IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 13, 2019, accepted May 10, 2019, date of publication May 14, 2019, date of current version June 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916586

Identifying Risk Stratification Associated With a Cancer for Overall Survival by Deep Learning-Based CoxPH

CHENG-HONG YANG[1,2], (Senior Member, IEEE), SIN-HUA MOI[1], FU OU-YANG[3,4], LI-YEH CHUANG[5,6], MING-FENG HOU[2,3,4], AND YU-DA LIN[1], (Member, IEEE)

[1]Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[2]Ph.D. Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[3]Division of Breast Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung 80708, Taiwan
[4]Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung 80708, Taiwan
[5]Department of Chemical Engineering, I-Shou University, Kaohsiung 84001, Taiwan
[6]Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung 84001, Taiwan

Corresponding authors: Li-Yeh Chuang (chuang@isu.edu.tw), Ming-Feng Hou (mifeho@kmu.edu.tw), and Yu-Da Lin (e0955767257@yahoo.com.tw)

This work was supported by the Ministry of Science and Technology, China, under Grant 106-2221-E-992-327-MY2, Grant 107-2811-E-992-500, Grant 105-2314-B-037-038-MY3, and Grant 105-2314-B-037-037-MY3.

ABSTRACT In a time-to-event model, Cox proportional hazards (CoxPH) analysis is the most frequently used method for estimating overall survival. However, the CoxPH analysis is limited to explaining only single or partial risk effects among clinicopathological factors. We introduced DeepCoxPH, a risk score estimation strategy based on deep learning (DL) and CoxPH, to improve the risk stratification for overall survival analysis. The abstracted weight from the DL and the hazard ratios from the CoxPH were transformed into the risk score estimation in the fully adjusted model. The DeepCoxPH exhibited more comprehensive risk weight estimation for overall survival. The DeepCoxPH was applied to predict ten-year overall survival in breast cancer. A Kaplan–Meier curve revealed that the DeepCoxPH improved discrimination of high- and low-risk stratification in both short- and long-term breast cancer for overall survival. To the best of our knowledge, this is the first report of the risk score estimation based on machine learning and parametric-statistical analysis aimed at identifying risk stratification for overall survival through the consideration of comprehensive risk effects among multiple clinicopathological factors.

INDEX TERMS Deep learning, CoxPH, risk stratification, overall survival.

I. INTRODUCTION

The identification of complex multifactor associations in diseases is among the crucial challenges facing human health [1]–[4]. Overall survival is a major primary endpoint for evaluating the outcome of a specific disease and determining its complex multifactor associations [5], [6]. Survival analysis based on a time-to-events model has been widely used to yield reliability models in biomedicine [7]–[9].

Breast cancer is the most commonly diagnosed malignancy among women worldwide. In Taiwan, the standardized breast cancer incidence of 2013 was 93 per 100,000 person-years (1.8 fold higher than that in 1997), and breast cancer has been the fourth leading cause of cancer deaths since 2005 (12.9–18.2 per 100,000 people) [10]. The disease burden of breast cancer is associated with human development [11]; thus, more precise methods for prognosis estimation are required. Several interpretations of associations between breast cancer prognosis and clinicopathological characteristics can be found in the literature [12]–[14].

Such survival analyses incorporate Cox models, Kaplan–Meier plots, log-rank tests, and survival tree analyses [15], [16]. Among Cox models, the Cox proportional hazards (CoxPH) model is the most commonly used survival analysis method in biostatistics [17], which examines time-varying effects in disease progression or mortality [15]. However, a CoxPH model can only singly or partially explain the risk effects of clinicopathological factors in a disease model. Comprehensive assessment of risk and interaction effects remains one of the greatest challenges for CoxPH analysis.

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

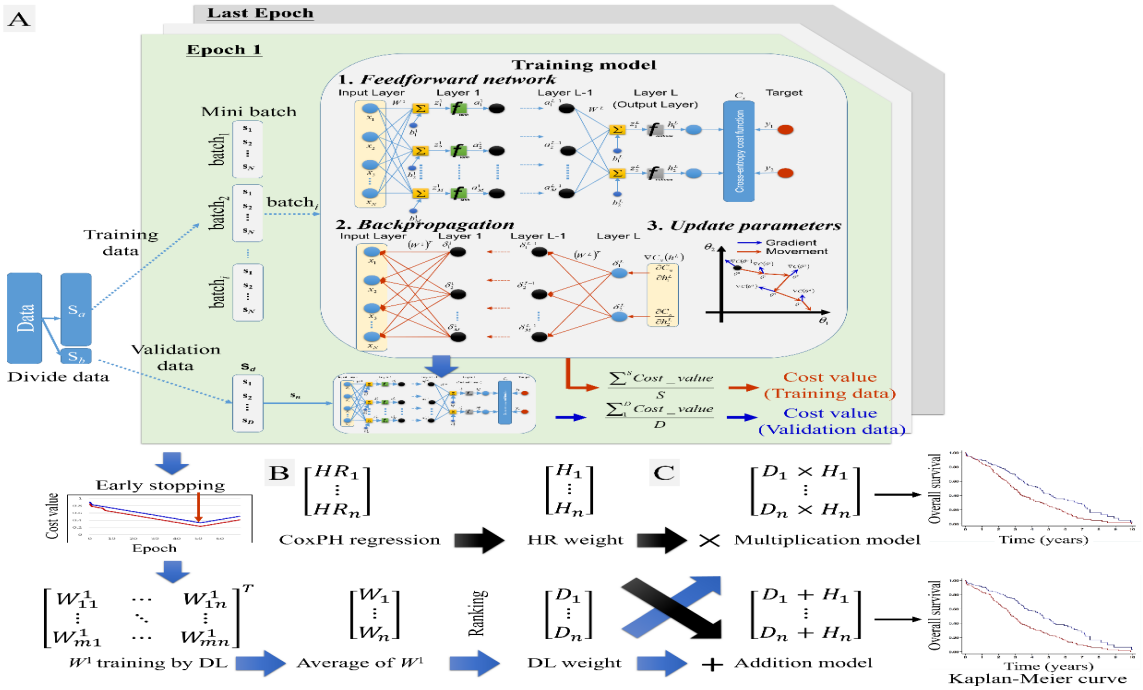

FIGURE 1. Illustration of deepCoxPH computation.

Deep learning (DL) is a promising machine-learning technology [18] that employs multilayered abstraction levels to learn input data representations. For classification problems, DL can discriminate and suppress irrelevant variations in the higher layers of representation to amplify characteristics of input data. DL has made major advances in bioinformatics [19] and numerous fields of science and engineering [20].

In this study, we proposed a comprehensive risk weight assessment strategy for overall survival estimation. A DeepCoxPH risk score estimation strategy based on DL and parametric statistical analysis was introduced to improve high-risk stratification for overall survival analysis. Abstracted weights were obtained from a network of DL and hazard ratios (HRs) were calculated by CoxPH. DeepCoxPH employed matrix operation to combine the abstracted weights with the hazard ratios and then transformed them into risk score estimation. We demonstrated that DeepCoxPH can improve risk stratification for overall survival through consideration of comprehensive risk effects among multiple clinicopathological factors. DeepCoxPH exhibited more comprehensive risk weight estimation for overall survival.

