IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 4, 2019, accepted August 26, 2019, date of publication September 16, 2019, date of current version September 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941543

Accurate Prediction of Neoadjuvant Chemotherapy Pathological Complete Remission (pCR) for the Four Sub-Types of Breast Cancer

XIN FENG[1], LELIAN SONG[2], SHAOFEI WANG[3], HAOQIU SONG[1,4], HANG CHEN[3], YUXUAN LIU[1,3], CHENWEI LOU[3], JIAN ZHAO[3], QUEWANG LIU[3], YANG LIU[3], RUIXUE ZHAO[3], KAI XING[3], SIJIE LI[2], YUNHE YU[2], ZHENYU LIU[2], CHENGYANG YIN[2], BING HAN[2], YE DU[2], RUIHAO XIN[5], LAN HUANG[3], ZHIMIN FAN[2], AND FENGFENG ZHOU[3], (Senior Member, IEEE)

[1]Cancer Systems Biology Center, China–Japan Union Hospital, Jilin University, Changchun 130033, China
[2]Department of Breast Surgery, The First Hospital of Jilin University, Changchun 130012, China
[3]Key Laboratory of Symbolic Computation and Knowledge Engineering, BioKnow Health Informatics Laboratory, College of Computer Science and Technology, Ministry of Education, Jilin University, Changchun 130012, China
[4]School of Computer Science, Hubei University of Technology, Wuhan, China
[5]College of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130012, China

Corresponding authors: Zhimin Fan (fanzhimn@163.com) and Fengfeng Zhou (fengfengzhou@gmail.com)

This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDB13040400, in part by the Jilin Provincial Key Laboratory of Big Data Intelligent Computing under Grant 20180622002JC, in part by the Education Department of Jilin Province under Grant JJKH20180145KJ, in part by the startup grant of Jilin University, in part by the Bioknow MedAI Institute under Grant BMCPP-2018-001, in part by the High Performance Computing Center, Jilin University, China, in part by the Multicenter Clinical Cohort Study of Breast Cancer under Grant 2016YFCO901302, and in part by the Breast Cancer biological Sample Database and Clinical Diagnosis and Treatment Information Database Construction under Grant 2016YFCO901304.

ABSTRACT Neoadjuvant chemotherapy (NAC) has become the main treatment option for breast cancer. Its adverse drug reactions (ADRs) make NAC painful both physiologically and psychologically. The factor pathological complete remission (pCR) describes how well a series of six or more chemotherapeutic treatments works on a patient. This study investigated the possibility of predicting pCR using only the nodal sizes of the first three treatments. A best feature combination for each breast cancer subtype was screened from the real nodal sizes of the first three treatments and the nodal sizes' of the next three treatments predicted from those of the first three ones. The prediction was evaluated by the metrics Avc = (sensitivity + specificity)/2. A triple-negative breast cancer (TN) patient may have an estimation of pCR Avc = 0.8696 after taking just three treatments. At least Avc = 0.7594 was achieved for all the four breast cancer subtypes investigated in this study.

INDEX TERMS Pathological complete response (pCR), breast cancer, neoadjuvant chemotherapy, biomarker detection, feature selection.

I. INTRODUCTION

Breast cancer is one of the most frequently occurred cancer type for females [1], [2] and exceeds the combined incidences of the next three top-ranked female cancer types in the United States, i.e., lung & bronchus, colon & rectum, and uterine corpus [3]. It is also the top-ranked cancer type for females in China [4]–[6] and its incidence rate keeps increasing since 2000 [7]. Despite the invention and deployment of many diagnosis and treatment techniques [8], [9], breast cancer is among the top two ranked cancer types for annual mortalities in both countries [3], [7]. Tumor size is one of the main factors positively correlated with the long-term mortality [10]–[13] and it is recommended to use various screening technologies to detect breast cancer smaller than 2 cm [14], [15].

Subtypes of breast cancer varied significantly in both pathological phenotypes and have different recommended treatment plans [16]–[18]. Although the incidence rates of all the breast cancer subtypes increase with the age, they tend to have different tumor sizes. The subtype luminal A tends to have a smaller tumor size than the subtype luminal-HER2. For example, 63% of the luminal A patients have tumor sizes smaller than 2 cm in the British Columbia cohort,

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

while 61% of the luminal-HER2 patients have tumors larger than 2 cm [16]. This supports the two previous observations that the subtype luminal A has a longer median survival duration than the subtype luminal-HER2 [19], [20], and there is a positive correlation between the tumor size and mortality [10], [21]–[23]. So a conservative radiation treatment is usually recommended for the subtype luminal A while luminal-HER2 patients need both mastectomy and radiotherapy [16]. Adjuvant chemotherapy is one of the main effective treatment options for breast cancer patients with or without mastectomy [24]. It was reported that a combination of multiple treatment technologies may significantly improve the survival durations of breast cancer patients [25]. While neoadjuvant chemotherapy demonstrated its effectiveness to a wide spectrum of breast cancer subtypes, the drug resistance may rapidly develop and the tumor recurred with a high rate [26], [27]. Pathologic complete response (pCR) was defined to describe the status of no invasive and no in situ residuals in the breast tissue [28], [29], and breast cancer patients with a positive pCR achieved a much better disease-free and overall survival rates [30].

