IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 July 2023, accepted 30 July 2023, date of publication 10 August 2023, date of current version 16 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3304242

RESEARCH ARTICLE

Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis

EHTISHAM KHAN JADOON[1], FIAZ GUL KHAN[1], SAJID SHAH[2], AHMAD KHAN[1], AND MUHAMMED ELAFFENDI[2]

[1]Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22010, Pakistan
[2]EIAS Data Science Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding author: Fiaz Gul Khan (fiazkhan@cuiatd.edu.pk)

This work was supported by the Emerging Intelligent Autonomous Systems (EIAS) Data Science Laboratory, Prince Sultan University, Saudi Arabia.

ABSTRACT Ensemble models based on deep learning have made significant contributions to the medical field, particularly in the area of disease prediction. Breast cancer is a highly aggressive disease with a high mortality rate. Timely and effective prediction of breast cancer can reduce the risk of it progressing to later stages and the need for unnecessary medications. While previous studies have focused on predicting breast cancer using single-modal datasets, multi-modal datasets that include gene expression (gene exp), clinical, and copy number variation (CNV) data have become available in recent years for predictive model development. However, despite multiple studies using multi-modal data for disease prediction, models designed for breast cancer are typically homogeneous neural networks. This article proposes a heterogeneous deep learning-based ensemble model for effective breast cancer prediction using multi-modal data. The model consists of three phases: feature extraction, stacked feature set creation, and using extracted features as input for a stacked-based model using a random forest algorithm for effective prediction. For feature extraction, convolutional neural networks (CNNs) are used for clinical and gene expression data, and deep neural networks (DNNs) are used for CNV data. The extracted features from CNNs and DNNs are stacked to create a comprehensive feature set. The simulation results demonstrate the superiority of the proposed framework in terms of accuracy compared to uni-modal and homogeneous model-multi-modal frameworks.

INDEX TERMS Breast cancer, deep learning, feature extraction, machine learning, prognosis prediction.

I. INTRODUCTION

Breast cancer is a highly lethal disease that predominantly affects women, with a significant mortality rate [1]. The condition is characterized by abnormal cell growth in the breast tissue, resulting in the formation of primary tumors, which can be either benign or malignant. Benign tumors are confined to the affected area and do not spread, while malignant tumors can invade and affect other body parts. Breast cancer is classified into two types: invasive and non-invasive [2]. Invasive breast cancer can spread to neighboring tissues, whereas non-invasive breast cancer is confined to the milk ducts or lobules of the breast [3], as illustrated in Figure 1. Although breast cancer is more commonly diagnosed in females, it can also affect males, but at a lower frequency [4].

According to the report of Cancer.Net (ASCO website) an estimated 2,261,419 new cases were diagnosed with breast cancer around the globe in 2020. According to the data, in the United States, approximately 287,850 women will receive a diagnosis of invasive breast cancer, while about 51,400 women will be impacted by non-invasive breast cancer. Approximately, 43,580 death cases were forecast for the year 2022 due to breast cancer in the United States. The

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

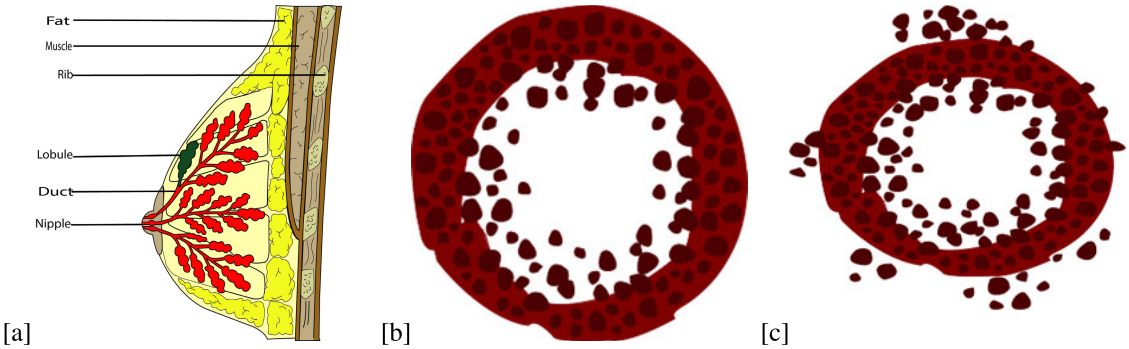

FIGURE 1. a) Anatomy of female breast b) Noninvasive cancer c) Invasive cancer.

survival rate is classified as long-term survival i-e more than 5 years and short-term survival i-e less than 5 years [3]. According to Cancer.Net1 report, the 5-year survival rate for non-invasive cases is 90 percent and the 10-year survival rate for women with non-invasive diagnosed breast cancer is 84 percent. If invasive breast cancer is confined to the breast, the 5-year survival rate is 99 percent. But if it is spread to lymph nodes the 5-year survival is 86 percent, and if it spread to other parts of the body it becomes 29 percent. It is arduous for physicians to effectively diagnose and prognoses invasive breast cancer due to its variety of clinical outcomes and complexity [3]. Initiating a prognosis is significant for the treatment of breast cancer. Firstly, a clear understanding of the disease allows physicians and patients to make informed decisions regarding treatment options tailored to the individual's specific needs [5].

In the past decade, rapid advances in machine learning and deep learning open up the ways for medical fields. Various techniques and tools were proposed by the researcher in the medical domain for the identification of health-related issues [6], [7]. Machine learning demonstrates superior performance in diagnosing and treating diseases that can lead to death. On the other hand, deep learning is specifically developed to identify the most relevant features for predicting disease [8], [9]. Research reports that deep learning is capable of diagnosing cancer before revealing its symptoms to patients with clinical standards [10].

