Received June 12, 2020, accepted June 21, 2020, date of publication June 26, 2020, date of current version July 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005228

Deep Learning Based Analysis of Breast Cancer Using Advanced Ensemble Classifier and Linear Discriminant Analysis

XINFENG ZHANG[1], DIANNING HE[2], YUE ZHENG[3], HUAIBI HUO[3], SIMIAO LI[3], RUIMEI CHAI[3], AND TING LIU[3]

[1]Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute Breast Surgery, Shenyang 110042, China
[2]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110001, China
[3]Department of Radiology, The First affiliated Hospital of China Medical University, Shenyang 110001, China

Corresponding author: Ting Liu (cmuliuting@sina.cn)

This work was supported by the Whole heart coronary plaque quantification and risk predictive model for plaque rupture based on deep learning neural network (NSFC 81871435).

ABSTRACT In the recent past, the Classifiers are based on genetic signatures in which many microarray studies are analyzed to predict medical results for cancer patients. However, the Signatures from different studies have been benefitted with low-intensity ratio during the classification of individual datasets has been considered as a significant point of research in the present scenario. Hence to overcome the above-discussed issue, this paper provides a Deep Learning Framework that combines an algorithm of necessary processing of Linear Discriminant Analysis (LDA) and Auto Encoder (AE) Neural Network to classify different features within the profile of gene expression. Hence, an advanced ensemble classification has been developed based on the Deep Learning (DL) algorithm to assess the clinical outcome of breast cancer. Furthermore, numerous independent breast cancer datasets and representations of the signature gene, including the primary method, have been evaluated for the optimization parameters. Finally, the experiment results show that the suggested deep learning frameworks achieve 98.27% accuracy than many other techniques such as genomic data and pathological images with multiple kernel learning (GPMKL), Multi-Layer Perception (MLP), Deep Learning Diagnosis (DLD), and Spatiotemporal Wavelet Kinetics (SWK).

INDEX TERMS Deep learning framework, breast cancer, advanced ensemble classifier, linear discriminant analysis.

I. INTRODUCTION

Breast cancer seems to be the most common type of cancer that women around the world, and it is found in developed countries after lung cancer [1]. In all over the world, 50% to 60% of cases of breast cancer occur in late stages, and patients have one of the lowest survival levels in the region [2], [3]. Hence there is a need to determine multiple factors that affect the survival rate of breast cancer patients.

Clinicians are using basic Software programs for analyzing factors influencing breast cancer survival rates. Such traditional statistical methods cannot be modified to identify new variables or create innovative and inclusive visualizations. Different approaches to machine learning (ML) [4] are used in this area as decision tree (DT) random forest (RF) approaches neural network, extreme boost, logistic reversal, and support vector machine (SVM) [5], [6]. The Decision Tree is a supervised learning algorithm representing the outcomes in a tree structure easily interpretable, where visualization is a significant factor in the analysis of large numbers of data. In both supervised and unsupervised mode [7], the Random forest (Breiman's algorithm) is a derivation of DT, which manages simultaneous and numerical data in classification or regression [8]. Neural networks are complex and are frequently represented as black boxes that model data through training with known results and maximizing weights for a better estimation in situations with undetermined results [9].

Gene Expression is an ensemble of classification and regression trees that can be paralleled and generated with successful prediction accuracy, which is simple to use in various machine learning functions [10]. Logistical regression assumes the distribution of Gaussian variables and

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

describes all types of variables, such as continuous, discrete, and dichotomous, without assuming normality. The support vector machine will be used for supervised classification and performs by identification of the best decision-making limit, which separates data points from different groups and then the prediction of new observations on the basis of that classification limits [11].

The machine learning ability is increasingly essential for the accuracy of clinical diagnostics. In addition to this, advanced state-of-the-art methods [12] with the technology of machine learning also provide the ability to extract information from statistically significant complex medical imaging data sets. Deep learning methods [13] are particularly successful in various medical image analysis tasks which have significant achievement in specific diagnostic tasks [14].

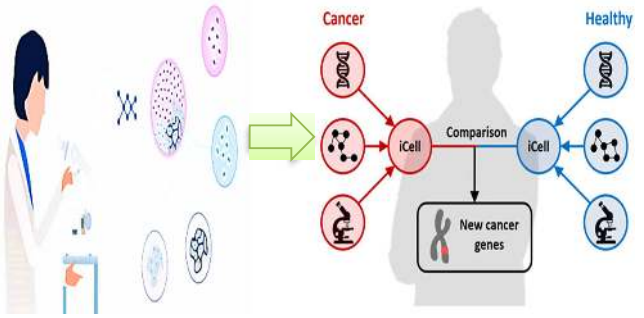

FIGURE 1. Deep learning for breast cancer identification.

Pathology has been used for digitizing histological samples correctly using a sub-micron resolution to perform a computerized analysis of the individual samples with machine learning algorithms (Figure 1) [15]. Mitosis detection methods, cell immune infiltration, and other tissue entities such as separation of tissue from epithelial or discrimination between are computerized. Recent reviews include extensive summaries of the methods for histological sample analysis [16].

The development of image classifications usually involves a specific type of deep learning process, convolutional neural networks (CNNs) [17]. CNN's consist of coherent, interconnected, and hierarchical steps— an architecture inspired by the organization of genetic neural networks that permit the visual collection of abstract patterns [18]. The use of CNNs in breast cancer metastases is recently detected in lymph node parts.

The purpose of this paper is to boost cancer prognosis prediction efficiency and to establish a more comprehensive classification of the outcomes. To achieve this (i), it introduces a more general learning approach by integrating feature selection and extraction with a few methods of deep learning. (ii) A boosting algorithm with an advanced ensemble classifier for the metastasis of distant cancer is being developed. (iii) This paper shows unsupervised feature learning and supervised classifications of learning in contrast with previous classification methods. The breast cancer images are taken from https://www.ncbi.nlm.nih.gov/gds.

