

Kernel PCA-Enhanced Deep Learning for Cancer Classification in High-Dimensional Microarray Gene Expression Data

Anil Kumar R.J.
Maharanis Science
College for Women,
Mysuru Karnataka, India

Veena M.N.
PES College for
Engineering, Mandya,
Karnataka India

Monica R.
DOS in CS, UOM,
Mysuru, Karnataka India

Nirmala M.S.
Government College for
Women, Mandya,
Karnataka, India

ABSTRACT

Gene expression datasets used for cancer analysis are frequently high-dimensional and complex, making accurate bracketing delicate. This work presents a harmonious machine learning technique for classification of cancer types using multiple reference gene expression datasets, including leukaemia, DLBCL, brain, breast cancer, Golub, and colon cancer. Originally, the datasets are pre-processed using standard point scaling to reduce variations in gene expression values. To address the dimensionality problem, KPCA with a RBF is employed to extract applicable nonlinear features. Latterly, the class markers are converted to a numerical format, and Min-Max normalization is used for enhancing learning effectiveness. The reused data is divided for training and testing the sets, and a feedforward deep neural network is trained for cancer prediction. The model's performance is estimated using bracket delicacy. The experimental results demonstrate the proposed frame effectively handles high-dimensional gene expression data and achieves harmonious bracket performance across five cancer datasets.

Keywords

RBF, DLBCL, KPCA, Gene expression datasets.

1. INTRODUCTION

Cancer continues to pose a serious threat to global health, and identifying it accurately at an early stage remains a major challenge [3], [4]. Recent advances in technologies such as microarray analysis and RNA sequencing have made it possible to measure the activity of thousands of genes at once, offering deeper insight into the biological mechanisms of different cancers [16][17]. Datasets related to leukemia (including the Golub dataset), diffuse large B-cell lymphoma (DLBCL), breast cancer, brain cancer, and colon cancer are used commonly to study these patterns [5][8]. However, these datasets are meant to be high-dimensional, containing many gene features but only a limited number of samples. This imbalance makes it difficult for machine learning models to perform effectively and often leads to issues such as overfitting and increased computational complexity [18][20].

To make this data more manageable, the techniques of dimensionality reduction methods are applied to retain the most relevant information while reducing noise and redundancy [15]. In this study, Kernel Principal Component Analysis (KPCA) is used for non-linear dimensionality reduction approach with a Radial Basis Function (RBF) kernel to capture complex relationships within micro array data elements [1][2]. The decreased features are then used to train classification models such as Artificial Neural Networks (ANN) to classify different cancer types [11], [12]. Accuracy, precision, recall,

F1-score, and confusion matrix analysis are used for evaluation of the performance of the model [3], [4].

The reduced set of features is obtained through KPCA is then provided as input to classification models such as Artificial Neural Networks (ANN) for cancer classification. The neural network learns complex patterns from the transformed data and predicts the corresponding cancer classes [11]. To improve learning efficiency and model stability, preprocessing techniques such as feature scaling and normalization process are also included into the workflow [18], [19].

The proposed approach's performance is evaluated using several widely accepted metrics, such as accuracy, precision, recall, F1-score, AUC-ROC, log-loss, and confusion matrix analysis [3][20]. These measures provide an understanding of the comprehensive model's classification capability and prediction reliability across different datasets.

Experimental analysis demonstrates that dimensionality reduction significantly improves classification performance by eliminating irrelevant and redundant genes while preserving informative features [13], [15]. The proposed framework also reduces training complexity and enhances the ability of the classifier to generalize on unseen samples.

The main aim of this work is to investigate the efficiency of KPCA in handling high-dimensional cancer datasets and to evaluate the performance of deep learning-based classification models across multiple cancer types. The study further aims to provide an efficient and reliable framework for cancer prediction using gene expression data, which may support future research in biomedical data analysis and computer-aided diagnosis systems [11], [18].

2. RELATED WORKS

Cancer is a class of diseases characterized by abnormal cell growth that leads to the formation of cancerous cells [3], [4]. In a healthy body, cell growth and cell death are naturally regulated over time. However, genetic damage caused by internal and environmental factors can result in cells that continue to divide uncontrollably and form tumours [17], [18]. Some of the major internal causes of cancer include abnormal cell division and DNA damage, while external factors include exposure to tobacco smoke, harmful chemicals, radiation, and ultraviolet rays from sunlight [18].