Optimizing hyper-parameters for deep learning models can improve training and prediction, leading to more effective therapy plans for patients and doctors. Breast cancer prognosis can be improved by using multi-modal data such as clinical data, gene expression data (gene-exp), copy number variation (CNV), and other genomics data [11] through the application of deep learning models. Currently, researchers have proposed a limited number of methods that use a homogeneous deep learning model applied to all relevant modalities. Ensemble models are considered the optimal tool for improving performance and generalization [12], [13]. keeping the point in view we proposed a model with the following key characteristics:

- The objective of this study is to create a breast cancer prognosis method that utilizes a variety of deep-learning models through stacking. Our study targets to improve breast cancer prognosis by identifying novel prognostic factors that can help in the treatment of the disease.
- The primary aim of the research is to develop heterogeneous stacking-based models that will enable the efficient prognosis of breast cancer. This is achieved through three main steps, namely feature extraction, feature stacking, and classification algorithm.
- The innovation of this proposed method lies in the use of multiple deep learning models to extract features from a vast array of data modalities.
- The heterogeneous models are designed to extract highly informative features and produce more generalized results. The findings show that the proposed model outperforms the current benchmarks for breast cancer prognosis.
- Our study has important clinical implications, as it can guide the development of new prognostic tools and treatment strategies for breast cancer patients.

The paper is organized in a way that Section II of the paper will provide an overview of existing methods and models, while Section III will detail the methods and materials used in the proposed approach. Section IV will describe the experimental setup for the proposed model, and Section V will present the results, discussions, and conclusions.

II. RELATED WORK

The related work can be studied in the following subsections:

A. BREAST CANCER PROGNOSIS WITH SELECTIVE FEATURES

During past decades, with the fast implementation of gene expression analysis, and microarray techniques, many contributions help to understand the molecular signatures of breast cancer based on gene expression patterns in previous literature. Yixin wang and his colleague identify 76 gene expression signatures from 115 tumor tissues that can be useable in the prediction of distant metastasis lymph nodes specifically negative breast cancer [14]. From the independent data of 171 lymph node-negative patients, the results show 48% specificity and 93% sensitivity. As Breast cancer is a genetic disease, so for prediction and prognosis various

TABLE 1. Overall METABRIC dataset information.

Renounce Years	5 Years
Patient Count	1980
Survivors Long Term	1489
Survivors Short Term	491
Diagnosis Median Age	61
Average Survival Month	125.1

researchers contribute with upgraded techniques that hold multi-modal data like clinical data, gene expr, and CNV to enhance the prediction and prognosis accuracy [15].

B. BREAST CANCER PROGNOSIS WITH MULTIMODAL DATA USING MACHINE LEARNING AND DEEP LEARNING

The multi-modal is becoming more prevalent in capturing the complication of biological process, as it can unveil the underlying interconnections [16]. To enhance breast cancer prognosis Sun. integrates multi-modalities (clinical and genomic data). In their work, they proposed a new method of GPMLK of multi-kernel learning. GPMKL was trained for 130 selected pathological features out of 1990 and 20 features of gene expression and it shows better results than the previous contribution. As it holds limited data so the performance of GPMKL is not efficient [17]. Whereas Shimizu. targeted to design a novel prognosis score mPS that applies to a wide range of breast cancer patients. In the contribution, researchers identify 184 prognosis-related genes by training Random Forest and Neural Network. mPS system is tuned to 23 expression status and it is applicable where long rank $P < 0.05$ specifically independent of platform. It shows better performance when the mPS system is integrated with clinical data and also provides value that can be used to avoid over-treatment [18].

Previous work relies on selected features of genomic data for breast cancer prognosis. But Sun et al. work on multi-dimensional data with a deep learning model. Where the dense neural network is applied to multi-modal data i-e CNV, gene expression, and clinical independently. The output of DNN is fused in the middle layer. At the output, layer modalities were integrated with the score level fusion at the final prediction results. It shows better results than the existing work with 79% accuracy [19]. Li Tong. Integrates the multi-omics data to predict breast cancer survival. In this work, researchers implement ConcatAE (concatenation autoencoder) that merges the feature that a model learns from each modality and CrossAE to gain an invariant representation of modality. Proposed models show remarkable performance on TCCA multi-omics data of breast cancer [20]. Han introduced a Model Class Structure Based Deep Convolutional Neural Network (CSDCNN) that uses a nonlinear representation to efficiently classify breast cancer into multiple categories. With the BreaKHis dataset, which contains eight subclasses of breast cancer, CSDCNN achieved an impressive accuracy of 93.2% [21]. In other work Wei Shao. Proposed a method for feature selection for multi-modal data that were useful to identify the relationship between prognosis and prediction. The proposed framework efficiently highlights the relationship between prognosis and prediction by identifying the related features from multi-modal data [22]. Breast cancer modeling has proven to be a useful tool in comprehending the intricate nature of tumor growth during treatment. In a similar vein, the authors have introduced a mathematical model for breast cancer that employs a system of differential equations with piecewise constant arguments to investigate tumor growth and the effects of chemotherapy treatment. The authors have validated their theoretical findings through numerical simulations [23]. Luis A. presented MultiSurv, a multi-modal method for predicting long-term survival in PAN cancer. MultiSurv employs separate sub-models for each modality, including imaging, clinical, and other omics data, to establish feature representations. The sub-model outputs are merged to generate conditional survival predictions. The proposed method can handle missing data and be evaluated on 33 types of cancer that show an accurate survival curve [24].

C. BREAST CANCER PROGNOSIS WITH MULTIMODAL DATA USING ENSEMBLE DEEP LEARNING

Ensemble deep learning models have become increasingly important in the medical field, particularly in disease diagnosis and prognosis. The EDLCDS-BCDC technique, as suggested by the authors [25], utilizes ultrasound images for the detection of breast cancer. To extract features, the VGG-16, VGG-19, and SqueezeNet models are employed in the proposed approach. The final classification is achieved through the application of Cat Swarm Optimization in combination with a Multilayer Perceptron for the identification of breast cancer tissue. In [26], The authors incorporated a CNN into their study to predict progression-free survival, pathological complete response, and residual cancer burden. They utilized MRI scans, demographic data, and molecular subtypes as inputs. The researchers evaluated three different procedures: stacking, concatenation, and integration. The results demonstrated that integration outperformed the other two methods in addressing the concerned problem. The authors of this article [27] have presented an approach that employs hybrid transfer learning with a modified VGG architecture, utilizing both 2D and 3D mammogram images in the dataset. The experimental findings demonstrate that this proposed method achieves a prediction accuracy of 89.8%, surpassing the performance of existing methods. In this study, the authors suggest two automated approaches for breast cancer (BC) classification that utilize a combination of the Whale Optimization Algorithm (WOA) and Dragonfly Algorithm (DA) along with Radial Basis Function Kernel Support Vector Machines (RBF-SVM). The proposed methods aim to enhance the accuracy of BC classification by identifying the optimal SVM parameters, validated on the WBCD dataset. The proposed model outperforms the benchmarks [28]. J. Gao introduced

