International Journal of Advanced Technology and Engineering Exploration, Vol 12(124)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2024.111101498

Research Article

Breast cancer survival rate prediction using multimodal deep learning with multigenetic features

Yasmine M. Tabra[1*] and Furat N. Tawfeeq[2]
Department of Information and Communication, College of Information Engineering, Al -Nahrain University, Jadriya, Baghdad, Iraq[1]
Website Division, University of Baghdad, Jadriya, Baghdad, Iraq[2]

Received: 17-August-2024; Revised: 16-March-2025; Accepted: 19-March-2025

Abstract
Breast cancer is a heterogeneous disease characterized by molecular complexity. This research utilized three genetic expression profiles—gene expression, deoxyribonucleic acid (DNA) methylation, and micro ribonucleic acid (miRNA) expression—to deepen the understanding of breast cancer biology and contribute to the development of a reliable survival rate prediction model. During the preprocessing phase, principal component analysis (PCA) was applied to reduce the dimensionality of each dataset before computing consensus features across the three omics datasets. By integrating these datasets with the consensus features, the model's ability to uncover deep connections within the data was significantly improved. The proposed multimodal deep learning multigenetic features (MDL-MG) architecture incorporates a custom attention mechanism (CAM), bidirectional long short-term memory (BLSTM), and convolutional neural networks (CNNs). Additionally, the model was optimized to handle contrastive loss by extracting distinguishing features using a Siamese network (SN) architecture with a Euclidean distance metric. To assess the effectiveness of this approach, various evaluation metrics were applied to the cancer genome atlas (TCGA-BREAST) dataset. The model achieved 100% accuracy and demonstrated improvements in recall (16.2%), area under the curve (AUC) (29.3%), and precision (10.4%) while reducing complexity. These results highlight the model's efficacy in accurately predicting cancer survival rates.

Keywords
Breast cancer, Genetic expression, Survival prediction, Multimodal deep learning, Siamese network, Principal component analysis.

1.Introduction
Breast cancer is one of the leading causes of mortality among women. In countries with a low to moderate human development index (HDI), it accounts for 48% of cancer-related deaths [1]. Breast cancer originates in the tissues of the breast and can develop into two types of tumors: benign and malignant. Benign tumors are non-cancerous, do not spread to other parts of the body, and are not considered life-threatening. In contrast, malignant tumors are cancerous, invade nearby organs, and have the potential to spread by proliferating abnormal cells [2]. The management of the disease is significantly affected by effective screening programs that facilitate the early detection of neoplastic growths in the body [3, 4].

Insufficient predictive outcomes frequently result from the issue of cancer heterogeneity, which generates indistinguishable patient samples exhibiting varying risks [5].

Therefore, accurately predicting survival time is crucial for doctors to offer suitable treatment recommendations aimed at enhancing the survival rate and quality of life for cancer patients [6]. Clinical studies indicate that the inclusion of supplementary molecular markers associated with cancer in genetic data suggests that integrating genetic information from multiple sources may effectively tackle cancer heterogeneity [7]. Deep learning (DL) is a subset of machine learning (ML) within artificial intelligence (AI) that leverages neural networks (NNs) with multiple layers to enable advanced pattern recognition, representation learning, and decision-making, contributing to AI's ability to perform tasks that mimic human cognitive functions

*Author for correspondence

[8]. It may reduce the probability of pathologists making erroneous judgments due to factors such as probable fatigue and negligence [9]. Many different types of DL are available nowadays among them convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), etc. [10]. The hybrid CNN-LSTM shown its versatility across different applications and domains [11]. Prior studies examined the application of different genetic datasets utilizing NN; however, there was a deficiency in feature reduction that impacted the learning process and, consequently, prediction accuracy.

This research utilized DL techniques to predict cancer survival rates using multi-genetic datasets combined with a feature reduction strategy. The proposed DL model integrates a hybrid CNN, bidirectional LSTM, and a custom attention mechanism (CAM). Additionally, an enhanced feature reduction technique is utilized, which integrates data fusion of consensus features with principal component analysis (PCA). Cancer survival rate prediction can be based on numerical data, such as patient age, tumor size, and various clinical and biochemical markers. Meanwhile, categorical data are often used for classification, including patient gender, clinical stage, and pathological type. When DL models incorporate extensive datasets, they enhance predictive accuracy by identifying meaningful associations with clinical outcomes [12].

This paper is structured as follows: Section 1 provides an introduction to the methods used in this research. Section 2 presents an overview of prior studies on cancer prediction. Section 3 details the methods and materials employed in this study. After training and testing the model, Section 4 presents the results and compares them with existing literature. Section 5 discusses the findings and highlights the limitations of this work. Finally, Section 6 summarizes the results and offers recommendations for future research.

2.Literature review

This section provides a detailed elaboration of the related work.

2.1Convolutional neural network (CNN)

CNNs are recognized as some of the most prominent architectures in the realm of deep neural networks (DNNs) [13, 14]. CNNs replicate the neural activities observed in both humans and various other species. CNN consists of multiple interconnected layers, each of which is designed to execute a particular function. Typically, CNNs comprise three layers: convolution, pooling, and fully connected layers, culminating in a nonlinear SoftMax output layer [15, 16]. The convolution layer is the most essential component among the three layers, serving as the model's core. The convolution layer identifies the information and produces a duplicate of the information while maintaining the original size of the input data [17]. The architecture of CNN is constructed through the integration of convolutional and pooling layers. The fully connected layers, which use a nonlinear activation function, are added at the end of the CNN structure to show the results of the classification process. *Figure 1* illustrates the CNN architecture. The convolution design enables the potential for millions of training parameters. This indicates that a substantial volume of data is necessary to train the ML models effectively for achieving high accuracy in photo categorization [18, 19].

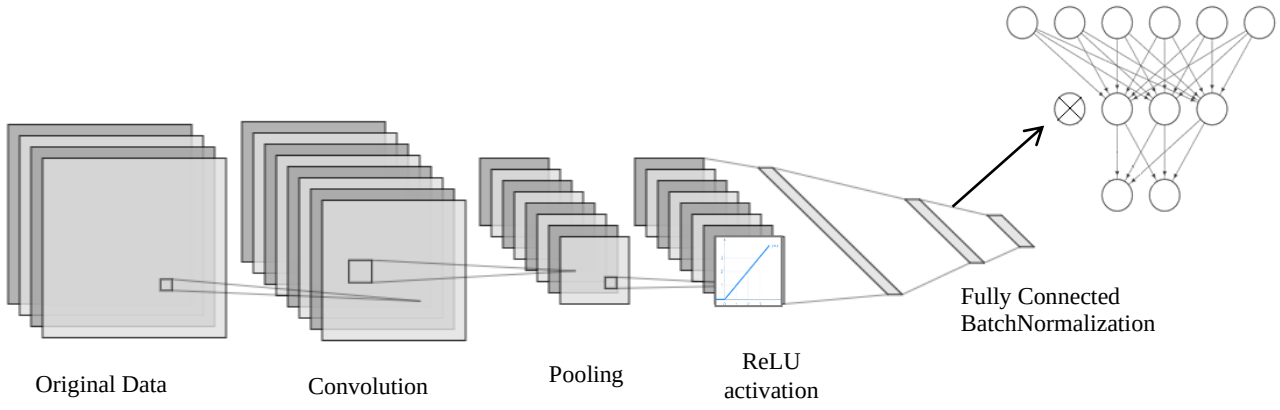


Figure 1 CNN architecture

Non-linearity is introduced into the NN by activation functions, which enables the NN to simulate complicated interactions between the data that is input and the data that is output. The non-linearity is necessary due to the fact that the majority of the data collected from the real world displays sophisticated patterns that cannot be accurately captured by linear functions [10]. Implementing rectified linear unit (ReLU) activation function involves a straightforward comparison of its input to zero, see Equation 1. Moreover, the derivative maybe 0 or 1, contingent upon the negativity of its input. The latter holds significant implications for backpropagation in the context of training. It demonstrates that calculating a neuron's gradient is relatively inexpensive [20, 21].

$$ReLU = \begin{cases} 0 & for\ x < 0 \\ 1 & for\ x \geq 0 \end{cases} \quad (1)$$

Another function that can be added to CNN is Batch Normalization. The Batch Normalization layer functions as a mask that maintains the integrity of all other neurons while deactivating the contributions of specific neurons to the subsequent layer [12, 22].

2.2Bidirectional long short-term memory (BLSTM) layers

BLSTM is a kind of LSTM network that enables improved learning at each data timestep by leveraging two units with the same input and coupled to the same output in the hidden layers to learn from both forward and backward time series data [23–25].

Assume the input vector x_t is considered by the forward hidden layer (h_t^f) in increasing order and by the backward hidden layer (h_t^b) in decreasing order. Finally, the results of h_t^f and h_t^b are combined to create the output y_t. The BLSTM model is implemented using the Equations 2-4: [25–27].

$$h_t^f = \tanh\left(W_{xh}^f x_t + W_{hh}^f h_{t-1}^f + b_h^f\right) \quad (2)$$

$$h_t^b = \tanh\left(W_{xh}^b x_t + W_{hh}^b h_{t+1}^b + b_h^b\right) \quad (3)$$

$$y_t = W_{hy}^f h_t^f + W_{hh}^b h_t^b + b_y \quad (4)$$

When a NN architecture incorporates an AM, it automatically determines when to examine data by prioritizing feature vectors with the most useful information over those with less valuable information. *Figure 2* contains the basic architecture of BLSTM [23].