According to molecular biology studies, Cancer types can be distinguished through their specific gene expression patterns [5][8]. Microarray classification helps map gene-phenotype relationships and predict disease conditions in untested samples[13], [16]. By constructing predictive models based on input feature patterns, classification problems can be

effectively solved [3], [20]. Using training and testing data, prediction models are developed to generate accurate classification results.

Analization of microarray gene expression data is the toughest job in the areas of research such as statistics, machine learning, bioinformatics, genomics, computational bio and pattern recognition [18][20]. Due to the curse of high dimensionality with the small samples, it contains irrelevant features which makes the analysis of microarray cancer dataset difficult [13][15]. In addition, the datasets related to medical field may contain noise, class imbalance which might produce the problem of overfitting and affect the accuracy of classification [19].

Microarray data analysis is therefore essential for understanding cancer-related biological mechanisms and improving cancer diagnosis, particularly during the early stages [16], [17]. Early detection supports medical professionals in designing effective treatment strategies and improving patient survival rates. Classification of labelled microarray cancer data involves several important stages, including data collection, feature scaling, feature selection, classification, and post-classification analysis [18], [19].

Feature selection plays a critical role in identifying the most informative genes from tens of thousands of highly correlated features [13]. The selected features are then provided to classifiers to improve prediction accuracy. Identifying an optimal subset of features increases computational efficiency, improves model stability, and enhances classification performance [15].

In this work, a deep learning algorithm is employed for microarray cancer classification [11], [12]. Since deep learning models require meaningful feature representations, dimensionality reduction techniques are applied before classification. Benchmark datasets including leukemia (Golub dataset), diffuse large B-cell lymphoma (DLBCL), breast cancer, brain cancer, and colon cancer are used to validate the proposed methodology [5][8]. The Min-Max scaling approach is applied to normalize feature values and reduce bias toward features with larger numerical values [18]. Experimental results demonstrate that the proposed model achieves accurate prediction performance when compared with recent related approaches [11], [13].

3. PROPOSED METHODOLOGY TO CLASSIFY BINARY CLASS DATASETS

3.1 Feature Scaling

This work unfolds the usage of different techniques such as reduction of dimensionality, feature scaling, deep feed forward classification as a classifier technique, which is well tuned parameterised configuration through neural networks. The provided approach will initially include loading of raw microarray cancer dataset followed by normalization with the technique called min-max normalization which is helpful for reducing dimensionality which aids in good classification of an algorithm.

This approach of feature scaling technique is widely used in machine learning and pattern recognition, mainly concentrates on standardising data. By scaling the data elements to certain range, the model improves its quality of production up to certain percentage. The microarray datasets contains a high variance of characteristics so we recommend that the feature-scaling for normalization of data is a necessary step to be followed. To use the activation function

called sigmoid, the min-max type of normalization is used, in which it brings the data elements between 0 to 1. While model training, we include 0.5 as a threshold for the binary classification.

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

Where,

“X” refers to normalised data. “Xi” is a original feature value “Xmin” denotes minimum value and “Xmax” is a maximum value.

3.2 Dimensionality Reduction-based Feature Selection using KPCA

In this study, dimensionality reduction is applied as an essential preprocessing step to transform high-dimensional gene expression data into a more compact and informative representation [15]. Instead of traditional Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) is employed to better capture the complex and non-linear relationships present in microarray cancer datasets [1], [2]. High-dimensional biological data is mapped into a higher-dimensional feature space using a kernel function, such as the Radial Basis Function (RBF), where linear separation becomes more feasible [1]. This transformation enables the extraction of meaningful patterns while reducing the overall dimensionality of the data.

Let X represent the original dataset consisting of n samples and d features, where each sample is defined as:

$$X = [X_1, X_2, \dots, X_n], X \in \mathbb{R}^{n \times d}$$

Here, each sample $X_i = [x_1, x_2, \dots, x_d]$ represents a high-dimensional gene expression profile. In KPCA, the data is implicitly mapped to a higher-dimensional feature space through a non-linear mapping function $\phi(X)$, and principal components are extracted in this transformed space. Unlike PCA, which maximizes variance in the original feature space, KPCA captures non-linear variance by computing eigenvectors of the kernel (Gram) matrix.