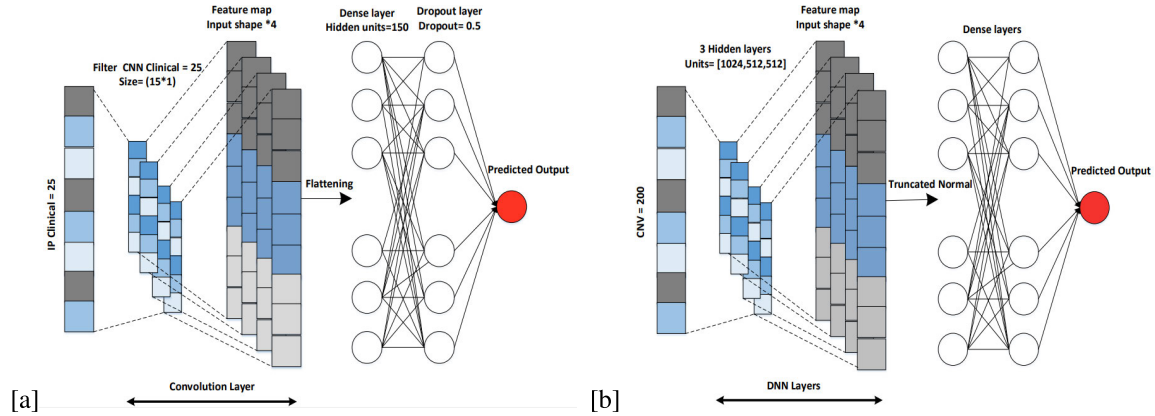

FIGURE 2. The detailed architecture a) CNN model architecture for feature extraction b) DNN model architecture for feature extractions.

TABLE 2. Feature selection detail.

Data set	Total Features	Selected Features
CNV	26298	200
Gene Expression	24368	400
Clinical	27	25

a novel framework called the Multimodal Graph Neural Network, which utilizes gene expression, CNV, and pathological data to classify short and long-term survivals. The framework constructs bipartite graphs between patients and multimodal data and fuses all features at the final layer. With an accuracy of 94% and an AUC of 0.97 [29]. In [30], the author has proposed for breast cancer diagnosis using a snapshot ensemble using neural networks with the feature reduction procedure using t-SNE that provides an efficient visualization and an accurate classification than the existing benchmarks. The authors proposed a multimodal adversarial representation learning approach for breast cancer prognosis prediction. The proposed approach integrates both clinical and imaging data to create a unified representation that can accurately forecast the prognosis of breast cancer patients. Based on the experimental findings, this approach performs better than existing methods in terms of predictive accuracy [31]. Nikhilanand Arya developed a model for prognosis that uses multi-modal data which were trained on CNN with stacking. Random forest is applied to stacked files for breast cancer prognosis that give AUC = 0.93 and ACC = 90% [3]. Liao et al. proposed a hybrid deep learning model that combines multiple modalities, including imaging and gene data. The researchers developed a fusion framework that utilized feature selection networks for each modality and then merged the outputs using a weighted linear aggregation. The fused features were then used to classify breast cancer subtypes, achieving an accuracy of 88.07% [32].

The literature has pointed out that several breast cancer prognosis models have been successful in making accurate predictions for patient survival, but they still have some limitations. Initially, researchers proposed models that focused on a single modality for breast cancer prognosis. Vijver [33] introduced the first prognosis model that was based on gene modality. However, relying solely on gene expression is insufficient to effectively predict breast cancer survivability. Incorporating additional sources of information, such as clinical data, copy number variations, and gene expression, can lead to more efficient prognosis models. To address the limitations of unimodal architectures, researchers have proposed multi-modal architectures. Sun et al. proposed one of the most effective multi-modal architectures in which authors designed a Multimodal Deep Neural Network by integrating Multi-dimensional Data (MDNNMD) [19]. The authors utilized clinical data, copy number variations, and gene expression modality for the proposed model. DNN neural network is applied to all concerned data for feature extraction and then performs fuse scoring. The result shows better performance than other prognosis models. N.Arya proposes another breast cancer prognosis model citearya2020multi, which is a dual-staged deep learning-based stacked ensemble model. This model utilizes clinical data, copy number variations, and gene expression datasets. In the first stage, CNNs are used for feature extraction, and in the second stage, the extracted features are inputted into the stacked ensemble model. The proposed model uses a homogeneous model-CNN for each modality to extract features. However, the literature suggests that heterogeneity in ensemble deep learning is more effective and provides better generalization [34]. Our proposed architecture utilizes a heterogeneous model for multi-modal datasets to ensure a more accurate survivability model for breast cancer prognosis.

The motivation behind the proposed work is to address the limitations of existing breast cancer prognosis models, which often rely on a single modality and may not provide accurate predictions for patient survivability. The research gaps in the field have highlighted the need for multi-modal approaches that incorporate diverse sources of information. The objective of the proposed work is to develop a heterogeneous model