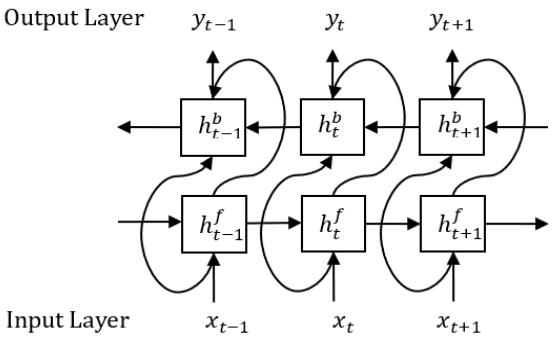


Figure 2 BLSTM architecture

2.3Attention mechanism (AM)

The ability of human perception to concentrate only on the relevant portion of information rather than processing it all at once [28] is a crucial characteristic. For instance, when human perception scans a scene, it concentrates on a specific area while disregarding extraneous information in order to extract the information it requires [29].

This led to the development of the custom attention model, which was first presented for the machine translation problem and is now a key idea in the literature on NNs for additional use cases [28]. While the fundamentals of attention models remain the same, advances in AM have allowed for better task-specific adaptation of AM [29]. Important stimuli with a high semantic level and a low saliency are gathered by the attention layer group for use in later architectural levels [28].

The principles of minimum redundancy and maximum relevance, abbreviated as min-redundancy and max-relevance (mRMR), are frequently utilized in the process of lowering dimensions or certain qualities. The objective is to choose characteristics that are extremely pertinent to the variable of interest but not excessively so when considered in isolation [30]. However, these concepts are not as directly applicable to NN architectures and consideration components.

Assume the final hidden state of i^{th} BLSTM as h_{it}, which is computed in Equation 5:

$$h_{it} = \left[h_t^f, h_t^b\right] \quad (5)$$

Next, the following formulas are used to calculate the AM in Equations 6 to 8:

$$e_{it} = \tanh(W_a h_{it} + b_a) \quad (6)$$

$$a_{it} = \frac{\exp(e_{it})}{\sum_{i=1}^{T} \exp(e_{jt})} \quad (7)$$

$$v_t = \sum_{t=1}^{T} a_{it}.h_{it} \quad (8)$$

The weight a_{it} is assigned to the i[th] output vector at time t. W_a is the weight and b_a is the bias. At last, the result is producing a vector v_t, is calculated as the weighted sum of the multiplication of the i[th] yield vector at time and the attention weight a_{it} [25]. *Figure 3* illustrate the placement of the Custom AM(CAM) layers between CNN and BLSTM layers.

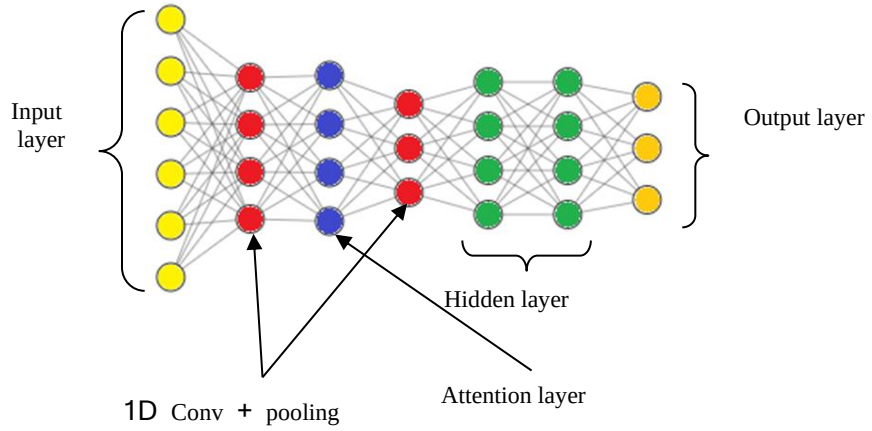


Figure 3 Custom Attention-based CNN

2.3.1PCA

PCA aims to reduce the data to just a few representative dimensions for visualization and analysis. This is accomplished by deriving a new set of variables which represent linear combinations of the original variables. The selection of these variables aims to preserve significant information within a small number of new variables. This succinctly summarizes the samples or observations [31]. The principal decomposition of X, an N-by-M matrix, illustrates the original dataset as in Equation 9 [32].

$$\mathrm{PCA} = \mathrm{X\,W} \quad (9)$$

where the columns of W, a M-by-M matrix of weights, are the X^TX eigenvectors. Given that the N rows represent the patients and each of the M columns gives a particular kind of feature.

2.3.2Siamese network (SN)

The SN comprises two identical networks that process distinct inputs and are integrated using a cost function. The primary objective is to enable the SN to assess the similarity between two inputs. Upon training, the networks can identify similarities among samples originating from various sources. So, one task is to find failures, but we can only see an example of a possible machine failure before we can guess what will happen with a test sample [33, 34]. The SN will compare between two input sample to identify similarity incorporating Euclidian distance (E_d) and contrastive loss (L_c).

Assume a training set X, where (x, x') represents a single sample, such that x, x' $\in X$. The (L_c) function usually conducts the SN learning process described in Equation 10 [35].

$$L_c = (1-y).{E_d}^2 + y.\max\left(0, m_g - E_d\right)^2 \quad (10)$$

The maximum value between m_g (the difference between the two data pairs (x, x')) and the E_d is multiplied by the real label y (i.e., 0 or 1) for dissimilarity or similarity respectively.

Various researchers employed ML or DL techniques utilizing genetic information to investigate breast cancer prediction. The studies encompass the subsequent research papers:

In their 2019 study, Tseng et al. [36] employed ML technology to assess the metastasis of breast cancer. The study identified that the most effective method for predicting breast cancer metastases three months in advance is a ML model utilizing random forests, which attained an accuracy of 91.7% with a limited sample size of 86 patients. Magna et al. (2020) [37] utilized patients' medical histories to investigate the application of DL, ML, and word insertion in breast cancer classification. An attempt was made to present a support system that assists in the decision-making of physicians.

Tong et al. (2020) [38] improved breast cancer survival estimates by integrating multi-omics data, including copy number variations, deoxyribonucleic acid (DNA) methylation, gene expression, and micro ribonucleic acid (miRNA) expression. The proposed model by the authors, LMIR, utilizes divergence-

based shared regularization techniques to improve performance by acquiring shared data from multiple modalities. Arya and Saha (2021) [39] integrated multiple data modalities to develop DL models for predicting breast cancer survival. It emphasizes the integration of heterogeneous data to enhance prediction accuracy. The approach known as SiGaAtCNNs may improve survival prognosis for breast cancer patients. In 2021, Vale-silva and Rohr [40] presented a multimodal DL approach for predicting cancer survival. It generates feature representations for various high-dimensional omics data modalities through the use of specialized submodels. The data fusion layer collects multimodal representations, while a prediction submodel generates conditional survival probabilities for various decades of follow-up time intervals.

In 2022, Chen et al. [41] integrated molecular profile data from 14 distinct cancer types with pathology whole-slide images through multimodal DL. Utilizing diverse modalities to weakly supervise outcome predictions, he offers comprehensive analyses related to patient prognosis for each of the 14 cancer types in an interactive, publicly accessible database. Hao, et al. in 2023 [5] deals with the task of forecasting cancer survival rate using DL with fused multi-omics data. The authors combined multi genetic expression data, which aims at capturing a holistic picture of cancer patients' survival probabilities. They employed DFRMGD model based autoencoder technique for feature representations. In 2023, Kayikci and Khoshgoftaar [3] developed an attention-based multimodal DL model for breast cancer prediction, referred to as GAMDL. A model is developed by integrating clinical data and gene expression data to evaluate mammography images and patient information. The authors utilized mRMR for feature representation.

Wu et al. in 2023 [42] created both modality-specific and modality-invariant representations, which are then fused to provide a comprehensive and holistic view of the data for more accurate survival prediction. The CAMR method predicts cancer survival by integrating data from multiple modalities and aims to overcome the issues related to the heterogeneous nature of multimodal data. AlSalman et al. in 2024 [43] used deep CNN (DCNN) in the identification of breast cancer, its implementation is frequently hindered by the limitations of training on small, localized datasets and privacy concerns related to sharing patient data. By using DCNNs across numerous, heterogeneous datasets without sacrificing data privacy, it presents an efficient federated learning strategy specifically designed for breast cancer diagnosis.

Wang et al. (2024) [44] enhanced the prediction of disease-free survival (DFS) in breast cancer patients by integrating multiple medical data modalities into a multi-modal DL model. Despite notable enhancements in predictive performance and clinical value, challenges related to biases, generalizability, and interpretability persist. Yang et al. (2024) [45] present a novel training method termed multimodal object-level contrast learning. This method seeks to develop a multimodal survival risk predictor by creating contrastive learning pairs that leverage the survival risk correlations among samples in the training set. The architecture employs a self-normalizing NN for genomic data and an attention based NN for medical images. Rintala et al. in 2024 [46] introduce a DL system designed to combine patient and omics data in order to investigate potential applications for precision medicine. In order to better preserve the patient heterogeneity that is relevant to survival, it incorporates patient survival prediction.