II. METHODS

The DeepCoxPH risk score estimation strategy consists of DL and a CoxPH model. DeepCoxPH—illustrated in Fig. 1—can be divided into the following three parts. 1) Deep neural networks (Symbol A in Fig. 1): the DL was used to train an effective network with a minimum classification error cost. DL weights were obtained from the trained network. 2) CoxPH was used to calculate HRs (Symbol B in Fig. 1). 3) The DL weights and HRs of all clinicopathological variables were ranked in ascending order and then divided into three risk degrees according to the ranking order. Thus, HR and DL weight risk degrees can be combined through matrix multiplication (i.e., the DeepCoxPH multiplication model) or matrix addition (i.e., the DeepCoxPH addition model) to identify high-risk stratification for overall survival. The flowchart of DeepCoxPH is shown in Fig. 2; all steps are detailed in the following subsection.

A. DEEP NEURAL NETWORKS

DL is multileveled approach to representation learning within a network composed of nonlinear models that transform a representation at one level into another at a higher, more abstract level. A complex function $y = f(x; \theta), f: R^N \rightarrow R^2$, where x denotes the input data and θ is the network parameters, can be learned through sufficient transformations of the representation at more abstract levels. The objective of DL is to produce the optimal parameter θ.

In this study, a deep neural network (DNN) was used to produce a classification and then obtain a deep abstraction to identify the factor importance associated with risk stratification (symbol A in Fig. 1). DNN consists of an input layer, numerous hidden layers, and an output layer, with the fully connection between all layers. DNN comprises three steps: (a) feedforward network, (b) backpropagation, and (c) update parameters. In feedforward network, each layer includes neurons that compose a dot product between the weight w and input vector x, to which a bias b is added. Its representation is $z = w \cdot x + b$, where w is a weight vector, x an input vector, and b a bias vector. Each neuron behavior is determined by a non-linearity (activation) function. We used a sigmoid-based

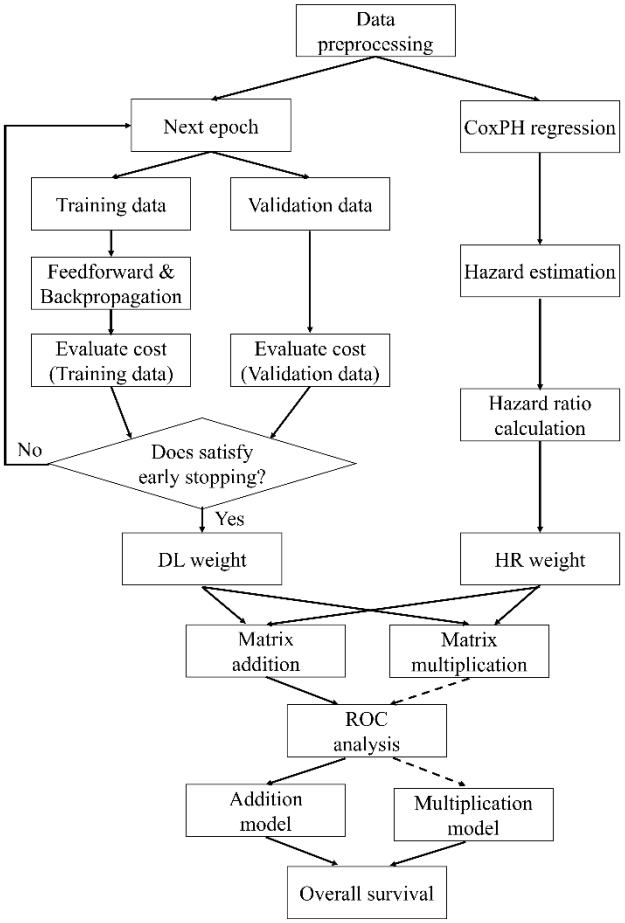

FIGURE 2. Flowchart of deepCoxPH.

logistic function (Eq. 1) as the activation function to obtain nonlinear mapping.

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{1}$$

In the output layer (denoted as L), a softmax function σ^L was used to handle binary classification (i.e., yes or no). The softmax function normalizes the outputs into the values between 0 and 1 (Eq. 2). The probability of this input being in a particular class can thus be obtained.

$$\sigma^L(z)_j = \frac{e^{z_j}}{\sum_k^K e^{z_k}} \quad for\ j = 1, \ldots, K \tag{2}$$

A vector y in the output layer is obtained through a recursion function:

$$\begin{aligned} y &= f(x) \\ &= \sigma^L\left(W^L \cdots \sigma\left(W^2 \cdot \sigma\left(W^1 x + b^1\right) + b^2\right) \cdots + b^L\right) \end{aligned} \tag{3}$$

where x is a vector in the input layer, W^l is a weight matrix in layer l, and b^l is a bias vector in layer l.

The cost can be evaluated by comparing outputs with the correct answer (target) to obtain error derivatives using the cross-entropy function $C = -logy_r$, where $r = 0$ indicates target as control group, and $r = 1$ indicates target as case group. A mini-batch stochastic gradient descent procedure with momentum was used for backpropagation and update parameters. Therefore, the cost based on a mini-batch update strategy can be formulated as follows:

$$C = -\frac{1}{m}\sum_{i=s}^{s+m} logy_{ir} \tag{4}$$

where m is the mini-batch size, i is the index of the training data sample, and s is the sample's start index according to the mini-batch. To backpropagate gradients in DNN, the weight W and bias b were updated using (5) and (6), respectively.

$$W = W - \eta\frac{\partial C}{\partial W} \tag{5}$$

$$b = b - \eta\frac{\partial C}{\partial b} \tag{6}$$

where η denotes the learning rate.