The transformed dataset is represented as:

$$Z = \phi(X)W = [z_1, z_2, \dots, z_k], Z \in \mathbb{R}^{n \times k}, k \ll d$$

where W denotes the eigenvectors obtained from the kernel matrix, and k represents the reduced number of features. The resulting feature set Z retains the most significant information from the original dataset while significantly reducing dimensionality. This reduction helps improve computational efficiency, minimize overfitting, and enhance classification performance [13][15].

During preprocessing, it is observed that applying dimensionality reduction separately to training and testing datasets may lead to inconsistencies in feature dimensions. Since classifiers require consistent input dimensions, it is necessary to ensure that both datasets share the same feature space. To address this, the reduced dimension P is selected based on the minimum number of features obtained from either the training set (T_r) or the testing set (T_s):

$$P = \min(T_r, T_s)$$

Additionally, only the common features present in both training and testing datasets are retained, while non-overlapping features are discarded. This ensures uniformity in both feature dimension and representation, allowing the classification model to be trained and evaluated effectively.

3.3 Deep Learning-based Classification

This section describes the proposed framework for microarray data classification, which combines preprocessing, dimensionality reduction using Kernel Principal Component Analysis (KPCA), and classification using a deep feed-forward neural network [1],[11],[12]. The overall workflow of the model is illustrated in Figure 1. The main objective is to handle high-dimensional gene expression data efficiently by reducing redundancy and improving classification performance.

The process begins with loading raw microarray datasets, which typically contain a large number of gene features and a comparatively small number of samples [5][8]. Due to this high dimensionality, preprocessing becomes essential. Initially, the data is normalized using StandardScaler, which transforms the features to have zero mean and unit variance. This step ensures that all features contribute equally during model training and prevents bias toward features with larger magnitudes [18], [19].

Following normalization, Kernel Principal Component Analysis (KPCA) is applied for dimensionality reduction. Unlike traditional PCA, KPCA is capable of capturing non-linear relationships in the data by mapping it into a higher-dimensional feature space using a kernel function [1], [2]. In this work, the Radial Basis Function (RBF) kernel is used due to its effectiveness in handling complex gene expression patterns.

Mathematically, KPCA transforms the input data x into a new feature space using a mapping function $\phi(x)$, and computes the kernel matrix:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

where γ is a kernel parameter controlling the spread of the RBF function. The transformed features are obtained by solving the eigenvalue problem of the kernel matrix, and only the most significant components are retained. This reduces the number of features while preserving important information.

After dimensionality reduction, the processed data is fed into a deep feed-forward neural network for classification [11][12].

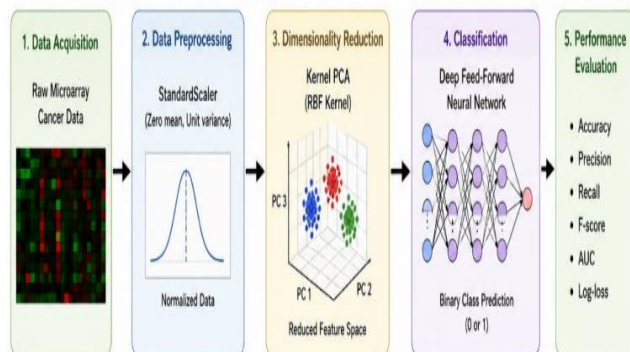


FIGURE 1: Structure of the Deep Learning Approach Proposal.

This section presents a clear analysis of the experimental results obtained using the proposed methodology. The model was evaluated across multiple datasets and assessed using different performance metrics to better understand its effectiveness and overall behaviour.

The deep feed-forward neural network is a layered architecture where information flows in a single direction—from input to output—without any feedback connections. Let the input feature vector be represented as $Z = \{z_1, z_2, \dots, z_n\}$. The network learns a mapping function:

$$y_p = f(Z)$$

where y_p is the predicted class label.

Each neuron computes a weighted sum of inputs followed by an activation function:

$$a = \sum_{i=1}^n w_i z_i + b$$

where w_i represents weights and b is the bias term. The result is passed through the sigmoid activation function, defined as:

$$f(a) = \frac{1}{1 + e^{-a}}$$

This function maps outputs into the range $[0, 1]$, making it suitable for binary classification problems.

The network consists of multiple hidden layers, each learning increasingly abstract representations of the input data. The final output layer contains a single neuron that predicts the class label based on a threshold value (typically 0.5).

