Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model

Nikhilanand Arya and Sriparna Saha

Abstract—reast Cancer is a highly aggressive type of cancer generally formed in the cells of the breast. Despite significant advances in the treatment of primary breast cancer in the last decade, there is a dire need to attempt of an accurate predictive model for breast cancer prognosis prediction.reast Cancer is a highly aggressive type of cancer generally formed in the cells of the breast. Despite significant advances in the treatment of primary breast cancer in the last decade, there is a dire need to attempt of an accurate predictive model for breast cancer prognosis prediction.B Researchers from various disciplines are working together to develop methods to save people from this fatal disease. A good predictive model can help in correct prognosis prediction of breast cancer. This accurate prediction can have several benefits like detection of cancer in the early stage, spare patients from getting unnecessary treatment and medical expenses related to it. Previous works rely mostly on uni-modal data (selected gene expression) for predictive model design. In recent years, however, multi-modal cancer data sets have become available (gene expression, copy number alteration and clinical). Motivated by the enhancement of deep-learning based models, in the current study, we propose to use some deep-learning based predictive models in a stacked ensemble framework to improve the prognosis prediction of breast cancer from available multi-modal data sets. One of the unique advantages of the proposed approach lies in the architecture of the model. It is a two-stage model. Stage one uses a convolutional neural network for feature extraction, while stage two uses the extracted features as input to the stack-based ensemble model. The predictive performance evaluated using different performance measures shows that this model produces a better result than already existing approaches. This model results in AUC value 0.93 and accuracy 90.2% at medium stringency level (Specificity = 95% and threshold = 0.45). Keras 2.2.1, along with Tensorflow 1.12, is used for implementing the source code of the model. The source code can be downloaded from Github: https://github.com/nikhilaryan92/BreastCancer.

Index Terms—Breast cancer prognosis prediction, Multi-modal learning, Stacked-based ensemble network, Random forest (RF), Support vector machine (SVM), Convolution neural network (CNN), Stacked features, Deep-learning.

1 Introduction

BREAST cancer is caused by the uncontrolled growth of the cells present inside the breast. This growth results in the formation of the sheet of cells or a mass, termer as tumor. Cancerous tumors are categorised into two classes: malignant and benign [1]. Malignant tumors can grow and spread, while benign tumors can grow but not spread across other body parts. Breast cancer can be divided into two categories as invasive and non-invasive. Invasive breast cancer [1] is cancer that spreads across adjacent tissues inside the body while non-invasive breast cancer [1] is limited to the milk ducts or within the lobules in the breast. Ductal carcinoma or lobular carcinoma, which originates in ducts or lobes is considered as highly aggressive kind of cancer and leads to significant health problems in the female, and is one of the major causes of cancer-related deaths worldwide [2] [3]. The data provided by **Cancer.Net (a website under ASCO)** suggests that invasive breast cancer has affected 266,120 women, while non-invasive breast cancer has affected 63,960 women in the United States. Approximately, 41,000 deaths (98.82% women and 1.18% men) due to breast cancer have been estimated for the year 2018 in the USA. 62% of breast cancer cases were diagnosed at non-invasive stage (present only in the breast), and 5-year survival rate for these cases is 99%. In the case of invasive breast cancer, the 5-year survival rate is 85% if it has spread to the regional lymph nodes, and 27% if cancer has spread to a distant part of the body. It is difficult for doctors to predict and treat invasive breast cancer cases due to its complexity and significantly varied clinical outcomes [4]. A more accurate predictive model may help patients who are suffering from breast cancer by predicting the chances of survival. It may also help the doctors to take appropriate therapy decision for the patients [5] [6]. The life expectancy of the cancer patient is further classified into long-time survivor (more than 5-year survival) and short-time survivor (less than 5-year survival). With the help of predictive models when patients are predicted as short time survivors, doctors may suggest patients **with personalized cancer treatments and spare them from receiving unnecessary adjuvant therapy and sufferings caused by the toxic side effects**. There are multiple sources containing breast cancer information such as genetic details (gene expression data and copy number variation data) and clinical data (age, early menstruation and late menopause, the timing of pregnancy, lifestyle factors and others). The integration of these multi-modal data may improve the predictive power of the model.

- *Nikhilanand Arya is pursuing M.tech in Mathematics and Computing, Indian Institute of Technology Patna, India.*
  E-mail: nikhilaryan92@gmail.com
- *Dr. Sriparna Saha is an Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India.*
  E-mail: sriparna@iitp.ac.in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3018467, IEEE/ACM Transactions on Computational Biology and Bioinformatics

1.1 Literature Survey

Past studies on breast cancer have suggested some gene expression patterns to understand the molecular signature of breast cancer. *Van de Vijer et al.* were the first to propose the predictive model of breast cancer prognosis using gene expression profiles [7]. They used gene expression of 98 primary breast cancer patients and selected 70 gene signatures with the help of supervised classification methods. The result was validated with unseen data of 19 young breast cancer patients. Regarding multi-modality, ***Sun et al.* proposed a model which uses hybrid signature consisting of three gene markers and two clinical markers generated from both 70 genetic signatures and clinical markers for the prediction of breast cancer prognosis.**[5]. A probabilistic model based on Bayesian Network (BN) for the prognosis of lymph-node negative breast cancer by combining clinical and genomic data was also proposed by *Gevaert et al.* [6]. Literature review suggests that most of the previous works infer that gene expression profile is conducive in breast cancer prognosis prediction, but **a few** studies were primarily based on the assumption that different genes of a particular patient were not related to each other.

Considering the fact that different genes of a particular patient may have significant relations amongst themselves, several works like an efficient feature selection method based on support vector machine (SVM) proposed by *Xu et al.* [8] and feature selection combined with random forest (RF) proposed by *Naguyen et al.* [9] outperformed the previous 70 gene signatures in terms of performance for breast cancer prognosis prediction. Later, a probabilistic graphical model (PGM) was proposed by *Khademi et al.* [10], which is the combination of two independent models of microarray data (high dimensional data consists of approx 25,000 genes per patient) and clinical data, for prognosis and diagnosis of breast cancer. It uses principal component analysis (PCA) for dimensionality reduction of microarray data and constructs a deep belief network to extract feature representation of data. It also uses the structure learning algorithm for clinical data. All these methods resulted in good performing models, but these models directly combine different modalities of data (e.g., gene expression and clinical) without taking into consideration the fact that different modalities may have different feature representations.

