysis usage as a preliminary stage in our study prior to implementing deep learning, a popular machine learning technique to increase the classification's accuracy. The following sections go into more detail about the details.

3. Proposed Methodology to Classify Binary Class Datasets

The proposed work, within this section, an elaborate framework unfolds, encompassing multiple stages such as reduction of dimensionality, feature scaling, and the employment of a sophisticated deep feed-forward classification technique, incorporating finely-tuned parameter configurations via neural networks. The proposed approach's primary responsibilities include loading the raw microarray cancer data, normalizing it with the Min-Max approach, reducing its dimensionality, and classifying it with deep learning, as illustrated in Figure 2.1.

3.1 Feature Scaling

The practice of exploring feature scaling techniques is widely acknowledged within the domains of machine learning and pattern recognition, primarily aimed at standardizing the data. In scaling down all data elements, this process helps to improve prediction quality by preventing outliers. Given the high variance of characteristics in cancer microarray datasets, we suggest investigating feature scaling for data normalization as a potential pre-processing approach. To accommodate the sigmoid activation function, the Min-Max approach is used to scaling features, which normalizes values to a range of 0 to 1 and during the model training procedure includes a 0.5 threshold for the classification of binary. We have taken into account the Min-Max feature scaling technique (See Equation 2.1) in our work.

$$X = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Here, "X" represents the normalized data, "Xi" corresponds to the original feature value, "Xmin" denotes the minimum value, and "Xmax" symbolises the highest value obtained from the original dataset prior to scaling.

3.2 Dimensionality Reduction-based Feature Selection

Within the context of this study, we recommend to apply Principal Component Analysis, a methodology named dimensionality reduction, as a prerequisite preliminary processing phase in order to establish a data representation that exhibits improved discriminative properties and superior conciseness. Microarray cancer data of high-dimensional is linearly transformed into a new reduced-

dimensional space by PCA in order to maximize the data's variance under a low dimensional space [51]. It is widely recognized that dimensionality reduction enhances accuracy, mitigates overfitting, and simplifies the model, collectively leading to improved classification performance. Equation 2.2, which considers X as any sample within the dataset having d dimensions, is employed to represent each individual sample in the original dataset.

$$X = [X_1, X_2, \cdots, X_n], \qquad X \epsilon R^{n \cdot d} \qquad (2.2)$$

Equation (2.2) serves to articulate an input dataset dimension, which is defined as Xi = x1, x2, • • •, xd, where Xi stands for any of the samples found in dataset (input). Principal Component Analysis (PCA) adeptly preserves the highest variance inherent among the features within the novel dataset, harnessing this variance as a potent lever to engender a trimmed feature set denoted as "Z," characterized by a reduced dimensionality of "k." This meticulous curation of "Z" maintains the utmost salient information, ensuring the retention of paramount value. Here, noticeably smaller than the original dataset's dimension, d, is the new matrix's dimension, k. To construct the new dataset, the PCA model's eigenvectors, denoted as W, are each sample vector multiplied.

$$Z = XW = [z_1, z_2, \cdots, z_k], Z \epsilon R^{n \cdot k}, W \epsilon R^{d \cdot k}, k << d \qquad (2.3)$$

Hence, the new matrix's dimension is significantly greater than the original dataset's size (d), with the new matrix's dimension Z is presented as $Z \epsilon R^{n \cdot k}$.

The new matrix produces a vector zn within a reduced-dimensional feature subspace of dimension k, which is lower in dimension compared to the original d-dimensional feature space, as illustrated in Equations 2.3. In this context, the original dataset's dimension, d, significantly surpasses the dimension of the new matrix, k. The 2.1 figure outlines the steps involved in dimensionality reduction employing PCA.

In the course of dimensionality reduction, given that pre-processing is conducted independently for both the training and test data, we've noted variations between the training as well as test datasets' dimensions. However, in order to train a classifier, it is imperative that both the training and test data share the same dimension. Hence to alleviate this problem, we have considered the selected features' of only minimal count. from either the training or test samples by defining Equation 2.4. Moreover, not only the dimension but the name of features also partially varies. To alleviate this, we have considered the common features and those features that are not common in both training and testing sets are not considered.

$$P(T_r, T_s) = argmin(T_r, T_s), T_r \neq T_s$$

where T_r and T_s stands for the dimension of training and test data and P is the dimension of the dataset after dimensionality reduction which takes either of the minimum numberof features Tr or Ts.

3.3 Deep Learning-based Classification

This section presents the proposed approach, which encompasses multiple stages such as feature scaling, dimensionality reduction, and a classification method based on deep feed-forward neural networks, along with parameter configurations. As depicted in Figure 2.1, the primary objectives of the recommended approach encompass loading the raw microarray cancer data, normalizing it via Min-Max approach, performing dimension reduction, and for classification, deep learning is utilized.

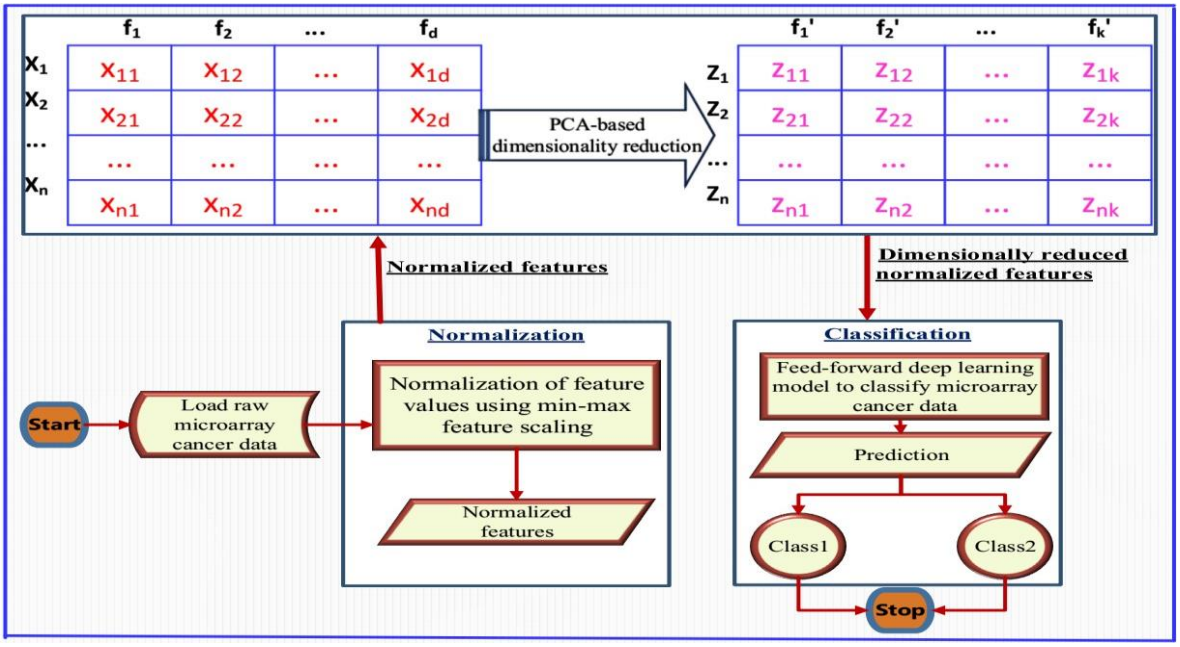


FIGURE 2.1: Structure of the Deep Learning Approach Proposal.

An experiment using deep feedforward neural networks to classify microarray data is proposed in our proposal. One essential deep learning model is the deep feedforward network. Because information moves from x, the input feature vector, through intermediary calculations that define f, and finally to Y output, whereas yi is predicted class lambda, the nomenclature assigned to the model is denoted as a "feedforward neural network". The model does not have any feedback between its outputs. If feedback connections from the same nodes are added to a feed-forward neural network, it no longer qualifies as a feed-forward model; instead, it transforms into a recurrent neural network model. A directed graph that explains how these functions are connected to one another is formed by a collection of interconnected directed functions that define a feed-forward neural network. The model represents the relationship between the functions as a directed graph. To define f (x) = (f1, f2, f3, (x)), for example, let us take three functions, f1, f2, and f3, that together form the network by being linked in a chain. These sequential structures are the prevalent configurations in neural networks. Here, the neural network's first layer is represented by f1, the second by f2, and the third by f3 and so forth [68].

We establish a fully connected neural network approach, where the input layers are initialized with the attributes of input. Moreover, the layers which are hidden is explicitly delineated, encompassing their respective parameters, which encompass log-loss function and activation functions. Each hidden layer’s output is fired with the sigmoid function's output, transforming it into the hidden layer’s output. The classifier can make predictions and estimate each sample's class label once the model converges into a single output layer. A single neuron that produces a class label makes up the output layer, which can be either class one or class two. The detailed workflow that shows feature scaling,dimensionality reduction and classification is shown in Figure 2.1. The outlined procedure is illustrated in Figure 2.2, commencing with the features (input), weight, activation function, and progressing through output computation, error assessment, and error back-propagation. To produce a single output, every feature of the input is multiplied by its weight separately, which serves as input for the subsequent layer, ultimately culminating in the prediction. In the assessment phase, the expected class label is subtracted from the actual class label to produce a discrepancy. This disparity is then used as the back propagation process' error term. This minimal error is noted since the objective of error back propagation is to adjust the weights, ultimately striving for the highest achievable prediction accuracy.