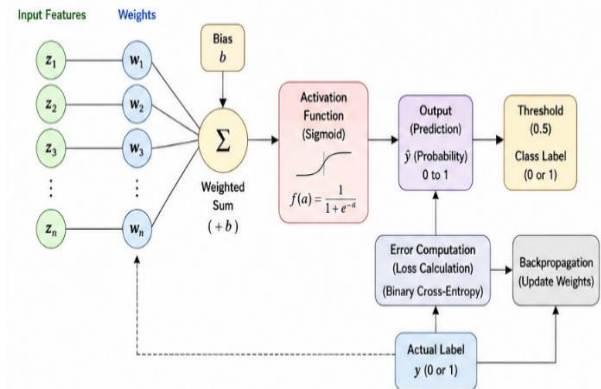


FIGURE 2: Structure of the Proposed Model.

During training, the predicted output is compared with the actual class label y_i , and the error is calculated using the **binary cross-entropy loss function**:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

where:

- N is the number of samples
- y_i is the true label

- p_i is the predicted probability

The error is minimized using the **backpropagation algorithm**, which updates the weights iteratively to improve model performance.

The sigmoid activation function is used in this model to transform the output into a probabilistic value between 0 and 1. It is defined as:

$$f(a) = \frac{1}{1 + e^{-a}}$$

The graphical representation of the sigmoid activation function is shown in Figure. 6.3.

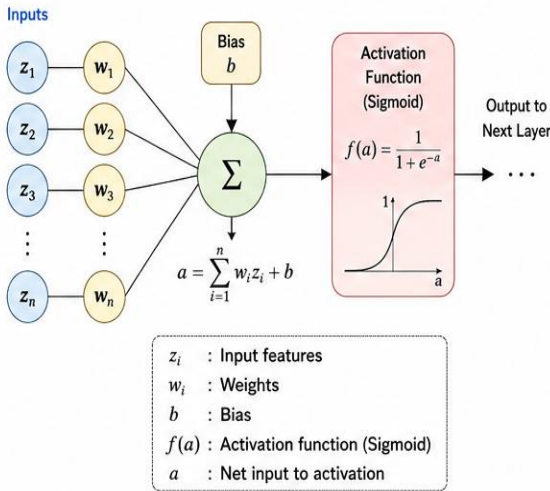


FIGURE 3: The Working Model of The Activation Function

The sigmoid function enables the model to interpret outputs as probabilities, making it suitable for binary classification tasks. A threshold value of 0.5 is used to assign class labels.

During training, the predicted output is compared with the actual class label, and the error is computed using the binary cross-entropy loss function:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

The model parameters are updated iteratively using the Adam optimizer to minimize the loss and improve prediction accuracy.

3.4 Parameter Settings

The proposed model uses a multi-layer deep feed-forward architecture consisting of several dense and dropout layers[11],[12]. The number of neurons decreases gradually across layers to enable hierarchical feature extraction.

TABLE 1: Proposed Deep Learning Model Configuration

Layer Type	Output Shape	Parameters
Dense Layer 1	(None, 200)	4600

Dropout Layer	(None, 200)	0
Dense Layer 2	(None, 100)	20100
Dropout Layer	(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		33,661

To optimize the model, the **Adam optimizer** is used, which adaptively adjusts learning rates during training. The default parameters are:

$$\beta_1 = 0.9, \beta_2 = 0.99, \epsilon = 10^{-8}$$

The model is trained for 100 epochs across all datasets. A threshold value of 0.5 is used to convert predicted probabilities into binary class labels.

4. DATASET DESCRIPTION

To evaluate the proposed model of multidimensional microarrays, five standard microarray datasets were obtained from <http://www.gems-system.org> and <http://www.ncbi.nlm.nih.gov/geo/> to assess high-dimensional biomedical datasets [5][8]. The DLBCL datasets consists of 7129 features which has 60 samples in which 39 are cancerous and 21 are non cancerous[6].The colon cancer's dataset has 2000 collections of genes and 62 samples, wherein 22 are of normal people and 40 are said to be cancered people. There are 15153 number of genes in Golub dataset[5]. It consists of 253 samples with 162 samples of tumour and 91 samples of non tumorous cases .On top of that, the Leukemia dataset has 7129 characteristics with 72 samples of data where 25 number of cases belongs to Acute Myeloid Leukemia(AML) and 47 samples comes under Acute Lymphoblastic Leukaemia(ALL)[5].Further, the colon cancer dataset will have 12600 features where there are 102 samples, under which 52 corresponds to abnormal cell condition 50 goes for normal gene collection. The Breast cancer dataset is made up of 24481of total genes and 97 samples. In which, 46 cases are of positive and 51 are of negative samples.