This study investigated the pCR prediction problem using only the data of the first three neoadjuvant chemotherapeutic treatments. Despite its effectiveness, neoadjuvant chemotherapy is notorious for the adverse drug reactions (ADRs), including neutropenia [31], memory retrieval impairment [32], muscle and joint pain [33], and cognitive impairment [34], [35], etc. This study firstly trained a classification model for the factor pCR using the clinical data of the first three treatments. Then the data of the next three treatments were predicted by a personalized regression model using those of the first three treatments. The pCR classification model was refined by these predicted data. The proposed algorithm demonstrated satisfying pCR prediction accuracies and achieved the best accuracy for the subtype TN.

II. MATERIALS AND METHODS

A. DATA COLLECTION

This study manually curated the database of the electronic medical records and retrospectively collected the clinical data of 495 breast cancer patients diagnosed and treated with neoadjuvant chemotherapy (NAC) during 2011-2016 at the Department of Breast Surgery in the First Hospital of the Jilin University. 2 male patients were excluded from further analysis, as summarized in Table 1.

NAC was usually taken to reduce the tumor size before a radical treatment and mostly consisted of sequentially six to eight treatments [36]–[38]. This study hypothesized that the data of the first six NAC treatments were significantly associated with the final pCR status. So this study utilized the clinical data of at least six NAC treatments for each patient. Only the patients with at least six treatments were kept for further analysis. 115 patients with missing clinical data were removed. So among the 378 female patients with at least six chemotherapy treatments, there are 5 sub-types of breast cancer. The sub-type luminal A has only 27 patients, and was excluded from further analysis due to the limited number of samples. The two sub-types of luminal B with HER2 positive and negative have 74 and 154 patients, respectively. They were denoted as "LBP" and "LBN". The sub-type HER2 (+) has 63 patients and was denoted as HER2. There are 60 triple-negative sub-type patients and this group of patients was denoted as "TN". The final cohort has 351 female breast cancer patients from the four sub-types, i.e., LBP, LBN, HER2 and TN. This study was approved by the Institutional Review Board (IRB) of the First Hospital of the Jilin University. This study analyzed the archived medical records and individual patient consent was waived by the IRB of the First Hospital of the Jilin University, including the approval for the waiver request of authorization to collect protected health information.

TABLE 1. Numbers of samples for the four breast cancer subtypes, i.e., TN, HER2, LBP, and LBN. The three columns for the dataset "Raw" gave the sample numbers within the original cohort, and the three columns for the dataset "6 + Treatments" gave the sample numbers of the final cohort investigated in this study.

Subtype	Raw			6+Treatments		
	pCR=1	pCR=0	Total	pCR=1	pCR=0	Total
TN	36	55	91	24	36	60
HER2	32	53	85	23	40	63
LBP	21	78	99	17	57	74
LBN	22	198	220	15	139	154

B. DATA PREPROCESSING

Each patient has the breast cancer nodal sizes for each consecutive chemotherapeutic treatment and the clinical pCR diagnosis. This study assumed that the tumor nodals were rectangles, and a nodal size was measured by the product of the largest diameter and its middle vertical diameter in millimeters. If a patient carried multiple nodals, the sum of the nodal sizes was calculated as the feature for that specific chemotherapeutic treatment of that patient. We observed that some nodal sizes were zero and added a pseudo-one to all the nodal sizes to avoid the division by zeros.

C. PROBLEM SETTING AND PREDICTION ALGORITHMS

The pCR prediction was modelled as a binary classification problem of samples with pCR = 1 or 0 [39], [40]. Samples were grouped as positive and negative ones, if the pCRs values were 1 and 0, respectively. That is to say, the value pCR is 1 if the specific patient achieved the pathological complete remission (pCR) after the six neoadjuvant chemotherapy (NAC) treatments. Otherwise pCR is 0 if the patient didn't achieve the pathological complete remission after all the six NAC treatments.

Six representative classifiers were evaluated for their prediction performances using sensitivity (Sn), specificity (Sp), Accuracy (Acc), balanced accuracy Avc = (Sn + Sp)/2, and Matthew's correlation coefficient (MCC) [41]–[45]. Sn and Sp were defined as the ratios of correctly predicted positive and negative samples, respectively. Acc was the ratio of

correctly predicted samples. MCC was calculated between the samples' predicted pCR values and their real pCR values, and was between −1 and +1. The higher MCC, the better a prediction was.

The six representative classifiers were Support Vector Machine (SVM), k nearest neighbors (KNN), Naïve Bayes (NBayes), Decision Tree (DTree), Random Forest (RF) and extreme gradient boosting (XGB).

The detailed description of the problem setting and prediction algorithms may be found in the Supplementary Materials at http://www.healthinformaticslab.org/supp/.