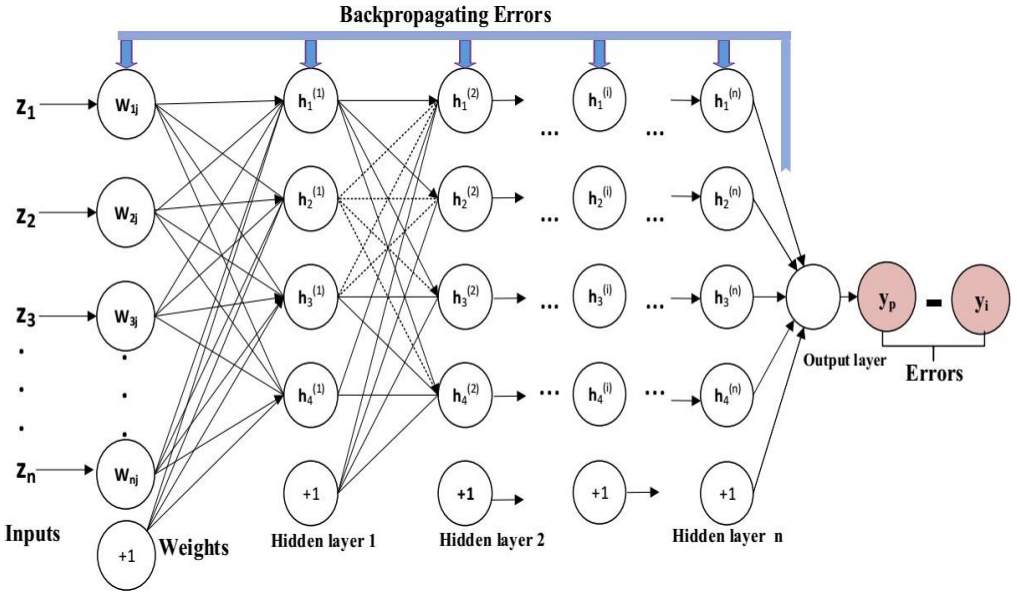


FIGURE 2.2: The Deep Learning Model.

The dot product outcome involving data z1, z2, and so on up to zn, along with their respective weights w1, w2, and so forth up to wn in the h1 (initial hidden layer), is determined by the activation function 'a,' as depicted in Figure 2.3. An activation's output moves on to the following hidden layer node before being triggered at output layer node. Backpropagating weights to preceding nodes is used to continually update them until an optimal prediction is achieved.

Z = z1, z2, z3, · · ·, zn, where Z∈Rn·k, is the input data format used by the suggested deep learning technique. The relevant weight of each input vector is multiplied, with w denoting w1, w2, w3, and so forth, representing the input data's weight vector. Additionally, within the weighted input vectors lies the bias, b. Equation 2.5 delineates the application of the sigmoid activation function to the weighted input vectors within each layer of the model. This process results in the generation of intermediate probabilistic outcomes within the hidden layers.

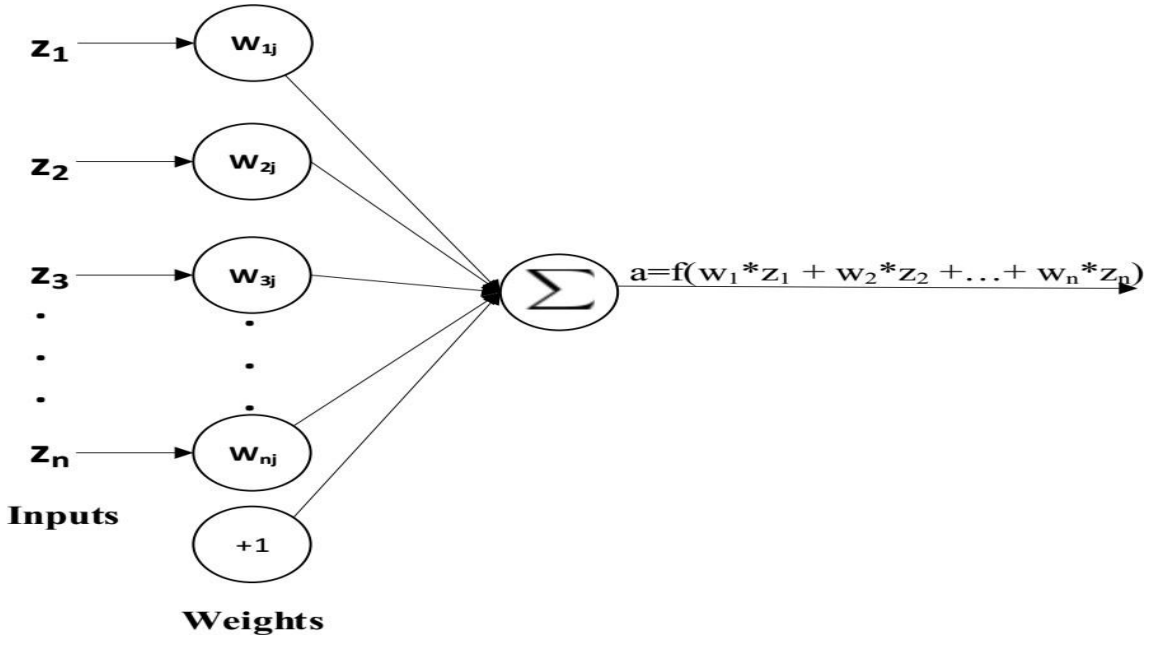


FIGURE 2.3: Activation Function.

$$y_p = f\left(\sum_{i=1}^{n} w_{n_i} \cdot z_{k_i} + b\right)$$

The target variable, yp, is the one we aim to forecast. The weight matrices, wn, and feature vectors, zk, are present. The sigmoid activation function, or f, is introduced in equation 2.6. In order to predict the class membership of a new data point, we recommend using the sigmoidal activation function for training the deep feed-forward neural network model. Our model's activation function is the sigmoid function, which produces limited output values between 0 and 1, facilitating the interpretation of class labels as probabilities. The sigmoid function is as follows:

1

$$f(z_i) = \frac{}{1 + e^{-(z_k \cdot w_k)}}$$

Equation 2.7 addresses the accommodation of multiple features across the hidden layers, as it captures the data's intriguing characteristics as it progresses toward the output layer.

$$y_p = \begin{cases} 1, & \textbf{if} \quad w_0 + w_1 \cdot z_1 + w_2 \cdot z_2 +, \cdots, +w_n \cdot z_n >= 0.5 \\ 0, & \textbf{otherwise} \end{cases} \tag{2.7}$$

Where yp represents the anticipated class label, zn stands for the feature vector, wn corresponds to the weight vector, and w0 signifies a bias value initialized to 1.

One useful method for determining the relationship between multiple independent features or target-dependent variables is Equation 2.7. Random and uniform weight initialization is followed by uniform weight updates during training. Equation 2.7 computes a class value by comparing the actual and predicted probabilities and assigns the prediction to either the actual or predicted class based on a threshold value. In a deep learning-based classification, The cross-entropy objective function, which is preset, is used for calculating the ultimate forecast error according to the difference between the predicted class label (yp) and class label (yi). Subsequently, the weights are fine-tuned to minimize the error, which is achieved by propagating errors throughout the entire network. [69].

Figure 2.2 illustrates the computational elements and the method of deep feed-forward to classify data. It comprises the error calculation, input features, activation function, weights, output, and error back-propagation. Each input feature's weight is multiplied to generate a single output, which, in turn, serves as input for the subsequent layer to make predictions. The assessment is carried out by deducting the estimated the label of class from the actual label of the class. The resulting difference is then employed as an error of back-propagation. The minimum error is logged, as the error back-propagation mechanism is engineered to adjust the weights, ultimately striving for the highest attainable prediction accuracy. The input characteristics [z1, z2, and so on, up to zn] and their accompanying weights [w1, w2, w3, and so on] are dot products in the first hidden layer, or "h1," which produces the output. Figure 2.3 provides an example of this approach. Activation function's output proceeds to the subsequent hidden layer node before eventually being activated at the output layer node. By propagating the updated weights backward to the nodes that came before it, the process of updating weights never ends until the model achieves the most accurate forecast.

3.4 Parameters settings

For 7 of the datasets, we've used a model of deep feed-forward of 7 layer, as shown in Table 2.1. Equation 2.8 is used to calculate the number of parameters in each of the layers, where Ci at layer 'I' is the

number of neurons, Pi in the preceding layer is the number of neurons, and 1 is the bias. Every layer in the methodology of deep feed-forward has a different batch size, as denoted by the "None" parameter. The quantity of neurons in the input layer determines the input shape for the subsequent layer. An epoch size of 1000 is consistent across all datasets. Summing all of the parameters for each layer yields the total number of parameters, or "$T_{parameters}$" for a particular model. As a result, after adding up all parameters from all layers, our study has 33,661 trainable parameters.

$$T_{parameters} = \sum_{i=1}^{7} C_i(P_i + 1)$$

TABLE 2.1: Proposed seven-layer deep feed-forward model

	Type of Layers	Output Shape	# Parameters
	Dense Layer 1	(None, 200)	4600
Dropout 1	(None, 200)	0	
Dense Layer 2	(None, 100)	20100	
Dropout 2	(None, 100)	0	
Dense Layer 3	(None, 50)	5050	
Dense Layer 4	(None, 40)	2040	
Dense Layer 5	(None, 30)	1230	
	Dense Layer 6	(None, 20)	620
	Dense Layer 7	(None, 1)	21
	Total number of trainable parameters		**33,661**

The selected parameters include the binary cross-entropy for loss computation as well as sigmoid activation function on test data and training, and the utilization of the Adaptive Moment Estimation (ADAM) optimizer [70]. Each parameter's learning rate is adjusted by the ADAM optimizer by figuring out and keeping up a preceding gradient momentum decaying average and an exponentially declining average of previous squared gradient variances. We refer to the ADAM optimizer using its name in our investigation, with the factors assuming their default values. These default ADAM optimizer parameter values are as follows:

optimizer parameters are β_1 =0.9, β_2 =0.99, $\epsilon = 10^{-8}$

where β_1 *and* β_2 are decay rates [70].