II. LITERATURE REVIEW

Breast cancer is a significant cause of cancer death for women. The high mortality rate of breast cancer is primarily due to its prevalence and significantly different clinical outcomes in invasive breast cancer. Improving the accuracy of the diagnosis of breast cancer is, therefore critical and becomes one of the key areas of research. Currently, many mathematical models are proposed for the prediction of breast cancer survival. Still, the majority of them produce predictive models only using genomics, and few take into consideration the extra information from pathological images. Sun *et al.*, [19] developed a new GPMKL approach focused on the use of heterogeneous data that includes genomic (gene expression, modification of copy numbers, gene methylation, preserving appearance), as well as disease imaging based on multiple kernels of learning (MKL). GPMKL is proposed to execute the feature fusion that is incorporated in the breast cancer classification to the above heterogeneous features.

Qiu *et al.* [20] introduced a new deep-learning approach to the short-term risk assessment model. A total of 270 negative evaluation instances were collected during this analysis. 135 cases were positive and 135 cases remained negative in the next sequential screening mammography. These cases were divided into 200-cases randomly and a 70-cases test set. A computer-aided diagnostic framework for deep analysis was then developed for the risk assessment consisting of two modules: the adaptive module for defining the features and the risk prediction module. A multi-layer perception (MLP) classification framework is introduced in the prediction of risk to estimate the likelihood of the woman acquiring short-term mammographic-detectable cancer. The results showed that the new CAD risk model provided a positive predictive value of 69.2% and a negative predictive value of 74.2%, with the overall predictive accuracy of 71.4%.

Vandenberghe *et al.*, [11] developed a mathematical method that immediately measures HER2, a biomarker that defines the patient's eligibility in mammograms for anti-HER2 directed therapies. Computerized scoring demonstrated 83% concordance with the physician in a study of 71 breast tumor biopsy studies. The 12 discordant cases were then examined individually, and a diagnostic change was made for the initial evaluation of both the pathologist in eight cases. The diagnostic discrepancy was mainly caused by perceptive differences for the assessment of HER2 expression due to the high variability of HER2 staining. This study indicates that Deep Learning Diagnosis (DLD) can make clinical decisions on breast cancer easier by recognizing cases with a significant risk of misdiagnosis.

Cancer heterogeneity can affect therapy response and the prognosis of patients. Histological tests have historically been used to quantify heterogeneity, but both the essential recurrence risk and the response monitor need accurate, non-invasive measurements. In this regard, they proposed [21] to quantify the intratumor heterogeneity in breast cancer by applying Spatiotemporal Wavelet

Kinetics (SWK) from dynamical comparison-enhanced magnetic resonance imaging. The pharmaco-kinetic measures are first divided into homogeneous sub-regions.

The methods mentioned above still have a problem with accurate analysis of breast cancer. Hence, This paper provides a Deep Learning Framework that combines an algorithm of crucial processing and an autoencoding neural network to classify different features within the profile of gene expression for effective diagnosis of breast cancer are discussed as follows,

III. BACKGROUND OF THE STUDY

The Linear Discrimination (LDA) analysis and the Autoencoder neural network used to reduce complexity. The basic structure is described as follows

A. STUDY OF THE MAJOR COMPONENT

LDA is one of the most common methods for the reduction of linear dimensions, and it is considered to be a multivariate technique. It finds the significant elements in records that are unrelated constants using the eigenvalue matrix and vectors, each representing a specific variation of the data. Let $X = \{x_i\}_{i=1}^{m}$ mark the training data collection. x_i represents a D dimension variable, and it has the genomic patterns in this report. The goals of the LDA are: 1) To collect the main thing of data from x_m; 2) Compressing the X dimension by maintaining only necessary information. LDA aims to minimize the variance of the predicted data, as shown in figure 2. This means that orthogonal representations of the existing data are called on the new k-dimensional space.

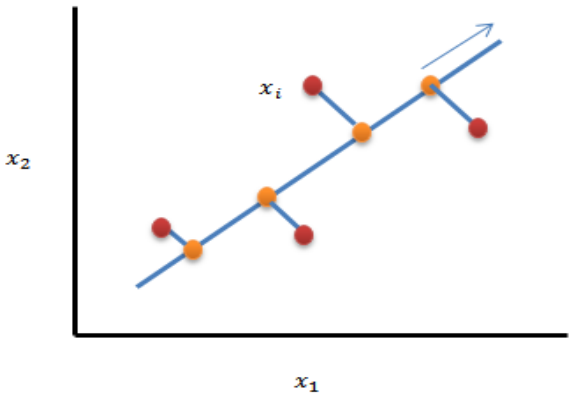

FIGURE 2. LDA process of orthogonal projection.

Taking the path of the image with a vector u_1 (dimensional D). Each x_i a data point is then transformed into a scalar value of $u_1^T x_i$. The average of the data predicted is equal to $u_1^T \overline{X}$, where $\overline{X}$ corresponds to the average sample range provided by $\overline{X} = \frac{1}{m}\sum_{i=1}^{m} x_i$. In addition, the variation of the results predicted will be: $\frac{1}{m}\sum_{i=1}^{m}\left(u_1^T x_i - u_1^T \overline{X}\right)^2 = u_1^T S u_1$, Where S belongs for all samples to a standard covariance matrix: $S = \frac{1}{m}\sum_{i=1}^{m}(x_i - \overline{X})(x_i - \overline{X})^T$. LDA now seeks to optimize the predicted variance $u_1^T S u_1$ in addition to u_1.