Recent advancement in deep learning methods shows that a model having multiple modalities of input data source performs well as compared to a model with a single source of input data. This fact has been validated in some studies of breast cancer prognosis and diagnosis, which are based on multi-modal data [11] [12] [13]. Sun et al. have also introduced GPMKL method [14] for breast cancer prognosis prediction by integrating genomic data and pathological images. Similarly, Zhang et al. presented a novel MKL (Multiple Kernel Learning) method known as HI-MKL [15] for Glioblastoma Multiforme (GBM) prognosis prediction by integrating histopathological image and multi-omics data. Considering the need of research for survival analysis in the medical domain and availability of histopathological images, Tang et al. have designed a Capsule Network termed as CapSurv [16]. This approach is based on a new loss function known as survival loss, specially designed for survival analysis of cancer patients. Multi-modal deep learning has also proved its superiority in the fields of computer visual recognition [17], multimedia analysis [18], speech recognition [19], sentiment analysis [20] and bioinformatics [13]. Models based on the single source of input data suffer from a dearth of non-universal, unique and noisy data, while multi-modal models overcome all these limitations by combining related information from multiple sources [21] [18].

1.2 Drawbacks of Existing Models

Existing models for breast cancer prognosis prediction have given proper predictive measures, but these models have certain limitations. The very first model designed by *Van de Vijer et al.* [7] is a uni-modal model, and it takes gene expression data as input for the model. As we know, gene signatures are not the only factor for breast cancer. Clinical data, copy number alteration and scanned image of cancerous cells also contain multiple information regarding breast cancer patients which might help in the prognosis and diagnosis prediction of the breast cancer. Hence, researchers proposed multi-modal architectures to overcome the limitations of uni-modal models. One of the best multi-modal models for breast cancer prognosis prediction is **Multi-modal Deep Neural Network by integrating Multi-dimensional Data (MDNNMD)** [13] proposed by *Sun et al.*. It uses score level fusion in which values of three coefficients, α, β and γ are decided manually. These coefficients are weight coefficients used to balance the contributions of Deep Neural Network - Experimental (DNN-Expr), Deep Neural Network - Copy Number Alteration (DNN-CNA) and Deep Neural Network - Clinical (DNN-Clinical) in score level fusion layer of the MDNNMD model, respectively. The final output of MDNNMD is given as follows:

$$O_{DNNMD}=\alpha * O_{DNN-Expr} + \beta * O_{DNN-CNA} + \gamma * O_{DNN-Clinical}$$

$$s.t. \quad \alpha + \beta + \gamma = 1, \alpha \geqslant 0, \beta \geqslant 0, \gamma \geqslant 0$$

$\alpha = 0.3$ $\beta = 0.1$ and $\gamma = 0.6$ have been finalized for METABRIC dataset by changing these values manually over a range with an increment factor of 0.1. The manual process for finding these coefficients generates a huge number of combinations for training the model again and again. Apart from this manual approach, MDNNMD uses simple machine learning technique which is a deep neural network (DNN) for feature extraction from an individual data source. A convolutional neural network (CNN) may perform in a better way for feature extraction from each modality of the data source. In general, CNNs are capable of identifying hidden features present in a data set [22]. While MDNNMD uses a simple neural network for fusing all the three modalities and for predicting the final output, our proposed model uses a stacked ensemble technique. Literature survey shows that stacked ensemble techniques, in general, outperform any simple machine learning architecture in terms of predictive performance [23]. Our proposed model completely removes this manual effort of setting the coefficients for the fusion of multiple modalities and also uses CNN for feature extraction from the individual data source. These changes

result in better predictive measures as compared to existing models for breast cancer prognosis prediction.

1.3 Motivation for the Proposed Model

Inspired by the multi-modal deep neural network by integrating multi-dimensional data (MDNNMD) proposed by Sun et al. which uses three different modalities of data (gene expression profile, copy number alteration (CNA) profile and clinical data) with a score level fusion to achieve the best accuracy in breast cancer prognosis prediction [13], we propose a stacked-based approach which uses all the three modalities of data along with a stacked-based ensemble network. The motivation for using the stacked-based approach arises with diving deep into the concepts of the stacked model. Stacked models improve predictive performance by decreasing variance and bias. In this approach, we use CNN [22] [24] for extracting the features from all of the different modalities and these features feed to the stacked layer, which itself is a machine learning model. The stacked layer implements ensemble machine learning method like the random forest (RF). The motivation of using an ensemble method comes from the fact that it performs well on imbalanced data [25] and the dataset available for the current study is imbalanced. Another motivation for developing our proposed model is to remove the manual determination of the score level fusion coefficients in MDNNMD and to improve the predictive performance of the model by a large extent. Here, we have used CNN over the deep neural network for feature extraction because CNN produces more hidden features from any input data as compared to deep neural network and results in a more expressive stacked feature set for the stacked layer of the proposed model. Ten-fold cross-validation has been performed to get the final results, and we have obtained overall better performance as compared to existing models like MDNNMD, SVM, RF, logistic regression (LR). We conclude this study by showing 8.5% improvement in AUC (Area under the curve) value and 7.6% improvement in accuracy of breast cancer prognosis prediction as compared to MDNNMD [13].

2 Datasets and Methods

2.1 Datasets

We have directly use the dataset available on Github: https://github.com/USTC-HIlab/MDNNMD. It is the pre-processed version of METABRIC [1] dataset.

TABLE 1
THE OVERALL INFORMATION OF METABRIC BREAST CANCER DATASET

Cut-off (years)	5
Total # of patients	1980
Long-time survivors	1489
Short-time survivors	491
Median age in diagnosis	61
Average survival (months)	125.1

The dataset is having 1,980 valid breast cancer patient's data of METABRIC trial [26]. Each patient has multi-modal data of breast cancer such as the gene expression profile, CNA profile and clinical data. The patients have been categorized into long-time survivors (more than 5-year survival) and short-time survivors (less than 5-year survival) with each category is having 1489 and 491 patients, respectively. The median age at diagnosis is 61 years and average survival time of all patients is 125.1 months. The long-time survivors are labelled as 0, and short time survivors are labelled as 1 for this binary classification model. The overall description of the dataset is given in Table 1. Weighted nearest neighbour algorithm [27] is used for estimating the missing values of gene expression profile data and CNA profile data. *Sun et al.* have normalized and further discretized the gene expression features into three categories: under-expression(-1), over-expression (1) and baseline (0) according to *Gevaert et al.* [6]. Clinical data is also normalized using min-max normalization [7] in the range [0,1]. The CNA features are used as available with five discrete values (-2,-1,0,1,2). **Out of 1980 patients, 64 patients are alive and censored (i.e., lost to follow up within five years), which is 3.23% of the total samples. Here, we can not decide whether the patient died within five years (short-time survivor) or after five years (long-time survivors). Considering the small size of data and very high survival estimate (90% to 99%) provided by METABRIC for these 64 patients, we assumed these patients as long-time survivors for our study.**