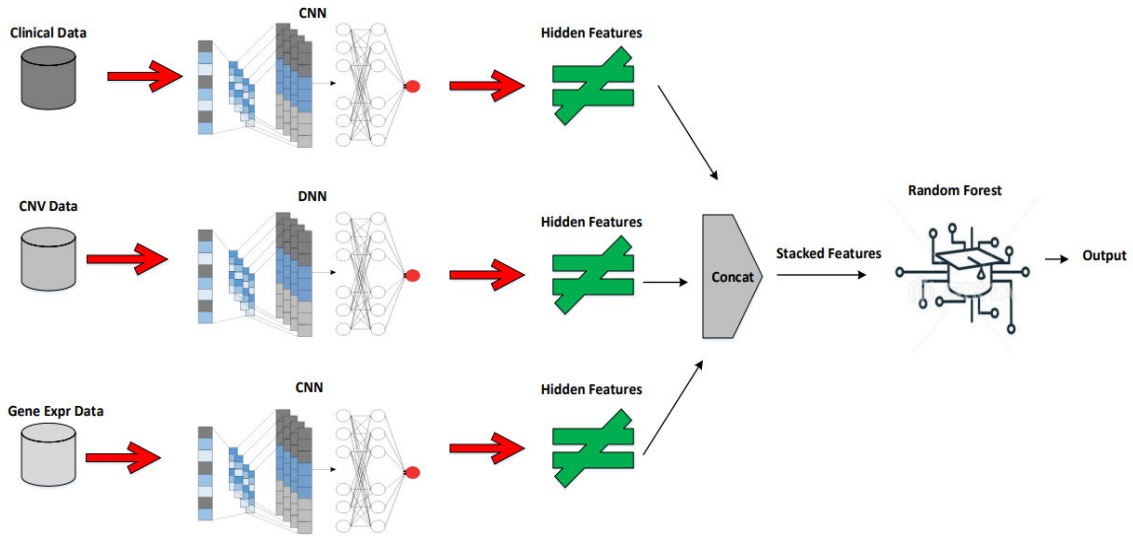

FIGURE 3. The systematic proposed framework.

TABLE 3. CNN tuned parameters for clinical and gene expression.

Initializer	Adam
Initializer Constant	0.1
No. of convolutional layers	1
No. of filters	25,4
Kernel Size	15
Stride	2
Padding	Same
Activation function	Tanh
Number of hidden layers	1
Hidden units	150
Training Epoch	20
Batch size	8
Activation Function	Tanh
Loss Function	L2 regularization, Binary cross entropy

that leverages multiple modalities to improve the accuracy of breast cancer prognosis models, ultimately leading to better patient outcomes. By justifying the research gaps and objectives, the proposed work aims to make a significant contribution to the field of breast cancer research.

III. MATERIAL AND METHODS

1) DATASET

The preprocessed METABRIC dataset has been directly used and is publicly available at: https://github.com/USTC-HIlab/MDNNMD. The METABRIC dataset comprises (n = 1980) valid breast cancer patient records [35]. Patients diagnosed with breast cancer have access to multi-modal data, including gene expr, clinical information, and CNV. The samples were categorized into two cohorts: long-term survivors, who lived for more than 5 years after diagnosis, and short-term survivors, who lived for less than 5 years after diagnosis. METABRIC consists of 1489 and 491 samples for long-term survival and short-term survival respectively. The diagnosis median age for the available samples is 61 years and the average survivability in months is 125.1 months. Long-term survivors were labeled as '0' and short-term survivors were labeled as '1'. The overall METABRIC data set information is given in Table 1. Out of 1980 patients, 64 patients were lost to follow-up for five years, representing 3.23% of the total samples. For our model, suspicious samples from METABRIC were considered long-term survivors. Any missing values in the gene expression and copy number variation dataset were imputed using a weighted nearest neighbor algorithm [36]. Sun et al. have discretized and normalized the gene expression data into three levels: under-expression (−1), baseline (0), and over-expression (1) in correspondence to Gevaert et al. [37]. The copy number variation-CNV dataset has been used with five values [−2, −1, 0, 1, 2]. Normalization of the clinical dataset is performed by using min-max algorithm [38] in the range of [0,1] by considering the eq.(1).

$$x'_{i,n} = \frac{x_{i,n} - min(x_i)}{max(x_i) - min(x_i)}(nMax - nMin) + nMin \quad (1)$$

2) FEATURE SELECTION

Considering the nature of the dataset having the high dimensional and low sample size (HDLSS) curse would cost the efficiency of deep learning models [39]. In the proposed architecture, multi-modal data consisting of copy number variation, gene expression, and clinical profiles are taken into account. These data have 26,000, and 24,000 features for CNV and gene expression modality. The clinical modality for each patient consists of 27 features, such as age at diagnosis, cellularity, size, lymph node positivity, and others that can be seen in the provided link.[1] To alleviate the curse of HDLSS, we used the well-known algorithm mRMR to reduce the dimensionality of our data [40], [41]. Feature extraction was

[1]https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric

TABLE 4. DNN tuned parameters for copy number variation.

Initializer	Adam
Initializer Constant	0.1
No. of hidden layers	4
Hidden units at 1,2,3,4 layer	[1024,512,512,512]
Batch size	64
Stride	2
Activation Function	Tanh
Initial Learning Rate	'10-3
Training Epoch	20
Hidden units	150
Training Epoch	20
Loss function	L2 regularization, Binary Cross entropy

done incrementally using the mRMR algorithm, and the final selection was based on AUC values obtained from various cohorts. The model's performance was evaluated using the top 100 features, which were selected in the first iteration. In subsequent iterations, the number of features increased to 200, 300, and finally, 500, and the model's performance was evaluated each time. The features with high AUC scores were selected for the heterogeneous stacked RF model, and Table 2 provides a detailed description of these selected features. The selection of features was based on their high AUC score, and for CNV data, Gene expression profile, and clinical data, 200, 400, and 25 features were selected, respectively.

IV. EXPERIMENTAL SETUP

For our experimental configuration, we utilized ten-fold cross-validation to assess the proposed framework, as in prior research [3]. Our dataset consisted of 1980 patients, which we randomly divided into ten subsets. We combined nine of these subsets to form a training set, while one was reserved as the testing set. Additionally, we further partitioned the merged training set into an 80% training subset and a 20% validation subset Figure 3 shows the systematic proposed framework where Keras 2.12.0, along with Tensorflow 2.12.0, is used for implementing the source code of the model.

1) A CNN AND DNN-BASED PREDICTION FOR A UNI-MODAL DATASET

The Heterogeneous stacked model proposed in this study utilizes both a convolutional neural network (CNN) and a deep neural network (DNN) for predicting breast cancer in humans and extracting features for the subsequent phase of the model. The proposed model comprises both CNN and DNN and incorporates learnable filters directly on the Multimodal sequences. The CNN is applied to clinical and gene expression data, while the DNN is applied to the copy number variation dataset. The CNN and DNN are used at phase one for feature extraction for the next stage of the model while having certain numbers of filters. It generates a feature map that is the output of the convolution and dense process. The feature mapping is achieved through element-wise multiplication followed by addition between

TABLE 5. Confusion metrics for the validation set.