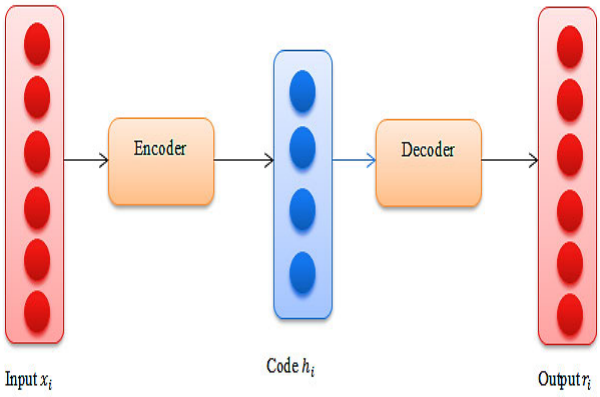

FIGURE 3. Autoencoder framework.

B. NEURAL NETWORK AUTOENCODER

As illustrated in figure 3, an Autoencoder is a feed-forward neural network that is often specialized in the processing of depictions or successful decoding of original data $X = \{x_i\}_{i=1}^{m}$. It allows the $g(f(x_i)) \approx r_i$ system to be understood as an indication for estimating and representing the input data produced from a finite amount of available activations. Then the overall Autoencoder architecture is broken down into the following parts: 1) The data tools x_i; 2) An encoder function f; 3) A hidden representation or ''code'' $h_i = f(x_i)$; 4) A role of decoder g; 5) Input modules called ''rebuilding'' $r_i = g(h_i) = g(f(x_i))$; 6) $\mathcal{L}(x_i, r_i)$ loss function Scalar $\|x_i - r_i\|_2$ computing that tests how successful the restoration r_i has originally from x_i data. The Autoencoder is designed to maximize $\mathcal{L}$ over Learning Examples X. The predicted values are reduced.

The Autoencoder is simpler, and the encoder functionality eliminates non-linear dimensions if the amount of the hidden layer is more significant than one. Authenticators are trained to identify the best input compression feature on these cached layers, where the measurements of the hidden layers are smaller than that of the data. Alternatively, the Autoencoder can be trained to map the feature to a larger space.

IV. METHODOLOGY

In this section, the proposed LDA and AE based deep learning framework have been discussed. Next, it defines five data sets for gene expression that uses a specific pre-processing method.

A. DATASETS AND PRE-PROCESSING OF GENE EXPRESSION

Information on gene expression has been downloaded from the GEO database. Every study includes 129,158 Genomic versions of the Affymetrix microarray platform, and each profile has *22,268* mutations, equivalent to *978* landmarks and *21,290* aim genes. Especially from the LINCS cloud, the alternative approach has tested in five different data sets for breast cancer. The information is given in-depth in Table 1.

TABLE 1. Samples have been removed if patients within 5 years are censored for treatment.

Records	Bad result	Better result	Final result	Samples were taken
GSE2990	84	154	238	12
GSE3494	32	101	133	125
GSE9195	25	87	112	186
GSE17705	34	168	202	15
GSE17907	27	189	216	61

Patients use various immune and physiological factors in the five samples to affect the prognosis of outcomes. For example, all ER-positive patients in GSE3494 include ER-positive and ER-negative patients in addition to other data sets.

In view of the classification mission, the pre-processing with the 5 data sets has carried out in two steps: The first argument is to follow the dataset partitioning method with a weak prognostic (set to 1) and reliable predictive (set to 0) division of all people with cancer. Data on aid has eliminated from the review of patients receiving adjuvant or screened for 5 years. In addition, it has been quantified the five datasets with a MAS 5.0 algorithm and converted all of the samples into an Expression gene ID, because the microarray models are used to calculate gene expression value.

B. GENERAL APPROACH

The aim is to integrate both feature selection and feature elimination with profound learning basics that it learns from genomic profiles quite concisely and establish a more accurate classification of cancer prognoses. The flowchart of the method is shown in Figure 4.

1) DEEP LEARNING ASSISTED UNSUPERVISED FEATURE LEARNING

Two stages are comprised of this deep learning approach are discussed as follows

- **LDA:** Considering that data on gene expression is extensive, containing repetitive, noisy data, the LDA program (as defined in section I-A) is utilized as the tool for selecting features to reduce genomic model complexity. LDA conducts a linear estimation of the existing data and, in the meantime, maintains essential information.
- **Autoencoder:** The result is simply a continuous representation of the source data since LDA has been applied. An optimized version of LDA characteristics is subsequently incorporated in a feature extraction system in addition to raw attributes, to capture non-linear interactions between the expressions of different genes. For feature extraction, it uses an Autoencoder neural network.

2) DEEP LEARNING ASSISTED SUPERVISED CLASSIFIER LEARNING

The features extracted from the proposed two-phase unregulated attribute learning methodology are ultimately applied to a set of labels for classifier learning to forecast clinical outcomes for cancer patients. The method of classification is subject to test labels that indicate directed classification preparation.

The key components for women with breast cancer therapies are outlined below.

- First, provided multiple genomic profiles $X = \{x_i\}_{i=1}^{m}$
- Second, the primary analysis component is used to learn compressed feature sets $\hat{X} = \{\hat{x}_i\}_{i=1}^{m}$, for $\frac{1}{m}\sum_{i=1}^{m}\left(u_1^T x_i - u_1^T \overline{X}\right)^2 = u_1^T S u_1$;
- Third, the raw expression of genes X and compressed feature $\hat{X}$ is combined into $\tilde{X}$, where $\tilde{X} = \left\{\tilde{x}_i \mid \tilde{x}_i = (\hat{x}_i, x_i)\right\}_{i=1}^{m}$. The inputs of the autoencoder neural network X are considered to allow deep feature learning techniques for more complex representation $h^{(2)}$.
- The compressed features $\hat{X}$ and profile $h^{(2)}$ are eventually combined to form an integrated classifier in a robust X' manner.