III. RESULTS

A. SUBTYPE TN: pCR PREDICTION BASED ON THE FIRST THREE TREATMENTS

This study firstly investigated whether a patient's pathological complete remission (pCR) may be predicted by taking only the first three neoadjuvant chemotherapeutic treatments. Each treatment was described by the total area of all the lesion nodes. Six different representative classifiers were utilized to calculate the stratified 5-fold cross validation performances [46]–[51] of the binary classification problem between the subtype TN breast cancer patients with and without pCR, i.e., pCR = 1 and 0, respectively. Three prediction performance metrics were used to compare the six classifiers on the given dataset, *i.e.*, balanced accuracy Avc = (Sn + Sp)/2 [44], [45], [52], sensitivity (Sn) and specificity (Sp), as shown in Figure 1 (a). XGB achieved the best classification Avc = 0.7990 for predicting pCR of the TN breast cancer patients. This suggested that the proposed model may accurately predict whether an TN breast cancer patient would achieve the pathological complete remission (pCR = 1). The next two best classification algorithms are DTree and RF, and both of these two algorithms achieved a slightly worse Avc, about 2% decrease. The second best classifier DTree achieved a slightly worse Sn = 0.7300 than that (Sn = 0.7933) of XGB, but a better Sp = 0.8631 than that (Sp = 0.8048) of XGB. The metrics Sn described how accurately those patients with pCR = 1 may be detected, while Sp described the detection accuracy of the patients with poor prognosis (pCR = 0).

Firstly, the exploratory experimental data suggested that the subtype-TN breast cancer patients may be split into the groups of good (pCR = 1) or poor (pCR = 0) prognosis with reasonable prediction performances, in the measurements of Avc, Sn and Sp. Secondly, the clinicians may choose a classification model based on whether they were more interested in Sn or Sp.

So this study conducted a comprehensive evaluation of the binary classification model for the other three breast cancer subtypes.

B. THE OTHER THREE SUBTYPES: pCR PREDICTION BASED ON THE FIRST THREE TREATMENTS

The pCR of the subtype HER2-positive patients can also be satisfyingly predicted using the data of the first three

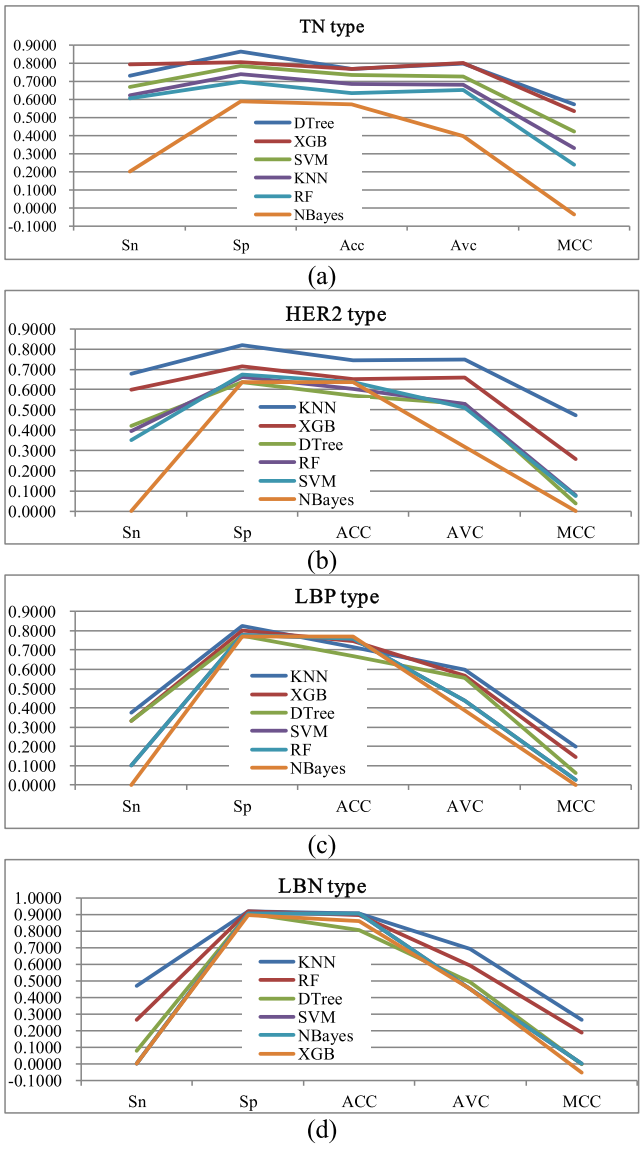


FIGURE 1. Predicting pCR by the first three chemotherapeutic treatments of breast cancer patients. The five classification performance measurements of all the six classifiers were calculated here for the subtypes (a) TN, (b) HER2, (c) LBP, and (d) LBN.

treatments, as shown in Figure 1 (b). The algorithm KNN achieved the best prediction Avc = 0.7493 and MCC = 0.4725. KNN also outperformed all the other five classifiers in the prediction metrics Sn and Sp. This suggested that after the first three chemotherapeutic treatments, a HER2-positive subtype patient may get an estimation of whether she can achieve the pathological complete remission (pCR = 1) by taking three or more treatments. The data in Figure 1 (b) also demonstrated that the other classification algorithms performed much worse than KNN, with at least a decrease 0.0765 in Avc.

Figure 1 (c) and (d) showed that the other two subtypes LBP and LBN had worse pCR prediction performances, which may be due to the limited number of positive samples. The algorithm KNN achieved the best pCR prediction Avc

0.5989 and 0.6930 for the two subtypes LBP and LBN, respectively. Although KNN outperformed all the other five classifiers in both Sn and Sp, its prediction sensitivities (Sn) were smaller than 0.5000, and required a reasonable improvement. So this study investigated how our pCR prediction models may be improved in the following sections.