	Long Term Survivor	Short Term Survivor
Long Term Survivor	368	0
Short Term Survivor	10	117

TABLE 6. Class evaluation of heterogeneous stacked model.

	Precision	Recall	F1 score
Long Term Survivor	1.00	0.97	0.98
Short Term Survivor	0.92	1.00	0.96

the corresponding values of the input matrix and the filter matrix.

In [3], The Glorot normal initializer is utilized to initialize the filter values. This approach involves selecting values that have a mean of zero and standard deviation within a specific range.

$$-\sqrt{\frac{2}{n_i, n_o}}, \sqrt{\frac{2}{n_i, n_o}} \tag{2}$$

The input and output values are denoted by ni and n0, respectively. A constant seed value of 0.1 is utilized by both the CNN and DNN models. The stride rate for the convolutional layer is 2 which shifts the filter with the difference of 2 to perform convolution over the input matrix. Padding is also added to convolution layers for the control of feature size. After that flattened layer is used to flatten the output of the convolutional layer then we passed the output from the dense layer with 150 hidden units. The regularization technique Dropout [42], was applied for DNN after each layer to prevent overfitting. Whereas L2 regularization is used for CNN as it has vast applications in deep learning [43], [44]. The activation function used at the convolutional layer is Tanh whereas the sigmoid activation function is used at the dense layer. The detailed architecture of CNN and DNN is shown in Figure 2.

For model training, a recently proposed optimizer Adam [45] was used. The parallel cross-entropy is used, and work was upgraded utilizing that enhancer since it is computationally proficient. It has required a little memory and is all around applied to those issues that are huge as far as boundaries and additional information. A loss function is used to measure or see how well a deep learning classifier fits or suites to empirical data. In our case, since the problem is a binary classification problem, the ground truth (y) can only have two states, which are one or zero. To evaluate the prediction performance of the deep learner, we used binary cross-entropy loss in this study.

Finally, the CNN model is composed of a single convolutional layer, a flattened layer, a dense layer, and an output layer. On the other hand, the DNN comprises four dense layers with varying filter sizes and neurons. Each layer is followed by a 50% dropout layer, and an output layer is present at the end. The AUC metric is evaluated for different subgroup sizes ranging from 8 to 128, and the results show

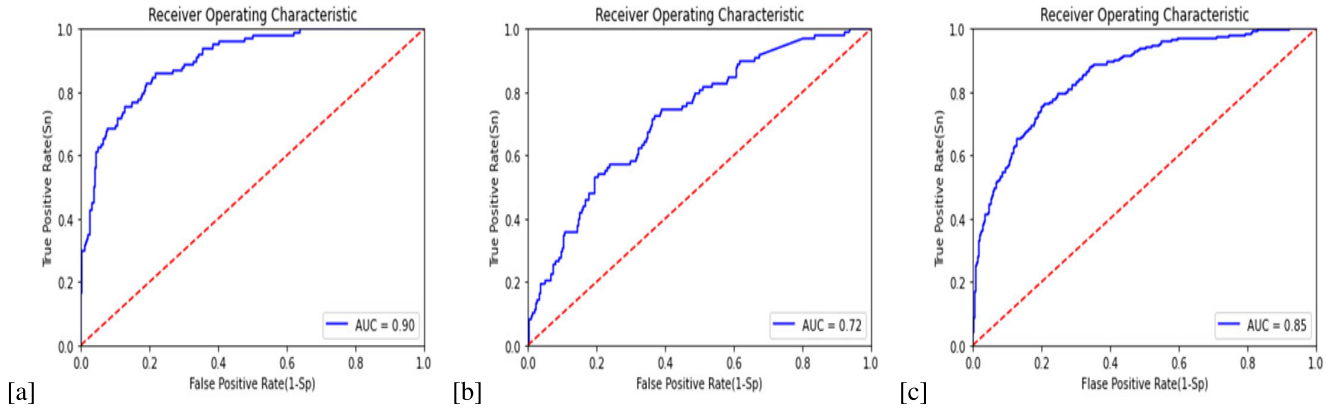

FIGURE 4. ROC curve for the heterogeneous stacked model having both CNN and DNN for stacked feature extraction for prognosis prediction a) AUC of CNN for Gene expression b) AUC of DNN for CNV c) AUC of CNN for clinical.

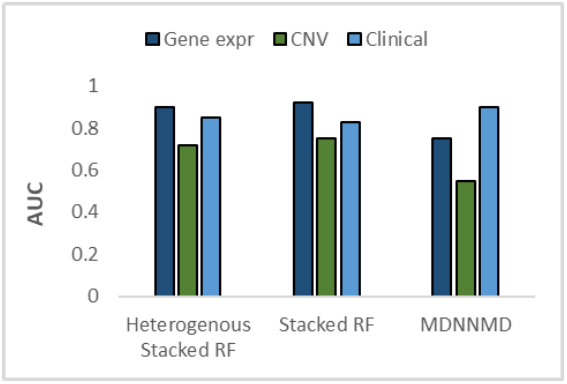

FIGURE 5. AUC comparison of individual modality.

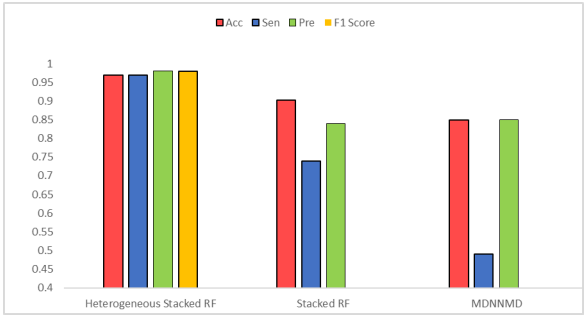

FIGURE 7. Result evaluation of heterogeneous stacked model with existing benchmarks.

FIGURE 6. Accuracy comparison of individual modality.