Therefore, an LDA-Ada compartment is designed to use LDA compressed results as input properties to deal with a two-stage classifying feature learning system.

V. DETAILS OF IMPLEMENTATION

A. ALIGNMENT OF DATA

For example, compact characteristic vectors in different dimensions can be accomplished when the LDA algorithm is used for the expression for dimension reduction tests because the scale and the source value of the neural network differ from one of the individual datasets. To fit the model, it has been built into data of different sizes; it has placed all the functional vectors with null values in a single direction without compromising performance.

B. FEATURE EXTRACTION TARGET FUNCTION

The second step has to develop a deep neural network by hierarchically stacking many Autoencoders in the proposed Feature Analysis.

1) THE NON-LINEAR METHOD

The autoencoder and decoder portion consists of several non-linear processing layers which are analyzed based on the mixture of the initial $\tilde{X}$ data that has been used as data entry to analyze the non-linear model and the corresponding mathematical formulation has been equated as follows in the Eq(1):

$$h^{(1)} = \sigma\left(\omega^1 \tilde{X} + b^{(1)}\right),$$
$$h^{(j)} = \sigma\left(\omega^j h^{(j-1)} + b^{(j)}\right), \quad j = 2, \ldots, n. \qquad (1)$$

As discussed in the Eq (1) n indicates the surface number, and σ specifies the aspect of activation. $h^{(j)}$, ω^j and $b^{(j)}$ represent

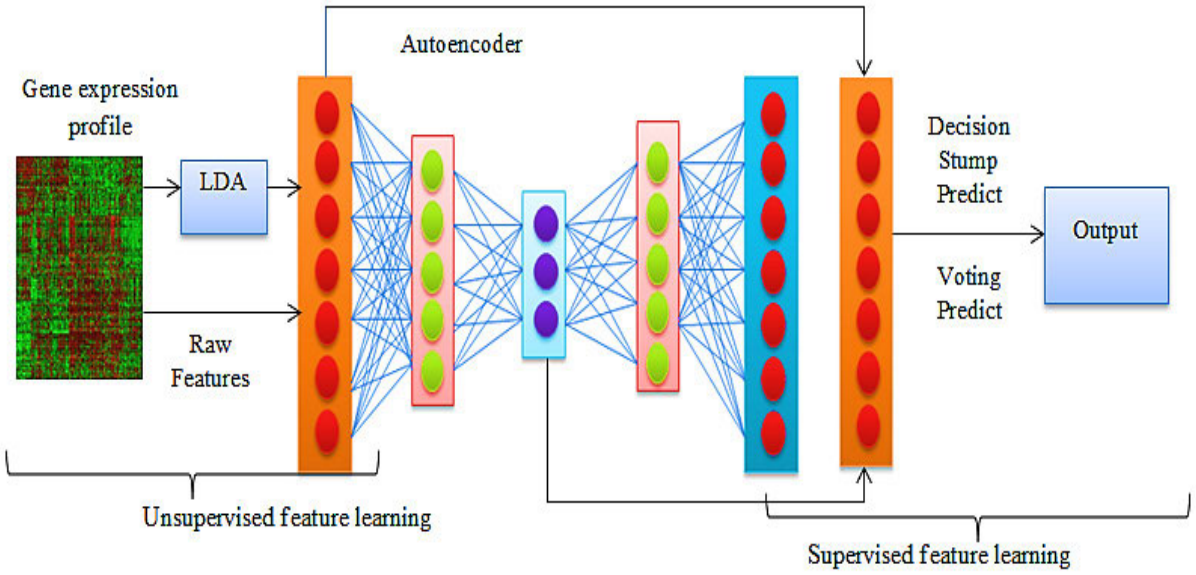

FIGURE 4. Flowchart of the proposed method.

in the j[th] layer as a hidden vector, mass index, and bias vector.

2) LOSS IN REBUILDING

Autoencoder tries to reduce the discrepancy in inputs $\tilde{X}$ to restored output $sh^{(n)}$. Despite the infinite number of parameters in the quantum system and the existence of small samples, the possibility of overfitting is challenging for training deep neural networks. It is placed some sparsity penalties on the hidden layers to mitigate this issue, thus the loss of reconstruction as follows in equation (2):

$$\mathcal{L}_{rec} = \left\| \tilde{X} - h^{(n)} \right\|_2^2 + \eta \sum_{j=1}^{n} \left\| b^{(j)} \right\|_2^2. \tag{2}$$

Here, $h^{(n)}$ are the effects, and η is a hyperparameter for balancing the biases of the individual parts. The network is defined heuristically as a single input layer, 3 exposed layers, and a single output layer.

3) AUTOENCODER ACTIVATION FUNCTION

The ELU (Exponential Linear Unit) has been used as the activator function to speed up formation in deeper neural networks, and to increase precise classification σ is shown in equation (3) and (4)

$$\sigma\left(h^{(j)}\right) = \begin{cases} h^{(j)} & if\ h^{(j)} > 0 \\ \alpha(\exp\left(h^{(j)}\right) - 1) & if\ h^{(j)} \leq 0 \end{cases} \tag{3}$$

$$\sigma\left(h^{(j)}\right) = \begin{cases} 1 & if\ h^{(j)} > 0 \\ \sigma\left(h^{(j)}\right) + \alpha & if\ h^{(j)} \leq 0 \end{cases} \tag{4}$$

In this case, the ELU hyperparameter α(set $\alpha = 1.0$) controls the value to which an ELU saturates for negative input.

4) CUSTOMIZATION

It uses a customization approach based on a stochastic gradient. In combination with RMSProp, which operates in a very positive fashion either in lines or in non-stationary processes with different parameter values from original and second instants, the Adaptive training scores measure the advantages of both AdaGrad and RMSProp. The conditions for this work are 64 lots, iteration times are 10k, and the limit for learning is 0.001.