C. ESTIMATING THE LESION AREAS AFTER THE FIRST THREE TREATMENTS

We hypothesized that a pCR prediction model may perform better using data of more treatments. Most patients in our cohort received 6 chemotherapeutic treatments. But chemotherapeutic treatment may induce various adverse drug reactions (ADRs) [31]–[34]. So we chose to estimate the nodal sizes of the next three treatments using the data of the first three ones.

A simple regression function $y = ax^b + c$ was chosen to predict nodal sizes of the treatments 4/5/6 using the data of the treatments 1/2/3. The three parameters a, b and c were determined by the training data of the first three treatments. The nodal sizes of some patients reached 0 before treatment 6, which caused program errors for the regression training. So a pseudo-one was added to all the nodal sizes during the training step and was subtracted in the final prediction results. A prediction of the nodal size was defined as a correct prediction if the predicted size has a difference to the real size smaller than 20%.

The regression model was trained on the data of three consecutive treatments and predicted the next treatment. The real data of only the first three treatments were used. So treatment 4 was predicted by the model trained on the real data of treatments 1/2/3. Treatment 5 was predicted by the model trained on the real data of treatments 2/3 and the predicted data of treatment 4. Treatment 6 was predicted by the model trained on the real data of treatment 3 and the predicted data of treatments 4/5.

D. OPTIMIZING THE NODAL SIZE PREDICTION MODELS

All the six binary classification algorithms were evaluated using the real data of treatments 1/2/3 and the regressed data of treatments 4/5/6, as shown in Figure 2. The classification algorithm XGB still achieved better than the other four classifiers (except for RF) on the subtype TN (Figure 2 (a)), and its Avc reached 0.7619. But XGB achieved Avc = 0.7990 for the subtype TN using only the real data of the first three treatments, as shown in Figure 2 (a). The best Avc = 0.7748 was achieved by the classifier RF, which also achieved the best Sn = 0.7933. But RF's specificity (Sp) was worse than the three classifiers SVM, XGB and DTree.

The classifiers RF and XGB achieved the best prediction performances in all the three metrics Avc/Sn/Sp for the subtypes HER2 and LBP, as shown in Figure 2 (b) and (c). The prediction performance measurement mAvc was decreased by 0.1026 and 0.0673 by adding the predicted data of treatments 4/5/6 to the training data. The classifier XGB achieved the best Avc = 0.6257 but its specificity (Sp = 0.9181) was slightly worse than that (Sp = 0.9188) of the classifier KNN, as shown in Figure 2 (d).

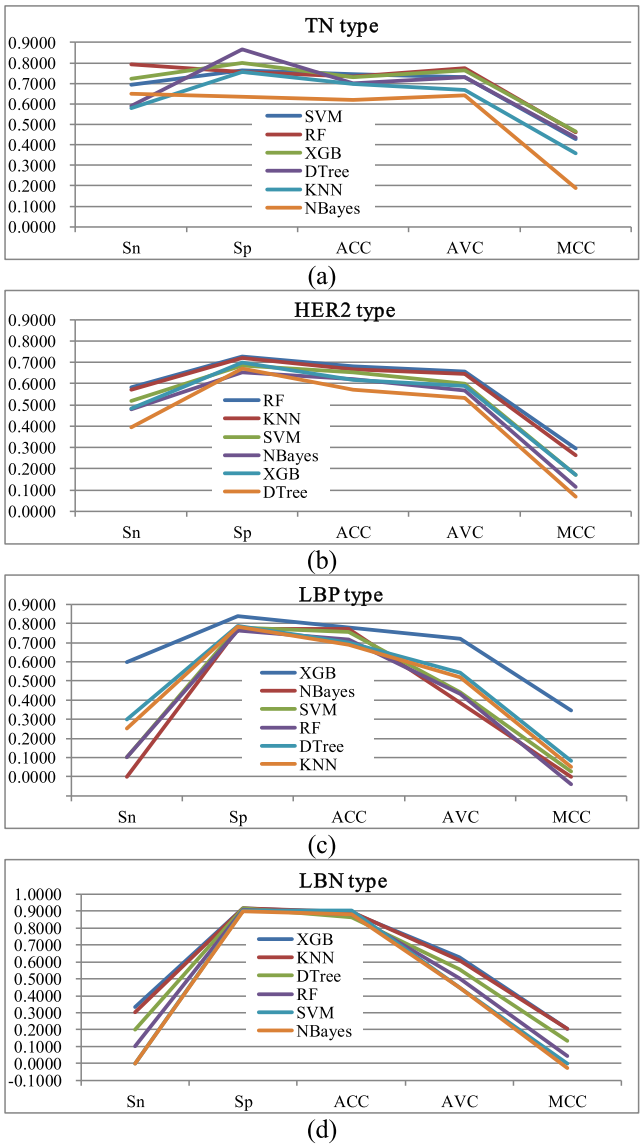


FIGURE 2. Predicting pCR by the first six chemotherapeutic treatments of breast cancer patients. The five classification performance measurements of all the six classifiers were calculated here for the subtypes (a) TN, (b) HER2, (c) LBP, and (d) LBN.