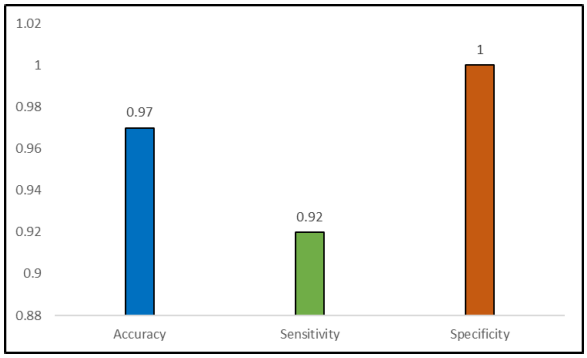

FIGURE 8. Heterogeneous model evaluation parameters.

that a batch size of 8 yields the best performance. Hence, a batch size of 8 is chosen for the final model. The detailed parameter configurations of CNN and DNN are presented in Figure 3.

2) HETEROGENEOUS STACKED-BASED MODEL FOR MULTI-MODAL PREDICTION

The proposed model is divided into three phases. These phases were described as under:

Phase 1: In phase 1, we train the CNN for clinical data at the first layer, secondly CNV data is used to train a DNN at the second layer, and finally we train the CNN for the modality of Gene Expr data.

Phase 2: In this phase, we shape a stacked feature set by using extracted features of the output of CNN and DNN.

Phase 3: Finally, we pass this stacked feature set to Random Forest (RF) algorithm for further classification. We evaluate the performance of our proposed framework with the parameters given as Sensitivity, Specificity, Precision, and Accuracy.

TABLE 7. Heterogeneous stacked model results along with existing benchmarks.

Model	AUC	ACC
Heterogeneous Stacked RF		
CNN-gene exp	0.90	0.80
DNN- cnv	0.72	0.80
CNN- Clinical	0.85	0.81
STACKED RF		
CNN-gene exp	0.92	0.80
CNN- cnv	0.75	0.74
CNN- Clinical	0.83	0.80
MDNNMD		
DNN-Gene Exp	0.761	0.74
DNN-CNV	0.614	0.76
DNN-Clinical	0.807	0.79

TABLE 8. Performance comparison of the proposed model with existing benchmark.

Model	Sen	Pre	F1 score	ACC
Heterogeneous Stacked RF	0.97	0.98	0.98	0.97
STACKED RF	0.74	0.84	-	0.90
MDNNMD	0.45	0.74	-	0.82

Mathematically, these evaluation parameters are described as:

$$Sensitivity = \frac{Tp}{Tp + Fn} \tag{3}$$

$$Specificity = \frac{Tn}{Tn + Fn} \tag{4}$$

$$Precision = \frac{Tp}{Tp + Fp} \tag{5}$$

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{6}$$

The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) is used to calculate the model's performance. Additionally, the area under the curve (AUC) value is computed using the receiver operating characteristic (ROC) curve to evaluate the model's efficiency.

V. RESULTS AND DISCUSSION

This section presents a detailed comparison of our proposed framework with existing benchmarks.

A. HETEROGENEOUS STACKED MODEL PERFORMANCE

A heterogeneous stacked model is designed to combine the final feature extractions from the CNN and DNN models. The AUC metric, calculated from the ROC curve, is utilized as the performance evaluation criterion for the feature extraction models, in addition to the accuracy metric. The model with a higher AUC value is deemed more effective compared to the one with a lower AUC value. In Figure 4, we present the AUC values obtained from the ROC curve analysis. As the graph depicted AUC for the CNN applied to gene exp and clinical modality is 0.90 and 0.85 respectively. AUC for DNN applied to CNV data modality is 0.72. Furthermore, the accuracies for CNN-gene Exp, DNN-CNV, and CNN-clinical are 80.56%, 80.45%, and 81.21%, respectively. The AUC results were not as anticipated, primarily due to a high false positive rate that carries potential implications for patients. To overcome the issue of variance caused by the limited size of the dataset, we employed ten-fold cross-validation to evaluate the proposed model. The dataset of 1980 patients was divided into 10 subsets, and in each iteration, 9 subsets were combined to form the training set while the remaining subset was used as the testing set.

After analyzing the AUC the features from the individual neural network are taken as stacked features. Where RF is applied to the stacked features, it is evident from the literature that RF outperforms stacked features as compared to other classifiers [17]. So we validate our model at the last level using an RF classifier. Where we determine the confusion metrics, Sensitivity, F1 score, Precision, Specificity, and accuracy of the proposed model. The confusion metrics for the validation set are shown in Table 3. The table presented below confirms that the proposed models are efficient in predicting the samples. Out of a total of 495 samples in the validation set, 485 samples were correctly predicted, as shown in the table. There were only 10 false negative predictions and no false positive predictions. The lack of false positives may be attributed to imbalanced classes. Whereas the class evaluation of the heterogeneous stacked model is shown in Table 4. It is clear from the table below that the heterogeneous stacked model shows better results. It shows the precision, recall, and F1 score for the binary classes. Where the long-term survival holds 1.00, 0.97, and 0.98 for precision, recall, and F1 score respectively. For short-term survival, the heterogeneous stacked model secures 0.92, 1.00, and 0.96 precision, recall, and F1 score respectively. The stacked features are passed to the random forest where the overall results of the model were evaluated with multiple threshold values. The overall result evaluation metrics were accuracy, sensitivity, and specificity, where it secure 0.97, 0.92, and 1.0 respectively as depicted in Figure 8. It is evident from our results that the heterogeneous stacked models show better results than the existing benchmarks as shown in Table 5.

B. COMPARISON OF HETEROGENEOUS STACKED MODEL WITH EXISTING BENCHMARKS

It is evident from the result above that the heterogeneous stacked model has improved the result compared to the current benchmarks. We compare our proposed heterogeneous stacked model with popular prediction methods such as MDNNMD, Stacked RF-based ensemble model [3], [17]. We compute the AUC values under the ROC for the heterogeneous stacked-based model, in comparison to stacked based ensemble model, MDNNMD. The literature demonstrates that stacked-based ensemble and Heterogeneous stacked-based methods outperform the other benchmarks. In comparison to the stacked RF model and MDNNMD, the heterogeneous stacked model had higher AUC and ACC values. In comparison, the stacked RF model gains 0.83, 0.92, and 0.70 AUC values for CNN-Gene Exp, CNN-Clinical,

and CNN-CNV respectively that are lesser than the heterogeneous stacked RF as shown in Figure 5(a). Whereas Figure 5(b) demonstrates the individual modality accuracy in comparison to the existing benchmarks. The graph shows that the heterogeneous stacked model shows better prediction on individual modality as compared to the existing prognosis model.