As this model generates predictions through a probabilistic framework, by applying a binarization process to the class level, producing Boolean outcomes for each class when the threshold value is set at 0.5. In this model, the cost function is defined using binary cross-entropy, as outlined in Equation 2.9, where a feature is represented by ‘zk’, target is ‘yi’, and the class (whether it's zero or one) is ‘c’. In the output layer, zk is used as an input vector to determine the class membership of a given class c. The proposed model estimates a probabilistic target value yi given an input vector zk.

$$f(z_i) = p(y_i = c|z_n) \quad (2.9)$$

In the case of binary classification problems, it is necessary to determine whether the predicted and actual probability vectors differ significantly from one another, by employing a distance function, as depicted in Equation 2.10.

$$D(f(z_i, y_p, y_i)) = \frac{1}{N}\sum_i((w_n.z_n + b), y_p, y_i)$$

D represents the distance function between the actual class labels, denoted as yi, and the predicted class labels, denoted as Yp. The vector distance that was previously indicated is averaged over all the values of input (zn) in the training dataset, represented by the letter N. In this context, b refers to the bias, and wn is the weight matrix. Equation 2.11, which aims to enhance the estimation of a particular sample zn's association with a class label yi, exemplifies the application of the negative log-likelihood minimization function to maximize the correct target yi probability, with regard to the 'zi' input features.

$$L(f(z_i), y_i) = -\sum_c 1_{(y_i=c)} \log f(z_n)_c = -\log f(z_n) \cdot y_i$$

In this case, y_i is labeled as dependent class that corresponds to all features, c represents class, feature vector 'zn', and L represents loss.

4. Experimental Setup and Results Analysis

The context here, includes an explanation of the datasets we used for our experiments, performance metrics we employed to evaluate the performance of the suggested model, and the analysis and results of our experiments. We developed our model using Anaconda Python 3.5 as a development tool, the backend is powered by the open-source Tensor Flow deep learning library, and the frontend is driven by the Keras Deep Learning Library. [71].

4.1 Dataset Description

In order to evaluate the proposed model of multidimensional microarrays, eight different standard microarray datasets were taken from the ELVIRA Biomedical Dataset Repository (http://leo.ugr.es/elvira/DBCRepository/index.html) sourced from the BIO ELVIRA Biomedical Data Repository for assessing high-dimensional biomedical data sets. The dataset related to Central Nervous System cancer comprises 7129 features and includes samples of 60, among which 39 deal with failures and 21 with survivors. There are 2000 genes and 62 samples in the colon cancer dataset. Of these samples, 22 are considered normal cases and 40 are classified as positive tumours. The Ovarian cancer dataset, on the other hand, features 15153 genes and consists of 253 samples, with 162 samples corresponding to cancer cases and 91 to normal cases. Additionally, the Leukaemia dataset, which includes 72 samples and 7129 characteristics, is related to bone marrow malignancy.

There are two different groups in the Leukaemia dataset: 25 samples indicate Acute Myeloid Leukaemia (AML) and 47 samples indicate Acute Lymphoblastic Leukaemia (ALL). In the case of the Prostate cancer dataset, it involves 12600 features and a total of 102 samples, out of which 52 correspond to tumor observations, and a further 50 cases are considered normal. 24481 genes and 97 samples make up the Breast Cancer dataset; 46 cases are classified as relapses and 51 samples are classified as non-relapses. The lung-Michigan cancer contains 96 samples that contains 86 adenocarcinomas and 10 non-neoplastic. It is significant to observe that the class distributions in the Lung-Harvard2 and Lung-Michigan cancer datasets are incredibly unbalanced. Adenocarcinoma (ADCA) cases (150) and malignant pleural mesothelioma (MPM) cases (32), together making up the Lung-Harvard2 dataset, total 181 samples. On the other hand, of the 96 samples in the Lung-Michigan dataset, 86 are categorized as Adenocarcinomas, and the other 10 are classed as non-neoplastic.

Table 2.2 provides an overview of these datasets, offering details on their initial feature count, PCA-selected features, and percentage of discarded features, number of classes, training and test data sizes, sample sizes. Through PCA-based dimensionality reduction, irrelevant features were eliminated, enhancing

classifier performance with the most informative features. It was observed that a substantial portion of the original features in these datasets lacked significance in predicting class labels.

TABLE 2.2: Dataset Description

Datasets	# Features	# Selected Features	% of discarded features	Sample size	Training size	Test size	#Classes
CNS	7129	108	98.49	60	36	24	2
Colon	2000	104	94.80	62	37	25	2
Leukemia	7129	53	99.26	72	39	33	2
Prostate	12600	76	99.40	102	61	41	2
Ovarian	15154	24	99.84	253	202	51	2
Breast	24481	60	99.75	97	78	19	2
Lung-Michigan	7129	45	99.37	96	57	39	2
Lung-Harvard2	12533	77	99.39	181	32	149	2

4.2 Performance Measures

Several recognized performance metrics were employed to verify the efficacy of our suggested model. These metrics included f-score, classification accuracy, recall, precision, the region under the curve (AUC) as represented by the confusion matrix, ROC curve, and log-loss. Equation 2.12 illustrates the use of accuracy as a measure to assess the overall predictive performance of the model. This metric considers four crucial parameters: False Negatives, False Positive, True Positive and true negative. This involves determining the ratio between the total number of test samples and the number of samples that have been accurately classified.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.12)$$

Equation 2.13 shows that true positive rate is represented by recall, which is sometimes referred to as sensitivity. According to the equation, it measures the proportion of true positives to the total of TP and false negatives (FN).

$$Recall = \frac{TP}{TP+FN} \quad (2.13)$$

Another performance metric employed in this study is precision, sometimes referred to as positive predictive value (PPV), as outlined in Equation 2.14.

$$Precision = \frac{TP}{TP+FP} \quad (2.14)$$

Equation 2.15 illustrates how the F-Measure, which considers the harmonic mean of both precision and recall metrics, is used to balance their respective effects.

$$F_{score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (2.15)$$

The proposed model's error score is determined through the utilization of the log-loss function, as depicted in Equation 2.16. In this equation, N stands for the samples' quantity, Actual class label is represented by y_i, and the likelihood that the i^{th} sample is a member of one of the classes is represented by pi. Log-loss evaluates a model's performance by measuring predictions as probability values within the range of zero to one. A reduced log-loss error value is the goal of an improved classifier, with the ideal scenario being zero, indicative of a perfect classifier.

$$LogLoss = \frac{-1}{N}\sum_{i=1}^{N} y_i \times (log(p_i) + (1-y) \times log(1-p_i)) \tag{2.16}$$

5. Experimental Results and Analysis

A detailed examination of the findings from experiments is presented in this subsection that were attained using the suggested methodology across a range of datasets and performance metrics.

TABLE 2.3: Experimental Results on all datasets

Dataset	Training set size	Test set size	Training Accuracy	Test Accuracy	Precision	Recall	F-Measure	AUC	Log-loss error
Breast cancer	78	19	1.00	0.95	0.95	0.95	0.95	0.96	0.41
CNS	36	24	0.96	0.66	0.66	0.66	0.66	0.67	0.29
Colon	37	25	1.00	0.96	0.97	0.96	0.96	0.97	0.18
Leukemia	39	33	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Ovarian	202	51	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Prostate	61	41	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Lung-Harvard2	32	149	1.00	0.99	0.99	0.99	0.99	0.99	0.03
Lung-Michigan	57	39	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Figures present the experimental outcomes showing the confusion matrix, training and test samples' classification accuracy, loss, ROC curves, and loss. Moreover, the results are also presented in Tables and comparison analysis using novel methods.

In the proposed microarray cancer classification model which is based of deep feed-forward neural network, each input layer collects information and transmits them for further processing in the next layer. The model's neurons are densely coupled. A single neuron in the fully connected neural network model generates a probabilistic outcome with a range of 0 to 1, inclusive. Upon binarization of this magnitude, it begets the prognosticated outcome contingent upon the designated threshold of 0.5, an oft-adopted default value for quandaries pertaining to binary classification. To predict a class on the output layer, the model takes the chosen features as input and processes them via its hierarchical hidden layers.

In order to perform pre-processing, such as dimensionality reduction, feature scaling, and classification, we used separate test samples and training. Since certain data sets, such as Lung-Harvard2, Leukemia and Breast Cancer have established criteria of test and training, that throughout our work, we apply these criteria. The utilization of a ratio of 60:40 for training and testing cases, prior to commencing the pre-processing phase, is employed for the purpose of segregating the datasets into training and test samples. This approach is especially pertinent when dealing with datasets, namely CNS, Colon, and prostate that do not possess discrete sets of test samples and training. Furthermore, in the specific case of data related to ovarian cancer, a ratio of 80:20 is employed for the same purpose of partitioning the dataset. Many machine learning algorithms use this as the benchmark, which is why the datasets were divided into the designated ratios for testing as well as training.