The overall pCR prediction of the subtype LBP was significantly improved by integrating the predicted data of treatments 4/5/6 in the training data, as shown in Figure 2 (c). The best pCR prediction model by the data of the first three treatments was achieved by the algorithm KNN, with mAvc = 0.5989. After the data integration of all the six treatments, the algorithm XGB achieved the best performance mAvc = 0.7183, with an increase of 0.1194.

E. SELECTING BEST FEATURES FOR DIFFERENT SUB-TYPES

We further hypothesized that we may need to select different feature combinations for the best pCR prediction

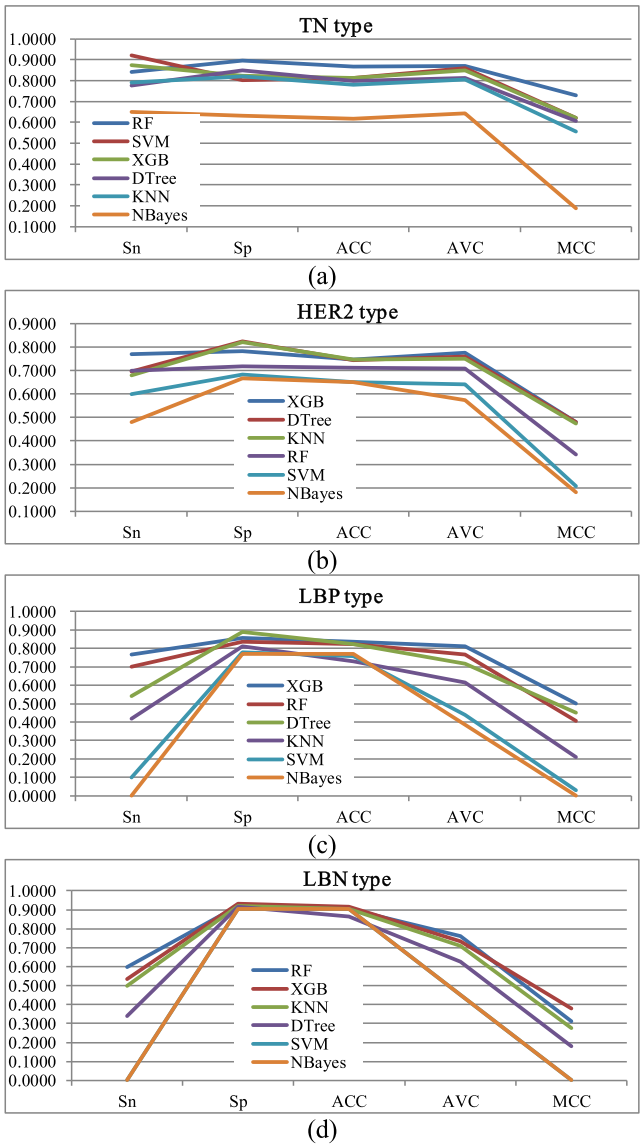


FIGURE 3. Prediction performances of pCR for the four breast cancer subtypes. (a) TN, (b) HER2, (c) LBP, and (d) LBN. The evaluated classification performance measurements are Sn, Sp, Acc, Avc, and MCC.

performances for the four subtypes [53]–[55]. So we collect the real nodal sizes of the first three treatments (S1, S2, S3) and the predicted nodal sizes of the next three treatments (S4, S5, S6) as the features for each participating patient. A combination of features with the best pCR prediction performance Avc was screened for each classification algorithm on a breast cancer subtype. There were $2^6 = 64$ different feature combinations in total for the six features and an exhaustive feature screening step was carried out to find the best feature subset [56].

The pCR prediction performance Avc reached the best value 0.8696 by the algorithm RF for the subtype TN patients, as shown in Figure 3 (a). Our aforementioned hypothesis was also supported by the observation that this best RF model used only the two features S3 and S4. This model achieved an improvement of 0.0706 for the best Avc = 0.7990 by the algorithm XGB using the three features S1, S2 and S3. Even the same algorithm XGB performed better on these two features S3 and S4, with an improvement 0.0498 in Avc. Both Sn and Sp were also improved compared with the cases in Figures 1 (a) and 2 (a).

An improved Avc was also obtained for the subtype HER2 patients, as shown in Figure 3 (b). The algorithm KNN achieved the best Avc = 0.7493 using the first three features S1, S2 and S3 in the above sections. After the step of feature selection, the algorithm XGB achieved a better Avc = 0.7763 using only two features S1 and S6. Another algorithm DTree also achieved a better Avc = 0.7543 using the features S2, S4 and S6. The first three features seem to be the optimal feature subset for the classifier KNN, since it achieved the same performance after the step of feature selection. The two classifiers DTree and KNN outperformed XGB in Sp but not in Sn, as shown in Figure 3 (b).

The step of feature selection significantly improved the pCR prediction performance for the subtype LBP, as shown in Figure 3 (c). The algorithm KNN achieved the best Avc = 0.5989 using the three original features S1, S2 and S3. After integrating the three predicted features S4, S5 and S6, XGB improved Avc to 0.7183 using all the six features. Figure 3 (c) demonstrated that not every feature positively contributed to the prediction model, and a significant improvement 0.0941 may be achieved by the classifier XGB using only the three features S1, S4 and S6. This best model also improved both Sn and Sp of all the previous models for the subtype LBP.