We used epochs to run the DNN multiple times and obtain the optimal parameters $\hat{\theta}$. One epoch is defined as when the training dataset passes forward and backward through the DNN only once. We used 80% of the dataset as training data for the DNN and 20% as validation data for early stopping. The optimal parameters $\widehat{\theta} = \left\{\widehat{W}^1, \widehat{b}^1, \widehat{W}^2, \widehat{b}^2, \ldots, \widehat{W}^L, \widehat{b}^L\right\}$ may be obtained when the DNN is stopped.

$$A = f\left(\widehat{W}^1\right) = \left[\frac{\sum_{i=1}^m \widehat{w}_{1i}^1}{m} \cdots \frac{\sum_{i=1}^m \widehat{w}_{ni}^1}{m}\right] \tag{7}$$

The DL-weights of all clinicopathological variables were ranked in ascending order and then categorized into three risk degrees according to this ranking order. Finally, the risk degree of DL-weight, expressed as the vector D_{risk_degree}, could be obtained through a ranking approach (Eq. 8).

$$\begin{aligned} D_{risk_{degree}} &= Ranking\left(A^T\right) \\ &= \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} \end{aligned} \tag{8}$$

B. CoxPH MODEL

Multifactor effects on overall survival are considered simultaneously in a CoxPH model, thus enabling the influence of each clinicopathological variables of disease progression and mortality at a particular observation time to be determined. A hazard rate is commonly used to determine the effect of individual clinicopathological variables. The associated clinicopathological variables in the survival analysis are usually termed covariates. In the CoxPH model, a hazard is estimated as follows:

$$h(t) = h_0(t)\, exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \tag{9}$$

where t represents the survival time, and $h(t)$ is the hazard function determined by a set of k covariates $(x_1, x_2, \ldots, x_p)$. The coefficients $(b_1, b_2, \ldots, b_p)$ measure the effect

(i.e., the effect size) of covariates. The term h_0 is called the baseline hazard and represents the hazard value if all x_i equal 0 (i.e., the quantity $exp(0)$ equals 1). Because a hazard may vary over time, the t in $h(t)$ represents time-varying effects. The CoxPH model provides estimates of $\beta_1, \ldots, \beta_k$ but no direct estimate of $h_0(t)$, that is, the baseline hazard. Formally, $h_0(t)$ is not estimated directly but can be obtained as an estimate of cumulative hazard $H_0(t)$ and, consequently, of the baseline survivor function $S_0(t)$.

Given two subjects k and k' that differ in their x-values, the corresponding hazard function can be formulated as the following hazard ratio for the two observations k: $h_k(t)$ and k': $h_{k'}(t)$.

$$\begin{aligned} hr &= \frac{h_k\,(t)}{h_{k'}\,(t)} \\ &= \frac{h_0\,(t)\,e^{\sum_{i=1}^{n}\beta x}}{h_0\,(t)\,e^{\sum_{i=1}^{n}\beta x'}} \\ &= \frac{e^{\sum_{i=1}^{n}\beta x}}{e^{\sum_{i=1}^{n}\beta x'}} \end{aligned} \tag{10}$$

The baseline hazard is an "intercept" term that varies with time, and the hazard logarithm of x_i variables is estimated using multiple linear regression of the CoxPH model according to the baseline hazard. The quantities e^{bi} are HRs. A b_i greater than 0 or, equivalently, a HR greater than 1 indicates an increasing hazard event and, therefore, a decreasing survival length as the i^{th} covariate increases. Therefore, a HR greater than 1 indicates a covariate highly associated with the event probability and thus negatively associated with survival length. Whereas an HR equal to 1 indicates no effect, an HR less than 1 indicates a reduced hazard and an HR greater than 1 indicates an increased hazard.

$$H = \begin{bmatrix} hr_1 & \cdots & hr_n \end{bmatrix} \tag{11}$$

The HRs of all clinicopathological variables are ranked in ascending order and then categorized into three risk degrees according to this ranking order. Finally, the risk degree of the CoxPH HR, expressed as the vector $H_{\text{risk}_\text{degree}}$, can be obtained through a ranking approach (Eq. 12; symbol B in Fig. 1).

$$\begin{aligned} H_{risk_{degree}} &= Ranking\left(H^T\right) \\ &= \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \end{aligned} \tag{12}$$

C. DeepCoxPH

The HR from CoxPH model and DL-weight from DL model were abstracted and ranked in ascending order. We then categorized the HRs and DL-weights into three risk degrees according to the ranking order. Each HR and DL-weight was converted into a risk degree and then entered in the risk score calculation. Using HR and DL-weight risk degrees, we calculated four risk scores by means of the CoxPH HR model, DL-weight model, DeepCoxPH multiplication model, and DeepCoxPH addition model. The DeepCoxPH multiplication model is calculated using Eq. 13, as follows:

$$\begin{aligned} &DeepCoxPH_{multiplication} \\ &= H_{risk_{degree}} \times D_{risk_{sdegree}} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \times \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} h_1 \times d_1 \\ \vdots \\ h_n \times d_n \end{bmatrix} \end{aligned} \tag{13}$$

The DeepCoxPH addition model is calculated using Eq. 14, as follows:

$$\begin{aligned} &DeepCoxPH_{addition} \\ &= H_{risk_{degree}} + D_{risk_{degree}} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} h_1 + d_1 \\ \vdots \\ h_n + d_n \end{bmatrix} \end{aligned} \tag{14}$$

The fully adjusted model included all the clinicopathological factors in the multivariate CoxPH model. We then constructed risk schemas by adding the risk degree for each clinicopathological factor in the fully adjusted model. Receiver operating characteristic curve (ROC) analysis was used to calculate the sensitivity and specificity corresponding to overall survival dichotomous for each score in the range of the risk score schema. The best cutoff point for each risk score model was chosen according to the highest sum of sensitivity and specificity in ROC analysis. Each risk schema was transformed into risk stratifications using the best cutoff point by the corresponding risk degree model (CoxPH HR, DL-weight, DeepCoxPH multiplication, or DeepCoxPH addition models). The performance of each risk stratification schema was evaluated using an overall survival function Kaplan–Meier curve between high- and low-risk groups, which was dichotomous according to different risk stratification schemas. The risk function between high and low risk in each risk stratification schema was tested using CoxPH analysis to ensure risk effects in high-risk strata.

D. STATISTICAL ANALYSIS

The distribution difference between cases and controls were evaluated using the chi-square test. Univariate and multivariate CoxPH models were used to estimate the association of clinicopathological factors in breast cancer overall survival. The relationship between the risk scores of each model were summarized and visualized using a matrix plot. The postestimation of each schema was determined using Akaike's information criterion (AIC) and Bayesian information criterion (BIC) values. Schemas with lower AIC and BIC values were considered more likely to be the true model. The level of significance was set at $P \leq 0.05$. All statistical analyses were performed using STATA (version 11.1).

III. RESULTS

A. DATASETS

All of the data were collected from the single-center Taiwan Breast Cancer Consortium (TBCC) database, which

is prospectively maintained by the Breast Surgery Division of Kaohsiung Medical University Hospital (IRB number: KMUIRB990174). Patients diagnosed with ductal carcinoma in situ or who had incomplete clinicopathological data were excluded. A total of 1646 breast cancer patients were included in an age-matching case–control procedure. A case was defined as a patient who expired or whose disease progressed within the follow-up period, whereas a control was defined as a patient who survived with no progression during the follow-up period. The case–control ratio was set at 1:1. After the age-matching procedure, a total of 458 patients were included, with 229 cases and 229 controls. All case and control patients were tracked from 2000 to 2016 using records in the TBCC database. The clinicopathological factors in this dataset were age, grade, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER 2), tumor size (AJCC (American Joint Committee on Cancer) stage), lymph node (AJCC stage), lymph vascular invasion (LVI) status, dermal invasion, perineural invasion, operation method, radiotherapy, chemotherapy, and hormone therapy. The 10-year overall survival of all breast cancer subjects was tracked from the date of first diagnosis to the date of death or the end of the study. Patients lost to follow-up before the end of the study were censored.