2.2 Feature Selection

Gene expression profile, CNA profile and clinical data in METABRIC dataset have approximately 24000, 26000 and 27 features, respectively, for each patient and our dataset, consists of 1980 patients. This high dimensionality and low sample size may produce bad results for deep learning methods [28]. Hence, *Sun et al.* [13] have used the well-known feature selection algorithm mRMR [29] [30] [31] for dimensionality reduction of our dataset. Final selection of features has been made by evaluating the AUC values of different sets of features. *Sun et al.* have used an incremental approach for feature selection. They executed the mRMR algorithm and selected the top 100 features and tested the model performance on the selected features. Then, they selected the top 200 features and so on till 500 features. In this way, they searched the best N features from 100 to 500 with a step size of 100 by evaluating the AUC values at each step. Finally, 400 genes from gene expression profile data, 200 genes from CNA profile data with the highest AUC value and 25 clinical features from clinical data are selected as features for the existing MDNNMD model [13] and our proposed stacked-based ensemble model. The detailed information of selected features is presented in Table 2.

TABLE 2
SELECTED FEATURES FOR THE MODEL

Data Category	Total Features	Selected Features
Clinical	27	25
Gene Expression	24368	400
CNA	26298	200

The clinical features consist of age at diagnosis, size, lymph nodes positive, grade, inferred menopausal state,

1. https://www.cbioportal.org/study/summary?id=brca_metabric

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3018467, IEEE/ACM Transactions on Computational Biology and Bioinformatics

type of breast surgery, type of therapy the patient has gone through (Chemotherapy, Hormone Therapy and Radio Therapy) and some pathological details of genes and tumours.

2.3 A Convolutional Neural Network Based Prediction Model For Uni-modal Dataset

The proposed stacked-based model uses a convolutional neural network (CNN) to predict the prognosis of human breast cancer and extract the features for the next stage of the model. A CNN takes inputs from the individual modality of data, and each input data passes through convolution layers having a certain number of filters or kernels. It produces a feature map which is the result of the convolution process. This feature map is the simple element-wise multiplication followed by addition between the filter matrix and corresponding values of the input matrix. The values of filter matrix are initialized using glorot normal initializer [32], which selects random numbers with mean equals to zero and standard deviation in the range of $[-\sqrt{\frac{2}{n_i+n_o}}, \sqrt{\frac{2}{n_i+n_o}}]$. Here n_i, n_o represent the number of input and output units for the selected layer, respectively. Convolution layer also uses biases which are initialized with constant value of 0.1. In this layer, we have used stride value 2 for shifting the filter by the specified number to perform convolution over entire input matrix. The convolution layer also provides the flexibility to control the size of the feature map using padding. With the help of padding, this CNN keeps the size of our feature map same as input data shape. Further, flatten layer flattens the output of convolution layer and then passes it through a fully connected dense layer having 150 hidden units. An output layer with sigmoid activation function follows the dense layer and produces the predicted class of breast cancer prognosis. Rectified linear (ReLU) and hyperbolic tangent (TANH) are used as activation functions to the convolution layer and dense layer, respectively. Glorot normal initializer is also used to initialize weights between hidden layers. The detailed architecture of CNN is shown in Fig. 1.

TABLE 3
DETAILED PARAMETER CONFIGURATION OF CNN MODEL

# of convolution layers	1
Filter size of convolution layer	15
# of filters in convolution layer	4
Stride size in convolution layer	2
Padding in convolution layer	same
Activation function	ReLU
# of hidden layers	1
# of hidden units in hidden layer	150
Mini-batch size	8
Training Epoch	20
Activation function	TANH
Loss function	binary cross-entropy + L2 regularization

Since breast cancer prognosis prediction is a binary classification task (classification of short-time survivor and long-time survivor), we have used binary cross-entropy as a loss function in our CNN model. L2 regularization is used with the loss function to prevent the model from over-fitting. L2 regularization is widely used regularization

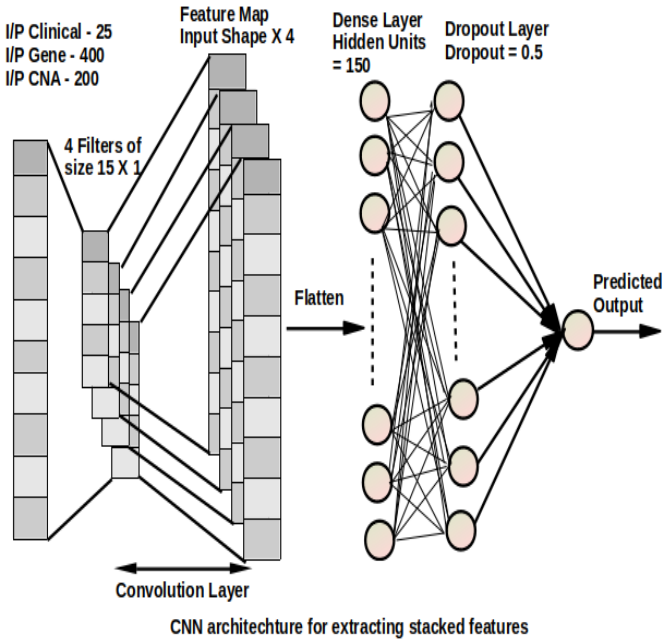

Fig. 1. CNN FOR UNI-MODAL DATASET

technique in deep learning [33] [34]. The loss function is given by

$$L(y_t,\hat{y_t}) = \frac{-1}{N}\sum_{i=0}^{N}[y_t(i)\log\hat{y}_t(i)-(1-y_t(i))\log(1-\hat{y}_t(i))] + \frac{1}{2}\lambda\sum_{k=1}^{K}\sum_{j=1}^{n_k}\sum_{i=1}^{m_k} {w_{ij}}^{k^2} \quad (1)$$

where L is the cost function that combines the cross-entropy and the regularization term. y_t and $\hat{y_t}$ are the actual label and predicted label, respectively. N is the batch size. $W^k = \left(w_{ij}^k\right)_{m_k \times n_k}$ is the k_{th} weight matrix and K is the number of weight matrices in the CNN model.

Finally, our CNN model has one input layer, one convolution layer, one flatten layer, one fully connected dense layer, and one output layer. The CNN architecture is not too complicated because a complex CNN on the tiny dataset can overfit. In the case of gene expression and CNA profile, our CNN uses additional dropout layer of 50% dropout [35]. AUC value of the model is evaluated at different mini-batch sizes ranging from 8 to 128, and the optimal value of 8 is selected, which resulted in the highest AUC value. Other hyper-parameters of the model have been optimized by setting different values, and the optimal value is selected from the model having the best AUC value. The detailed parameter configurations of CNN are provided in Table 3. The learning rate of our model is 10^{-3}.