It is determined from the simulations that our heterogeneous stacking of neural networks achieves upto-mark results as compared to existing frameworks such as stacked RF and MDNNMD as shown in table 7. The heterogeneous stacked model gains up-to-the-mark results for each modality trained under different neural networks. The overall model shows better results than the existing models. Compared to these benchmarks, our proposed framework achieves 0.97, 0.97, 0.98, 0.98 for accuracy, sensitivity, precision, and F1 score respectively as shown in Figure 7. While comparing the different evaluation parameters it is validated that the proposed model earned better results than the existing benchmarks as shown in Table 8.

VI. CONCLUSION

This article presents a novel approach to predicting breast cancer prognosis using ensemble deep learning. Unlike previous research, we utilized a stacked ensemble model with individual neural networks for feature extraction from different data modalities. Our model incorporated CNN for feature extraction of clinical data and gene expression data, and DNN for CNV data. In the second phase, the extracted features were stacked and used as input for the RF for the classification of short-term and long-term survivals. Our results demonstrate that our heterogeneous framework outperformed homogeneous stacking and other benchmarks, achieving an accuracy of 97%. It is important to note that further validation or integration with other sources of information is necessary before utilizing the predicted output for clinical decision-making. Additionally, our approach can be extended to predict other diseases with different algorithms and can include more data modalities for prognosis such as gene methylation and miRNA.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

REFERENCES

[1] G. Chugh, S. Kumar, and N. Singh, "Survey on machine learning and deep learning applications in breast cancer diagnosis," *Cognit. Comput.*, vol. 13, no. 6, pp. 1451–1470, Nov. 2021.

[2] S. Kumar and S. Sharma, "Sub-classification of invasive and non-invasive cancer from magnification independent histopathological images using hybrid neural networks," *Evol. Intell.*, vol. 15, no. 3, pp. 1531–1543, Sep. 2022.

[3] N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 2, pp. 1032–1041, Mar. 2022.

[4] A. Gucalp, T. A. Traina, J. R. Eisner, J. S. Parker, S. R. Selitsky, B. H. Park, A. D. Elias, E. S. Baskin-Bey, and F. Cardoso, "Male breast cancer: A disease distinct from female breast cancer," *Breast Cancer Res. Treatment*, vol. 173, no. 1, pp. 37–48, Jan. 2019.

[5] C. Siotos, M. McColl, K. Psoter, R. C. Gilmore, M. E. Sebai, K. P. Broderick, L. K. Jacobs, S. Irwin, G. D. Rosson, and M. Habibi, "Tumor site and breast cancer prognosis," *Clin. Breast Cancer*, vol. 18, no. 5, pp. 1045–1052, Oct. 2018.

[6] K. Rautela, D. Kumar, and V. Kumar, "A systematic review on breast cancer detection using deep learning techniques," *Arch. Comput. Methods Eng.*, vol. 29, pp. 4599–4629, Apr. 2022.

[7] F. M. Calisto, N. Nunes, and J. C. Nascimento, "Modeling adoption of intelligent agents in medical imaging," *Int. J. Hum.-Comput. Stud.*, vol. 168, Dec. 2022, Art. no. 102922.

[8] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," *Artif. Intell. Med.*, vol. 127, May 2022, Art. no. 102276.

[9] W. El-Shafai, A. A. Mahmoud, A. M. Ali, E.-S. M. El-Rabaie, T. E. Taha, O. F. Zahran, A. S. El-Fishawy, N. F. Soliman, A. A. Alhussan, and F. E. Abd El-Samie, "Deep CNN model for multimodal medical image denoising," *Comput., Mater. Continua*, vol. 73, no. 2, pp. 3795–3814, 2022.

[10] E. Landhuis, "Deep learning takes on tumours," *Nature*, vol. 580, no. 7804, pp. 551–554, 2020.

[11] W. Zhu, L. Xie, J. Han, and X. Guo, "The application of deep learning in cancer prognosis prediction," *Cancers*, vol. 12, no. 3, p. 603, Mar. 2020.

[12] A. Petrakova, M. Affenzeller, and G. Merkurjeva, "Heterogeneous versus homogeneous machine learning ensembles," *Inf. Technol. Manag. Sci.*, vol. 18, no. 1, pp. 135–140, Jan. 2015.

[13] E. Mustafa, E. K. Jadoon, S. Khaliq-Uz-Zaman, M. A. Humayun, and M. Maray, "An ensembled framework for human breast cancer survivability prediction using deep learning," *Diagnostics*, vol. 13, no. 10, p. 1688, May 2023.

[14] Y. Wang, J. G. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Talantov, M. Timmermans, M. E. Meijer-van Gelder, J. Yu, T. Jatkoe, E. M. Berns, D. Atkins, and J. A. Foekens, "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *Lancet*, vol. 365, no. 9460, pp. 671–679, Feb. 2005.

[15] M. Khademi and N. S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 727–732.

[16] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: A review," *Briefings Bioinf.*, vol. 23, no. 2, Mar. 2022, Art. no. bbab569.

[17] D. Sun, A. Li, B. Tang, and M. Wang, "Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome," *Comput. Methods Programs Biomed.*, vol. 161, pp. 45–53, Jul. 2018.

[18] H. Shimizu and K. I. Nakayama, "A 23 gene-based molecular prognostic score precisely predicts overall survival of breast cancer patients," *EBioMedicine*, vol. 46, pp. 150–159, Aug. 2019.

[19] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 3, pp. 841–850, May 2019.

[20] L. Tong, J. Mitchel, K. Chatlin, and M. D. Wang, "Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–12, Dec. 2020.

[21] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Jun. 2017.

[22] W. Shao, T. Wang, L. Sun, T. Dong, Z. Han, Z. Huang, J. Zhang, D. Zhang, and K. Huang, "Multi-task multi-modal learning for joint diagnosis and prognosis of human cancers," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101795.