The distribution of all clinicopathological characteristics between the case and control groups is summarized in Table 1. The results indicated that the case group had a significantly higher proportion of the following factors: HER2 positive, ER positive, PR positive, tumor size with AJCC stage T3–T4, positive lymph node invasion, positive dermal invasion, positive neural invasion, with chemotherapy, without hormone therapy, and without target therapy.

B. CoxPH MODEL FOR OVERALL SURVIVAL IN BREAST CANCER

The association between each clinicopathological factor and overall survival in breast cancer is shown in Table 2. The results of both univariate and multivariate Cox regression analysis indicated that patients with the following characteristics had a higher risk of all-cause mortality: grade III (crude HR = 1.64, 95% CI = 1.25–2.16, $P < 0.001$; adjusted HR = 1.50, 95% CI = 1.12–2.02, $P < 0.001$); PR positive (crude HR = 1.57, 95% CI = 1.21–2.05, $P = 0.001$; adjusted HR = 1.79, 95% CI = 1.17–2.73, $P = 0.007$); T3–T4 tumor size (crude HR = 1.74, 95% CI = 1.22–2.49, $P = 0.002$; adjusted HR = 1.60, 95% CI = 1.04–2.46, $P = 0.007$); perineural invasion (crude HR = 1.45, 95% CI = 1.06–2.45, $P = 0.021$; adjusted HR = 1.45, 95% CI = 1.02–2.05, $P = 0.038$).

C. ESTIMATED RISK SCORE SCHEMAS BASED ON DL AND Cox MULTIVARIATE REGRESSION MODEL

The optimal DL model, identified using a grid search technique, had the following characteristics: hidden layer size = 15 (the layer unit sizes were 30, 30, 30, 20, 20, 20, 20, 20, 20, 15, 15, 15, 10, and 10), learning rate (η) = 0.01, and mini-batch = 5. Learning in the optimal model was learning at 1460 epochs through early stopping on the validation and training sets. The estimated risk scores and degrees were derived from HRs by CoxPH regression analysis, DL-weight value, and multiplication and addition of both risk degrees (Table 3). According to the ranking order of CoxPH HRs, grade III, no target therapy, lymph node invasion, age over 50, and perineural invasion were assigned a risk degree of 1. The clinicopathological factors assigned a risk degree of 2 were ER negative, no radiotherapy, lymph node invasion, no hormone therapy, and HER2 negative. The remaining clinicopathological factors, including PR negative, total mastectomy, dermal invasion, larger tumor size, and no chemotherapy, were assigned a risk degree of 3. Furthermore, the DL-weight ranking order assigned a risk degree of 1 to age over 50, dermal invasion, HER2 negative, no target therapy, and no chemotherapy, whereas larger tumor size, lymph node invasion, perineural invasion, no radiotherapy, and total mastectomy were assigned a risk degree of 2. The remaining clinicopathological factors, including PR negative, lymph node invasion, no hormone therapy, ER negative, and grade III were assigned a risk degree of 3.

D. COMPARISON OF FOUR RISK STRATIFICATION SCHEMAS FOR OVERALL SURVIVAL IN BREAST CANCER

For each model, the summation of the risk degrees in the fully adjusted model generated a risk score for each patient. The relationship between the risk scores from each model are summarized in Fig. 3. The slopes from the lower left to the upper right indicate the positive correlation between the risk scores. All the models showed a positive and non-conflicting mutual correlation. When the tighter together the points are clustered, the correlation between the risk scores is strong, reveling DeepCoxPH addition model has the strongest positive correlations between the risk scores. Next, four risk stratification schemas were devised according to the best cutoff value from ROC analysis using the risk degree summation in the fully adjusted model for each schema. The risk stratification of the DeepCoxPH multiplication model (risk degree multiplication) showed the highest area under the curve (AUC) in ROC (AUC = 0.672, 95% CI = 0.623–0.721), followed by the DL-weight model (DL-weight risk degree; AUC = 0.671, 95% CI = 0.621–0.720), the DeepCoxPH addition model of (risk degree addition; AUC = 0.669, 95% CI = 0.620–0.718), and the CoxPH HR model (CoxPH HR risk degree model; AUC = 0.660, 85% CI = 0.611–0.710). As shown in Fig. 4, although the risk stratification of the CoxPH HR model (sensitivity = 79.5%, specificity = 44.5%) and the DL-weight model (sensitivity = 83.8%, specificity = 45.4%) achieved a higher sensitivity, the DeepCoxPH multiplication model (sensitivity = 64.6%, specificity = 63.3%) and addition model (sensitivity = 64.6%, specificity = 63.3%) performed better for specificity in both high- and low-risk stratification. Our results demonstrated

TABLE 1. Clinicopathological factors of age-matched case-control patients in breast cancer.

Variable	Total		Control		Case		*P*
	n	%	n	%	n	%	
Age (years)							1.000
≤ 50	260	56.8%	130	56.8%	130	56.8%	
> 50	198	43.2%	99	43.2%	99	43.2%	
Grade							0.077
I-II	300	65.5%	159	69.4%	141	61.6%	
III	158	34.5%	70	30.6%	88	38.4%	
HER2							<0.001
Positive	183	40.0%	72	31.4%	111	48.5%	
Negative	275	60.0%	157	68.6%	118	51.5%	
ER							0.007
Positive	305	66.6%	166	72.5%	139	60.7%	
Negative	153	33.4%	63	27.5%	90	39.3%	
PR							0.001
Positive	265	57.9%	150	65.5%	115	50.2%	
Negative	193	42.1%	79	34.5%	114	49.8%	
Triple negative							<0.001
No	408	89.1%	218	95.2%	190	83.0%	
Yes	50	10.9%	11	4.8%	39	17.0%	
Tumor size (AJCC stage)							<0.001
T1-T2	408	89.1%	218	95.2%	190	83.0%	
T3-T4	50	10.9%	11	4.8%	39	17.0%	
Lymph node (AJCC stage)							<0.001
Negative	242	52.8%	156	68.1%	86	37.6%	
Positive	216	47.2%	73	31.9%	143	62.4%	
Lymphovascular invasion							<0.001
Negative	280	61.1%	159	69.4%	121	52.8%	
Positive	178	38.9%	70	30.6%	108	47.2%	
Dermal invasion							<0.001
Negative	420	91.7%	222	96.9%	198	86.5%	
Positive	38	8.3%	7	3.1%	31	13.5%	
Perineural invasion							0.021
Negative	375	81.9%	197	86.0%	178	77.7%	
Positive	83	18.1%	32	14.0%	51	22.3%	
Operation							<0.001
Partial mastectomy	214	46.7%	134	58.5%	80	34.9%	
Total mastectomy	244	53.3%	95	41.5%	149	65.1%	
Radiotherapy							0.084
Yes	280	61.1%	149	65.1%	131	57.2%	
No	178	38.9%	80	34.9%	98	42.8%	
Chemotherapy							0.005
Yes	329	71.8%	151	65.9%	178	77.7%	
No	129	28.2%	78	34.1%	51	22.3%	
Hormone therapy							0.003
Yes	281	61.4%	156	68.1%	125	54.6%	
No	177	38.6%	73	31.9%	104	45.4%	
Target therapy							0.351
Yes	92	20.1%	42	18.3%	50	21.8%	
No	366	79.9%	187	81.7%	179	78.2%	

ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2

that the DeepCoxPH model (multiplication and addition) can provide better recognition in both low- and high-risk stratification for breast cancer overall survival.