2.4 Stacked-based Prediction Model For Multi-modal Data

The proposed framework is divided into two stages. These stages are described below:

- **Stage One:** Considering the fact that different modalities of data may have different feature representations and directly combining the multiple sources of data as an input to any deep neural model may not

be efficient [10], the stacked-based ensemble model uses separate CNN for each uni-modal data (clinical, gene expression and CNA) in stage one of the model. Then, we extract the features from the hidden layer of the CNNs and integrate these features all together to form the stacked features.

- **Stage Two:** The stacked features are the input for stage two of the stacked-based ensemble model. Different machine learning classifiers are utilized for final classification considering the stacked feature set. Note that the proposed framework is very generic; any machine learning classifier can be utilized for the final binary classification.

Fig. 2 depicts the architecture of the final proposed model. Stacked-based model is implemented using Keras 2.2.1 with TensorFlow 1.12 as backend.

Following are the stepwise details of the proposed stacked-based ensemble architecture for breast cancer prognosis prediction:-

- **Stage One**
  - We use the METABRIC dataset, which is already preprocessed in work done by *Sun et al.* [13].
  - We design three different CNNs for all the three modalities of data. These CNNs are termed as CNN-Clinical, CNN-Expr, and CNN-CNA for clinical, gene expression and CNA dataset, respectively.
  - The CNNs are trained on each uni-modal dataset individually by setting the AUC value as the performance measurement criteria.
  - We extract either hidden features or the predicted output from each of the trained CNN and concatenate these features to form the stacked feature set.
- **Stage Two**
  - We pass stacked features extracted from stage one as the input to stage two of the model.
  - SVM (support vector machine), RF (random forest), NB(Naive Bayes) and LR(Logistic Regression) are used in stage two of the model one by one and performance is measured for each one of them.
  - The predicted output from machine learning classifier is the final output of our proposed model.

3 Experimental Design

To overcome the variance problem raised due to the limited size of the dataset, we performed ten-fold cross-validation to evaluate the proposed model. The 1980 patients in our dataset are randomly divided into 10 subsets, out of which 9 subsets are merged and considered as the training set while remaining one subset is regarded as the testing set one after the other. The further training set is divided into 80% training set and 20% validation set. In our model, stage one and stage two do not execute simultaneously. Firstly, we train individual CNN for each modality of data and fix the optimal parameters by using AUC value as the criterion. Secondly, we extract the outputs from either the hidden layer or output layer of each CNN and combine them to form the stacked feature set. At last, these stacked features are passed to stage two of the stacked-based model, which itself is a machine learning model such as SVM, RF, Naive Bayes and LR. For performance evaluation, we have used ROC curve (Receiver Operating Characteristics curve) [36] which is plotted between false positive rate (1-Specificity) vs true positive rate (Sensitivity) by varying the decision threshold. With the help of the ROC curve, we calculate AUC value as a measure of model efficiency. The evaluation metrics, sensitivity (Sn) [36], specificity (Sp) [36], accuracy (Acc) [36], precision (Pre) [36] and matthew's correlation coefficient (Mcc) [36] are also used for performance evaluation and are defined in the following equations:

$$Sn = \frac{TP}{TP+FN} \tag{2}$$

$$Sp = \frac{TN}{TN+FP} \tag{3}$$

$$Pre = \frac{TP}{TP+FP} \tag{4}$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN) \times (TP+FP) \times (TN+FN) \times (TN+FP)}} \tag{6}$$

where TP, FP, TN and FN stand for true positive, false positive, true negative and false negative in a confusion matrix, respectively.

3.1 Other Prediction Methods For Comparison

Before finalising the architecture in this work, we have implemented multiple approaches to obtain the best model for breast cancer prognosis prediction. The very first approach used is motivated by the MDNNMD work where *Sun et al.* have used different deep neural networks (DNNs) for each modality of data. We have also trained three separate DNNs like DNN-Clinical, DNN-CNA and DNN-Expr having the same configuration as proposed by *Sun et al.* [13]. Later, we extract stacked features from DNNs in multiple ways and pass these features to different machine learning classifiers for final prediction. The approaches are listed below:-

- The weights from the hidden layer (the layer before the output layer) of each DNN model are extracted and concatenated together to form stacked features.
- The predicted outputs of different DNN models are concatenated together to form stacked features.
- The weights from the hidden layer (the layer before the output layer) of each DNN model are multiplied by their respective AUC values and then concatenated together to form stacked features.
- The predicted outputs of different DNN models are multiplied by their respective AUC values and then concatenated together to form stacked features.

The comparative study of the above approaches is presented in Table 4 of the results section.

To support the statement that CNN is the excellent choice for stacked feature extraction instead of DNN, we

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3018467, IEEE/ACM Transactions on Computational Biology and Bioinformatics