In 2024, Gomaa et al. [47] propose and evaluate a non-linear, non-proportional survival prediction model utilizing transformers. The model employs self-supervised learning techniques to efficiently encode the magnetic resonance imaging (MRI) input for integration through cross-attention. Zhang et al. (2024) [48] conducted an investigation of relevant literature through forward and backward citation analysis. The researchers employed their datasets and methodologies to conduct a comprehensive analysis and review of the research. Previous research indicates that various authors have been trying to enhance cancer prediction and survival rate forecasting through the use of multimodal approaches involving multiple genetic databases. Despite its apparent advantages, heterogeneity introduced higher levels of uncertainty into the process. Authors utilizing prediction based on slides through image processing techniques face limitations due to the availability of dataset-based slide samples and the significant processing time required. Consequently, incorporating slide-clinical based multimodal prediction will increase both complexity and processing time. The initial step in addressing this challenge involves utilizing clinical-based datasets, such as gene expression and blood tests. We integrated various clinical datasets and standardized their dimensionality through suitable techniques,

ensuring the preservation of critical information. A shared feature set will be developed by integrating these datasets with relevant designated features. These features are trained with shared characteristics through a hybrid DL method that utilizes a learning algorithm based on BLSTM, leveraging its ability to learn from both current and previous samples over long- and short-time frames. *Table 1* provides a summary of survival analysis techniques presented in this section

Table 1 Summery of compared models

Model	Dataset based	Learning strategy	Dimension engineering	Employed strategies
SBC-ML[36]	Slide	Random Forest	--	--
ML&WE [37]	Slide	ML	TF-IDF	--
FLA[43]	Slide	DCNN	CLAHE	Fusing features
MOCL[45]	Slide	AMIL-SNN	-	Fusing features
MMF[41]	Slide-Clinical	AMIL-SNN	--	Fusing features
DeepClinMed-PGM[44]	Slide-Clinical	Random Forest	CLAM	Fusing features
MDL-MSG[47]	Slide-Clinical	3D-CNN	--	Fusing features
GAMDL[3]	Clinical	Bi-AM	mRMR	Fusing features with consensus information
DFRMGD[5]	Clinical	Fully connected neural network (FCNN)	Autoencoder	Fusing features
LMIR[38]	Clinical	FCNN	PCA	Fusing features
SiGaAtCNNs[39]	Clinical	CNN-AM	mRMR	Fusing features
MultiSurv[40]	Clinical	NN	Autoencoder	Fusing features
CAMR[42]	Clinical	FCNN	Autoencoder	Fusing information only
MODAE[46]	Clinical	SVM	Autoencoder	--
Proposed multimodal deep learning -multigenetic features (MDL-MG)	Clinical	CNN-BLSTM-AM	PCA	Fusing features with consensus features

3.Methods

Figure 4 illustrates the proposed MDL-MG system for predicting cancer survival rates using multimodal DL and various genetic characteristics. The MDL architecture integrates a hybrid CNN, BLSTM, and CAM following the preprocessing of multigenic datasets to extract essential features. An orthogonal constraint is applied to aggregate similar samples, while the Euclidean distance facilitates the grouping of related data. Contrastive loss is utilized to distinguish samples from different classes and cluster samples from the same class, thereby enhancing data separability. The subsequent sections provide a detailed account of the proposed system.

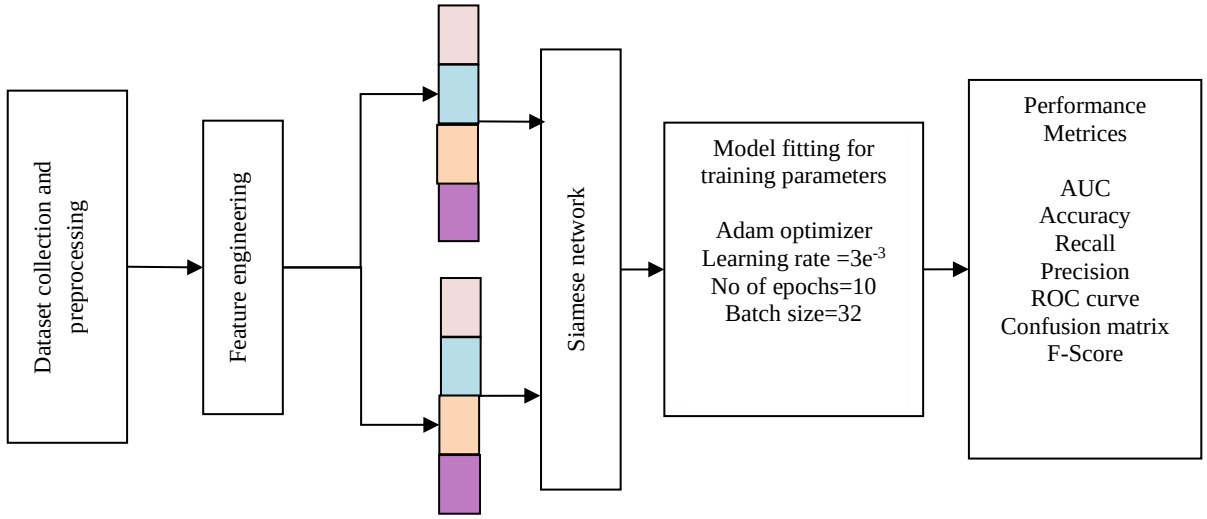


Figure 4 Proposed MDL-MG main steps

3.1Data collection and preprocessing

Three types of genetic data are utilized in this study: miRNA expression, DNA methylation, and mRNA. The breast cancer datasets employed were obtained and preprocessed based on the work referenced in [49].

Following the download of data from the cancer genome atlas (TCGA) website [50], three preprocessing steps are performed:

1. Sample Selection: Patient samples with more than 20% missing data for a specific data type are removed. For missing data imputation, genes with more than 20% missing values are filtered out, while those with fewer missing values are completed using k-nearest interpolation.
2. Normalization: The z-score transformation is applied to normalize the data.

The dataset contains records from 109,767 patient samples, categorized into 105 main groups based on feature similarity. Patients with survival durations exceeding 300 days are classified as having medium or high survival rates. A low survival rate is assigned to patients with survival durations below 300 days, particularly when deaths are recorded. Survival rates are labeled as follows: 1 for high survival and 0 for low survival. Each dataset varies in the number of features, as detailed in *Table 2*.

Table 2 genetic dataset dimensionality

Dataset	Feature size
mRNA	17814
miRNA	354
DNA	23094

3.2Feature engineering

The proposed feature mapping method ensures that the number of features in the analyzed datasets is aligned. All datasets must be of equal size to compute the consensus feature. Consequently, dimensionality reduction is applied, as increasing the dimensionality of a dataset with limited features—such as miRNA with 354 dimensions—would be impractical. Afterward, distinct features from the datasets are combined with the consensus features before being utilized. The proposed hybrid CNN-BLSTM-CAM architecture accepts input derived from these fused features. *Figure 5* illustrates the proposed steps for feature engineering.

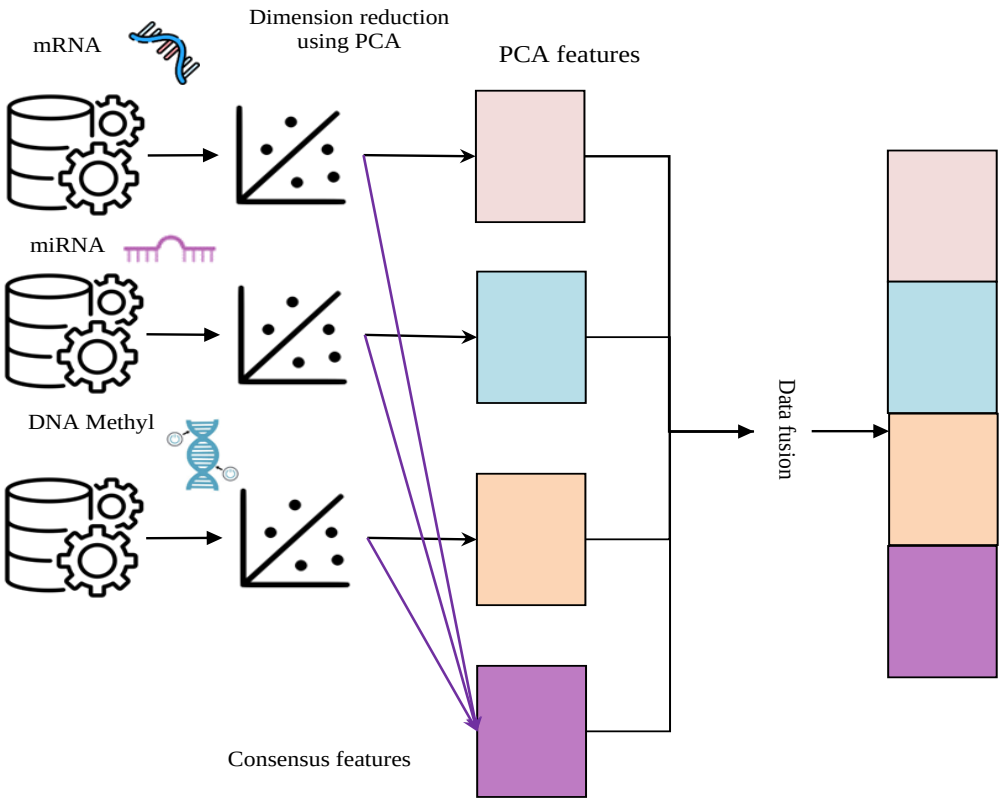


Figure 5 Proposed feature engineering

3.2.1Dimensionality reduction

The three utilized datasets contain varying numbers of features; therefore, PCA [51, 52] was employed to compress the information from the large datasets into a collection of summary indices. This approach facilitates easier presentation and analysis while minimizing the loss of critical information. The objective is to extract essential information from the three datasets. This aids the next step, which involves utilizing common features from all three datasets for fair decision-making.