The Kaplan–Meier curves in Fig. 5 illustrate consistent results for the four risk stratification schemas between high- and low-risk strata. According to these curves, the DL-weight model achieved suitably dichotomous results for breast cancer overall survival. The hazard function analysis for high- and low-risk strata in Fig. 5 displays the HR risk ratio between both strata. All the high-risk stratas which is stratified using the four risk stratification schema, obtained higher hazard function in breast cancer overall survival compared to low-risk stratas. The DL-weight model achieved the highest HR [HR (95% CI) = 1.96 (1.38–2.79), $P < 0.001$], followed by the DeepCoxPH addition model [HR (95% CI) = 1.85 (1.41–2.44), $P < 0.001$], the CoxPH HR model [HR (95% CI) = 1.75 (1.27–2.42), $P = 0.001$], and the DeepCoxPH multiplication model [HR (95% CI) = 1.67 (1.28–2.19), $P < 0.001$]. The postestimation results showed that the DeepCoxPH addition model obtained lower AIC and BIC values (AIC = 2160.110, BIC = 2164.206). Analysis with the Kaplan–Meier method indicated that the DeepCoxPH addition model has superior risk stratification ability than the CoxPH HR and DL-weight schemas. Overall, the DeepCoxPH addition model has superior risk stratification ability for 10-year overall survival in breast cancer patients through indication of the risk degrees of 15 clinicopathological factors.

TABLE 2. CoxPH regression model for overall survival in breast cancer.

Variable	Comparison	Univariate			Multivariate		
		Crude-HR	95% CI	*P*	Adjust-HR	95% CI	*P*
Age	> 50 *vs.* ≤ 50	0.84	0.64-1.09	0.190	0.99	0.74-1.32	0.941
Grade	III *vs.* I-II	1.64	1.25-2.16	<0.001	1.50	1.12-2.02	<0.001
HER2	Negative *vs.* Positive	0.90	0.69-1.17	0.417	1.30	0.93-1.81	0.131
ER	Negative *vs.* Positive	1.36	1.04-1.78	0.026	0.78	0.43-1.40	0.405
PR	Negative *vs.* Positive	1.57	1.21-2.05	0.001	1.79	1.17-2.73	0.007
Tumor size (AJCC stage)	T3/T4 *vs.* T1/T2	1.74	1.22-2.49	0.002	1.60	1.04-2.46	0.034
Lymph node (AJCC stage)	Positive *vs.* Negative	1.46	1.11-1.91	0.006	1.09	0.78-1.51	0.626
Lymphovascular invasion	Positive *vs.* Negative	1.48	1.14-1.93	0.003	1.07	0.78-1.46	0.680
Dermal invasion	Positive *vs.* Negative	1.57	1.06-2.31	0.023	1.08	0.68-1.73	0.748
Perineural invasion	Positive *vs.* Negative	1.45	1.06-2.00	0.021	1.45	1.02-2.05	0.038
Operation	Total mastectomy *vs.* Partial mastectomy	1.49	1.13-1.97	0.004	1.36	0.95-1.94	0.092
Radiotherapy (RT)	Without RT *vs.* With RT	0.91	0.70-1.19	0.491	0.83	0.60-1.16	0.270
Chemotherapy (CT)	Without CT *vs.* With CT	0.84	0.61-1.16	0.283	1.11	0.78-1.59	0.553
Hormone therapy (HT)	Without HT *vs.* With HT	1.33	1.02-1.73	0.034	1.07	0.65-1.76	0.802
Target therapy (TT)	Without TT *vs.* With TT	0.61	0.44-0.85	0.003	0.55	0.36-0.84	0.005

ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2.

TABLE 3. Estimated risk score schemas based on DL and Cox multivariate regression model.

Variable	Comparison	Risk score schema			
		CoxPH HR model	DL-weight model	Multiplication model	Addition model
Age	< 50 *vs.* ≥ 50	0.55 (R.D. = 1)	0.14 (R.D. = 1)	1	2
Grade	II *vs.* I	0.78 (R.D. = 1)	0.15 (R.D. = 3)	3	4
HER2	Negative *vs.* Positive	0.83 (R.D. = 2)	0.02 (R.D. = 1)	2	3
ER	Negative *vs.* Positive	0.99 (R.D. = 2)	0.53 (R.D. = 3)	6	5
PR	Negative *vs.* Positive	1.07 (R.D. = 3)	0.07 (R.D. = 3)	9	6
Tumor size	T3/T4 *vs.* T1/T2	1.07 (R.D. = 3)	0.12 (R.D. = 2)	6	5
Lymph node	Positive *vs.* Negative	1.08 (R.D. = 2)	0.24 (R.D. = 2)	4	4
Lymphovascular invasion	Positive *vs.* Negative	1.09 (R.D. = 1)	0.03 (R.D. = 3)	3	4
Dermal invasion	Positive *vs.* Negative	1.11 (R.D. = 3)	0.16 (R.D. = 1)	3	4
Perineural invasion	Positive *vs.* Negative	1.30 (R.D. = 1)	0.19 (R.D. = 2)	2	3
Operation	Total mastectomy *vs.* Partial mastectomy	1.36 (R.D. = 3)	0.04 (R.D. = 2)	6	5
Radiotherapy (RT)	Without RT *vs.* With RT	1.45 (R.D. = 2)	0.02 (R.D. = 2)	4	4
Chemotherapy (CT)	Without CT *vs.* With CT	1.50 (R.D. = 1)	0.28 (R.D. = 1)	1	2
Hormone therapy (HT)	Without HT *vs.* With HT	1.60 (R.D. = 3)	0.06 (R.D. = 1)	3	4
Target therapy (TT)	Without TT *vs.* With TT	1.79 (R.D. = 2)	0.09 (R.D. = 3)	6	5

ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2; CoxPH HR, Hazard ratio derived from CoxPH (multivariate model); DL-weight, weight derived from deep learning approach; R.D., risk degree; Multiplication model, CoxPH HR (R.D.) × DL-weight (R.D.); Addition model, CoxPH HR (R.D.) + DL-weight (R.D.).