Later, KPCA is used for dimensionality reduction via this technique, we can ensure that the vast dimensional features which are not relevant can be removed by enabling the model to be performed good and give good prediction rate.

It is said that the most features in our datasets are negligible portions, so this process helps in predicting class labels more efficiently.

4.1 Performance Measures

To assess the performance of the proposed model, a set of widely accepted evaluation metrics was used [3][4][20]. These include classification accuracy, precision, recall, F-score, and log-loss, all of which can be derived or interpreted using the confusion matrix. Together, these measures provide a comprehensive understanding of how well the model performs across different aspects of prediction.

Accuracy, defined in Equation, reflects the overall correctness of the model by considering four key outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN)[3][20]. It is calculated as the proportion of correctly classified samples out of the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall, given in Equation, measures the model's ability to correctly identify positive instances. It focuses on how many actual positive cases are captured by the model:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision, also known as Positive Predictive Value (PPV) and defined in Equation, evaluates how many of the predicted positive cases are truly positive:

$$\text{Precision} = \frac{TP}{TP + FP}$$

The F-score, presented in Equation, combines precision and recall into a single metric by taking their harmonic mean. This helps in balancing both false positives and false negatives:

$$\text{F-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

In addition to these metrics, log-loss (Equation) was used to measure the error of the model based on predicted probabilities. It evaluates how close the predicted probabilities are to the actual class labels. In this equation, N represents the total number of samples, y_i denotes the true class label, and p_i indicates the predicted probability for the i^{th} sample:

$$\text{Log-loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

Log-loss values range from zero to infinity, where a value closer to zero indicates better model performance. A perfect classifier would achieve a log-loss of zero, meaning its predicted probabilities exactly match the true class labels.

A lower log-loss value indicates better model performance, with zero representing a perfect classifier.

Furthermore, the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (AUC) were used to evaluate the model's discriminative ability. The ROC curve plots the True Positive Rate (TPR), also known as recall, against the False Positive Rate (FPR) at different threshold values:

$$\text{TPR} = \frac{TP}{TP + FN}, \text{ FPR} = \frac{FP}{FP + TN}$$

The AUC value summarizes the ROC curve into a single scalar, ranging from 0 to 1. A value closer to 1 indicates strong classification performance, while a value around 0.5 suggests random guessing. The AUC-ROC metric is particularly useful when evaluating models on datasets with class imbalance, as it considers performance across all classification thresholds.

5. RESULTS ANALYSIS

This subsection provides a thorough analysis of the experimental results obtained using the proposed methodology. The model's performance is evaluated across multiple datasets using various performance metrics to gain a clear understanding of its effectiveness.

TABLE 2: Experimental Results on all datasets

Dataset	Training set size	Test set size	Accuracy	Precision	Recall	F-Measure	AUC	Log-loss error
Breast cancer	120	31	0.74	0.66	0.74	0.67	0.97	0.96
Leukemia	51	13	0.84	0.75	0.84	0.78	0.96	0.36
Brain	33	9	0.55	0.70	0.55	0.56	0.74	1.57
Golub	56	15	0.80	0.84	0.8	0.77	0.96	0.38
DLBCL	60	16	0.81	0.80	0.81	0.80	0.87	0.38
Colon	49	13	0.69	0.80	0.69	0.64	0.83	0.66

The performance of the proposed model across different datasets is summarized in Table 2. The evaluation was carried out using multiple metrics, including accuracy, precision, recall, F-score, AUC, and log-loss.

For the **breast cancer dataset**, the model achieved an accuracy of 0.74, with balanced precision (0.66) and recall (0.74), resulting in an F-score of 0.67. The AUC value of 0.97 indicates excellent classification capability, while the low log-loss (0.96) suggests reliable probability predictions.

In the case of the **leukemia dataset**, the model performed better in terms of accuracy (0.84) and recall (0.84), with a moderate precision of 0.75. The AUC value (0.96) confirms strong discriminative performance, although the log-loss (0.36) indicates some variability in probability estimation.