3.2.2Consensus feature

The consensus feature representation involved calculating the average of the PCA features across the three datasets. In comparison to the PCA features derived from merged datasets, these shared features provide a greater depth of information.

3.2.3Data fusion

The consensus features and significant PCA features from the three genetic datasets have been combined to form a composite collection of features. When used together, these features provide a full picture of the model training process.

3.3Proposed hybrid CNN-BLSTM-CAM model architecture

The hybrid CNN-BLSTM-CAM model is designed with two configurations.

1. 5CNN-2BLSTM-2CAM Model (Configuration A): This architecture consists of two main phases. In the first phase, batch normalization layers are positioned after five convolutional layers that utilize the ReLU activation function for regularization. In the second phase, two BLSTM layers analyze spatial sequences, while two CAM layers enhance feature recognition. Finally, global max pooling is applied, and the output is flattened before being passed through two FCNN layers. The Softmax activation function is then used to generate the predicted classes.
2. 2CNN-BLSTM-CAM Model (Configuration B): This configuration includes two CNN layers with ReLU activation, followed by a single BLSTM layer integrated with a CAM layer. The remaining layers and processing steps are identical to those in the first configuration.

Figure 6 illustrates the layer architecture for both configurations of the proposed model

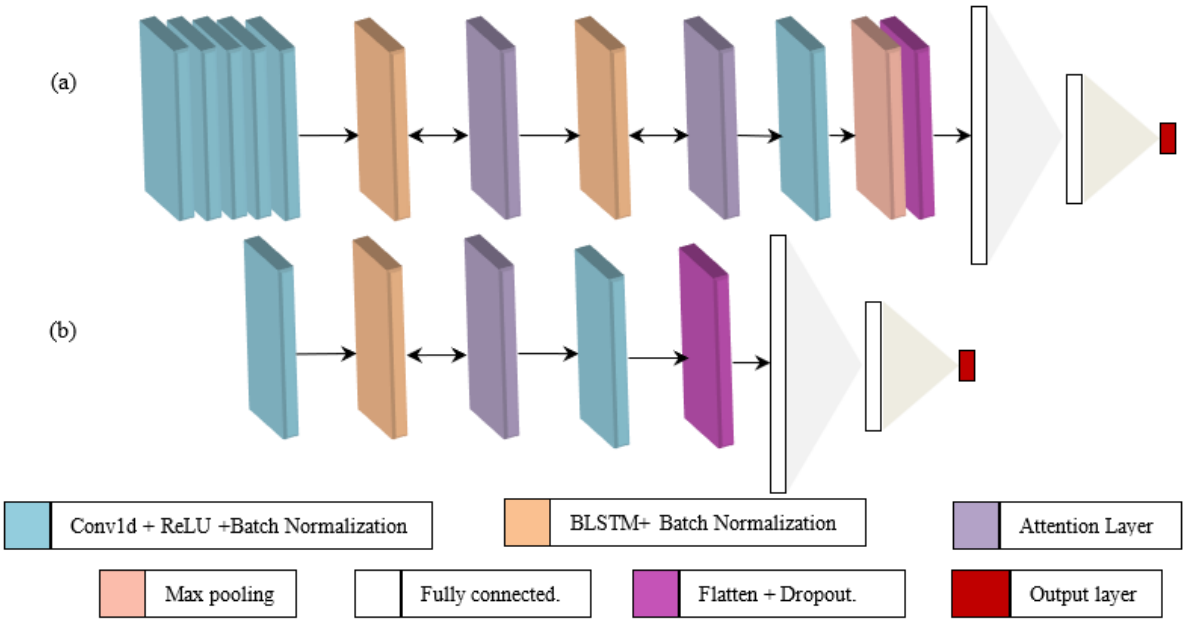


Figure 6 Proposed Architectures (a) Configuration of 5CNN-2BLSTM-2CAM Model (b) Configuration of 2CNN-BLSTM-CAM model

The parameters used and the order of the layers are shown in *Table 3*. The setup for the two configurations of the hybrid CNN-BLSTM-CAM system began with one-dimensional convolutional layers, each with a distinctive filter number. 'He_Uniform' was the method that initiated the first layer. The activation function used by each convolution layer was ReLU. Bach Normalization layers are added after each convolutional layer and the BLSTM layers to stop the training dataset from being too well fitted. The absence of batch normalization considerably affects the learning

process by impacting the first set of training samples. Before the last dense layer, there is a dropout layer with a 50% rate, this makes the model more stable and less reliant on any one node. As a result, characteristics that are unique to later samples or batches may become undetectable. The second part of the CNN-BLSTM-CAM architecture is the BLSTM with 128 filter in by mean of 64 filter from each direction.

Table 3 Proposed CNN-BLSTM-CAM layers and parameters for two configurations (padding='same', stride='1')

Layer	Parameters (Number of filters, filter size, initializer)	Activation Funtion	Layer	Parameters (Number of filters, filter size, initializer)	Activation Function
	Configuration A			**Configuration B**	
Conv1	(256, 2, He_Uniform)	ReLU	Conv1	(256, 2, He_Uniform)	ReLU
BN1			BN1		ReLU
Conv2	(128,2, None)	ReLU	BLSTM	128	ReLU
BN2			CAM	(128,2, Glorot_uniform)	ReLU
Conv3	(64, 2, None)	ReLU	Conv2	(64, 2, None)	SoftMax
BN3			G-Maxpooling		
Conv4	(32, 2, None)	ReLU	Flatten		
BN4			Hidden layer1	128	
Conv5	(64, 2, None)	ReLU	BN2		
BN5			Hidden layer2	64	
BLSTM	128		Dropout	50%	
BN6			Output layer	2	
CAM	(128,2, Glorot_uniform)				
BLSTM	128				
BN6					
CAM	(128,2, Glorot_uniform)				
Conv6	(64, 2, None)	ReLU			
BN8					
G-Maxpooling					
Flatten					
Hidden layer1	128	ReLU			
BN9					
Hidden layer2	64	ReLU			
Dropout	50%				
Output layer	2	SoftMax			

In the third part of the architecture following each BLSTM layer is the CAM. This research requires a large NN to analyze methylation, mRNA, and miRNA expressions. Utilizing the attention layer enables the model to progressively focus on the most relevant portions of large inputs, enhancing its ability to extract critical features and generate more accurate predictions.

Based on the previously discussed points, implementing CAM helps the model highlight the most significant data points, indirectly maximizing relevance and reducing redundancy in the learned features. Additionally, the model employs 'Glorot_uniform' for weight initialization in all attention layers and utilizes the Softmax activation function in its final layer.

The fourth part of the CNN-BLSTM-CAM architecture is applying global max pooling and flattening the results. Global pooling layers can decrease the size of the network and may be less prone to overfitting due to their lack of learnable parameters. Moreover, these networks may withstand input data spatial translations with higher robustness. The entire resulting matrix from the pooled feature maps is then flattened to produce a single long continuous linear vector. To classify the output, the flattened matrix is fed into the FCNN layer.

After flattening, two NN hidden layers are employed with different hidden nodes activated by ReLU function. The number of nodes in each layer are 128, and 64 respectively. Each is followed by batch normalization layer. The final layer, the output layer,

uses the Softmax activation function to categorize the output into two classes. These two classes represent the final labels of either long or short survival time. The proposed CNN-BLSTM-CAM network is assembled into SN assigning the concatenated PCA features of the three datasets. The SN employs the use of contrastive loss function to calculate the Euclidean distance between its two outputs. *Figure 7* illustrates the SN architecture used in the proposed model.

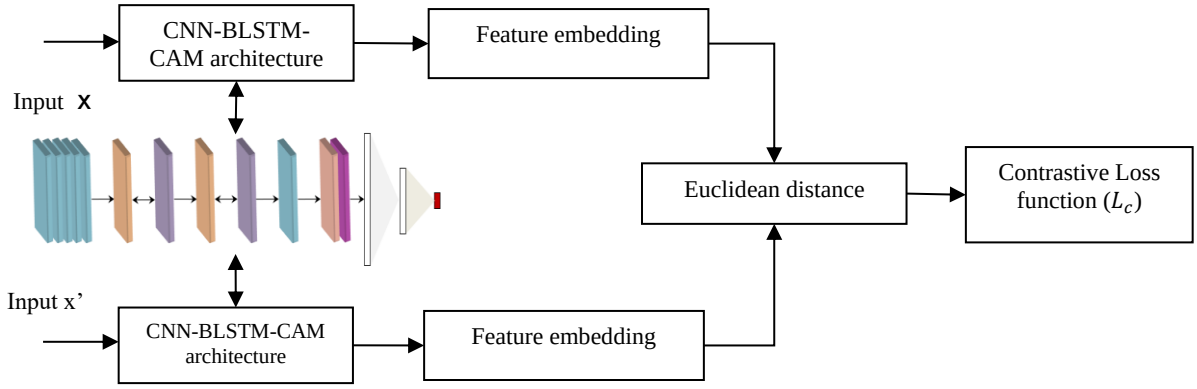


Figure 7 Proposed SN-based CNN-BLSTM-CAM architecture

4.Results

4.1System specifications

The proposed system was developed and evaluated on a 12th-generation local computer equipped with a Core i7 central processing unit (CPU), a 3.8GB Iris Xe graphics processing unit (GPU), and 8GB of shared random-access memory (RAM). The system was implemented in a Python 3.11.4 environment using the TensorFlow library for ML, with Keras serving as the high-level application programming interface (API).

4.2Parameter selection

The dataset was preprocessed and then randomly split using the train-test-validation method. Eighty percent of the data was allocated for training, 10% for validation, and 10% for testing. With 84 PCA-derived features from each dataset, plus the consensus features, a total of 336 unique characteristics was used for each patient. The distribution of the dataset split is illustrated in *Figure 8*.