5) STANDARDIZED STARTUP

It has initialized the distinctions as 0 and the matrix of weight ω^j in each layer with the standard uniform distribution as described in the neural network training.

$$\omega^{(j)} \sim U\left[-\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}\right] \tag{5}$$

If $U[-a, a]$ is an interval uniform distribution -1×10^{-4}, 1×10^{-4} and k is the clock layer scale (number of columns ω).

C. CURRENT TECHNICAL REVIEW

The most frequently used IHC markers are Ki-67, estrogen receptors, progesterone receptor, P53 protein, and the human epidermal factor-2. One of the most important diagnostic and predicting indicators in the growth of breast cancer is a non-histone protein. Radiation and chemotherapy are incredibly susceptible to patients with high Ki-67. For breast cancer, Ki-67 expression has a substantial predictive and prognostic benefit.

1) REVIEW OF DEEP LEARNING METHODS

Automated Ki-67 scoring and identification of points via a deep learning approach have not yet been tried. This paper contains significant contributions:

- Design of a revolutionary in-depth thought method for the recognition and evaluation of stained hotspots.
- Inclusion of the validation framework in the latest deep learning system.
- This is an additional benefit for the vital quantification of Ki-67 scoring techniques already developed.

2) EXPERIMENTAL CONFIGURATION

- Preparing slides and acquiring images.
- The slide with its histological parts has stained with a monoclonal antibody Ki-67. In compliance with the institutional protocols, all procedures such as slide planning, image processing, etc. are carried out.

3) CONVOLUTIONAL NETWORK LAYERS

A CN consists of multi-sub-sampling, sub-sampling, non-linear, and entirely interconnected layers. Let f is a CN with N number layers serial composition or function $(f_1, f_2, \ldots, f_N)$. Mappings can be expressed between input (w) and output (u) as shown in the following equation (6):

$$u = f\ (w; X_1, X_2, X_3, \ldots, X_N) = f_1\ (w; X_1)\ \ of_2(.; X_2) \ldots of_{N-1}(.; X_{N-1})\ of_N(.; X_N) \quad (6)$$

f_N has historically been delegated to carry out spatial convolution or non-linear activation or classification. Where X_N shows bias and weight vector for the nth layer f_N. According to the range of η training data $\left\{(w^{(i)}, u^{(i)})\right\}_{i=1}^{\eta}$, vectors $(X_1, X_2, X_3, \ldots, X_N)$ can be determined as follows:

$$\arg\min_{X_1, X_2, X_3, \ldots, X_N} \frac{1}{\eta} \sum_{i=1}^{\eta} f_{Loss}(f\left(w^{(i)}; X_1, X_2, X_3, \ldots, X_N\right), u^{(i)} \quad (7)$$

where f_{Loss} implies loss function. Stochastic reduction and reverse propagation strategies can accomplish Equation 7. In the computation of a feature map, a convolutional layer typically utilizes convolutional filters. The feature map FM_m^h the equation at m level is shown in equation (8)

$$FM_m^h = f\left(\alpha_m^h + \sum\nolimits_j FM_j^{h-1} \times G_{jm}^h\right) \quad (8)$$

FM_{in}^{h-1} and FM_{out}^h are some characteristics of input and output. Biases and kernels respectively are G_{jm}^h and α_m^h. Two components create feature maps for each convolution layer. The local receptive area is the first element, and mutual weights are the second part. The benefits of this method are the ability to measure the image size of the data and to create a positive difference in local regions. The following equation is used to determine the function.

$$\Psi_j = \max(\psi_i^{n \times n} z(n, n)) \quad (9)$$

Here ψ is the input image, z indicates the role of the window and n listed simply the size of the input patch. Rectified Linear Units (ReLUs) are used as a tool for activation and descent gradient.

$$q\ (r) = \max(0, r) \quad (10)$$

where q shows the output component of the model with the r information, each layer has the same size for input and output. Let us consider X and Y, respectively, for the data and the finite output spaces. The decision tree also applies to decision nodes as internal branch nodes, indexing them with D. Similarly, nodes of prediction are suggested by P as terminal nodes. The decision-making feature f_d has allocated for each decision node $0 \in Df_d\ (.; \ominus) : X \rightarrow [0, 1]$. The probability distribution π_p over Y is available in each $p \in P$ projection node. If the reference $x \in X$ enters node of judgment d, it will propagate to the right or the left of the substratum based on a $f_d(x; \ominus)$. The final results for sample x of the tree T with decisions parameterized with oscillating are shown by the following equation

$$P_T\ [y\,|x, \ominus, \pi] = \sum\nolimits_{p \in P} \pi_{py} \mu_p(x \mid \ominus) \quad (11)$$

In this scenario, π_{py} and $\pi = (\pi_p)p \in P$ is a likelihood that the sample would enter leaf p on class y and identify by $\mu_p\ (x \mid \ominus)$ when $x \in X, \sum_p \mu_p\ (x \mid \ominus) = 1$. Decision nodes are based on the stochastic routine and have been described as:

$$f_d\ (x; \ominus) = \sigma(f_r\ (x; \ominus)) \quad (12)$$

The sigmoid function $\sigma(x)$ in this case is set to $\sigma\ (x) = \frac{1}{(1+e^{-x})}$. the decision forest is called a group of decision-making trees and is defined by the following equation (13)

$$F = \{T_1, T_2, \ldots, T_z\} \quad (13)$$

Let I is a $= I_1, I_2, I_3, \ldots, I_Q$, in which Q displays a set of pixels and I_Q means the size of the gray-level of a pixel L, $K = (K_1, K_2, K_3, \ldots, K_Q$ where $K_Q \in LL = \{0, 1\}$ can be generalized to a set with positive labels.