For the **brain cancer dataset**, the model showed comparatively lower performance, with an accuracy of 0.55. While the recall (0.70) is relatively higher, the precision (0.55) and F-score (0.56) suggest inconsistency in predictions. The AUC value of 0.74 indicates moderate classification ability, and the higher log-loss (1.57) reflects less confident predictions.

The **Golub dataset** produced strong results, with an accuracy of 0.80, precision of 0.84, and recall of 0.80. The AUC score of 0.96 demonstrates excellent separability between classes, supported by a low log-loss value of 0.38.

Similarly, for the **DLBCL dataset**, the model achieved balanced performance, with accuracy, precision, recall, and F-score all around 0.80. The AUC value of 0.87 indicates good classification performance, and the log-loss (0.38) suggests stable probability predictions.

Similarly, the **Colon dataset**, the model achieved better performance, with accuracy, precision, recall, and F-score all around 0.80. The AUC value of 0.83 indicates a moderate classification performance, and the log-loss (0.66) suggests medium probability predictions.

Overall, the result shows that the proposed model performed effectively with most datasets, particularly in high-dimensional gene expression data. However, performance variations across datasets highlight the impact of data complexity and feature characteristics on model behavior.

5.1 Visual Representation of the Results

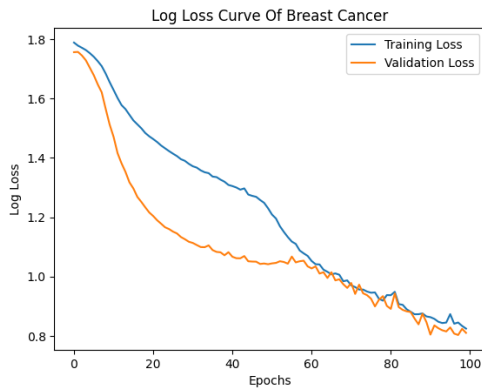
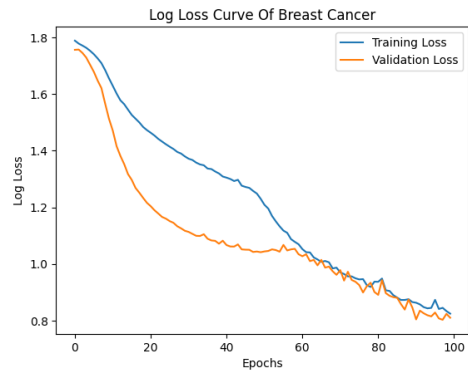


FIGURE 4 Loss Less Curve of Brain and Breast Cancer Cells

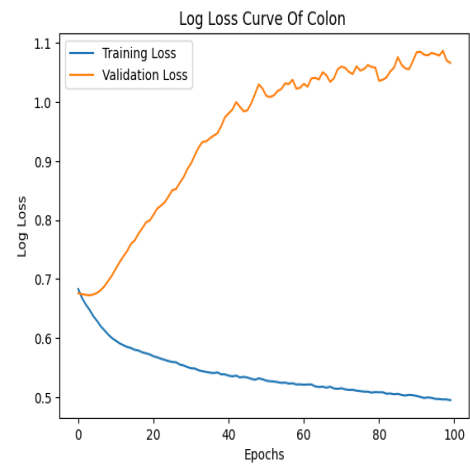
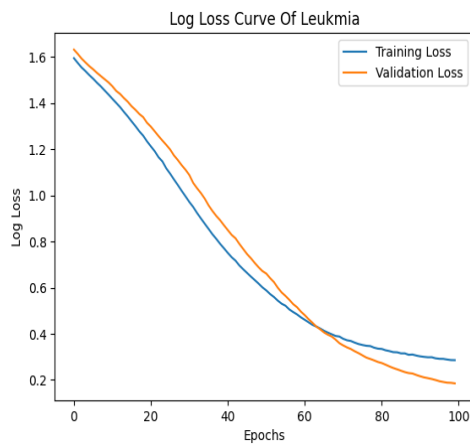