Table 2.3 displays the experimental outcome of the suggested approach along with the sample sizes for each dataset's training and testing stages. Additionally provided are the log-loss error, f-score, AUC, recall, precision, and accuracy of the classification. The suggested approach demonstrates flawless classification performance, obtaining a 1.00 rating across 4 datasets: the Prostate, Lung-Michigan cancer, Leukemia and Ovarian. The proposed methodology has yielded noteworthy classification accuracies, with an impressive score of 0.99 achieved on the Lung-Harvard dataset, a commendable 0.95 on the Breast cancer dataset, and a substantial 0.96 on colon and CNS cancer datasets. These results were obtained when applied to the remaining four datasets, underscoring the effectiveness of the approach.

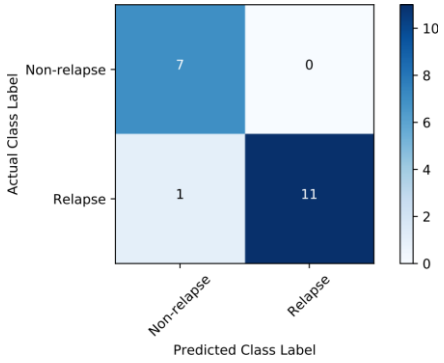
To confirm the proposed method’s efficacy, a confusion matrix is utilized that shows test samples that have been identified properly or erroneously. Along the diagonals, Accurately categorised samples are displayed in the confusion matrix; test samples that are off-diagonal are misclassified. It has been shown that out of 19 samples for breast cancer, one case had an error in its classification, with the relapse category assigned to the non-relapse category depicted Figure 2.5(a). In a similar vein, 1 test sample from the fails class out of 24 is incorrectly categorised as a member of the surviving class in the dataset of CNS which can be seen in Figure 2.5(b).

Furthermore, as demonstrated in Figure 2.5(c) 1 negative case out of 25 test cases in the Colon dataset is incorrectly identified as a positive case. Among the 149 test cases meticulously evaluated within the Lung-Harvard dataset, a singular instance belonging to the ADCA class was observed to be erroneously misclassified as a member of the MPM class, as vividly demonstrated in Figure 2.5(e). As demonstrated in Figures 2.5 (d, f, g, h) for the Leukemia, Lung-Michigan, Ovarian, and Prostate cancer datasets, accordingly, the model exhibits flawless classification accuracy in the other datasets.

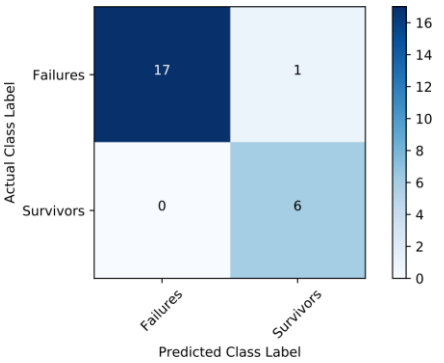
To evaluate the efficacy of the proposed methodology, we've calculated the accuracy of the categorization. The training and test cases' accuracy for the suggested method using data on breast cancer is displayed in Figure. Comparably, the suggested method's classification accuracy on the Lung-Harvard2, CNS and Colon datasets—which display the smallest gap between the lines of training and test cases—is displayed in Figures 2.7(b), 2.7(c), and 2.7(e). Figures 2.7(d), 2.7(f), 2.7(g) and 2.7(h) exhibit the suggested technique's classification accuracy on the datasets such as Leukemia, Lung-Michigan, Ovarian, Prostate, respectively. There is no appreciable difference between the test samples and lines of training since these datasets have 100% classification accuracy.

The log-loss error function is another metric used to assess the suggested approach. Evidently, as discerned from Figure 2.9(a), the loss function associated with the training set approaches near-zero values. Nonetheless, It is noteworthy that the test cases' stated loss, quantified at 0.410 (as detailed in Table 2.3),

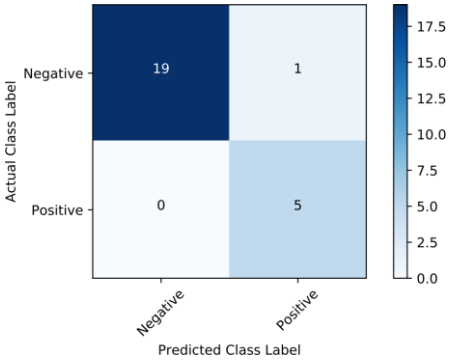
underscores the necessity for additional comprehensive investigations pertaining to this specific dataset. Such further studies are warranted in order to mitigate and minimize the prevailing errors and enhance the overall performance. As seen in Figures 2.9(b), 2.9(c), and 2.9(e), where the suggested technique scores 0.219, 0.189, and 0.032 (Check out the final column of Table 2.3), it performs the best in datasets like Lung Harvard, CNS, and Colon. Additionally, it is crucial to highlight that there is no obvious loss on the datasets related to prostate cancer, lung-michigan, and leukemia, as clearly seen in Figures 2.9(h), 2.9(f), and 2.9(d), respectively. Alternatively, as Figure 2.9(g) clearly illustrates, a small and essentially insignificant error of 0.003 is observed within the framework of the ovarian cancer dataset. It is worth highlighting that the disparity or divergence observed between the trajectories of the training and test cases is an indicative measure of whether the model may be suffering from overfitting, as it reflects the extent to which the model may have specialized excessively on the training data.