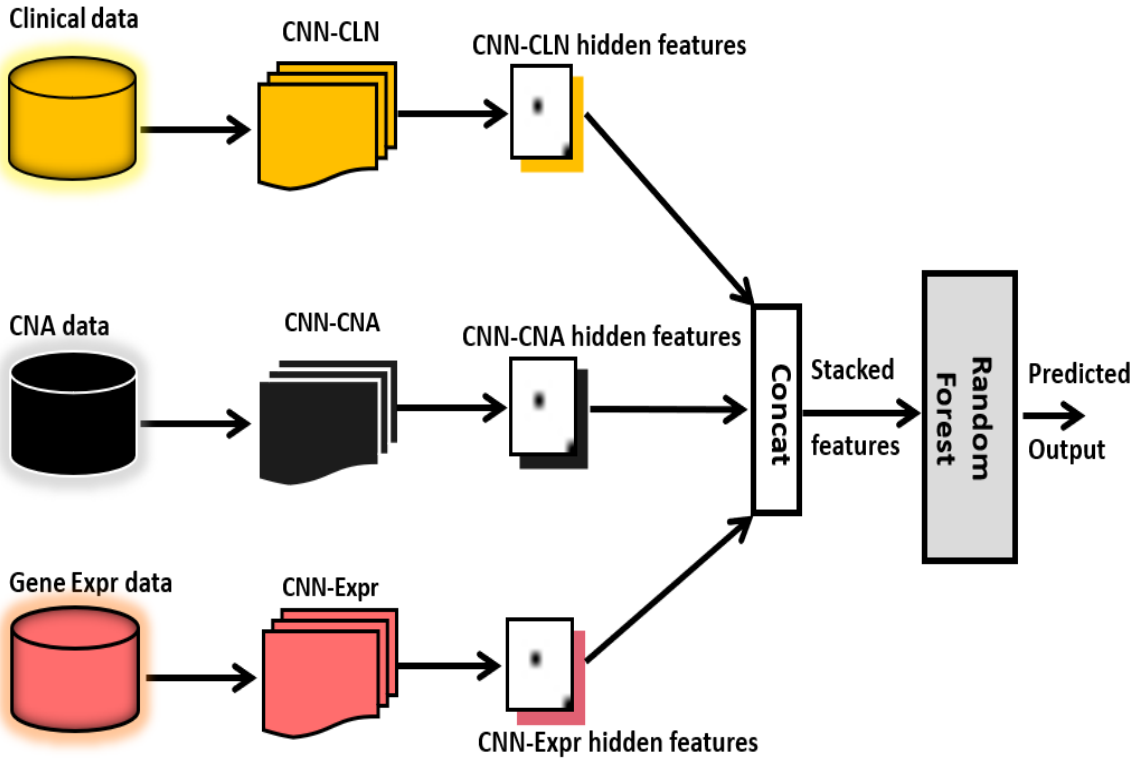

Fig. 2. STACKED-BASED ENSEMBLE MODEL ARCHITECTURE FOR BREAST CANCER PROGNOSIS PREDICTION

have compared the performances of CNN-Clinical, CNN-Expr and CNN-CNA with those by DNN based methods [13] such as DNN-Clinical, DNN-Expr and DNN-CNA, respectively. To validate the superiority of stacked-based ensemble model for breast cancer prognosis prediction, the AUC values of CNN-clinical, CNN-Expr and CNN-CNA are compared with the AUC values of our proposed model. We have also performed a comparative study of stacked-based ensemble model by using two different approaches in terms of stacked feature selection. In the first approach, predicted outputs of CNNs are selected as stacked features, while hidden features are selected as stacked features in the second approach. The stage two of the stacked model is also varied and uses either SVM, RF, Naive Bayes and LR as the underlying machine learning classifier.

To show the power of the proposed stacked model in predicting breast cancer prognosis, we have used results [13] from some of the popular methods such as MDNNDM [13], SVM [8], RF [9] and logistic regression (LR) [37] for comparison.

4 Results

In this section, we have discussed results of different methods.

4.1 Comparison of DNN Based Stacked Methods and MDNNMD

On the way to finalize CNN models as final feature extraction models, we have designed stacked-based models with stacked features extracted from DNNs. AUC values under ROC curves are used as the primary criterion to decide the predictive performance of different feature extraction models. If a model has lesser AUC value, this implies that the model is not good enough in comparison with the model having higher AUC value. We have also evaluated these models using accuracy, precision, sensitivity and matthew's correlation coefficient as secondary criteria. The AUC values and other performance metrics of DNN based stacked models and MDNNMD model are summarised in Table 4. Here, STACKED SVM (Hidden features) and STACKED SVM (Output features), STACKED RF (Hidden features) and STACKED RF (Output features) represent models having SVM and RF at stage two of the stacked-based approach defined in section 2.4, respectively. STACKED SVM (Hidden features) $\times$ AUC, STACKED SVM (Output features) $\times$ AUC, STACKED RF (Hidden features) $\times$ AUC and STACKED RF (Output features) $\times$ AUC are same as defined above but the only difference is that the stacked features are multiplied by the AUC values of corresponding DNN models and then passed to stage two of the stacked-based model.

It is evident from the results that DNN based stacked ensemble methods have improved the breast cancer prognosis prediction as compared to previous work, MDNNMD. STACKED RF (Hidden features x AUC) has achieved an AUC value of 0.855, which is approximately equal to AUC value achieved by MDNNMD. STACKED SVM (Output features) has shown better performance with $Acc = 0.851$, $Sn = 0.631$ and $Mcc = 0.584$ which are 2.5%, 18.1% and 9.8% higher in comparison with MDNNMD, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3018467, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 4
COMPARISON OF PERFORMANCE METRICS OF STACKED MODELS HAVING DNN AS STACKED FEATURE EXTRACTOR

Model	AUC	Acc	Pre	Sn	Mcc
Sp = 95.0%					
MDNNMD	0.845	0.826	0.749	0.450	0.486
STACKED SVM (Hidden features)	0.753	0.832	0.688	0.594	0.532
STACKED SVM (Hidden features x AUC)	0.761	0.841	0.714	0.600	0.554
STACKED SVM (Output features)	0.777	**0.851**	0.731	**0.631**	**0.584**
STACKED SVM (Output features x AUC)	0.763	0.835	0.687	0.619	0.545
STACKED RF (Hidden features)	0.850	0.840	0.771	0.509	0.536
STACKED RF (Hidden features x AUC)	**0.855**	0.840	0.773	0.507	0.536
STACKED RF (Output features)	0.837	0.841	**0.773**	0.509	0.537
STACKED RF (Output features x AUC)	0.834	0.836	0.763	0.492	0.521

4.2 Comparison of CNN Based Methods with DNN Based Methods on Uni-modal Data

To confirm the effectiveness of CNN over DNN in breast cancer prognosis prediction, we have compared the performances of CNN vs DNN with the help of ROC curves in Fig. 3 and different performance metrics of CNN-Clinical, CNN-Expr, CNN-CNA, DNN-Clinical, DNN-Expr and DNN-CNA in Table 5.

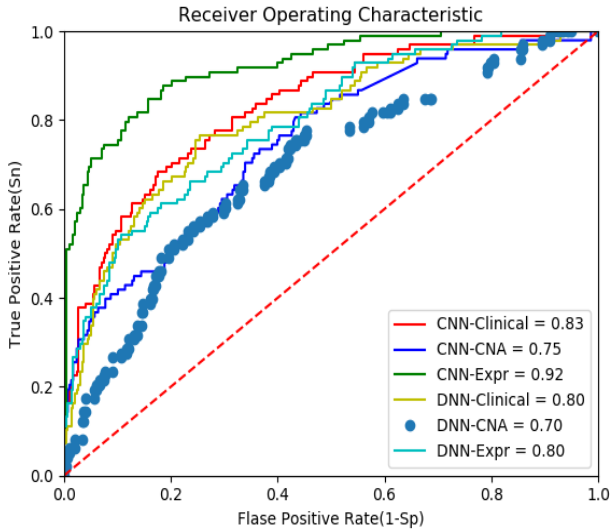

Fig. 3. ROC CURVES AND AUC VALUES OF CNN AND DNN USED FOR STACKED FEATURE SELECTION FROM EACH MODALITY OF DATA

From Table 5, we can see that CNN-Clinical, CNN-Expr, CNN-CNA attain 3%, 12%, 5% improvements in AUC values as compared to DNN-Clinical, DNN-Expr, DNN-CNA models, respectively. Hence, hidden features or outputs of CNNs are selected as the stacked feature set for stacked-based model. Other performance metrics of CNNs also show improvements as compared to DNNs.