FIGURE 5 Loss Less curve of Leukemia and Colon Cancer

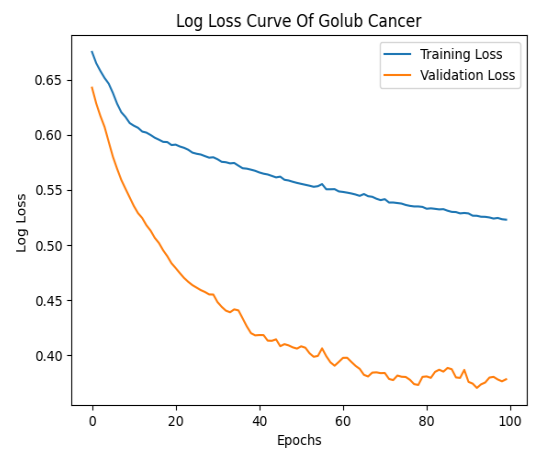
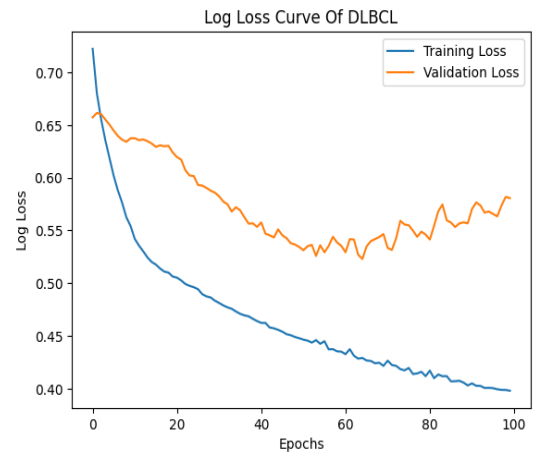