IV. DISCUSSION

In this study, we introduced a DeepCoxPH method for identifying overall breast cancer survival risk stratification. For this, CoxPH can estimate the HRs of all clinicopathological variables for overall survival. The HRs can determine the effect of individual clinicopathological variables. DL produced a classification and then obtained a deep abstraction to identify the factor importance associated with risk stratification. The HRs from CoxPH and deep abstraction from DL were ranked in ascending order to generate the HR and DL weights. DeepCoxPH employed matrix operations to combine the HR and DL weights and then transformed them into risk score estimations. The DeepCoxPH method was based on a DL approach to address weak assessment of risk and interaction effects associated with CoxPH methods. The DL weights potentially include the interaction effects between molecular subtypes and treatment effects that are generally neglected by CoxPH. Applying DeepCoxPH, parameters derived from CoxPH are improved by abstracted DL weights, enabling DeepCoxPH to achieve more comprehensive risk weight estimation for overall survival. Performance evaluation based on 10-year overall survival using breast cancer datasets confirmed that DeepCoxPH satisfactorily identified risk effects among multiple clinicopathological factors.

We employed sensitivity and specificity to compare the performance of the DeepCoxPH model, DL model, and CoxPH model. Sensitivity and specificity are statistical classification functions of the performance of a binary classification test. Sensitivity measures the proportion of actual progression correctly identified as high-risk stratification—also known as the true positive rate. Specificity measures the proportion of actual survival correctly identified as low-risk stratification—also known as the true negative rate.

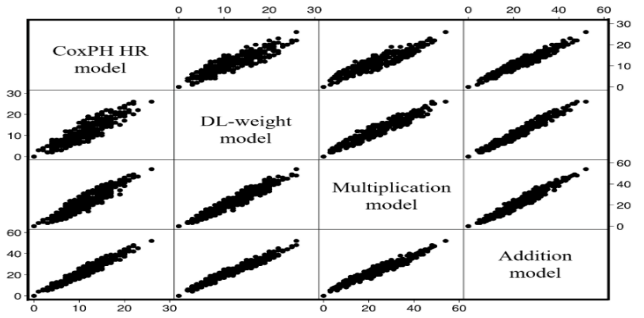

FIGURE 3. The relationship between risk scores from CoxPH HR, DL-weight, deepCoxPH (multiplication), and deepCoxPH (addition) models.

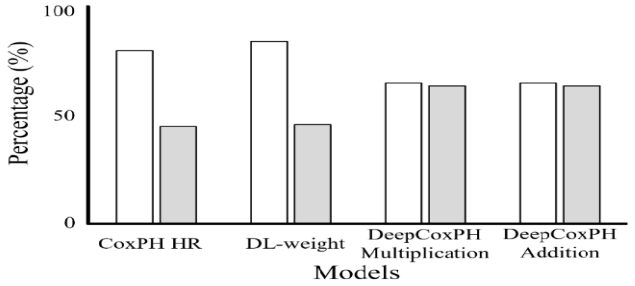

FIGURE 4. Comparison of risk score schema for risk stratification by sensitivity and specificity. The white bars indicate the sensitivities of the four models, and the gray bars indicate their specificities.

The high sensitivity results in the DL and CoxPH models may have contributed to the higher precision in high-risk subject recognition; however, the low specificity may have led to more false positive results. Although results with high sensitivity are desired, low specificity can lead to clinical burden. The DeepCoxPH model obtained higher sensitivity and specificity than did the CoxPH HR model. Hence, we combined the DL and CoxPH approach to improve the stratification ability for breast cancer overall survival risk classification. Our results revealed that DeepCoxPH performed effectively as a stratification approach and provided more comprehensive risk weighting in overall survival estimation. Moreover, the DeepCoxPH model results provided superior recognition in both low- and high-risk stratification for breast cancer overall survival.

All risk stratification schema used a fully adjusted model to achieve more comprehensive estimation of risk for all clinicopathological factors. Alongside advances in oncology research, inclusion of genomes and other biological factors must be verified through similar approaches in further studies. DeepCoxPH retained the benefits of DL and CoxPH models. First, combined use of CoxPH and DL in a fully adjusted model could provide comprehensive risk estimation owing to the advantages of both methods. An alternative risk stratification schema could provide superior alternatives in clinical practice and could be used as an assisted alert management tool for clinical follow-up risk assessment. Using an appropriate risk stratification schema for breast cancer characteristics could assist oncologists and clinicians in follow-up and making treatment decisions. Second, DeepCoxPH is a model-free

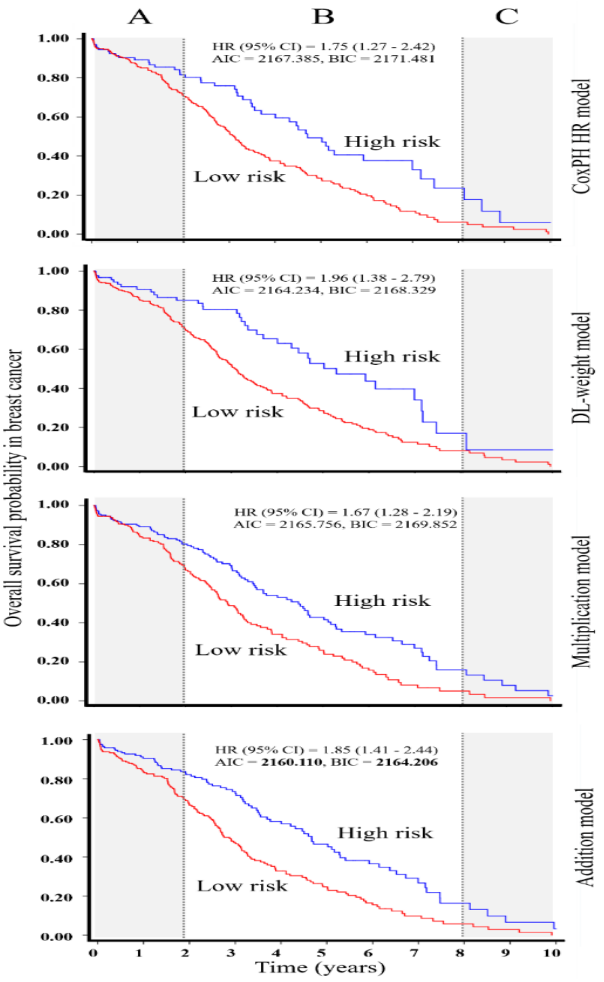

FIGURE 5. Comparison of Kaplan-Meier curves based on four models for high-risk and low-risk strata based on risk stratification.

method that does not require a specific mode of inheritance. Designing suitable feature extractors by using the conventional CoxPH model requires considerable domain expertise related to the disease in question; however, this requirement is unnecessary when suitable features are learned automatically through a DL procedure. Third, DL deep networks have nonparametric advantages over conventional learning algorithms that do not use distributed representations. A nonparametric method does not require that the data distribution be assumed before statistical analysis; this prevents problems associated with the use of parametric statistics for overall survival.