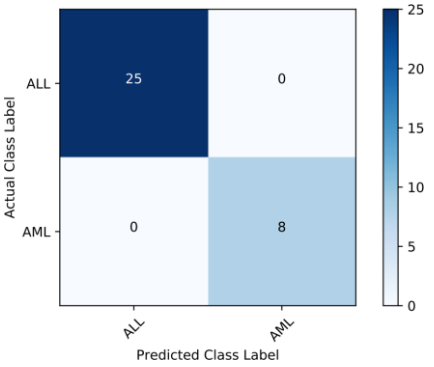
(a) CM for Breast cancer



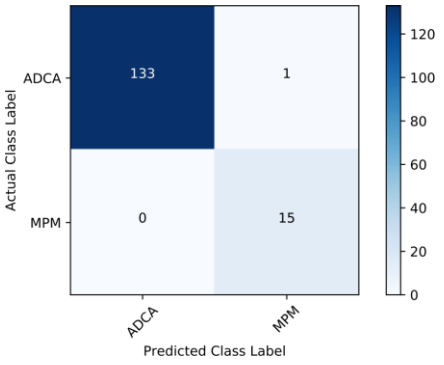
(b) CM for CNS



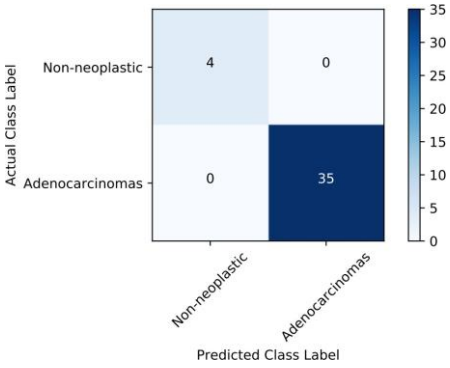
(c) CM for Colon cancer



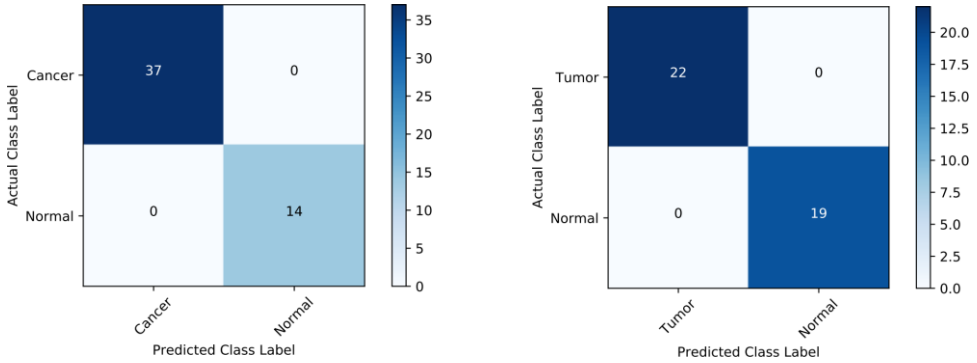
(d) CM for Leukemia



(e) CM for Lung cancer

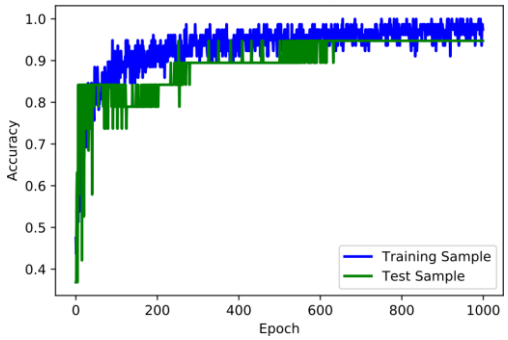


(f) CM for Lung-Michigan

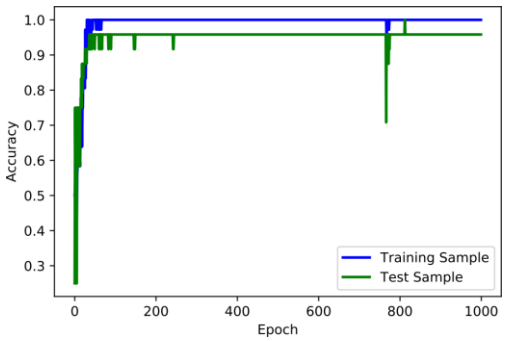


(g) CM for Ovarian cancer (h) CM for Prostate cancer

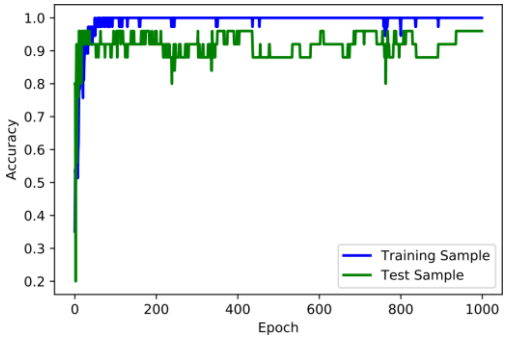
FIGURE 2.5: The confusion matrix due to the proposed Deep Learning method on (a) Breast cancer; (b) CNS Cancer; (c) Colon Cancer; (d) Leukemia Cancer; (e) Lung Cancer; (f) Lung-Michigan; (g) Ovarian cancer; and (h) Prostate cancer datasets.



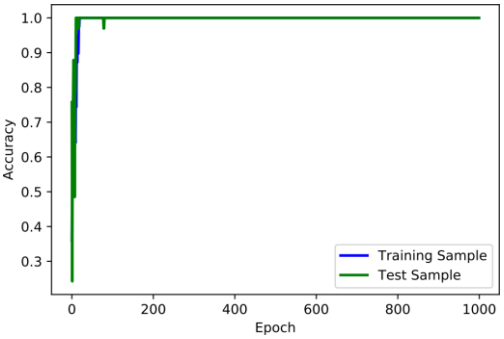
(a) Accuracy: Breast-Cancer



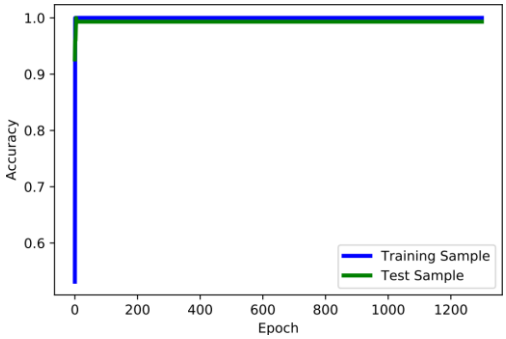
(b) Accuracy: CNS-Cancer



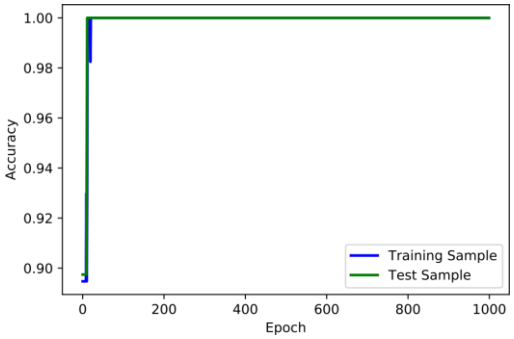
(c) Accuracy: Colon-Cancer



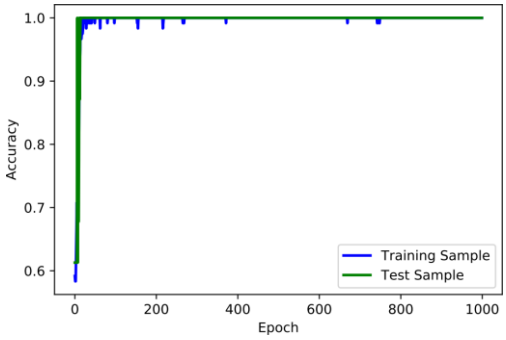
(d) Accuracy: Leukemia-Cancer



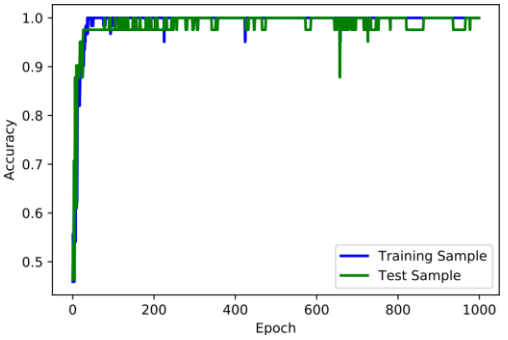
(e) Accuracy: Lung-Cancer



(f) Accuracy: Lung-Michigan

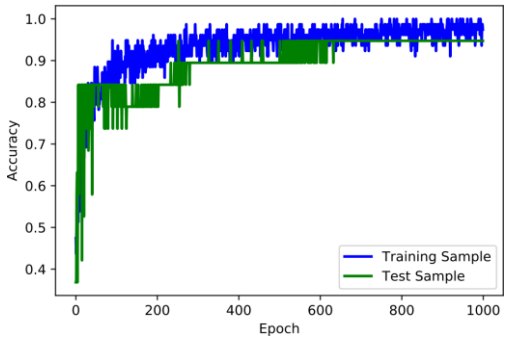


(g) Accuracy: Ovarian-Cancer

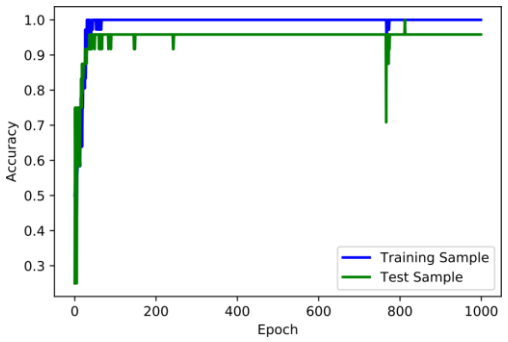


(h) Accuracy: Prostate-Cancer

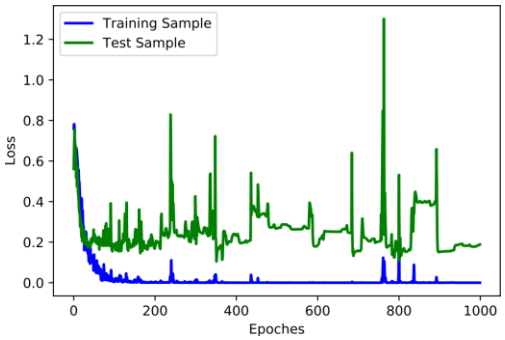
FIGURE 2.7: The classification accuracy due to the proposed Deep Learning method on (a) Breast cancer; (b)CNS Cancer; (c) Colon Cancer; (d) Leukemia Cancer; (e) Lung Cancer; (f) Lung-Michigan; (g) Ovarian cancer; and (h) Prostate cancer datasets.