TABLE 5
COMPARISON OF PERFORMANCE METRICS OF CNN & DNN USED FOR STACKED FEATURE SELECTION FROM EACH MODALITY OF DATA

Model	AUC	Acc	Pre	Sn	Mcc
Sp = 95.0%					
CNN-Clinical	**0.830**	**0.806**	**0.658**	**0.349**	**0.376**
DNN-Clinical	**0.800**	**0.802**	**0.620**	**0.314**	**0.366**
CNN-Expr	**0.920**	**0.806**	**0.596**	**0.430**	**0.382**
DNN-Expr	0.800	0.759	0.320	0.135	0.086
CNN-CNA	0.750	0.748	0.409	0.202	0.119
DNN-CNA	0.700	0.757	0.258	0.113	0.070

4.3 Effect of Stacked Feature Selection

We compare the performances of stacked-based ensemble method designed using two different ways of stacked feature selection. In the first approach, predicted outputs of individual CNNs are combined and passed to the second stage as features for breast cancer prognosis prediction. In the second approach, the hidden layer features of individual CNNs are combined and given to the second stage as features. The AUC values and other performance metrics are measured and presented in Table 6. Here, STACKED RF (Output features), STACKED RF (Hidden features), STACKED SVM (Output features), STACKED SVM (Hidden features), STACKED Naive Bayes (Hidden features), STACKED Naive Bayes (Output features), STACKED LR (Hidden features) and STACKED LR (Output features) are similar to the models defined in section 2.4. We have selected predicted outputs or hidden layer features of individual uni-modal CNNs and concatenated them as stacked features.

TABLE 6
PERFORMANCE METRICS OF STACKED MODEL RF/SVM/NB/LR WITH FEATURES EXTRACTED FROM HIDDEN LAYER OR OUTPUT LAYER OF CNNs

Sp = 95%					
Model	**AUC**	**Acc**	**Pre**	**Sn**	**Mcc**
STACKED RF (Output features)	0.890	0.880	0.814	0.672	0.665
STACKED RF (Hidden features)	**0.930**	**0.902**	**0.841**	**0.747**	**0.730**
STACKED SVM (Output features)	0.800	0.884	0.835	0.653	0.674
STACKED SVM (Hidden features)	0.810	0.873	0.762	0.710	0.674
STACKED (Naive Bayes) (Output features)	0.909	0.876	0.813	0.649	0.630
STACKED (Naive Bayes) (Hidden features)	0.873	0.866	0.742	0.708	0.612
STACKED (LR) (Output features)	0.921	0.886	0.830	0.680	0.677
STACKED (LR) (Hidden features)	0.782	0.806	0.631	0.529	0.461

After analysing the results reported in table 6, we select hidden layer features as the stacked feature set for our proposed model. Stacked SVM does not perform well as compared to stacked RF. Note that the random forest is an ensemble classifier. In general, it has been shown in the literature that ensemble classifier performs better than the individual classifier on the imbalanced dataset. SVM is a strong classifier. However, as in RF, the prediction capabilities of multiple classifiers are combined, it performs better

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3018467, IEEE/ACM Transactions on Computational Biology and Bioinformatics

than SVM. So, we have selected STACKED RF(Hidden features) with features extracted from CNNs as our proposed architecture.

4.4 Comparison of CNN Based Methods with Uni-modal Data and Stacked-based RF Method with Multi-modal Data

Stacked-based ensemble method with multi-modal data achieves overall better performance than those of uni-modal CNN methods. The AUC values of the methods are presented in Table 7.

TABLE 7
COMPARISON OF PERFORMANCE METRICS OF CNNs & STACKED (RF)

Model	AUC	Acc	Pre	Sn	Mcc
Sp = 95.0%					
STACKED RF	**0.930**	**0.902**	**0.841**	**0.747**	**0.730**
CNN-Clinical	0.830	0.806	0.658	0.349	0.376
CNN-Expr	0.920	0.806	0.596	0.430	0.382
CNN-CNA	0.750	0.748	0.409	0.202	0.119

From the above table, it is clear that stacked-based RF model outperforms the CNNs with uni-modal data. The AUC value of the stacked model is **0.93**, which is 10%, 1% and 18% higher than CNN-Clinical, CNN-Expr and CNN-CNA models, respectively.Other performance metrics of stacked-based RF model also have higher values in comparison with uni-modal CNNs.

4.5 Comparison with Other Prediction Methods

We have compared the stacked-based ensemble model with four popular methods for breast cancer prognosis prediction: MDNNMD [13], SVM [8], RF [9] and LR [37]. We have calculated the AUC values under the ROC curves for stacked-based ensemble method, MDNNMD, SVM, RF and LR. Stacked-based ensemble method achieves overall better AUC value as compared to other popular prediction methods. The AUC value of stacked-based ensemble method is 0.930 while the corresponding AUC values of MDNNMD, SVM, RF and LR are 0.845, 0.810, 0.801 and 0.663, respectively. The ROC curve of STACKED RF (Hidden features) is plotted in Fig. 4.

Some additional performance measures, *Acc, Pre, Sn* and *Mcc*, are used for comparison of stacked-based ensemble model with other popular methods. Two stringency levels: high (*Sp* = 99% with corresponding threshold 0.72) and medium (*Sp* = 95% with corresponding threshold 0.45) are used to compare the models [38]. The results are shown in Table 8. From the comparative study presented in this table, we can state that stacked-based ensemble method is better than all other methods for breast cancer prognosis prediction. For example, when *Sp* = 95%, the corresponding *Sn* values obtained by STACKED RF (Hidden features), MDNNMD, SVM, RF and LR are 0.747, 0.450, 0.365, 0.226 and 0.183, respectively. Precision value of STACKED RF (Hidden features) at *Sp* = 95% is 0.841, which is 9.2%, 13.3%, 7.5% and 2.9% higher than those obtained by MDNNMD, SVM, RF and LR, respectively. When *Sp* changes from 95% to 99%, the *Acc, Pre, Sn* and *Mcc* values increase by