$$K^* = \arg\min_k \{Y(I\,|K, \ominus) Y(K)\} \quad (14)$$

$Y(K)$ is a distribution of Gibbs. Equation 15 can be written as in the Expect-Maximization algorithm.

$$K^* = \arg\min_{K \in k} \{U(I\,|K, \ominus) U(K)\} \quad (15)$$

Here U corresponds to urinary potential or energy of chance and is indicated by

$$U\ (I\,|K,\ \ominus) = \sum\nolimits_Q \left[\frac{(I_Q - \mu_{KQ})^2}{2\sigma_K^2} + In\ \sigma_K\right] \quad (16)$$

According to the theorem, it assumes a Gaussian distribution with parameters $\sigma_{xi} = \mu_{xi}, \sigma_{xi}$ follows the segmented area strength. This theory cannot model events in real life. (Gamma Mixture Model)GMM is, therefore, the best choice for engineers for dynamic delivery. Below are the calculations of a GMM with c elements.

$$\sigma_i = \{\left(\mu_{i,1}, \sigma_{i,1}, w_{i,1}\right), \ldots, (\mu_{i,c}, \sigma_{i,c}, w_{i,c}) \quad (17)$$

The Gaussian parameter distribution can be written as

$$G(z;\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{(z-\mu_i)^2}{2\sigma_i^2}\right) \quad (18)$$

To get the estimated probability by combining Equation 18 with Equation 19 and it is described as follows

$$G_{mix}(z;\alpha_i) = \sum_{c=1}^{h} w_{i,c}G(z;\mu_{i,c},\sigma_{i,c}) \quad (19)$$

The pixel intensity is a three-dimensional vector for an RGB signal. Hence, the GMM as its criteria and it is described as follows in equation (20)

$$\alpha_{xi} = (\mu_{i,1}, \sum_{i,1}, w_{i,1}) \ldots (\mu_{i,c}, \sum_{i,c}, w_{i,c}) \quad (20)$$

According to equation 16 with equation 20, the energy index is estimated as follows,

$$U(I|K, \ominus) = \sum_Q \left[\frac{1}{2}(I_Q-\mu_{KQ})^T \sum_{KQ}^{-1}(I_Q-\mu_{KQ}) + \ln|\Sigma_{KQ}|\frac{1}{2}\right] \quad (21)$$

The proposed Deep Learning framework with the autoencoding neural network has been evaluated as below.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

Compared to the LDA&AE-DL classification performance with those of other classifiers, the LDA-AE works better than other methods such as GPMKL, MLP, DLD, SWK, which are built with LDA compressed features for all the evaluation metrics. LA-AE-DL shows a better performance, while LDA-DL works well only in positive cases. Hence, Deep learning has been effectively eliminating the problem of uneven distribution of training and improves classifier's capacity to generalize breast cancer.

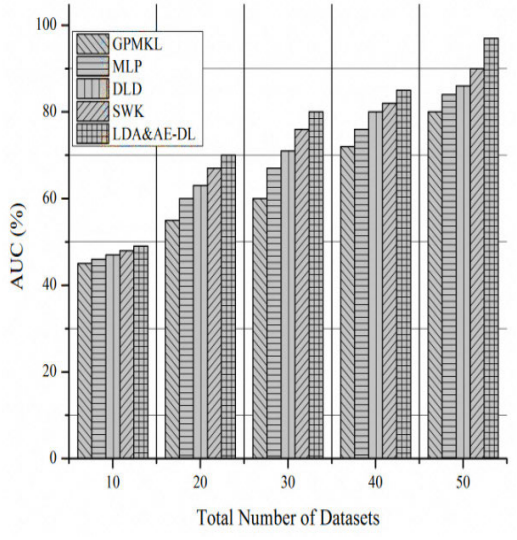

FIGURE 5. AUC evaluation.

Figure 5 shows that this advanced ensemble classification provides the best AUC(Area Under Curve) performance over several datasets, although the gene set methods are not based on the datasets. Hence, the gene-set ways that achieve higher AUC efficiency than that of the two gene classifiers. The MCC scores indicate a similar phenomenon. Table 2 shows the evaluation of the Matthews correlation coefficient of proposed LDA&AE-DL in comparison with GPMKL, MLP, DLD, SWK.

TABLE 2. MCC evaluation.

Total Number of Datasets	GPMKL	MLP	DLD	SWK	LDA&AE-DL
10	50.6	53.3	56.8	58.4	60.3
20	66.5	69.1	72.8	78.5	82.2
30	76.8	78.2	82.1	84.3	89.9
40	70.5	74.3	80.2	86.1	92.7
50	78.4	83.6	85.1	89.7	96.8

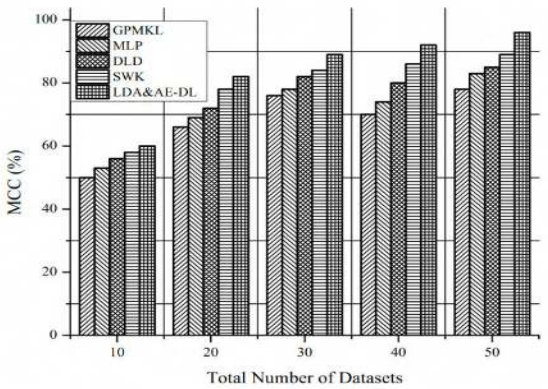

FIGURE 6. MCC evaluation.

From Figure 6, the proposed classification system (especially those based on physical signature), performed well on a dataset. This is because these two datasets have both a negative lymph node and a positive lymph node, whereas the other datasets have only lymph node-negative patients. Remarkably, the high performance of specific advanced ensemble classifiers [22], [23] achieves 96.7% MCC, which outperforms the others substantially.