An adaptive learning rate was implemented, starting at 1e-3, to adjust the rate for the Adam optimizer. Reduced learning rate scheduling was applied, decreasing the learning rate after three epochs of no improvement in validation loss. The new learning rate was set to 1% of the previous rate, but not lower than 1e-7. An adaptive learning rate helps the model overcome local minima. Four distinct batch sizes (16, 32, 50, and 64) were evaluated, along with other parameters, to determine the configuration that yields the highest accuracy. Utilizing the early stopping mechanism, training automatically terminates after seven consecutive iterations without an improvement in validation accuracy. *Table 4* provides a detailed tabular representation of the final parameter selections.

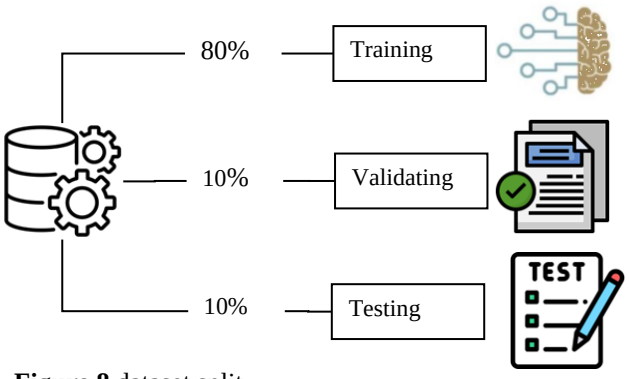


Figure 8 dataset split

Table 4 Hyper-parameters selected for model learning

Parameters	Value
Batch size	16
Learning rate	1e-3 to 1e-7 with step size 1e-2
Epochs	Max 100 with Early Stopping

4.3Experimental results and evaluation metrics

The performance of the proposed MDL-MG model was evaluated using various measurement metrics. Among these, recall measures the proportion of actual relevant values correctly identified, while precision assesses the relevance of the predicted

values. Accuracy is also considered as a key evaluation metric [53]. Another crucial metric is system complexity, which is quantified using floating-point operations per second (FLOP/s). The breakdown of FLOP calculations for each layer is detailed in Equations 11 to 15 [54, 55].

$$FLOP_{CNN} = 2 \times H_{out} \times W_{out} \times C_{out} \times (C_{in} \times k_h \times k_w) \quad (11)$$

$$FLOPs_{FCNN} = 2 \times (Input\ Units \times Output\ Units) \quad (12)$$

$$FLOPs_{BLSTM} = 2 \times 8h(d + h) \times T \quad (13)$$

$$FLOPs_{CAM} = 3 \times 2 \times (batch \times seq_{len} \times d^2) + 4 \times (Batch \times {seq_{len}}^2 \times d) + 5 \times Batch \times {seq_{len}}^2 \quad (14)$$

$$Total\ FLOPs = FLOP_{CNN} + FLOPs_{BLSTM} + FLOPs_{CAM} + FLOPs_{FCNN} \quad (15)$$

Where

H_{out} and W_{out} denote the output height and width.

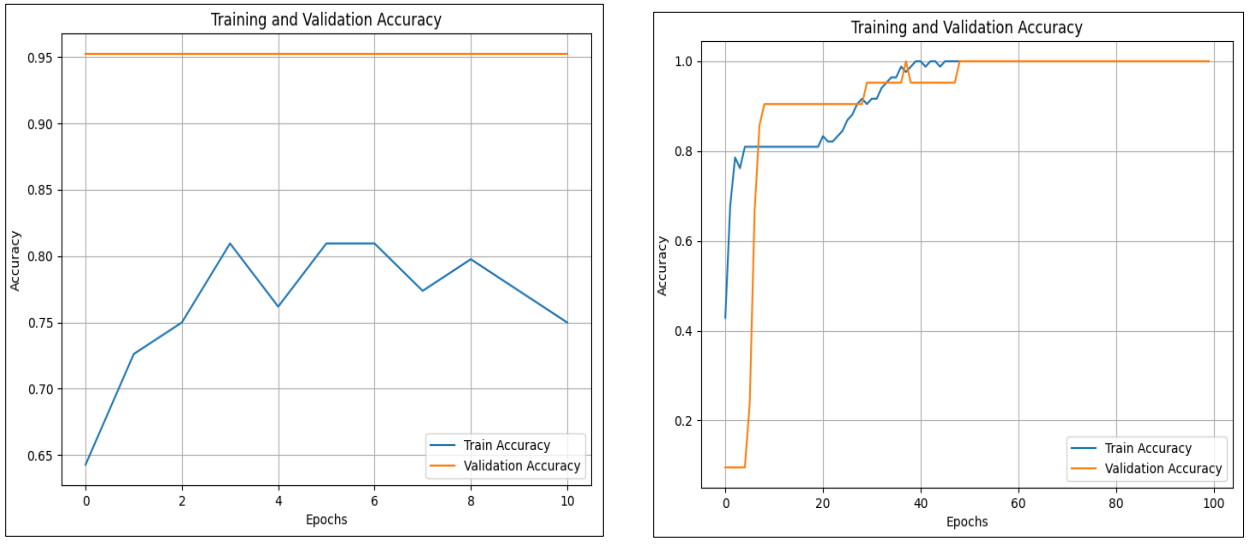
C_{in} and C_{out} represent the channel input and output, respectively.

k_h and k_w are kernel height and width, respectively.

d represents the input size, h denotes the hidden size, and T indicates the sequence length.

Figure 9 shows the training and validating accuracy of proposed model in each epoch. while comparison was performed between the proposed MDL-MG and five other models utilizing clinical datasets from the literature (*Table 1*), specifically: GAMDL [3], DFRMGD[5], LMIR [38], SiGaAtCNNs [39], and CAMR [42]. *Figure 10* presents a comparison of these methods based on the performance metrics of precision, recall, and accuracy. The proposed MDL-MG model achieved 95% and 100% accuracy for Configuration A and Configuration B, respectively, demonstrating improvements of 4.2% and 9% over other models. Configuration B further enhanced precision and recall by 16% and 21% compared to other models and by 5% relative to Configuration A.

A test's diagnostic success is determined by the receiver operating characteristic (ROC), which uses high values for specificity and sensitivity. It is also employed in the comparison of these values acquired at various cut-off positions. True positive values (sensitivity) are plotted on the ROC curve as a function of false positive values (1-specificity) with respect to various intersection points. The sensitivity/specificity ratio in relation to a specific threshold is shown by each point on the curve. How well a parameter separates the two groups is indicated by the area under the curve (AUC). As the curve rises, the number of accurately detected positive results also rises and -indicate higher separation between classes. Any AUC that is less than 0.5 suggests that the model is behaving erratically and incorrectly. The ideal range for an AUC score in Pecan is generally greater or equal 0.65.

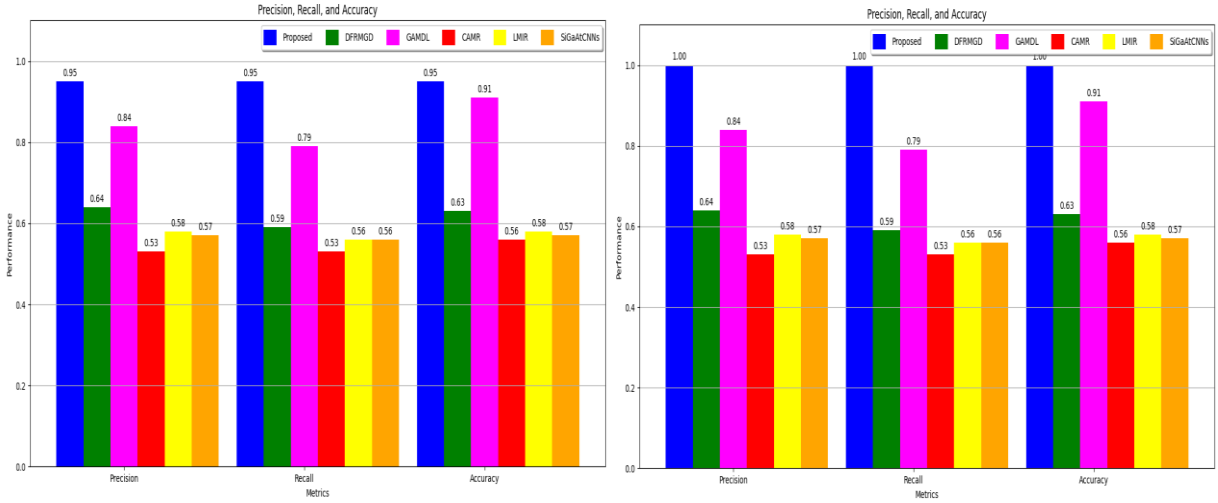


(a) (b)

Figure 9 Training and validation accuracy per epoch (a)Configuration A(b) Configuration B