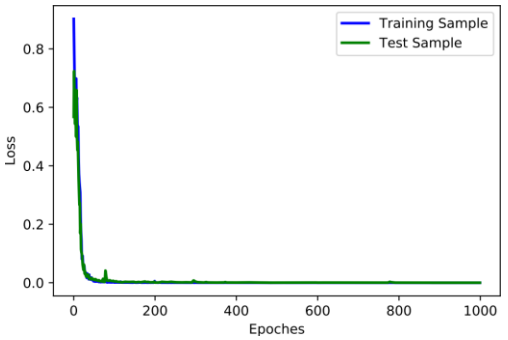
(a) Loss: Breast-Cancer



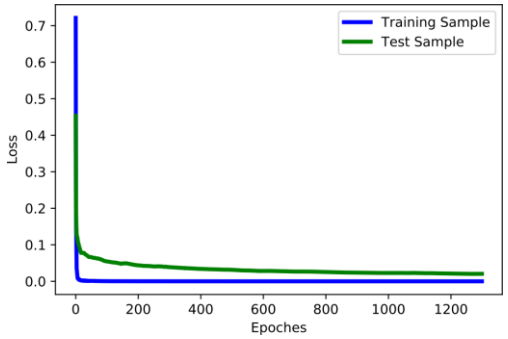
(b) Loss: CNS-Cancer



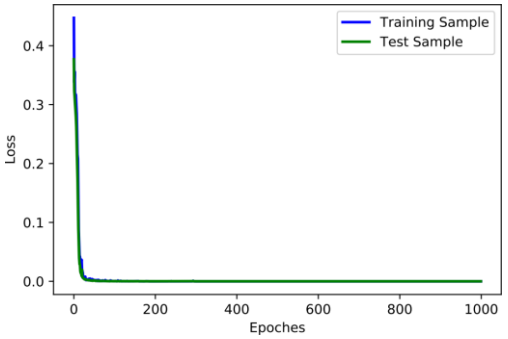
(c) Loss: Colon-Cancer



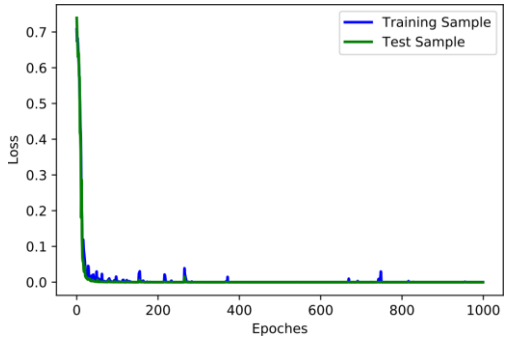
(d) Loss: Leukemia-Cancer



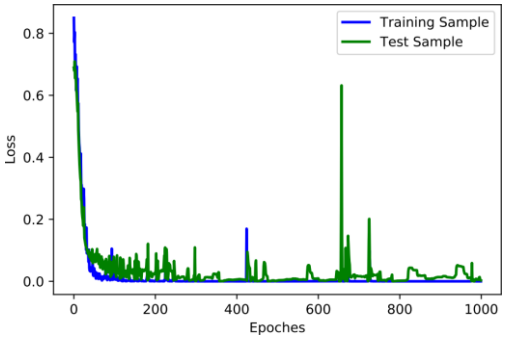
(e) Loss: Lung-Cancer



(f) Loss: Lung-Michigan



(g) Loss: Ovarian



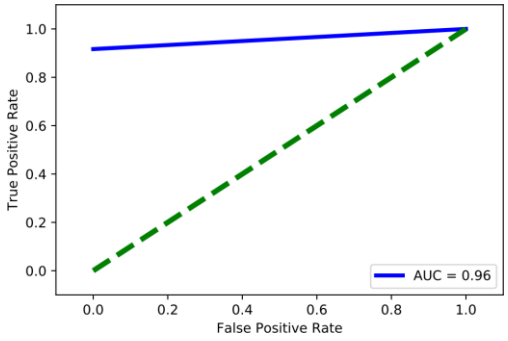
(h) Loss: Prostate-Cancer

FIGURE 2.9: The loss due to the proposed Deep Learning method on (a) Breast cancer; (b) CNS Cancer; (c) Colon Cancer; (d) Leukemia Cancer; (e) Lung Cancer; (f) Lung-Michigan; (g) Ovarian cancer; and (h) Prostate cancer datasets.

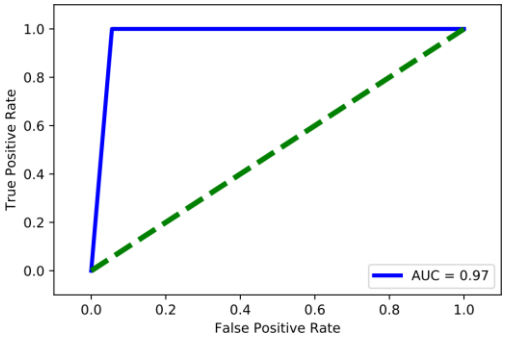
Generally, for classification problems, the curve ROC is often utilized to evaluate the model's performance. The AUC is calculated using this method at various threshold values. It has a high tolerance level and is one of the measures employed to evaluate a classifier's performance when it comes to classifying data with less class imbalance. ROC curve compares area coverage to accuracy and other performance indicators and shows the area coverage in terms of AUC. By shifting the curve to the left upper corner, ROCs and AUCs improve the classification accuracy. Applying the recommended approaches to the Breast Cancer dataset yields results from ROC analysis as shown in Figure 2.11(a) based on acceptability ratings found in the literature. The ROC curve scores an AUC of 0.96. The ROC curves corresponding to colon cancer and CNS datasets are displayed in Figures 2.11(b) and 2.11(c), with an AUC of 0.97 for each. This level of area coverage is regarded as highly commendable, aligning with the prevailing standards of result evaluation within the pertinent academic literature [75, 76]. In a similar vein, the curve of ROC engendered by the implemented methodology on the dataset Lung-Harvard2 attains a notably elevated value of 0.99, as vividly portrayed in Figure 2.11(e). Based on the AUC score of 1.00 for each of the datasets in Figures 2.11(d), 2.11(f), 2.11(g), and 2.11(h), It seems that the suggested approach is working exactly as expected on the Leukemia, Lung-Michigan, Ovarian, and Prostate datasets.

5.1 Discussion and Comparative Analysis

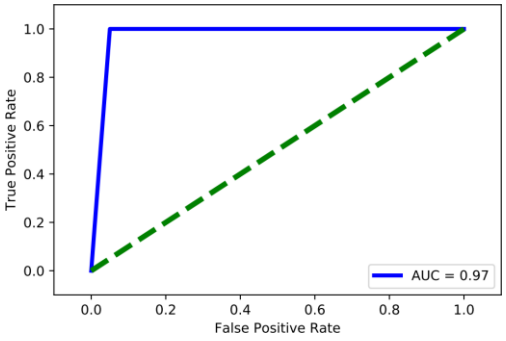
A thorough analysis of the suggested method's performance and comparisons with cutting-edge techniques are given in this section. The suggested deep learning-based classifier performs best when informative features are obtained using a PCA-based dimension reduction technique.



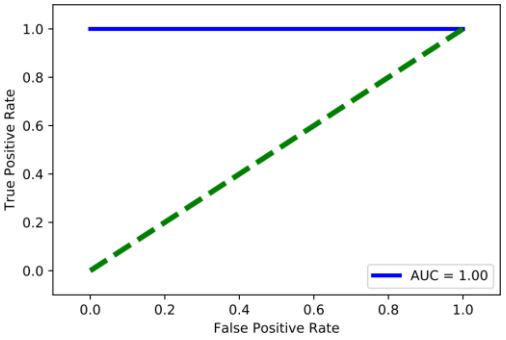
(a) ROC: Breast-Cancer



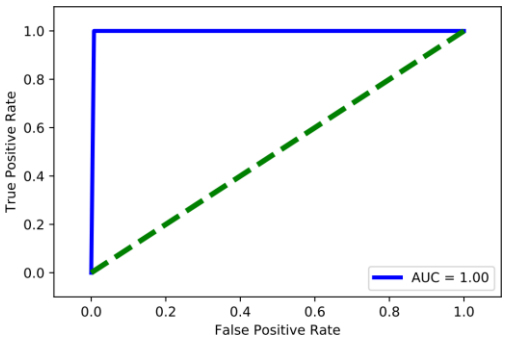
(b) ROC: CNS-Cancer



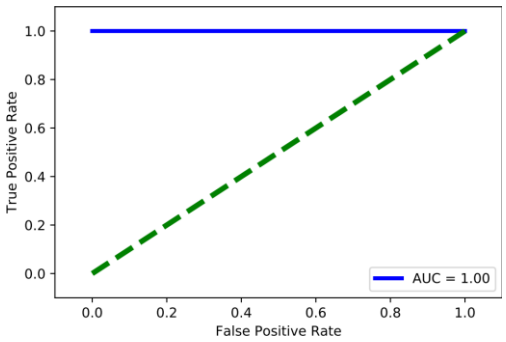
(c) ROC: Colon-Cancer



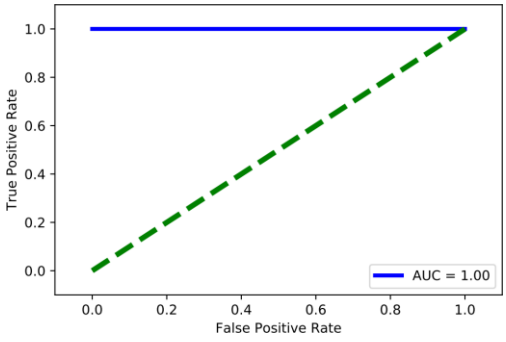
(d) ROC: Leukemia-Cancer



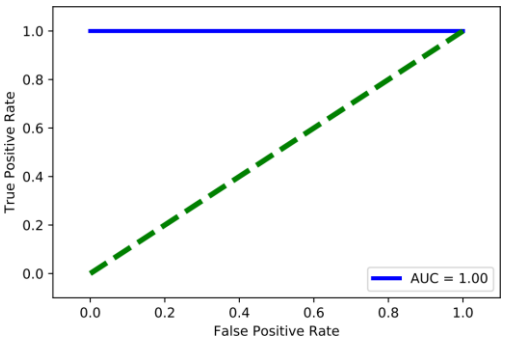
(e) ROC: Lung-Cancer



(f) ROC: Lung-Michigan-Cancer



(g) ROC: Ovarian-Cancer



(h) ROC: Prostate-Cancer

FIGURE 2.11: The ROC curve due to the proposed Deep Learning method on (a)Breast cancer; (b) CNS Cancer; (c) Colon Cancer; (d) Leukemia Cancer; (e) LungCancer; (f) Lung-Michigan; (g) Ovarian cancer; and (h) Prostate cancer datasets.