An increase 0.0399 was achieved for the subtype LBN by the algorithm XGB for Avc by using the two features S3 and S6, as shown in Figure 3 (d). Even the previous best algorithm KNN was improved with 0.0166 in Avc by using fewer features. And the best model for the subtype LBN was achieved by the classifier RF with Avc = 0.7594 using the features S5 and S6. This best model performed slightly worse in Sp = 0.9188 than that (Sp = 0.9192) of the classifier KNN using the first three features (S1/S2/S3), but it significantly improved the sensitivity (Sn = 0.6000) of the previous best model (Sn = 0.4667).

IV. DISCUSSION

This study explored the possibility of predicting the pathological complete remission (pCR) for breast cancer patients using only the data of the first three chemotherapeutic treatments. At least Avc = 0.7594 (Sn = 0.6000 and Sp = 0.9188) was achieved for the four subtypes of the breast cancer patients, and the algorithm RF may predict pCR for the triple-negative breast cancer (TN) patients with Avc = 0.8696 (Sn = 0.8429 and Sp = 0.8964). It's interesting to observe that the classifier XGB achieved the best accuracies for the two breast cancer subtypes, i.e., HER2 and LBP. And XGB performed the third (Sn = 0.8762 and Sp = 0.8233) and second (Sn = 0.5333 and Sp = 0.9323) best Avc for the other two subtypes TN and LBN, respectively.

The best models utilized the area data (S3, S4), (S1, S6), (S1, S4, S6), (S5, S6) for the subtypes TN, HER2, LBP and

LBN, respectively. So the tumor sizes after treatments 1, 3, 4 and 6 were important to predict the final pCR rates, even when S4/S5/S6 were calculated by the regression models of S1/S2/S3.

Our experimental data also demonstrated the importance of selecting a good feature subset for the prediction models. The pCR prediction Avc of the best model was improved from 0.7990 (Sn = 0.7933 and Sp = 0.8048) to 0.8696 (Sn = 0.8429 and Sp = 0.8964) for the subtype TN, with an improvement 0.0706 in Avc. Only two features S3/S4 as the model inputs also reduced the model calculation time. The pCR prediction Avc for the two subtypes LBP and LBN were improved by 0.0941 and 0.0664. The least improvement 0.0270 in Avc was achieved for the subtype HER2.

The subtype-specific pCR prediction models for breast cancers proposed in this study had the potential for clinical applications. Our models focused on simplifying the required input features and calculated the pCR prediction using only the tumor sizes of the first three neoadjuvant chemotherapies. A patient does not have to take a complete series of 6-8 neoadjuvant chemotherapy treatments to determine the prognosis (pCR = 1 or 0). After the first three treatments, the patient may make his or her own decision based on whether this patient was predicted to be able to achieve the pathological complete remission (pCR = 1) or not (pCR = 0) after three additional NAC treatments.

A multi-center clinical trial is being planned to carry out the cross-center validation of our models, before their applications in the clinical practice. Due to the clinical regulations, our models had to take all the procedures to officially get involved in the clinical decision process. We released our models and an easy-to-use pipeline for scientific research purpose only at http://www.healthinformaticslab.org/supp/.

Due to the difficulty in collecting clinical data under a study design, most of the clinical investigations recruited fewer than 300 participants [57]–[63]. This study established the initial pCR prediction models by recruiting 351 breast cancer patients who received at least 6 neoadjuvant chemotherapy treatments. The above-mentioned plan of a multi-center clinical trial will collect more categories of data entries and recruit more breast cancer patients to cross validate our models. This is anticipated to at least partially avoid the overfitting problem of our models and to cover a population in more habitats.

In conclusion, the pathological complete remission (pCR) rate of breast cancer patients undergoing neoadjuvant chemotherapy treatments may be predicted based on the tumor sizes after only the first three treatments. The pCR prediction models may be further improved by describing the heterogeneous status of breast cancer patients from the molecular and imaging perspectives.

ACKNOWLEDGMENT

The authors would like to thank the constructive comments from the three anonymous reviewers to greatly improved this study.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

[1] S. Tanini, A. D. Fisher, I. Meattini, S. Bianchi, J. Ristori, M. Maggi, and G. L. Russo, "Testosterone and breast cancer in transmen: Case reports, review of the literature, and clinical observation," *Clin. Breast Cancer*, vol. 19, no. 2, pp. e271–e275, Dec. 2018. doi: 10.1016/j.clbc.2018.12.006.

[2] R. Wang, J. Li, Y. Zhao, Y. Li, and L. Yin, "Investigating the therapeutic potential and mechanism of curcumin in breast cancer based on RNA sequencing and bioinformatics analysis," *Breast Cancer*, vol. 25, no. 2, pp. 206–212, Mar. 2018. doi: 10.1007/s12282-017-0816-6.

[3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA, Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018. doi: 10.3322/caac.21456.

[4] X. Jiang, H. Tang, and T. Chen, "Epidemiology of gynecologic cancers in China," *J. Gynecol. Oncol.*, vol. 29, no. 1, p. e7, Jan. 2018. doi: 10.3802/jgo.2018.29.e7.