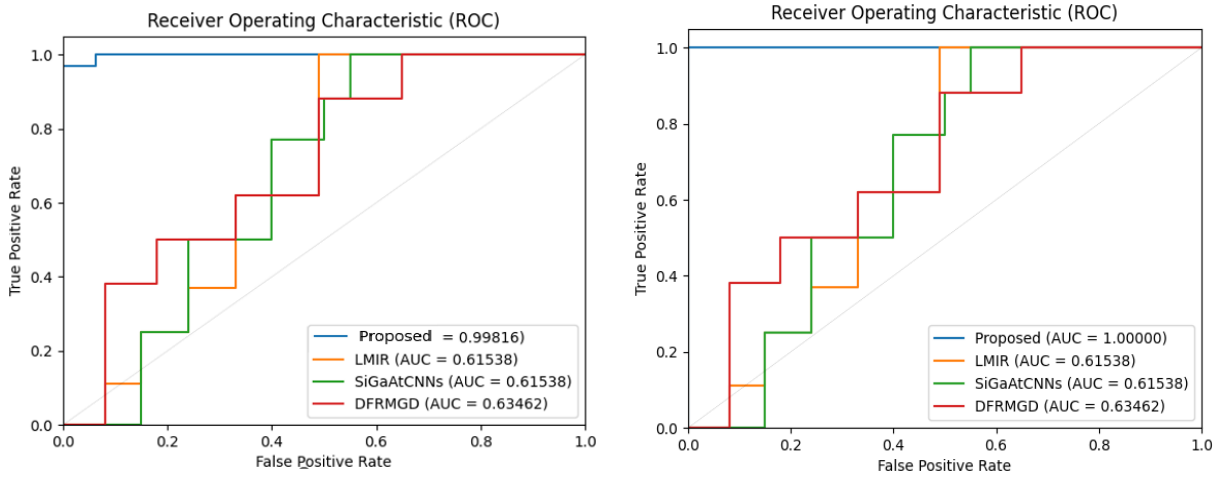
Figure 11 presents the ROC curves of the competing models on the same datasets. The figure also shows the AUC value for each model. Configuration A of the proposed model achieved an AUC of 0.99816 and

configuration B achieved 1.0 with an increased by 36.5%, over other comparable models. The AUC value is an indication for the overall performance of the system. This indicates that the model can distinguish between the two classes, and only one individual with a medium survival rate will be classified as having low survival rates.



(a) (b)

Figure 10 Performance comparison for breast cancers survival prediction of several models (a) Configuration A (b) Configuration B

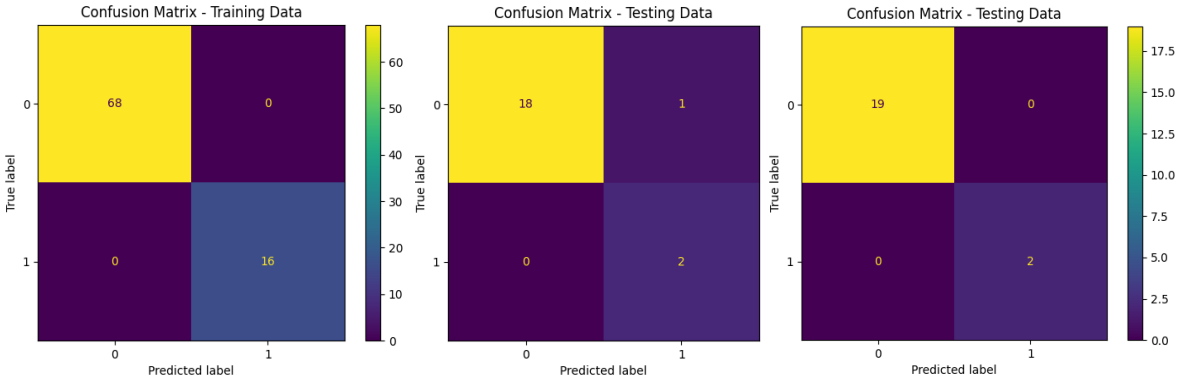


(a) (b)

Figure 11 ROC curves for Breast cancers survival rate prediction of several models (a) Configuration A (b) Configuration B

Figure 12 presents the confusion matrix of the proposed model. A total of 87 patient cases with medium or high survival rates and 18 patients with low survival rates were used for model training and testing, respectively.

During the training phase, all samples were accurately classified; however, Configuration A exhibited one false positive case. Medium survival rates that are close to the threshold may be classified as high survival rates, and this false positive case falls within that category. Since no patients with poor survival rates were misclassified as low-risk cases, this error rate is considered acceptable. Configuration B showed no falsely classified cases.



(a) (b) (c)

Figure 12 ROC curves for breast cancers survival rate prediction of several models (a) training of configurations A and B (b) testing Configuration A (c) testing Configuration B

Model architectures were assessed to illustrate the influence of each component on the attained accuracy. *Table 5* presents a comparison of CNN solely, CNN BLSTM, configuration A, and configuration B of the proposed model for accuracy, precision, recall, AUC, F1-score, FLOPs, training, and testing time. Configurations A and B are compared both with and without the influence of the SN. This comparison also includes the multi-omics bidirectional encoder representations from transformers (MoBERT) [56], a DL model adapted to predict cancer survival rates instead of disease prognoses for subtype categorization, employing the same dataset.

Table 5 comparison of different model scenarios for 100 epochs

Case No	Model	SN	Accuracy	Precision	Recall	AUC	F1-score	FLOP	Train time (sec)	Test time (sec)
1	2CNN-Only	Y	0.7143	0.6553	0.8095	0.7368	0.7243	5,809,740	11.548	0.055
2	5CNN-Only		0.8095	0.8603	0.7143	0.4211	0.7689	16,384,620	18.167	0.058
3	2CNN-BLSTM		0.9048	0.8186	0.9048	0.8737	0.8595	5,916,108	23.682	0.066
4	5CNN-2BLSTM		0.9048	0.9524	0.9048	0.5818	0.9181	17,859,180	61.296	0.082
5	Configuration A		0.9524	0.9571	0.9523	0.98816	0.9451	39,844,406	54.918	0.076
6	Configuration B		1.0	1.0	1.0	1.0	1.0	17,431,584	45.121	0.166
7	Configuration A	N	0.7143	0.7983	0.7143	0.6737	0.7541	39,844,406	51.765	0.073
8	Configuration B		0.8095	0.8553	0.8095	0.6035	0.7243	17,431,584	43.524	0.066
9	MoBERT [57]		1.0	1.0	1.0	1.0	1.0	157,229,188	1.8525	0.011

The optimal model is the one that delivers the maximum accuracy, precision, recall, and AUC, while also exhibiting minimal FLOPs, training time, and testing time. An analysis of the results presented in *Table 5* indicates that the incorporation of a BLSTM layer built on the CNN layer improved the first categories of performance metrics, yet at the cost of increased FLOPs and extended training/testing time. With the superiority of case 3 compared to case 4. In cases 5 and 6, the addition of the CAM layer demonstrates an improvement over prior cases, except in terms of FLOP count, which signifies an increase in system complexity. It is noteworthy that configuration B in instance 6 exhibits fewer FLOPs than case 4, while demonstrating superior performance metrics and reduced training/testing duration. Cases 1 to 6 employed SN, whereas cases 7 to 9 did not. The impact of SN is evident when comparing the results of cases 5 and 7 as well as cases 6 and 8, demonstrating that the incorporation of SN significantly improved accuracy while causing minimal variation in training and testing duration. Specifically, accuracy increased by 23.8% for Configuration A and 19% for Configuration B.

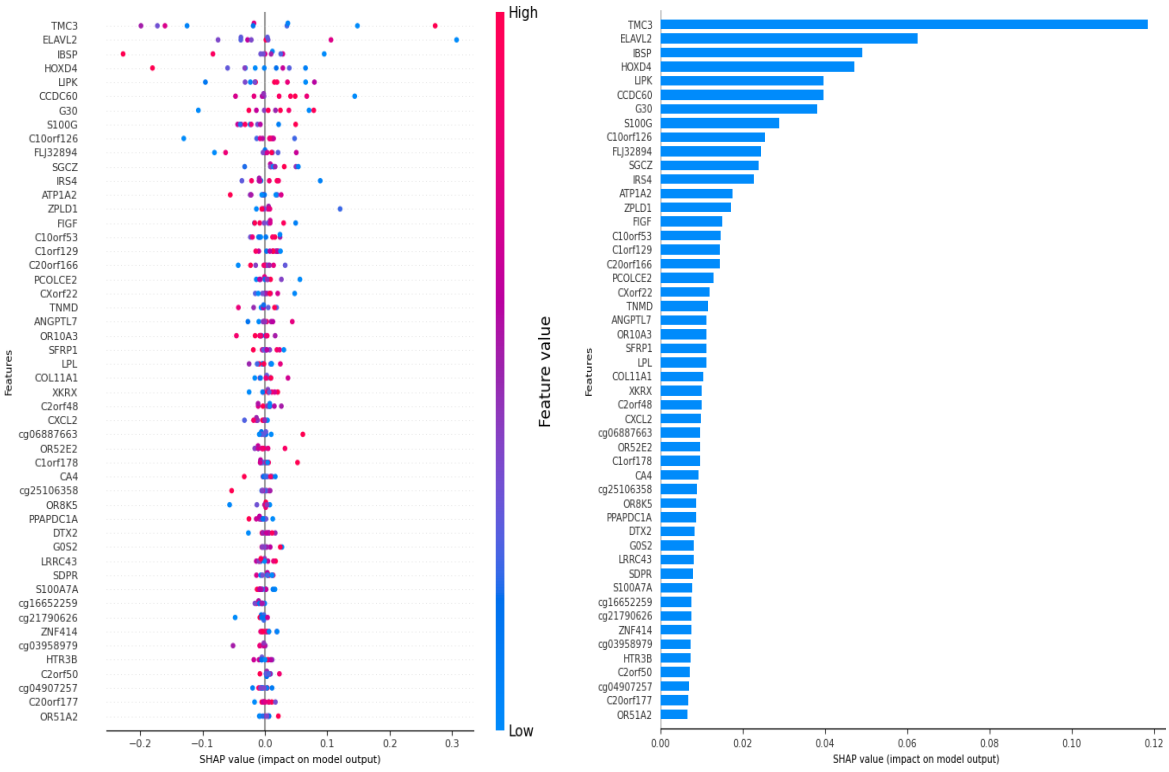
The comparison in *Table 5* indicates that Configuration B achieves the highest values across various metrics, while the MOBERT model exhibits similar performance in the absence of the SN effect. Configuration B also demonstrates the lowest FLOPs compared to MOBERT, although MOBERT achieves the shortest training and testing durations among all models.