We evaluated overall survival in breast cancer by means of the common clinicopathological factors through both a machine learning and a parametric-statistical approach, making use of the TBCC database [21]. Other studies have proposed various prognostic factors for breast cancer overall survival, in which these clinicopathological factors usually played a crucial role in clinical treatment decisions [22]. The most common clinicopathological factors are the molecular subtypes of breast cancer, including ER, PR, HER2, Ki-67 level, and *p53* gene mutation [23], [24]. Apart from molecular weight, tumor burden factors in breast cancer survival

prognosis such as tumor size, lymph node metastases, dermal invasion, and LVI are also considered major discriminant factors in breast cancer overall survival [25].

Advances in oncological medicine have generally prolonged overall survival in breast cancer, resulting in long-term follow-up of patients with breast cancer. Higher breast cancer proliferation grade generally show poor survival [26], [27]. Similarly, our results indicated that grade III subjects had a poor survival outcome in terms of 10-year breast cancer overall survival. Our findings for tumor size and lymph node invasion were also consistent with those of other studies [28]. Breast cancer molecular subtype played a major role in breast cancer prognosis which is mainly based on ER, PR, and HER2 expression in tumor tissue [29]–[31]. Target therapy has been the main treatment strategy, especially for HER2 overexpression in breast cancer [32]. The CoxPH HR model gives a risk degree of 2 and DL-weight model gives a risk degree of 3 for patients without target therapy. We assume the DL-weight model might make a comprehensive risk consideration which is potentially including the interaction effect between molecular subtype and treatment effect that are generally neglected by CoxPH model [33]. Total mastectomy showed a higher mortality risk in our results, possibly because most total mastectomy patients would have a larger tumor size or more malignant characteristics according to the total mastectomy indications. Whereas LVI and dermal invasion showed poor survival in breast cancer, consistent with other findings [28], [34], hormone therapy showed a good overall survival in breast cancer [35], [36].

V. CONCLUSION

We proposed a new DeepCoxPH method for breast cancer overall survival risk stratification based on common clinicopathological factors. The matrix operations in DeepCoxPH were able to combine the abstracted risk weights from both the CoxPH and DL models and then exhibit more comprehensive risk weight estimation for overall survival to yield enhanced risk stratification performance compared with either the single CoxPH model or the DL model. Overall, survival estimation usually neglects the potential interaction between clinicopathological factors in a CoxPH model and does not consider the time effects in a DL model. Thus, the DeepCoxPH method offers the advantage of simultaneously combining both risk weights in the CoxPH and DL models and of providing more precise risk stratification results for overall survival using common breast cancer clinicopathological factors.

REFERENCES

[1] C.-H. Yang, L.-Y. Chuang, and Y.-D. Lin, "Multiobjective multifactor dimensionality reduction to detect SNP–SNP interactions," *Bioinformatics*, vol. 34, pp. 2228–2236, Feb. 2018.

[2] C.-H. Yang, L.-Y. Chuang, and Y.-D. Lin, "CMDR based differential evolution identifies the epistatic interaction in genome-wide association studies," *Bioinformatics*, vol. 33, pp. 2354–2362, May 2017.

[3] T. L. M. Lee and J. H. Chiang, "A systems biology approach to solving the puzzle of unknown genomic gene-function association using grid-ready SVM committee machines," *IEEE Comput. Intell. Mag.*, vol. 7, no. 4, pp. 46–54, Nov. 2012.

[4] O.-Y. Fu, H.-W. Chang, Y.-D. Lin, L.-Y. Chuang, M.-F. Hou, and C.-H. Yang, "Breast cancer-associated high-order SNP-SNP interaction of CXCL12/CXCR4-related genes by an improved multifactor dimensionality reduction (MDR-ER)," *Oncol. Rep.*, vol. 36, no. 3, pp. 1739–1747, 2016.

[5] J. J. Driscoll and O. Rixe, "Overall survival: Still the gold standard: Why overall survival remains the definitive end point in cancer clinical trials," *Cancer J.*, vol. 15, pp. 401–405, Sep. 2009.

[6] E. Basch *et al.*, "Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment," *JAMA*, vol. 318, no. 2, pp. 197–198, 2017.

[7] C. H. Yang *et al.*, "Interaction of MRE11 and clinicopathologic characteristics in recurrence of breast cancer: Individual and cumulated receiver operating characteristic analyses," *Biomed. Res. Int.*, vol. 2017, Jan. 2017, Art. no. 2563910.

[8] S. H. Moi *et al.*, "Cumulative receiver operating characteristics for analyzing interaction between tissue visfatin and clinicopathologic factors in breast cancer progression," *Cancer Cell Int.*, vol. 18, no. 1, 2018, Art. no. 19.

[9] A. Athreya *et al.*, "Augmentation of physician assessments with multi-omics enhances predictability of drug response: A case study of major depressive disorder," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 20–31, Aug. 2018.

[10] F.-C. Liu, H.-T. Lin, C.-F. Kuo, L.-C. See, M.-J. Chiou, and H.-P. Yu, "Epidemiology and survival outcome of breast cancer in a nationwide study," *Oncotarget*, vol. 8, no. 10, pp. 16939–16950, 2017.

[11] R. Sharma, "Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: Evidence from global burden of disease study 2016," *Breast Cancer*, to be published. doi: 10.1007/s12282-018-00941-4.

[12] M. Abubakar *et al.*, "Breast cancer risk factors, survival and recurrence, and tumor molecular subtype: Analysis of 3012 women from an indigenous Asian population," *Breast Cancer Res.*, vol. 20, no. 1, 2018, Art. no. 114.

[13] L. L. Visser *et al.*, "Clinicopathological risk factors for an invasive breast cancer recurrence after ductal carcinoma *in situ*—A nested case–control study," *Clin. Cancer Res.*, vol. 24, pp. 3593–3601, Aug. 2018.

[14] S. Park *et al.*, "Risk factors associated with distant metastasis and survival outcomes in breast cancer patients with locoregional recurrence," *J. Breast Cancer*, vol. 18, no. 2, pp. 160–166, 2015.

[15] D. R. Cox, *Analysis of Survival Data*. New York, NY, USA: Routledge, 2018.

[16] A. Linden and P. R. Yarnold, "Modeling time-to-event (survival) data using classification tree analysis," *J. Eval. Clin. Pract.*, vol. 23, no. 6, pp. 1299–1308, 2017.

[17] N. Ternes, F. Rotolo, and S. Michiels, "Biospear: An R package for biomarker selection in penalized Cox regression," *Bioinformatics*, vol. 34, no. 1, pp. 112–113, 2018.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, May 2015, Art. no. 436.

[19] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief Bioinform.*, vol. 18, no. 5, pp. 851–869, 2016.

[20] D. Guo and Y. Zhang, "Novel recurrent neural network for time-varying problems solving," *IEEE Comput. Intell. Mag.*, vol. 7, no. 4, pp. 61–65, Nov. 2012.