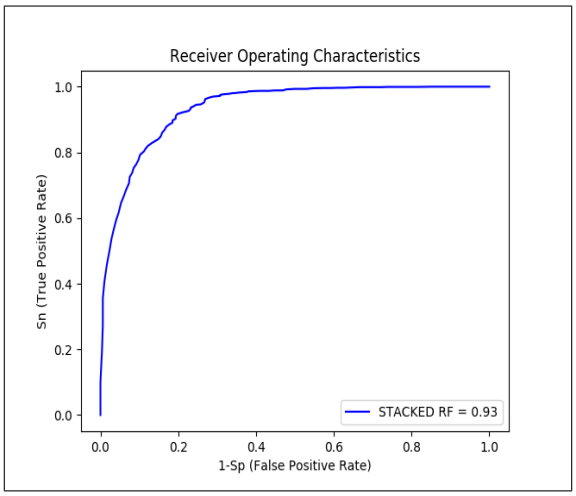

Fig. 4. ROC CURVE FOR THE **STACKED RF** MODEL HAVING **CNN** FOR STACKED FEATURE SELECTION AND **RF** FOR FINAL PREDICTION

7.7%, 9.1%, 30%, and 27.7%, respectively, as compared to MDNNMD [13]; improve by 9.6%, 15.5%, 38.1% and 37.6%, respectively, as compared to SVM [8]; improve by 10.1%, 17.9%, 40.5% and 41.0%, respectively, as compared to RF [9]; and have improvements of 11.7%, 40.3%, 46.6% and 54.0%, respectively, as compared to LR [37]. After going through all the above analysis, we conclude that STACKED RF (Hidden features) method with stacked features extracted from hidden layers of CNNs achieves better performance compared to all other methods for breast cancer prognosis prediction.

In order to show the performance of the proposed approach on true gold standard training data, we removed the 64 censored samples (as discussed in section 2.1) from our data and observed the performance of the proposed model on updated data. At $Sp = 95\%$ **&** $Threshold = 0.45$ **the outcomes are as follows :** $AUC = 0.926, Acc = 0.884, Pre = 0.819, Sn = 0.684$ **and** $Mcc = 0.676$

TABLE 8
COMPARISON OF PERFORMANCE METRICS OF STACKED RF (Hidden features), MDNNMD, SVM, RF & LR MODELS

Model	Acc	Pre	Sn	Mcc
Sp = 99.0% (threshold = 0.72)				
STACKED RF (Hidden features)	**0.870**	**0.957**	**0.501**	**0.622**
MDNNMD	0.794	0.875	0.200	0.356
SVM	0.775	0.811	0.122	0.257
RF	0.770	0.787	0.098	0.223
LR	0.754	0.563	0.037	0.093
Sp = 95.0% (threshold = 0.45)				
STACKED RF (Hidden features)	**0.902**	**0.841**	**0.747**	**0.730**
MDNNMD	0.826	0.749	0.450	0.486
SVM	0.805	0.708	0.365	0.407
RF	0.791	0.766	0.226	0.337
LR	0.760	0.549	0.183	0.209

4.6 Validation

To validate our proposed approach's performance, we have used breast cancer dataset from The Cancer Genome Atlas (TCGA) project [11]. The dataset is named as TCGA-BRCA, and it consists of gene expression data, CNA data and

clinical details of 752 patients. We have removed the male patient's data to preserve the unbiasedness in the dataset. The preprocessing and feature selection steps are similar to METABRIC dataset. Here, we have classified our patients into two classes, *dead* as 1 and *alive* as 0. The preprocessed dataset is passed to our proposed model via 10-fold cross-validation. The performance metrics of our proposed model and other existing methods are presented in Table 9. For

TABLE 9
COMPARISON OF PERFORMANCE METRICS OF STACKED RF (Hidden features), MDNNMD, SVM, RF, LR & CNN MODELS ON TCGA-BRCA DATASET

Model	AUC	Acc	Pre	Sn
STACKED RF (Hidden features)	**0.968**	**0.881**	**0.949**	**0.950**
CNN-Clinical	0.761	0.843	0.116	0.047
CNN-Expr	0.941	0.837	0.317	0.360
CNN-CNA	0.811	0.848	0.278	0.158
SVM	0.709	0.840	0.846	0.844
RF	0.835	0.843	0.713	0.843
LR	0.683	0.735	0.788	0.721

TCGA-BRCA dataset, stacked-based ensemble framework is better than CNN-Clinical, CNN-Expr, CNN-CNA, SVM, RF and LR. The AUC value of our model is 20.7%, 2.7%, 15.7%, 25.9%, 13.3% and 28.5% higher than those of CNN-Clinical, CNN-Expr, CNN-CNA, SVM, RF and LR, respectively.

4.7 Statistical Significance Test

To confirm the statistical significance of the predicted results for breast cancer prognosis by our proposed stacked-based ensemble model, we have performed t-test over the performance metrics of MDNNMD and our proposed model. t-test follows the Student's t-distribution and is applicable for a small-sized sample. So, we have used performance metric *Acc* of MDNNMD and our proposed model over METABRIC dataset of 1980 patients in our t-test. **t-value** and **p-value** for the performed t-test are **-14.93** and **0.00**, respectively. **The differences in performances by the proposed method and existing approaches are statistically significant, which imply that results are statistically significant.** We have also performed the t-test on validation dataset (TCGA-BRCA), the t-value and p-value are **-109.54** and **0.00**, respectively. For this test, we have used inbuilt library function $stats.ttest\_ind$ present within scipy library.

5 Discussion and Conclusion

Breast cancer is the most common disease and the reason for several million deaths worldwide. Its prognosis is also weak because of the gene signature complexity and other factors of breast cancer. So, it is imperative to design a fast and effective model for breast cancer prognosis prediction. In this work, we have developed a stacked-based ensemble machine learning model with the help of CNN and RF to predict the life expectancy of breast cancer patients. This model uses multi-modal inputs such as gene expression, copy number alteration and clinical data. Three independent CNNs are designed for different sources of input, and hidden layer features are extracted from each of the CNNs. Features extracted from CNNs are passed through a random forest classifier for the final prediction of the survival of breast cancer patients. This model achieves better performance than other existing prediction methods based on uni-modality as well as multi-modality. This architecture can also be applied for other similar diseases which require multi-modal inputs.

Even though our proposed model has outperformed all existing prediction models, we need to validate this model over different breast cancer data sets. The METABRIC dataset used in this study has only 1980 samples which are very small for the machine learning study. So, we can expect more improvement in the results if large dataset would be available in future. Apart from this, we can also extend our model by integrating images of breast cancer tissues as an additional modality in the dataset. Some other modalities can also be added to the dataset such as gene methylation, miRNA expression values.

Acknowledgment

Dr. Sriparna Saha would like to acknowledge the support of SERB Women in Engineering Award 2017 (SB/WEA-08/2017) for conducting this current research.