The above observation shows that the method being proposed is less susceptible to unbalanced data sets and is indeed more stable, compared with the four other categories which show dramatic changes in the various datasets.

However, the AUCs and MCCs are compared with further four combined methods in public datasets. Figure 7 shows the outcomes to demonstrate the efficiency of a particular way more in detail. Complete analysis indicates that this AUC classification reaches more than 75.3%, while the two genetic classification devices have a significantly worse outcome of 75.8%. However, both genetic classification devices can only achieve AUC 97.4%. The MCC indexes will demonstrate a similar phenomenon.

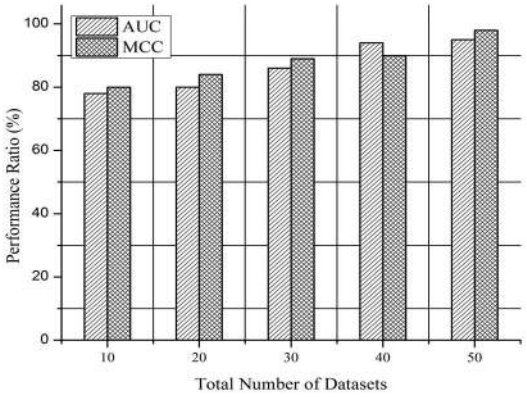

FIGURE 7. Performance analysis.

Deep learning in various machine learning tasks, including object detection, and classification, has proved breakthrough technology of recent years. Detailed learning methods from the input information concerning the target output following classical Machine Learning methods that involve a hand-crafted feature extraction level. The highest performance networks achieve marginally higher efficiency and exceed all other tasks in terms of accuracy when measured using the same data set. The architecture, however, offers detailed lesion segmentation as an input to extract from features. The performance of the system is, therefore, highly dependent on the quality of the segmentation provided, which can be a task that takes by the user with a great deal of time and effort. In the meantime, the lack of any CNN results in an enhanced end-to-end clinical classification system. Figure 8 shows the accuracy of the proposed LDA&AE-DL when compared to GPMKL, MLP, DLD, and SWK.

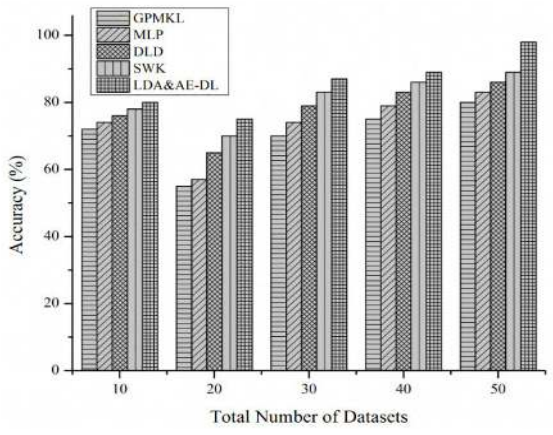

FIGURE 8. Accuracy evaluation.

Although several improvements in the training process and network architecture have been proposed in response to the difficulties caused by the increasing potential and complexity of models, there is a still significant amount of data required to provide adequate training which is not available for most medically focused applications, such as the present problem, i.e., the diagnosis of breast cancer.

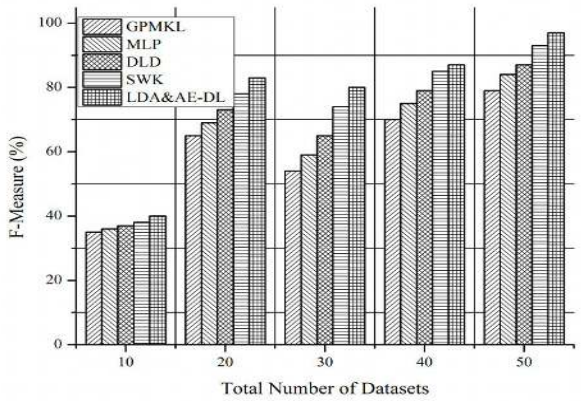

FIGURE 9. F-Measure evaluation.

The F measure is known as the harmonic mean of accuracy and recall of classification. In comparison to other examples, both FP and FN are included in the calculation of F-measure. The meaningful indicator of the consistency of binary classifications is the Matthews correlation coefficient. MCC calculates a correlation of the classification prediction with a more balanced index. Figure 9 shows the F-Measure of proposed LDA&AE-DL when compared to GPMKL, MLP, DLD, SWK. According to the results and discussion section, the proposed method has high AUC, MCC, and accuracy for classifying breast cancer.

VII. CONCLUSION

This paper incorporates a linear discriminant analysis (LDA) with an Autoencoder neural network with deep learning techniques to learn from the gene expression information with most characteristic features. This uses the deep learning algorithm at the stage of classification to create an advanced ensemble classification for the prediction. Hence, the suggested system has more prediction capacity with deep learning classification compared to other techniques, which are shown in evaluation results. This analysis showed excellent ability to generalize quickly and explicitly improve the performance of the prediction of the results with 98.27% of accuracy, which has been automatically obtained from the network. This approach has great potential for generalization, and it must be further enhanced with more public data sets.

REFERENCES

[1] J. Xie, R. Liu, J. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers Genet.*, vol. 10, p. 80, Feb. 2019.

[2] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*. [Online]. Available: http://arxiv.org/abs/1606.05718

[3] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Sci. Rep.*, vol. 6, no. 1, pp. 1–9, Jun. 2016.

[4] B. Ehteshami Bejnordi, J. Lin, B. Glass, M. Mullooly, G. L. Gierach, M. E. Sherman, N. Karssemeijer, J. van der Laak, and A. H. Beck, "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 929–932.