TABLE 2.4: Comparison of classification accuracy of the proposed method with someof the latest related research works on CNS, Colon, Ovarian, Prostate, Leukaemia, Lung-Harvard2, Lung-Michigan, and Breast cancer datasets

References	Datasets							
	CNS	Colon	Ovarian	Prostate	Leukaemi	Lung-Harvard2	Lung-Michigan	Breast
Salem *et al.* [1]	0.87	0.85	-	**1.00**	0.97	-	**1.00**	-
Mohapatra *et al.* [2]	-	0.93	-	0.99	0.99	-	-	0. 76
Medjahed *et al.* [3]	-	**0.97**	0.98	-	0.96	**0.99**	-	0.86
Chen *et al.* [14]	-	-	-	0.94	-	-	-	-
Kar *et al.* [25]	-	-	-	-	0.97	-	-	-
Moayedikia *et al.* [32]	0.79	0.72	-	-	**1.00**	**1.00**	-	-
Nguyen *et al.* [38]	-	0.88	-	0.91	0.94	-	-	-
Garcia *et al.* [51]	0.73	0.84	0.99	0.90	-	0.98	0.98	0.65
Zeebaree *et al.* [63]	-	0.65	-	0.92	**1.00**	-	0.72	-
Proposed Method	**0.96**	**0.96**	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**0.95**

TABLE 2.5: Comparison of Recall of the proposed method with IG/SGA [1]

Ref.	CNS	Colon	Ovar.	Prost.	Leuk.	Lung-Har.	Lung-Mich.	Breast
Salem *et al.* [1].	-	0.83	-	**1.00**	0.97	-	**1.00**	-
Our Method	**0.96**	**0.96**	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**0.95**

The suggested approach achieves flawless classification with 1.00 accuracy on 4 datasets: the Prostate, Leukemia, Ovarian, and Lung-Michigan. Additionally, the accuracy of the dataset Lung-Harvard is 0.99. In two dataset's case, namely CNS and colon, along with the breast cancer dataset, we have ascertained classification accuracies of 0.96 and 0.95, respectively. This demonstrates that the suggested approach outperforms numerous cutting-edge techniques.

This section of the work compares the suggested approach's classification accuracy to a few recent, carefully chosen works. The proposed method's accuracy in classification is compared with nine of the most recent approaches in Table 2.4. The hyphen (-) in those specific table cells indicates that the authors' work did not take the dataset into account.

TABLE 2.6: Comparison of the proposed method with Mohapatra *et al*. [2] on datadimensionality, Training size, Test Size, Classification Accuracy (CA) and AUC pa-rameters.

Method Authors and	Dataset	Dimension	Train. size	Test size	Acc.	AUC
Mohapatra *et al*. [2] (WKRR)	Breast	97 * 24481	70	27	0.81	0.89
	Colon	62*2000	40	22	0.95	0.79
	Leukemia	76*7129	50	26	0.95	0.87
Proposed method (Deep learning)	Breast	97*24481	78	19	**0.95**	**0.96**
	CNS	60*7129	36	24	**0.96**	**0.97**
	Colon Tumor	62*2000	37	25	**0.96**	**0.97**
	Leukemia	72*7129	39	33	**1.00**	**1.00**
	Ovarian	253*15154	202	51	**1.00**	**1.00**
	Prostate	102*12600	61	41	**1.00**	**1.00**
	Lun-Michigan	96*7129	57	39	**1.00**	**1.00**
	Lung Harvard	181*12533	32	149	**0.99**	**1.00**

Table 2.4 illustrates that the suggested approach attains superior classification accuracy of 1.00 across four datasets: the Prostate, Leukemia, Ovarian and Lung-Michigan. When compared to previous works, our results for the CNS and Colon datasets are 0.96, which is better. Accuracy values of 0.95 and 0.99 are obtained for Lung-Harvard2 and Breast Cancer, respectively. In comparison to other methods, the suggested method generally performs better. The suggested approach and the IG/SGA method are contrasted in Table 2.5 [1]. The evidence presented establishes that the proposed model exhibits superior recall performance when compared to the IG/SGA method, specifically in the context of the Colon and Leukemia datasets. This favorable outcome is replicated with parity on the Prostate and Lung-Michigan datasets. As indicated in the table, the hyphen (-) indicates that no datasets along that column are considered by the authors.

5. Conclusion

Within the framework of this study, we introduce a sophisticated approach called deep feed-forward neural network designed for the purpose of binary classification of microarray datasets. We validate the proposed method using a standard microarray cancer dataset comprising 8 binary classes: the Lung-Michigan, CNS, Leukaemia, Colon, Breast cancers, Prostate, and Ovarian, Lung-Harvard2. In addition, six multiclass microarray datasets namely 3-class Leukemia, 4-class Leukemia, 4-class SRBCT, 3-class MLL, 5-class Lung cancer and 11-class Tumor datasets are also considered. In the context of datasets characterized by binary class distributions, Principal Component Analysis (PCA) is harnessed as a dimensionality reduction technique, strategically employed to mitigate the challenges posed by the curse of dimensionality. Parameters like batch size and epoch count are created along with an architecture of fully connected neural network. And the networks hidden and input layers' activation function, which is initially set to sigmoid. In the output layer, the softmax

activation function is carefully used and initialized to account for the multiclass behaviour that is inherent in the proposed technique. The min-max technique is used to scale features. To calculate the method's error magnitude, binary cross-entropy, and categorical cross-entropy are used on the binary and multi- class datasets and the ADAM optimizer is for optimization. A study is conducted to compare the suggested approach with the most advanced techniques available. The suggested approach performs well when compared to state-of-the-art techniques, according to experimental results on several common microarray datasets.

References:

[1] Alonso-Betanzos, A.; Bolón-Canedo, V.; Morán-Fernández, L.; Sánchez-Marono, N. A Review of Microarray Datasets: Where to Find Them and Specific Characteristics. *Methods Mol. Biol.* **2019**, *1986*, 65–85. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[2] Bishop, C. *Neural Networks for Pattern Recognition*; Oxford University: Oxford, UK, 1995. [**Google Scholar**]

[3] Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [**Google Scholar**] [**CrossRef**][**Green Version**]

[4] Nogueira, A.; Ferreira, A.; Figueiredo, M. A Step Towards the Explainability of Microarray Data for Cancer Diagnosis with Machine Learning Techniques. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Online, 3–5 February 2022; pp. 362–369. [**Google Scholar**] [**CrossRef**]

[5] Garcia, S.; Luengo, J.; Saez, J.; Lopez, V.; Herrera, F. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 734–750. [**Google Scholar**] [**CrossRef**]

[6] Duda, R.; Hart, P.; Stork, D. *Pattern Classification*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2001. [**Google Scholar**]

[7] Escolano, F.; Suau, P.; Bonev, B. *Information Theory in Computer Vision and Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2009. [**Google Scholar**]

[8] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009. [**Google Scholar**]

[9] Guyon, I.; Gunn, S.; Nikravesh, M.; Zadeh, L. *Feature Extraction: Foundations and Applications*; Springer: Berlin/Heidelberg, Germany, 2006. [**Google Scholar**]

[10] Simon, R.; Korn, E.; McShane, L.; Radmacher, M.; Wright, G.; Zhao, Y. *Design and Analysis of DNA Microarray Investigations*; Springer: New York, NY, USA, 2003. [**Google Scholar**]

[11] Ferreira, A.; Figueiredo, M. Exploiting the bin-class histograms for feature selection on discrete data. In Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis, Santiago de Compostela, Spain, 17–19 June 2015; Springer: Cham, Switzerland, 2015; pp. 345–353. [**Google Scholar**]

[12] Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [**Google Scholar**] [**CrossRef**][**Green Version**]

[13] Dougherty, J.; Kohavi, R.; Sahami, M. Supervised and unsupervised discretization of continuous features. In *Machine Learning Proceedings 1995*; Elsevier: Amsterdam, The Netherlands, 1995; pp. 194–202. [**Google Scholar**]

[14] Fayyad, U.; Irani, K. Multi-interval discretization of continuous-valued attributes for classification learning. In Proceedings of the International Joint Conference on Uncertainty in AI, Washington, DC, USA, 9–11 July 1993; pp. 1022–1027. [**Google Scholar**]

[15] Alpaydin, E. *Introduction to Machine Learning*, 3rd ed.; The MIT Press: Cambridge, MA, USA, 2014. [**Google Scholar**]

[16] He, X.; Cai, D.; Niyogi, P. Laplacian score for feature selection. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 5–8 December 2005; MIT Press: Cambridge, MA, USA; Volume 18, pp. 507–514. [**Google Scholar**]

[17] Zhao, Z.; Liu, H. Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 1151–1157. [**Google Scholar**]

[18] Liu, L.; Kang, J.; Yu, J.; Wang, Z. A comparative study on unsupervised feature selection methods for text clustering. In Proceedings of the 2005 International Conference on Natural Language Processing and Knowledge Engineering, Wuhan, China, 30 October–1 November 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 597–601. [**Google Scholar**] [**CrossRef**]

[19] Fisher, R. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [**Google Scholar**] [**CrossRef**]

[20] Yu, L.; Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the International Conference on Machine Learning (ICML), Washington, DC, USA, 21–24 August 2003; pp. 856–863. [**Google Scholar**]

[21] Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **2005**, *27*, 1226–1238. [**Google Scholar**] [**CrossRef**]

[22] Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European Conference on Machine Learning, Catania, Italy, 6–8 April 1994; Springer: Berlin/Heidelberg, Germany, 1994; pp. 171–182. [**Google Scholar**]

[23] Ferreira, A.; Figueiredo, M. Efficient feature selection filters for high-dimensional data. *Pattern Recognit. Lett.* **2012**, *33*, 1794–1804. [**Google Scholar**] [**CrossRef**][**Green Version**]

[24] Zhao, Z.; Morstatter, F.; Sharma, S.; Alelyani, S.; Anand, A.; Liu, H. *Advancing Feature Selection Research—ASU Feature Selection Repository*; Technical Report; Computer Science & Engineering, Arizona State University: Tempe, AZ, USA, 2010. [**Google Scholar**]