[23] F. B. Yousef, A. Yousef, T. Abdeljawad, and A. Kalinli, "Mathematical modeling of breast cancer in a mixed immune-chemotherapy treatment considering the effect of ketogenic diet," *Eur. Phys. J. Plus*, vol. 135, no. 12, pp. 1–23, Dec. 2020.

[24] L. A. Vale-Silva and K. Rohr, "MultiSurv: Long-term cancer survival prediction using multimodal deep learning," *MedRxiv*, Aug. 2020.

[25] M. Ragab, A. Albukhari, J. Alyami, and R. F. Mansour, "Ensemble deep-learning-enabled clinical decision support system for breast cancer diagnosis and classification on ultrasound images," *Biology*, vol. 11, no. 3, p. 439, Mar. 2022.
[26] H. Dammu, T. Ren, and T. Q. Duong, "Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients," *PLoS ONE*, vol. 18, no. 1, Jan. 2023, Art. no. e0280148.
[27] A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, T. K. Phung, and D. N. H. Thanh, "Diagnosis of breast cancer based on modern mammography using hybrid transfer learning," *Multidimensional Syst. Signal Process.*, vol. 32, no. 2, pp. 747–765, Apr. 2021.
[28] A. S. Elkorany, M. Marey, K. M. Almustafa, and Z. F. Elsharkawy, "Breast cancer diagnosis using support vector machines optimized by whale optimization and dragonfly algorithms," *IEEE Access*, vol. 10, pp. 69688–69699, 2022.
[29] J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li, "MGNN: A multimodal graph neural network for predicting the survival of cancer patients," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2020, pp. 1697–1700.
[30] N. Sharma, K. P. Sharma, M. Mangla, and R. Rani, "Breast cancer classification using snapshot ensemble deep learning model and t-distributed stochastic neighbor embedding," *Multimedia Tools Appl.*, vol. 82, no. 3, pp. 4011–4029, Jan. 2023.
[31] X. Du and Y. Zhao, "Multimodal adversarial representation learning for breast cancer prognosis prediction," *Comput. Biol. Med.*, vol. 157, May 2023, Art. no. 106765.
[32] T. Liu, J. Huang, T. Liao, R. Pu, S. Liu, and Y. Peng, "A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data," *IRBM*, vol. 43, no. 1, pp. 62–74, Feb. 2022.
[33] M. J. Van De Vijver, Y. D. He, L. J. Van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, and M. J. Marton, "A gene-expression signature as a predictor of survival in breast cancer," *New England J. Med.*, vol. 347, no. 25, pp. 1999–2009, 2002.
[34] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," 2021, *arXiv:2104.02395*.
[35] A. Mukherjee, R. Russell, S.-F. Chin, B. Liu, O. M. Rueda, H. R. Ali, G. Turashvili, B. Mahler-Araujo, I. O. Ellis, S. Aparicio, C. Caldas, and E. Provenzano, "Associations between genomic stratification of breast cancer and centrally reviewed tumour pathology in the METABRIC cohort," *Npj Breast Cancer*, vol. 4, no. 1, pp. 1–9, Mar. 2018.
[36] B. Al-Helali, Q. Chen, B. Xue, and M. Zhang, "A new imputation method based on genetic programming and weighted KNN for symbolic regression with incomplete data," *Soft Comput.*, vol. 25, no. 8, pp. 5993–6012, Apr. 2021.
[37] O. Gevaert, F. D. Smet, D. Timmerman, Y. Moreau, and B. D. Moor, "Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks," *Bioinformatics*, vol. 22, no. 14, pp. 184–190, Jul. 2006.
[38] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524.
[39] S. L. Verghese, I. Y. Liao, T. H. Maul, and S. Y. Chong, "An empirical study of several information theoretic based feature extraction methods for classifying high dimensional low sample size data," *IEEE Access*, vol. 9, pp. 69157–69172, 2021.
[40] M. Van Gerven and S. Bohte, "Artificial neural networks as models of neural information processing," Frontiers Media Ltd., London, U.K., Tech. Rep., 2017.
[41] Y. Eroğlu, M. Yildirim, and A. Çinar, "Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104407.
[42] J. Roder, C. Oliveira, L. Net, M. Tsypin, B. Linstid, and H. Roder, "A dropout-regularized classifier development approach optimized for precision medicine test discovery from omics data," *BMC Bioinf.*, vol. 20, no. 1, pp. 1–14, Dec. 2019.
[43] V. A. Kumar, S. Gupta, S. S. Chandra, S. Raman, and S. S. Channappayya, "No-reference quality assessment of tone mapped high dynamic range (HDR) images using transfer learning," in *Proc. 9th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2017, pp. 1–3.
[44] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

EHTISHAM KHAN JADOON received the B.S. degree from the Abbottabad University of Science and Technology, and the M.S. degree from COMSATS University Islamabad, Abbottabad Campus. His research interests include deep learning, machine learning, and neural networks.

FIAZ GUL KHAN received the master's and Ph.D. degrees from Politecnico di Torino, Italy, in 2013. He is currently an Associate Professor with the Computer Science Department, COMSATS University Islamabad, Abbottabad Campus, Pakistan. His research interests include concurrent computing, machine learning, artificial intelligence, and GPU computing.

SAJID SHAH received the M.S. and Ph.D. degrees from Politecnico di Torino, Italy. He was an Assistant Professor with COMSATS University Islamabad, Abbottabad Campus, Pakistan. He is currently a Postdoctoral Researcher with the EIAS Laboratory, Prince Sultan University, Riyadh, Saudi Arabia. His research interests include data mining, text mining, machine learning, bioinformatics, and image processing.

AHMAD KHAN received the Ph.D. degree from the National University of Computer and Emerging Sciences (FAST-NU), Islamabad, Pakistan, in 2015. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI), Abbottabad Campus. His research interests include computer vision, machine learning, and evolutionary algorithms.

MUHAMMED ELAFFENDI is currently a Professor of computer science with the Department of Computer Science, College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh, Saudi Arabia. He is the Founder and the Director of the Center of Excellence in CyberSecurity (CYBEX) and the EIAS Data Science Research Laboratory, AIDE to the Rector, the Director of the Institutional Policy and Development Unit, and the Former Dean of CCIS. For the past few years, he has been performing research in the areas of machine learning, natural language processing, emerging distributed architectures, next generation computer systems, and complexity science.