[21] F. O. Yang *et al.*, "Efficacy and toxicity of pegylated liposomal doxorubicin-based chemotherapy in early-stage breast cancer: A multicenter retrospective case–control study," *Asia–Pacific J. Clin. Oncol.*, vol. 14, no. 3, pp. 198–203, 2018.

[22] N. J. Bundred, "Prognostic and predictive factors in breast cancer," *Cancer Treat. Rev.*, vol. 27, pp. 137–142, Jun. 2001.

[23] E. M. Ibrahim, M. E. Al-Foheidi, M. M. Al-Mansour, and G. A. Kazkaz, "The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: A meta-analysis," *Breast Cancer Res. Treat.*, vol. 148, pp. 467–476, Dec. 2014.

[24] A. M. Rosales, M. C. Molina, J. G. Moo, P. R. Morelos, and C. Bandala, "Breast cancer metastasis associations with clinicopathological characteristics in mexican women younger than 40 years of age," *Asian Pacific J. Cancer Prevention*, vol. 17, no. 11, pp. 5019–5023, 2016.

[25] H. A. El-mezayen, F. M. Metwally, and H. Darwish, "A novel discriminant score based on tumor-associated trypsin inhibitor for accurate diagnosis of metastasis in patients with breast cancer," *Tumor Biol.*, vol. 35, pp. 2759–2767, Mar. 2014.

[26] D. E. Henson, L. Ries, L. S. Freedman, and M. Carriaga, "Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index," *Cancer*, vol. 68, no. 10, pp. 2142–2149, 1991.
[27] G. Contesso, H. Mouriesse, S. Friedman, J. Genin, D. Sarrazin, and J. Rouesse, "The importance of histologic grade in long-term prognosis of breast cancer: A study of 1,010 patients, uniformly treated at the Institut Gustave-Roussy," *J. Clin. Oncol.*, vol. 5, no. 9, pp. 1378–1386, 1987.
[28] T. J. A. Dekker *et al.*, "Quantitative assessment of lymph vascular space invasion (LVSI) provides important prognostic information in node-negative breast cancer," *Ann. Oncol.*, vol. 24, no. 12, pp. 2994–2998, 2013.
[29] M. H. Li, C. L. Hou, C. Wang, and A. J. Sun, "HER-2, ER, PR status concordance in primary breast cancer and corresponding metastatic lesion in lymph node in Chinese women," *Pathol.-Res. Pract.*, vol. 212, no. 4, pp. 252–257, 2016.
[30] E. J. Campbell, M. Tesson, F. Doogan, Z. M. A. Mohammed, E. Mallon, and J. Edwards, "The combined endocrine receptor in breast cancer, a novel approach to traditional hormone receptor interpretation and a better discriminator of outcome than ER and PR alone," *Brit. J. Cancer*, vol. 115, pp. 967–973, Oct. 2016.
[31] J. Poorolajal, N. Nafissi, M. E. Akbari, H. Mahjub, N. Esmailnasab, and E. Babaee, "Breast cancer survival analysis based on immunohistochemistry subtypes (ER/PR/HER2): A retrospective cohort study," *Arch. Iranian Med.*, vol. 19, no. 10, pp. 680–686, 2016.
[32] R. Amodio *et al.*, "Target therapy in HER2-overexpressing breast cancer patients," *OMICS, A J. Integr. Biol.*, vol. 15, no. 6, pp. 363–367, 2011.
[33] D. J. Ferguson and P. Meier, "Lumpectomy or total mastectomy for breast cancer," *New England J. Med.*, vol. 321, no. 10, pp. 689–690, 1989.
[34] G. Gruber, M. Ciriolo, H. J. Altermatt, S. Aebi, G. Berclaz, and R. H. Greiner, "Prognosis of dermal lymphatic invasion with or without clinical signs of inflammatory breast cancer," *Int. J. Cancer*, vol. 109, no. 1, pp. 144–148, 2004.
[35] T. S. Mikkola *et al.*, "Reduced risk of breast cancer mortality in women using postmenopausal hormone therapy: A Finnish nationwide comparative study," *Menopause-J. North Amer. Menopause Soc.*, vol. 23, no. 11, pp. 1199–1203, 2016.
[36] N. Obi *et al.*, "Relationship between menopausal hormone therapy and mortality after breast cancer The MARIE*plus* study, a prospective case cohort," *Int. J. Cancer*, vol. 138, pp. 2098–2108, May 2016.

CHENG-HONG YANG (M'00–SM'03) received the M.S. and Ph.D. degrees in computer engineering from North Dakota State University, in 1988 and 1992, respectively. He is currently a Chair Professor with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. He is also with the Ph.D. Program in Biomedical Engineering, Kaohsiung Medical University, Taiwan. He has authored or coauthored over 380 refereed publications and a number of book chapters. His research interests include evolutionary computation, optimization, bioinformatics, data analysis, and their applications. He is an Editorial Board Member of other international journals. He is a Fellow of the Institution of Engineering and Technology and the American Biographical Institute.

SIN-HUA MOI received the M.S. degree from the Department of Oral Hygiene, Kaohsiung Medical University, Taiwan, in 2014. She is currently pursuing the Ph.D. degree with the National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan. She is specialized in biomedicine statistical analysis, database management, and machine learning analysis. She has authored or coauthored over 20 refereed publications. Her research interests include bioinformatics and biostatistics.

FU OU-YANG received the M.D. and Ph.D. degrees from Kaohsiung Medical University, Kaohsiung, Taiwan. He is currently the Director and a Visiting Staff with the Division of Breast Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung Medical University. He has authored or coauthored over 40 refereed publications. His research interests include bioinformatics, molecular oncology, breast surgery, and microscopy surgery. He is an Editorial Board Member of other international journals.

LI-YEH CHUANG received the M.S. degree from the Department of Chemistry, The University of North Carolina, in 1989, and the Ph.D. degree from the Department of Biochemistry, North Dakota State University, in 1994. She is currently a Professor with the Department of Chemical Engineering and the Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan. She has authored or coauthored over 300 refereed publications. Her research interests include bioinformatics, biochemistry, and genetic engineering.

MING-FENG HOU received the M.D. degree from Kaohsiung Medical University, Kaohsiung, Taiwan. He is currently the Superintendent of Kaohsiung Medical University Hospital and a Professor with the Department of Surgery, College of Medicine, Kaohsiung Medical University. He has authored or coauthored over 100 refereed publications, including Nature and Lancet. His research interests include clinical oncology and breast surgery. He is an Editorial Board Member of other international journals.

YU-DA LIN (M'17) received the M.S. and Ph.D. degrees in information science from the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan, in 2011 and 2015, respectively. He is currently a Postdoctoral Fellow with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology. He is also a Software Engineer and an Adjunct Assistant Professor. He has authored or coauthored over 80 refereed publications. His main research interests include artificial intelligence, biomedical informatics, bioinformatics, and computational biology. He is a member of the IEEE Tainan Section, the IEEE Young Professionals, and the IEEE Computational Intelligence Society Membership.