Since different dimensionality reduction methods are used in model comparisons, these methods influence training and testing times. *Table 6* presents a comparison of the training and testing durations for PCA, mRMR, and Autoencoder under identical training conditions.

Table 6 training and testing time of different methods

Method	Training time(seconds)	Testing time(seconds)
PCA	22.4630	0.0664
mRMR	33.4773	0.1452
Autoencoder	181.2867	0.9947

SHapley Additive exPlanations (SHAP) [57] values were computed for the PCA features used in training to illustrate their influence on model decisions. A summary plot is presented in *Figure 13.* The *Figure 13(a)* shows 50 of the most utilized features, with their values indicating their influence on decision-making. The higher the value the more significant impact on system prediction. The blue dots signify a relatively low value for the feature, whereas the red dots denote greater values. While in *Figure 13(b*) the SHAP bar representation further show the feature that highly impact the prediction. Detailed explanations of the impact that each multi-omics component has on model predictions are provided in *Figure 14*. The data reveals that methylation then gene expression have a substantially greater influence on survival rate prediction then comes the consensus and finally miRNA, which highlights the usefulness of this method in gathering comprehensive omics patterns. This leads us to suggest that the model is defensive and does not correspond with survival risk, while the data from transcriptomics is a reliable predictor of survival in breast cancer, with epigenetic regulation having a major influence on survival outcomes.



(a) (b)

Figure 13 The highest 50 SHAP value impact on the proposed model (a) SHAP summary(b) SHAP bar representation

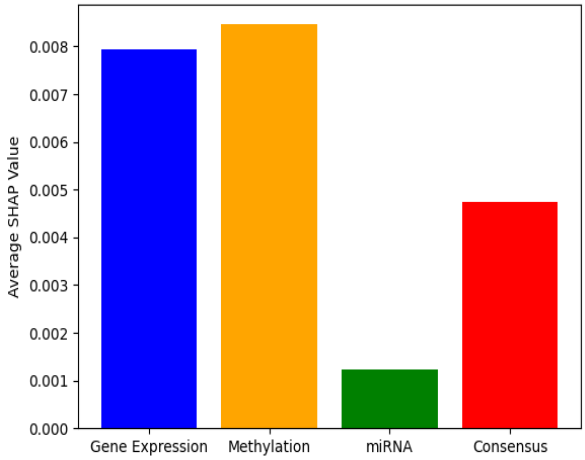


Figure 14 SHAP value by Omics type

5.Discussion

This paper aims to enhance the predictability of cancer survival rates by leveraging prior studies that utilized multi-omics approaches for this purpose. A major challenge in using three databases is the large number of features that must be processed by the DL algorithm. The initial task involved reducing the number of features while preserving the most informative and valuable ones. After thoroughly evaluating different feature reduction techniques, PCA was identified as the most efficient method for dimensionality reduction. PCA effectively retains informative features while minimizing training and testing time, outperforming mRMR and Autoencoders used in previous research.

In this research, various DL techniques were implemented, including CNNs and BLSTM networks. CNNs are widely recognized for their ability to extract meaningful features using multiple filters, while BLSTM facilitates learning from both past and future data over time. The CAM enhances feature focus during the learning process.

Two architectures of the hybrid CNN-BLSTM-CAM model were examined:

1. The first architecture consists of five CNN layers, two BLSTM layers, and two CAM layers.
2. The second architecture includes two CNN layers, one BLSTM layer, and one CAM layer.

In both architectures, SN were implemented, utilizing Euclidean distance and a contrastive loss function to identify similarities between Configuration A and Configuration B. The inclusion of SN improved precision by 23.8% in Configuration A and 19% in Configuration B.

AUC, another key evaluation metric, demonstrated an improvement in class separability, aiding in early identification of patients requiring urgent medical intervention. A confusion matrix was used to detect misclassified instances. While Configuration B exhibited no misclassification, Configuration A presented one false positive case, where a patient labeled with high or medium survival rates was misclassified as low survival rate. Although this misclassification slightly reduced overall system accuracy, it is reasonable, as survival rates sometimes fall close to the threshold dividing different classes. The primary concern is avoiding false negatives, as they pose a greater risk than false positives.

The hybrid learning approach demonstrated its effectiveness by enhancing accuracy, precision, recall, ROC, and AUC. Configuration B consistently outperformed Configuration A and individual model components, while achieving performance comparable to MoBERT. MoBERT had the shortest training/testing duration among all models; however, Configuration B surpassed MoBERT by reducing system complexity by 139.8 GFLOPs. The results indicated that the proposed model relies more heavily on PCA features from mRNA and DNA datasets for prediction than on PCA features from consensus and miRNA datasets.

Limitations

One limitation of this study is the size constraint of the TCGA dataset, particularly for rare breast cancer subtypes. The imbalance in the dataset (e.g., a higher proportion of long-term survivors) may introduce a bias toward the majority class. Additionally, due to variations in dataset sizes, PCA processing is limited by the smallest dataset. While PCA effectively identifies the most relevant features, incorporating additional features could further enhance survival rate predictions.

A complete list of abbreviations is listed in *Appendix I.*

6.Conclusion and future work

This research developed a DL-based algorithm for predicting cancer survival rates. The proposed approach aimed to enhance survival rate prediction compared to previous studies in multiple aspects. The first improvement focused on reducing the number of features in the multi-omics dataset using PCA for dimensionality reduction. The second enhancement involved generating shared features from PCA-

derived features across all datasets and using them as training inputs.

Two distinct configurations of the hybrid CNN-BLSTM-CAM learning algorithm were developed. Both configurations employed a contrastive loss function within a SN architecture to compute the Euclidean distance between the two models. The proposed model was evaluated against other models to demonstrate its efficiency, achieving equivalent or superior performance in accuracy, precision, recall, and AUC compared to state-of-the-art models. Additionally, it effectively managed computational complexity while maintaining high accuracy.

To gain deeper insights into feature importance, the research incorporated the SHAP measure. The findings revealed that features derived from miRNA and DNA expressions had the most significant impact on predictions, followed by the generated consensus features.

For future studies, a scalable network capable of handling large datasets while maintaining high prediction accuracy could be developed.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

Data availability
The dataset utilized in this study is publicly accessible and can be found at http://compbio.cs.toronto.edu/SNF/SNF/Software.html.

Author's contribution statement
Yasmine M. Tabra: Study conception, design, conceptualization, data creation, investigation on challenges, writing – original draft, writing – review and editing. **Furat N. Tawfeeq**: Data collection, conceptualization, writing – original draft, analysis, and interpretation of results.

References
[1] https://www.nomisweb.co.uk/datasets/mortsa. Accessed 22 February 2025.
[2] Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, et al. Breast cancer statistics 2024. CA: a Cancer Journal for Clinicians. 2024; 74(6):477-95.
[3] Kayikci S, Khoshgoftaar TM. Breast cancer prediction using gated attentive multimodal deep learning. Journal of Big Data. 2023; 10(1):1-11.
[4] Hernandez-cruz N, Saha P, Sarker MM, Noble JA. Review of federated learning and machine learning-based methods for medical image analysis. Big Data and Cognitive Computing. 2024; 8(9):1-37.
[5] Hao Y, Jing XY, Sun Q. Cancer survival prediction by learning comprehensive deep feature representation for multiple types of genetic data. BMC Bioinformatics. 2023; 24(1):1-16.
[6] American Cancer Society. Cancer prevention & early detection facts & figures 2023-2024.
[7] Cardoso MJ, Poortmans P, Senkus E, Gentilini OD, Houssami N. Breast cancer highlights from 2023: knowledge to guide practice and future research. The Breast. 2024; 74:1-11.
[8] Yu P. The future prospects of deep learning and neural networks: artificial intelligence's impact on education. Applied and Computational Engineering. 2024; 33:94-101.
[9] Zhang H, Xi Q, Zhang F, Li Q, Jiao Z, Ni X. Application of deep learning in cancer prognosis prediction model. Technology in Cancer Research & Treatment. 2023; 22:1-10.
[10] Jain R, Sarvakar K, Patel C, Mishra S. An exhaustive examination of deep learning algorithms: present patterns and prospects for the future. GRENZE International Journal of Engineering and Technology. 2024; 10(1):105-11.
[11] Masoud TD, Abdulazeez AM. Deep learning classification algorithms applications: a review. Indonesian Journal of Computer Science. 2024; 13(3):4287-11.
[12] Dandekar AR, Sharma A, Mishra JK. Optimized federated learning algorithm for breast cancer detection using the marine predators algorithm. Journal of Electrical Systems. 2024; 20(1):911-20.
[13] Zhu Z, Wang SH, Zhang YD. A survey of convolutional neural network in breast cancer. Computer Modeling in Engineering & Sciences: CMES. 2023; 136(3):2127-72.
[14] Karuppasamy A, Abdesselam A, Hedjam R, Zidoum H, Al-bahri M. Recent CNN-based techniques for breast cancer histology image classification. The Journal of Engineering Research [TJER]. 2022; 19(1):41-53.
[15] Elwahsh H, Tawfeek MA, Abdel-aziz AA, Mahmood MA, Alsabaan M, El-shafeiy E. A new approach for cancer prediction based on deep neural learning. Journal of King Saud University-Computer and Information Sciences. 2023; 35(6):1-12.
[16] Vasanthakumari RK, Nair RV, Krishnappa VG. Improved learning by using a modified activation function of a convolutional neural network in multi-spectral image classification. Machine Learning with Applications. 2023; 14:1-13.
[17] Taye MM. Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. Computation. 2023; 11(3):1-23.
[18] Rakhshan P. Breast cancer detection basedon CNN and federated learningusing embedded devices. School of Innovation Design and Engineering, Mälardalens University, Sweden.2023.