[25] Furey, T.; Cristianini, N.; Duffy, N.; Bednarski, D.; Schummer, M.; Haussler, D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **2000**, *16*, 906–914. [**Google Scholar**] [**CrossRef**][**Green Version**]

[26] Remeseiro, B.; Bolon-Canedo, V. A review of feature selection methods in medical applications. *Comput. Biol. Med.* **2019**, *112*, 103375. [**Google Scholar**] [**CrossRef**]

[27] Pudjihartono, N.; Fadason, T.; Kempa-Liehr, A.; O'Sullivan, J. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front. Bioinform.* **2022**, *2*, 927312. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[28] Dhal, P.; Azad, C. A comprehensive survey on feature selection in the various fields of machine learning. *Appl. Intell.* **2022**, *52*, 4543–4581. [**Google Scholar**] [**CrossRef**]

[29] Lazar, C.; Taminau, J.; Meganck, S.; Steenhoff, D.; Coletta, A.; Molter, C.; Schaetzen, V.; Duque, R.; Bersini, H.; Nowé, A. A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2012**, *9*, 1106–1119. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[30] Manikandan, G.; Abirami, S. A Survey on Feature Selection and Extraction Techniques for High-Dimensional Microarray Datasets. In *Knowledge Computing and its Applications: Knowledge Computing in Specific Domains: Volume II*; Springer: Singapore, 2018; pp. 311–333. [**Google Scholar**] [**CrossRef**]

[31] Almugren, N.; Alshamlan, H. A Survey on Hybrid Feature Selection Methods in Microarray Gene Expression Data for Cancer Classification. *IEEE Access* **2019**, *7*, 78533–78548. [**Google Scholar**] [**CrossRef**]

[32] Arowolo, M.; Adebiyi, M.; Aremu, C.; Adebiyi, A. A survey of dimension reduction and classification methods for RNA-Seq data on malaria vector. *J. Big Data* **2021**, *8*, 50. [**Google Scholar**] [**CrossRef**]

[33] Alpaydin, E. *Introduction to Machine Learning*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2010. [**Google Scholar**]

[34] Boser, B.; Guyon, I.; Vapnik, V. A training algorithm for optimal margin classifiers. In Proceedings of the Annual ACM Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; ACM Press: New York, NY, USA, 1992; pp. 144–152. [**Google Scholar**]

[35] Burges, C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [**Google Scholar**] [**CrossRef**]
[36] Vapnik, V. *The Nature of Statistical Learning Theory*; Springe: New York, NY, USA, 1999. [**Google Scholar**]
[37] Hsu, C.; Lin, C. A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [**Google Scholar**] [**CrossRef**][**Green Version**]
[38] Weston, J.; Watkins, C. *Multi-Class Support Vector Machines*; Technical Report; Department of Computer Science, Royal Holloway, University of London: London, UK, 1998. [**Google Scholar**]
[39] Breiman, L. *Classification and Regression Trees*, 1st ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 1984. [**Google Scholar**]
[40] Quinlan, J. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [**Google Scholar**] [**CrossRef**][**Green Version**]
[41] Quinlan, J. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993. [**Google Scholar**]
[42] Quinlan, J. Bagging, boosting, and C4.5. In Proceedings of the National Conference on Artificial Intelligence, Portland, OR, USA, 4–8 August 1996; AAAI Press: Washington, DA, USA, 1996; pp. 725–730. [**Google Scholar**]
[43] Rokach, L.; Maimon, O. Top-down induction of decision trees classifiers—A survey. *IEEE Trans. Syst. Man, Cybern. Part C Appl. Rev.* **2005**, *35*, 476–487. [**Google Scholar**] [**CrossRef**][**Green Version**]
[44] Yip, W.; Amin, S.; Li, C. A Survey of Classification Techniques for Microarray Data Analysis. In *Handbook of Statistical Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 193–223. [**Google Scholar**] [**CrossRef**]
[45] Statnikov, A.; Tsamardinos, I.; Dosbayev, Y.; Aliferis, C. GEMS: A system for automated cancer diagnosis and biomarker discovery from microarray gene expression data. *Int. J. Med. Inform.* **2005**, *74*, 491–503. [**Google Scholar**] [**CrossRef**] [**PubMed**]
[46] Witten, I.; Frank, E.; Hall, M.; Pal, C. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Morgan Kauffmann: Mateo, CA, USA, 2016. [**Google Scholar**]
[47] Meyer, P.; Schretter, C.; Bontempi, G. Information-theoretic feature selection in microarray data using variable complementarity. *IEEE J. Sel. Top. Signal Process.* **2008**, *2*, 261–274. [**Google Scholar**] [**CrossRef**]
[48] Statnikov, A.; Aliferis, C.; Tsamardinos, I.; Hardin, D.; Levy, S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics* **2005**, *21*, 631–643. [**Google Scholar**] [**CrossRef**][**Green Version**]
[49] Diaz-Uriarte, R.; Andres, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [**Google Scholar**] [**CrossRef**][**Green Version**]
[50] Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [**Google Scholar**] [**CrossRef**][**Green Version**]
[51] Li, Z.; Xie, W.; Liu, T. Efficient feature selection and classification for microarray data. *PLoS ONE* **2018**, *13*, 0202167. [**Google Scholar**] [**CrossRef**] [**PubMed**][**Green Version**]
[52] Consiglio, A.; Casalino, G.; Castellano, G.; Grillo, G.; Perlino, E.; Vessio, G.; Licciulli, F. Explaining Ovarian Cancer Gene Expression Profiles with Fuzzy Rules and Genetic Algorithms. *Electronics* **2021**, *10*, 375. [**Google Scholar**] [**CrossRef**]
[53] Saeys, Y.; Inza, I.; naga, P.L. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [**Google Scholar**] [**CrossRef**][**Green Version**]
[54] AbdElNabi, M.L.R.; Wajeeh Jasim, M.; El-Bakry, H.M.; Hamed, N.; Taha, M.; Khalifa, N.E.M. Breast and Colon Cancer Classification from Gene Expression Profiles Using Data Mining Techniques. *Symmetry* **2020**, *12*, 408. [**Google Scholar**] [**CrossRef**][**Green Version**]
[55] Alonso-González, C.J.; Moro-Sancho, Q.I.; Simon-Hurtado, A.; Varela-Arrabal, R. Microarray gene expression classification with few genes: Criteria to combine attribute selection and classification methods. *Expert Syst. Appl.* **2012**, *39*, 7270–7280. [**Google Scholar**] [**CrossRef**]

[56] Jirapech-Umpai, T.; Aitken, S. Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes. *BMC Bioinform.* **2005**, *6*, 148. [**Google Scholar**] [**CrossRef**][**Green Version**]

[57] Zhu, Z.; Ong, Y.; Dash, M. Markov blanket-embedded genetic algorithm for gene selection. *Pattern Recognit.* **2007**, *40*, 3236–3248. [**Google Scholar**] [**CrossRef**]

[58] Van't Veer, L.J.; Dai, H.; Van De Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; Van Der Kooy, K.; Marton, M.J.; Witteveen, A.T.; et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **2002**, *415*, 530–536. [**Google Scholar**] [**CrossRef**] [**PubMed**][**Green Version**]

[59] Pomeroy, S.L.; Tamayo, P.; Gaasenbeek, M.; Sturla, L.M.; Angelo, M.; McLaughlin, M.E.; Kim, J.Y.; Goumnerova, L.C.; Black, P.M.; Lau, C.; et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **2002**, *415*, 436–442. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[60] Alon, U.; Barkai, N.; Notterman, D.A.; Gish, K.; Ybarra, S.; Mack, D.; Levine, A.J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6745–6750. [**Google Scholar**] [**CrossRef**]

[61] Golub, T.; Slonim, D.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J.; Coller, H.; Loh, M.; Downing, J.; Caligiuri, M.; et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **1999**, *286*, 531–537. [**Google Scholar**] [**CrossRef**][**Green Version**]

[62] Bhattacharjee, A.; Richards, W.; Staunton, J.; Li, C.; Monti, S.; Vasa, P.; Ladd, C.; Beheshti, J.; Bueno, R.; Gillette, M.; et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Natl. Acad. Sci. USA* **2001**, *98*, 13790–13795. [**Google Scholar**] [**CrossRef**]

[63] Alizadeh, A.; Eisen, M.; Davis, R.; Ma, C.; Lossos, I.; Rosenwald, A.; Boldrick, J.; Sabet, H.; Tran, T.; Yu, X.; et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **2000**, *403*, 503–511. [**Google Scholar**] [**CrossRef**]

[64] Armstrong, S.A.; Staunton, J.E.; Silverman, L.B.; Pieters, R.; den Boer, M.L.; Minden, M.D.; Sallan, S.E.; Lander, E.S.; Golub, T.R.; Korsmeyer, S.J. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat. Genet.* **2002**, *30*, 41–47. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[65] Basegmez, H.; Sezer, E.; Erol, C. Optimization for Gene Selection and Cancer Classification. *Proceedings* **2021**, *74*, 21. [**Google Scholar**] [**CrossRef**]

[66] Khan, J.; Wei, J.; Ringner, M.; Saal, L.; Ladanyi, M.; Westermann, F.; Berthold, F.; Schwab, M.; Antonescu, C.; Peterson, C.; et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **2001**, *7*, 673–679. [**Google Scholar**] [**CrossRef**] [**PubMed**]