[5] R. Zheng, H. Zeng, S. Zhang, and W. Q. Chen, "Estimates of cancer incidence and mortality in China, 2013," *Chin. J. Cancer*, vol. 36, no. 1, p. 66, Aug. 2017. doi: 10.1186/s40880-017-0234-3.

[6] M. Anaya-Ruiz and M. Perez-Santos, "Innovation status of gene therapy for breast cancer," *Asian–Pacific J. Cancer Prevention*, vol. 16, no. 9, pp. 4133–4136, 2015. doi: 10.7314/apjcp.2015.16.9.4133.

[7] W. Chen, R. Zheng, P. D. Baade, S. Z. BMedSc, H. Zeng, F. Bray, A. Jemal, X. Q. Yu, and J. He, "Cancer statistics in China, 2015," *CA A. Cancer J. Clinicians*, vol. 66, no. 2, pp. 115–132, Mar./Apr. 2016. doi: 10.3322/caac.21338.

[8] S. H. Jafari, Z. Saadatpour, A. Salmaninejad, F. Momeni, M. Mokhtari, J. S. Nahand,M. Rahmati, H. Mirzaei, M. Kianmehr, "Breast cancer diagnosis: Imaging techniques and biochemical markers," *J Cell Physiol*, vol. 233, no. 7, pp. 5200–5213, Jul. 2018. doi: 10.1002/jcp.26379.

[9] M. C. L. Peek and M. Douek, "Ablative techniques for the treatment of benign and malignant breast tumours," *J. Therapeutic Ultrasound*, vol. 5, p. 18, Jul. 2017. doi: 10.1186/s40349-017-0097-8.

[10] S. A. Narod, "Tumour size predicts long-term survival among women with lymph node-positive breast cancer," *Current Oncol.*, vol. 19, no. 5, pp. 249–253, Oct. 2012. doi: 10.3747/co.19.1043.

[11] S. Tuminello, B. Liu, A. Wolf, N. Alpert, E. Taioli, and R. M. Flores, "Comparison of in-hospital and long-term outcomes of sublobar lung cancer surgery by VATS and open techniques," *Amer. J. Clin. Oncol.*, vol. 41, no. 12, pp. 1149–1153, Dec. 2018. doi: 10.1097/COC.0000000000000440.

[12] M.-B. Jensen, T. O. Nielsen, A. S. Knoop, A. V. Laenkholm, E. Balslev, and B. Ejlertsen, "Mortality and recurrence rates among systemically untreated high risk breast cancer patients included in the DBCG 77 trials," *Acta Oncol.*, vol. 57, no. 1, pp. 135–140, Jan. 2018. doi: 10.1080/0284186X.2017.1400181.

[13] B. Swearingen, F. G. Barker, L. Katznelson, B. M. K. Biller, S. Grinspoon, A. Klibanski, N. Moayeri, P. L. Black, and N. T. Zervas, "Long-Term mortality after transsphenoidal surgery and adjunctive therapy for acromegaly," *J. Clin. Endocrinol. Metabolism*, vol. 83, no. 10, pp. 3419–3426, Oct. 1998. doi: 10.1210/jcem.83.10.5222.

[14] A. B. Miller, C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod, "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: Randomised screening trial," *BMJ*, vol. 348, p. 366, Feb. 2014.

[15] C. L. Carter, C. Allen, and D. E. Henson, "Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases," *Cancer*, vol. 63, no. 1, pp. 181–187, Jan. 1989. doi: 10.1002/1097-0142(19890101)63:1<181::AID-CNCR2820630129>3.0.CO;2-H.

[16] K. D. Voduc, M. C. Cheang, S. Tyldesley, K. Gelmon, T. O. Nielsen, and H. Kennecke, "Breast cancer subtypes and the risk of local and regional relapse," *J. Clin. Oncol.*, vol. 28, no. 10, pp. 1684–1691, Mar. 2010. doi: 10.1200/JCO.2009.24.9284.

[17] Y. Li, Y. Zhou, F. Mao, J. Guan, Y. Lin, X. Wang, Y. Zhang, X. Zhang, S. Shen, and Q. Sun, "The influence on survival of delay in the treatment initiation of screening detected non-symptomatic breast cancer," *Sci. Rep.*, vol. 9, no. 1, p. 10158, Jul. 2019. doi: 10.1038/s41598-019-46736-1.

[18] J. Kim, S. Lee, S. Y. Bae, M.-Y. Choi, J. Lee, S. P. Jung, S. Kim, J.-H. Choe, J.-H. Kim, J. S. Kim, J. E. Lee, S. J. Nam, and J.-H. Yang, "Comparison between screen-detected and symptomatic breast cancers according to molecular subtypes," *Breast Cancer Res. Treat.*, vol. 131, no. 2, pp. 527–540, Jan. 2012. doi: 10.1007/s10549-011-1836-0.

[19] H. Kennecke, R. Yerushalmi, R. Woods, M. C. U. Cheang, D. Voduc, C. H. Speers, T. O. Nielsen, and K. Gelmon, "Metastatic behavior of breast cancer subtypes," *J. Clin. Oncol.*, vol. 28, no. 20, pp. 3271–3277, May 2010. doi: 10.1200/JCO.2009.25.9820.