[19] Salim AT, Khammas BM. Performance evaluation of deep learning techniques in the detection of IOT malware. Iraqi Journal of Information and Communication Technology. 2023; 6(3):12-25.
[20] Maurya R, Aggarwal D, Gopalakrishnan T, Pandey NN. Enhancing deep neural network convergence and performance: a hybrid activation function approach by combining ReLU and ELU activation function. In second international conference on informatics 2023 (pp. 1-5). IEEE.
[21] Akgül İ. Activation functions used in artificial neural networks. Book: Academic Studies in Engineering. Geçe Kitaplılğı: Turkey. 2023:41-58.
[22] Salehin I, Kang DK. A review on dropout regularization approaches for deep neural networks within the scholarly domain. Electronics. 2023; 12(14):1-23.
[23] Sinha J, Manollas M. Efficient deep CNN-BiLSTM model for network intrusion detection. In proceedings of the 3rd international conference on artificial intelligence and pattern recognition 2020 (pp. 223-31). ACM.
[24] Yılmaz RE, Serbes G. Breast cancer detection using transformer and BiLSTM based ensemble learning. In 31st signal processing and communications applications conference 2023 (pp. 1-4). IEEE.
[25] Yousaf K, Nawaz T. A deep learning-based approach for inappropriate content detection and classification of YouTube videos. IEEE Access. 2022; 10:16283-98.
[26] Farhan BI, Jasim AD. Improving detection for intrusion using deep LSTM with hybrid feature selection method. Iraqi Journal of Information and Communication Technology. 2023; 6(1):40-50.
[27] Agana MA, Agwu CO, Ukpoho NA. Breast cancer prediction and control using BiLSTM and two-dimensional convolutional neural network. International Journal of Software Innovation (IJSI). 2023; 11(1):1-9.
[28] Manshahia MS, Litvinchev IS, Weber GW, Thomas JJ, Vasant P. Human-assisted Intelligent Computing: Modeling, Simulations and Applications. IOP Publishing; 2023.
[29] De SCA, Colombini EL. Attention, please! a survey of neural attention models in deep learning. Artificial Intelligence Review. 2022; 55(8):6037-124.
[30] Xie S, Zhang Y, Lv D, Chen X, Lu J, Liu J. A new improved maximal relevance and minimal redundancy method based on feature subset. The Journal of Supercomputing. 2023; 79(3):3157-80.
[31] Lu M, Yin J, Zhu Q, Lin G, Mou M, Liu F, et al. Artificial intelligence in pharmaceutical sciences. Engineering. 2023; 27:37-69.
[32] Syms C. Principal components analysis. Elsevier. 2008: 2940-9.
[33] Rahbar K, Taheri F. Enhancing image retrieval through entropy-based deep metric learning. Multimedia Tools and Applications. 2024:1-27.
[34] Caliskan A, O'brien C, Panduru K, Walsh J, Riordan D. An efficient siamese network and transfer learning-based predictive maintenance system for more sustainable manufacturing. Sustainability. 2023; 15(12):1-23.
[35] De RGH, Papa JP. Learning to weight similarity measures with siamese networks: a case study on optimum-path forest. Optimum-Path Forest. 2022:155-73.
[36] Tseng YJ, Huang CE, Wen CN, Lai PY, Wu MH, Sun YC, et al. Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. International Journal of Medical Informatics. 2019; 128:79-86.
[37] Magna AA, Allende-cid H, Taramasco C, Becerra C, Figueroa RL. Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis. IEEE Access. 2020; 8:106198-213.
[38] Tong L, Wu H, Wang MD. Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer. Methods. 2021; 189:74-85.
[39] Arya N, Saha S. Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowledge-Based Systems. 2021; 221:106965.
[40] Vale-silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Scientific Reports. 2021; 11(1):1-12.
[41] Chen RJ, Lu MY, Williamson DF, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell. 2022; 40(8):865-78.
[42] Wu X, Shi Y, Wang M, Li A. CAMR: cross-aligned multimodal representation learning for cancer survival prediction. Bioinformatics. 2023; 39(1):1-8.
[43] Alsalman H, Al-rakhami MS, Alfakih T, Hassan MM. Federated learning approach for breast cancer detection based on DCNN. IEEE Access. 2024; 12:40114-38
[44] Wang Z, Lin R, Li Y, Zeng J, Chen Y, Ouyang W, et al. Deep learning-based multi-modal data integration enhancing breast cancer disease-free survival prediction. Precision Clinical Medicine. 2024; 7(2):1-20.
[45] Yang Z, Liu H, Wang X. A multimodal object-level contrast learning method for cancer survival risk prediction. In proceedings of the 1st international workshop on multimedia computing for health and medicine. 2024(pp.64-69).
[46] Rintala TJ, Napolitano F, Fortino V. Multi-task deep latent spaces for cancer survival and drug sensitivity prediction. Bioinformatics. 2024; 40(Supplement_2):182-9.
[47] Gomaa A, Huang Y, Hagag A, Schmitter C, Höfler D, Weissmann T, et al. Comprehensive multimodal deep learning survival prediction enabled by a transformer architecture: a multicenter study in glioblastoma. Neuro-Oncology Advances. 2024; 6(1):1-14.
[48] Zhang G, Ma C, Yan C, Luo H, Wang J, Liang W, et al. Multimodal deep learning for cancer survival prediction: a review. Current Bioinformatics. 2025; 20(4):299-22.

[49] Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nature Methods. 2014; 11(3):333-7.
[50] https://www.cancer.gov/tcga. Accessed 22 February 2025.
[51] Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010; 2(4):433-59.
[52] Ghojogh B, Crowley M, Karray F, Ghodsi A. Principal component analysis. In elements of dimensionality reduction and manifold learning 2023 (pp. 123-54). Cham: Springer International Publishing.
[53] Istighosah M, Sunyoto A, Hidayat T. Breast cancer detection in histopathology images using ResNet101 architecture. Sinkron: Jurnal Dan Penelitian Teknik Informatika. 2023; 7(4):2138-49.
[54] Paoletti ME, Haut JM, Tao X, Plaza J, Plaza A. FLOP-reduction through memory allocations within CNN for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2020; 59(7):5938-52.
[55] Patel C, Bhatt D, Sharma U, Patel R, Pandya S, Modi K et al. DBGC: dimension-based generic convolution block for object recognition. Sensors. 2022; 22(5):1-25.
[56] Wagner J. Generalised model-independent characterisation of strong gravitational lenses. Astronomy & Astrophysics. 2022; 663:1-13.
[57] Hamilton RI, Papadopoulos PN. Using SHAP values and machine learning to understand trends in the transient stability limit. IEEE Transactions on Power Systems. 2023; 39(1):1384-97.

Yasmine M. Tabra was born in 1984, baghdad, Iraq. She is an Information and Communication Engineer and an expert who works in multiple domains, including DSP, Multimedia, Beamforming, DoA, ML, and 5G Applications. Her educational background is Ph.D. in ICE. She is working as a lecturer in Al Nahrain University/College of Information Engineering for over 15 years. Having over than 10 research papers in different fields.
Email: yasminetabra@nahrainuniv.edu.iq

Furat N. Tawfeeq was born in 1985, Baghdad, Iraq. He is an information Engineer and an expert who works in multiple domains, including ML, Computer programming, image processing, and AI. His educational background is M.Sc. in Information Engineering. He is working as a responsible for website unit\ website division\ University of Baghdad. Having over 20 research papers in different fields.
Email: furat@bccru.uobaghdad.edu.iq

Appendix I

S. No.	Abbreviation	Description
1	AI	Artificial Intelligence
2	AM	Attention Mechanism
3	API	Application Programming Interface
4	AUC	Area Under the Curve
5	BLSTM	Bidirectional Long Short-Term Memory
6	CAM	Custom Attention Mechanism
7	CNN	Convolutional Neural Networks
8	CPU	Central Processing Unit
9	DCNN	Deep CNN
10	DFS	Disease-Free Survival
11	DL	Deep Learning
12	DNA	Deoxyribonucleic Acid
13	FCNN	Fully Connected Neural Network
14	FLOP	Floating-Point Operations
15	GPU	Graphics Processing Unit
16	HDI	Human Development Index
17	MDL-MG	Multimodal Deep Learning - Multigenetic Features
18	miRNA	Micro Ribonucleic Acid
19	ML	Machine Learning
20	MoBERT	Multi-Omics Bidirectional Encoder Representations from Transformers
21	MRI	Magnetic Resonance Imaging
22	mRMR	min-Redundancy and Max-Relevance
23	NN	Neural Network
24	PCA	Principal Component Analysis
25	RAM	Random-Access Memory
26	ReLU	Rectified Linear Unit
27	RNN	Recurrent Neural Network
28	ROC	Receiver Operating Characteristic
29	SHAP	SHapley Additive exPlanations
30	SN	Siamese Network
31	TCGA	The Cancer Genome Atlas