References

[1] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. Sharma, "Various types and management of breast cancer: an overview," *Journal of advanced pharmaceutical technology & research*, vol. 1, no. 2, p. 109, 2010.

[2] J. Ferlay, C. Héry, P. Autier, and R. Sankaranarayanan, *Global Burden of Breast Cancer*. New York, NY: Springer New York, 2010, pp. 1–19. [Online]. Available: https://doi.org/10.1007/978-1-4419-0685-4_1

[3] A. DORR, "Book reviews," *JNCI: Journal of the National Cancer Institute*, vol. 83, no. 14, pp. 1029–1030, 1991. [Online]. Available: http://dx.doi.org/10.1093/jnci/83.14.1029

[4] L. R. Martin, S. L. Williams, K. B. Haskard, and M. R. DiMatteo, "The challenge of patient adherence," *Therapeutics and clinical risk management*, vol. 1, no. 3, p. 189, 2005.

[5] Y. Sun, S. Goodison, J. Li, L. Liu, and W. Farmerie, "Improved breast cancer prognosis through the combination of clinical and genetic markers," *Bioinformatics*, vol. 23, no. 1, pp. 30–37, 2006.

[6] 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. e184–e190, 2006.

[7] 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, M. J. Marton *et al.*, "A gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.

[8] X. Xu, Y. Zhang, L. Zou, M. Wang, and A. Li, "A gene signature for breast cancer prognosis using support vector machine," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*. IEEE, 2012, pp. 928–931.

[9] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *Journal of Biomedical Science and Engineering*, vol. 6, no. 05, p. 551, 2013.

[10] M. Khademi and N. S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*. IEEE, 2015, pp. 727–732.

[11] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary oncology*, vol. 19, no. 1A, p. A68, 2015.

[12] J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson *et al.*, "Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal," *Sci. Signal.*, vol. 6, no. 269, pp. pl1–pl1, 2013.

[13] 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 Transactions on Computational Biology and Bioinformatics*, 2018.

[14] D. Sun, A. Li, B. Tang, and M. Wang, "Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome," *Computer Methods and Programs in Biomedicine*, vol. 161, 04 2018.

[15] Y. Zhang, A. Li, J. He, and M. Wang, "A novel mkl method for gbm prognosis prediction by integrating histopathological image and multi-omics data," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 02 2019.

[16] B. Tang, A. Li, B. Li, and M. Wang, "Capsurv: Capsule network for survival analysis with whole slide pathological images," *IEEE Access*, vol. PP, pp. 1–1, 02 2019.

[17] K. Sohn, W. Shang, and H. Lee, "Improved multimodal deep learning with variation of information," in *Advances in Neural Information Processing Systems*, 2014, pp. 2141–2149.

[18] A. K. Jain and A. Ross, "Multibiometric systems," *Commun. ACM*, vol. 47, no. 1, pp. 34–40, Jan. 2004. [Online]. Available: http://doi.acm.org/10.1145/962081.962102

[19] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689–696.

[20] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98 – 125, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566253517300738

[21] F. Wang and J. Han, "Multimodal biometric authentication based on score level fusion using support vector machine," *Opto-electronics review*, vol. 17, no. 1, pp. 59–64, 2009.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, June 2015, pp. 1–9. [Online]. Available: doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594

[23] C. Ju, A. Bibaut, and M. J. van der Laan, "The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification," *arXiv e-prints*, p. arXiv:1704.01664, Apr. 2017.

[24] M. Tan, S. Yuan, and Y. Su, "Content-based similar document image retrieval using fusion of cnn features," *Internet Multimedia Computing and Service*, p. 260–270, 2018. [Online]. Available: http://dx.doi.org/10.1007/978-981-10-8530-7_25

[25] T. M. Khoshgoftaar, M. Golawala, and J. V. Hulse, "An empirical study of learning from imbalanced data using random forest," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence - Volume 02*, ser. ICTAI '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 310–317. [Online]. Available: http://dx.doi.org/10.1109/ICTAI.2007.49

[26] C. Curtis, S. Shah, S.-F. Chin, G. Turashvili, O. Rueda, M. Dunning, D. Speed, A. Lynch, S. Samarajiwa, Y. Yuan, S. Graef, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowetz, L. Murphy, I. Ellis, A. Purushotham, A.-L. Borresen-Dale, J. Brenton, S. Tavare, C. Caldas, S. Aparicio, and METABRIC Grp, "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, no. 7403, pp. 346–352, 6 2012.

[27] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/17.6.520

[28] J. Tan, J. H. Hammond, D. A. Hogan, and C. S. Greene, "Adage analysis of publicly available gene expression data collections illuminates pseudomonas aeruginosa-host interactions," *bioRxiv*, 2015. [Online]. Available: https://www.biorxiv.org/content/early/2015/11/05/030650

[29] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.

[30] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, ser. CSB '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 523–. [Online]. Available: http://dl.acm.org/citation.cfm?id=937976.938050

[31] Y. Cai, T. Huang, L. Hu, X. Shi, L. Xie, and Y. Li, "Prediction of lysine ubiquitination with mrmr feature selection and analysis," *Amino Acids*, vol. 42, no. 4, pp. 1387–1395, Apr 2012. [Online]. Available: https://doi.org/10.1007/s00726-011-0835-0

[32] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterington, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. [Online]. Available: http://proceedings.mlr.press/v9/glorot10a.html

[33] 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," *Medical Image Analysis*, vol. 35, pp. 18 – 31, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841516300330

[34] 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 *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–3.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[36] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210832718301546

[37] M. Jefferson, N. Pendleton, S. Lucas, and M. Horan, "Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma," *Cancer*, vol. 79, no. 7, pp. 1338–1342, 4 1997.

[38] W. Fan, X. Xu, Y. Shen, H. Feng, A. Li, and M. Wang, "Prediction of protein kinase-specific phosphorylation sites in hierarchical structure using functional information and random forest," *Amino Acids*, vol. 46, no. 4, pp. 1069–1078, Apr 2014. [Online]. Available: https://doi.org/10.1007/s00726-014-1669-3

Nikhilanand Arya received the B.Tech degree in information technology from the Maulana Abul Kalam Azad University of Technology, West Bengal, India, in 2015. He worked at Mintree India as a software engineer till 2017. He is currently pursuing M.Tech degree at Indian Institute of Technology, Patna, India.

Dr. Sriparna Saha received the M.Tech and Ph.D. degrees in computer science from Indian Statistical Institute Kolkata, Kolkata, India, in 2005 and 2011, respectively. She is currently an Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. She is the author of a book published by Springer-Verlag. She has authored or coauthored more than 120 papers. Her current research interests include pattern recognition, multiobjective optimization and biomedical information extraction.