[5] S. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, Jul. 2019.
[6] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar, "Deep learning in mammography and breast histology, an overview and future trends," *Med. Image Anal.*, vol. 47, pp. 45–67, Jul. 2018.
[7] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, and H. Zheng, "Deep learning based classification of breast tumors with shear-wave elastography," *Ultrasonics*, vol. 72, pp. 150–157, Dec. 2016.
[8] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. 1, Jun. 2017, Art. no. 46450.
[9] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2440–2445.
[10] D. Bychkov, N. Linder, R. Turkki, S. Nordling, P. E. Kovanen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, and J. Lundin, "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, Dec. 2018.
[11] M. E. Vandenberghe, M. L. J. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker, "Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer," *Sci. Rep.*, vol. 7, no. 1, May 2017, Art. no. 45938.
[12] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, and O. Geessink, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *J. Amer. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.
[13] A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, "Deep learning in mammography: Diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer," *Investigative Radiol.*, vol. 52, no. 7, pp. 434–440, Jul. 2017.
[14] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Syst. Appl.*, vol. 46, pp. 139–144, Mar. 2016.
[15] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, Jun. 2018, pp. 737–744.
[16] E. Deniz, A. engür, Z. Kadiro lu, Y. Guo, V. Bajaj, and Ü. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health Inf. Sci. Syst.*, vol. 6, no. 1, p. 18, Dec. 2018.
[17] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Comput. Methods Programs Biomed.*, vol. 153, pp. 1–9, Jan. 2018.
[18] C. K. Ahn, C. Heo, H. Jin, and J. H. Kim, "A novel deep learning-based approach to high accuracy breast density estimation in digital mammography," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101342O.
[19] 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.
[20] Y. Qiu, Y. Wang, S. Yan, M. Tan, S. Cheng, H. Liu, and B. Zheng, "An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology," *Proc. SPIE*, vol. 9785, Mar. 2016, Art. no. 978521.
[21] M. Mahrooghy, A. B. Ashraf, D. Daye, E. S. McDonald, M. Rosen, C. Mies, M. Feldman, and D. Kontos, "Pharmacokinetic tumor heterogeneity as a prognostic biomarker for classifying breast cancer recurrence risk," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1585–1594, Jun. 2015.
[22] M. To açar, B. Ergen, and Z. Cömert, "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders," *Med. Hypotheses*, vol. 135, Feb. 2020, Art. no. 109503.
[23] B. Wei, Z. Han, X. He, and Y. Yin, "Deep learning model based breast cancer histopathological image classification," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Big Data Anal. (ICCCBDA)*, Apr. 2017, pp. 348–353.

XINFENG ZHANG received the master's degree from the China Medical University of Surgical Oncology, in 2008. He worked with the Liaoning Cancer Hospital and the Institute, as a Vice Chief Physician. His research interest includes breast oncology.

DIANNING HE received the Ph.D. degree in bio engineering from the Northwestern University, in 2018. He worked with the College of Medicine and Biological Information Engineering, Northeastern University. His research interest is machine learning with MR and CT.

YUE ZHENG received the M.D. degree from Henan Medical University, in 2019. She is currently pursuing the degree with the Department of Radiology, The First Affiliated Hospital of China Medical University. Her research interests include breast and cardiac imaging.

HUAIBI HUO is currently pursuing the degree with the Department of Radiology, The First Affiliated Hospital of China Medical University. Her research interests include breast and cardiac imaging.

SIMIAO LI is currently pursuing the master's degree with the Department of Radiology, The First Affiliated Hospital of China Medical University. Her research interests include breast and cardiac imaging.

RUIMEI CHAI received the master's and Ph.D. degrees from the China Medical University of Radiology in 2009 and 2018, respectively. Since 2009, she has been working with the Department of Radiology, The First affiliated hospital of China Medical University. Her research interest is machine learning in breast cancer.

TING LIU received the Ph.D. degree in medical imaging from The First Affiliated Hospital of China Medical University, Shenyang, China, in 2016. She is currently a FSCCT, the Director, a Cardiac Imaging Program, The First Affiliated Hospital of China Medical University Deputy Chief, a Cardiovascular Imaging Associated Professor of Radiology, and the M.D. with China Medical University, in 2005, China Medical University, Shenyang, China.. From 2012 to 2014, she was a Postdoctoral Researcher with the Massachusetts General Hospital, Harvard Medical School Biography. She went to Medical School, Shenyang, China, and completed Radiology Residency with The First Affiliated Hospital of CMU. In 2012, She finished her Fellowship in cardiovascular imaging with the Massachusetts General Hospital, Boston, MA, USA. Since 2017, she has been the Deputy Chief of cardiovascular imaging and the Director of the Cardiac imaging Program with The First Affiliated Hospital of CMU. She played an important role in the hallmark clinical trials related projects in cardiovascular imaging, such as ROMICAT II and Framingham. Her work has resulted in multiple peer-reviewed publications of high impact and has more than 20,000-case experience in coronary CT. She was the Secretary of the Chinese Society of Radiology and has plenty of experiences on international collaboration between China and other counties. She has been contributed as a member of FiRST committee of SCCT since 2016 and has served as the Deputy President of the Youth Committee of SCCT China IRC since 2017 and has also been instrumental in attracting early career radiologist and cardiologist in China to join SCCT. she is currently an Associated Professor of radiology with China Medical University and also leads the Multidisciplinary Group focused on utilizing advanced imaging to improve prevention, diagnosis, and treatment of disease. She is also the Director of the National Science Grant Program in cardiac imaging and has trained more than 40 researchers and clinicians. Her research interests include imaging of high-risk plaque imaging to improve efficiency of management of patients with acute and stable chest pain biomarkers of inflammation and injury and imaging to decipher the pathophysiology of acute myocardial infarction and the effects of treatments.