FIGURE 6 Loss Less curve of DLBCL and Golub Cancer

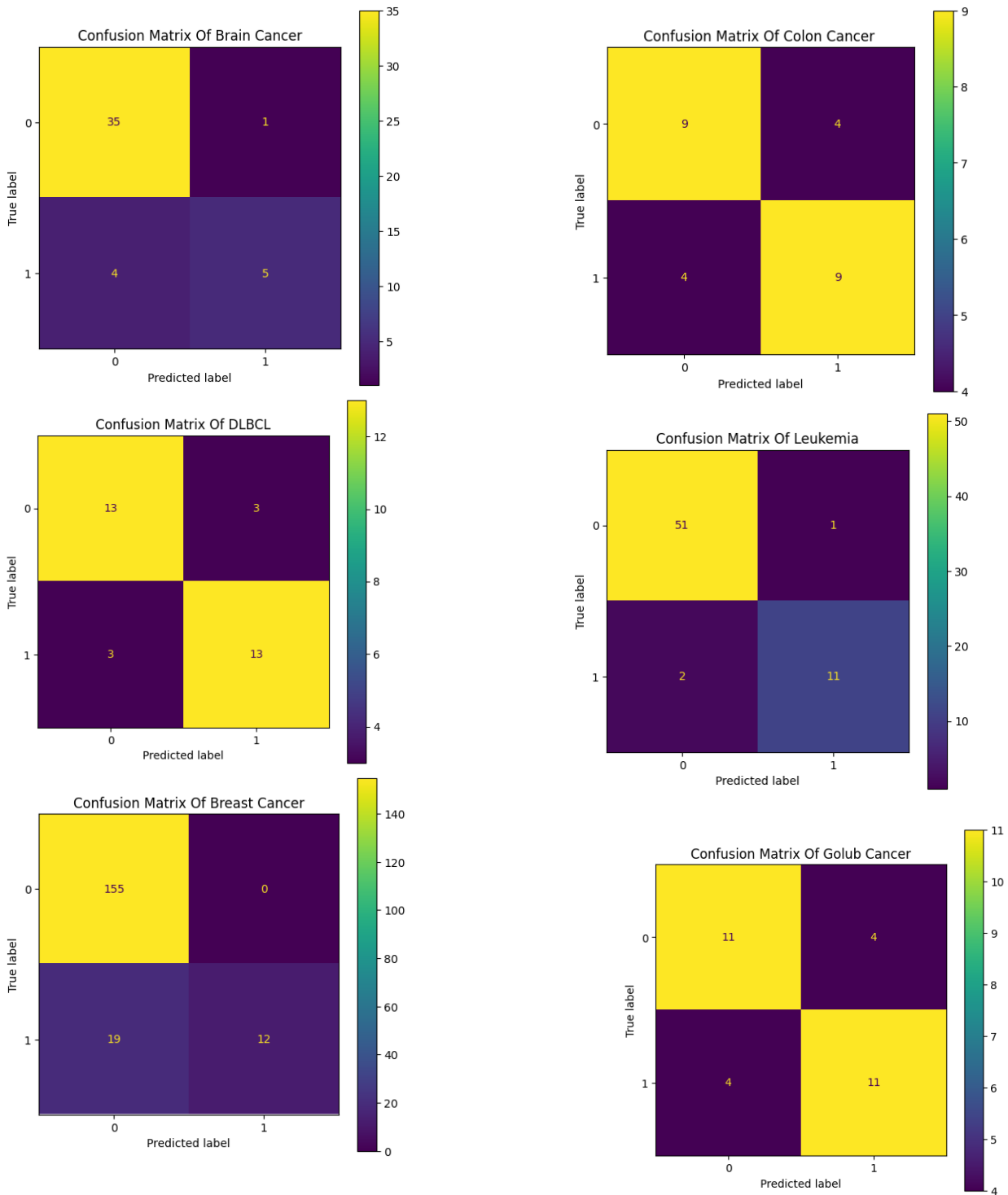


FIGURE 7 Confusion Matrix of Brain Cancer DLBCL Breast Cancer, Colon Cancer, Golub and Leukaemia

The visualised representation of the confusion matrix provides a good understanding of the classification performance of the proposed model by illustrating the distribution of correctly and incorrectly classified samples. A higher number of true positive and true negative predictions compared to false predictions indicates that the model is capable of effectively distinguishing between different cancer classes. The confusion matrix also helps in identifying misclassification patterns and evaluating the reliability of the classifier across various datasets.

Similarly, the log-loss visualization demonstrates the convergence behavior of the deep learning model during training. A gradual reduction in log-loss values over epochs indicates that the model learns efficiently and improves its prediction capability with training. Lower log-loss values represent better probability estimation and higher confidence in predictions. The obtained results show that the proposed model achieves relatively low error values for most datasets, reflecting stable learning and improved generalization performance.

Overall, both the confusion matrix and log-loss analysis confirm the effectiveness of the proposed KPCA-based deep learning framework in classifying high-dimensional microarray cancer datasets with reliable predictive performance.

6. CONCLUSION

The experimental results shows that the proposed KPCA-based deep learning framework is effective in handling high-dimensional microarray cancer datasets. By applying Kernel Principal Component Analysis for dimensionality reduction, the model successfully removed redundant and less informative gene features while preserving the most relevant characteristics required for classification. This reduction in dimensionality improved computational efficiency and enhanced the overall learning capability of the classifier.

Among the evaluated datasets, the leukemia dataset achieved the highest classification accuracy of 0.84 along with a strong AUC value of 0.96, indicating excellent discriminative performance. Similarly, the Golub and DLBCL datasets also produced stable and reliable results with balanced precision, recall, and F-score values. The breast cancer dataset succeeds with a high AUC score of 0.97, showing strong classification capability despite moderate accuracy. In contrast, the brain cancer dataset exhibited comparatively lower performance, which may be due to the complexity of the dataset and limited sample size. The colon cancer dataset produced satisfactory results with balanced evaluation metrics and acceptable prediction reliability.

Overall, the findings indicate that integrating KPCA with deep feed-forward neural networks can effectively improves classification performance on gene expression datasets. The proposed model demonstrates the ability to capture complex nonlinear patterns within microarray data and achieve reliable predictions across multiple cancer types. This framework can serve as a useful approach for biomedical data analysis and may provide to the development of computer-aided cancer diagnosis systems in future research.

7. REFERENCES

- [1] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [2] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proc. IEEE Neural Networks for Signal Processing*, 1999.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.
- [5] T. R. Golub et al., "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.
- [6] A. A. Alizadeh et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, pp. 503–511, 2000.
- [7] D. Singh et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.
- [8] S. Ramaswamy et al., "Multiclass cancer diagnosis using tumor gene expression signatures," *PNAS*, vol. 98, no. 26, pp. 15149–15154, 2001.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [12] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.
- [13] I. Guyon et al., "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.
- [14] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. ICML*, 2004.
- [15] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2002.
- [16] M. P. S. Brown et al., "Knowledge-based analysis of microarray gene expression data," *PNAS*, vol. 97, no. 1, pp. 262–267, 2000.
- [17] D. K. Slonim, "From patterns to pathways: Gene expression data analysis comes of age," *Nature Genetics*, vol. 32, pp. 502–508, 2002.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Burlington, MA, USA: Morgan Kaufmann, 2011.
- [19] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Pearson, 2006.
- [20] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.