[20] C. Denkert *et al.*, "Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: A pooled analysis of 3771 patients treated with neoadjuvant therapy," *Lancet Oncol.*, vol. 19, no. 1, pp. 40–50, Jan. 2018. doi: 10.1016/S1470-2045(17)30904-X.

[21] C. Palumbo, S.-J. Cyr, E. Mazzone, F. A. Mistretta, S. Knipper, A. Pecoraro, Z. Tian, S. F. Shariat, F. Saad, C. Simeone, A. Briganti, A. Kapoor, A. Antonelli, and P. I. Karakiewicz, "Impact of tumor size on cancer-specific mortality rate after local tumor ablation in T1a renal-cell carcinoma," *J. Endourol.*, vol. 33, no. 7, pp. 606–613, Jul. 2019. doi: 10.1089/end.2019.0179.

[22] V. Muralidhar, R. D. Nipp, D. P. Ryan, T. S. Hong, P. L. Nguyen, and J. Y. Wo, "Association between very small tumor size and increased cancer-specific mortality in node-positive colon cancer," *Diseases Colon Rectum*, vol. 59, no. 3, pp. 187–193, Mar. 2016. doi: 10.1097/DCR.0000000000000532.

[23] K.-D. Yu, Y.-Z. Jiang, S. Chen, Z.-G. Cao, J. Wu, Z.-Z. Shen, and Z.-M. Shao, "Effect of large tumor size on cancer-specific mortality in node-negative breast cancer," *Mayo Clinic Proc.*, vol. 87, no. 12, pp. 1171–1180, Dec. 2012. doi: 10.1016/j.mayocp.2012.07.023.

[24] A. Valachis, A. D. Nearchou, and P. Lind, "Surgical management of breast cancer in BRCA-mutation carriers: A systematic review and meta-analysis," *Breast Cancer Res. Treat.*, vol. 144, no. 3, pp. 443–455, 2014.

[25] B. A. M. Heemskerk-Gerritsen, "Efficacy of risk-reducing mastectomy (RRM) on overall survival (OS) in BRCA1/2-associated breast cancer (BC) patients," Amer. Soc. Clin. Oncol., Alexandria, VA, USA, Tech. Rep.,2013.

[26] M. T. McKenna, J. A. Weis, A. Brock, V. Quaranta, and T. E. Yankeelov, "Precision medicine with imprecise therapy: Computational modeling for chemotherapy in breast cancer," *Transl. Oncol.*, vol. 11, no. 3, pp. 732–742, Jun. 2018. doi: 10.1016/j.tranon.2018.03.009.

[27] B. Davidson, "Biomarkers of drug resistance in ovarian cancer—An update," *Expert Rev. Mol. Diagnostics*, vol. 19, no. 6, pp. 469–476, Jun. 2019. doi: 10.1080/14737159.2019.1618187.

[28] G. von Minckwitz, M. Untch, J. U. Blohmer, S. D. Costa, H. Eidtmann, P. A. Fasching, B. Gerber, W. Eiermann, J. Hilfrich, J. Huober, C. Jackisch, M. Kaufmann, G. E. Konecny, C. Denkert, V. Nekljudova, K. Mehta, and S. Loibl, "Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes," *J Clin Oncol*, vol. 30, no. 15, pp. 804–1796, May 2012. doi: 10.1200/JCO.2011.38.8595.

[29] F. M. Pinel-Giroux, M. M. El Khoury, I. Trop, C. Bernier, J. David, and L. Lalonde, "Breast reconstruction: Review of surgical methods and spectrum of imaging findings," *Radiographics*, vol. 33, no. 2, pp. 435–453, Mar. 2013. doi: 10.1148/rg.332125108.

[30] H. Li, L. Yao, P. Jin, L. Hu, X. Li, T. Guo, and K. Yang, "MRI and PET/CT for evaluation of the pathological response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis," *Breast*, vol. 40, pp. 106–115, Aug. 2018. doi: 10.1016/j.breast.2018.04.018.

[31] W. Li, C. C.-H. Kwok, D. C.-W. Chan, F. Wang, and L. A. Tse, "Weak circadian rhythm increases neutropenia risk among breast cancer patients undergoing adjuvant chemotherapy," *Breast Cancer Res. Treat.*, vol. 168, no. 2, pp. 483–493, Apr. 2018. doi: 10.1007/s10549-017-4614-9.

[32] A. Tajik, A. Rezayof, Z. Ghasemzadeh, and M. Sardari, "Activation of the dorsal hippocampal nicotinic acetylcholine receptors improves tamoxifen-induced memory retrieval impairment in adult female rats," *Neuroscience*, vol. 327, pp. 1–9, Jul. 2016. doi: 10.1016/j.neuroscience.2016.04.008.

[33] C. Andersen, M. Rørth, B. Ejlertsen, and L. Adamsen, "Exercise despite pain—Breast cancer patient experiences of muscle and joint pain during adjuvant chemotherapy and concurrent participation in an exercise intervention," *Eur. J. Cancer Care*, vol. 23, no. 5, pp. 653–667, Sep. 2014. doi: 10.1111/ecc.12192.

[34] M. B. de Ruiter, L. Reneman, W. Boogerd, D. J. Veltman, F. S. A. M. van Dam, A. J. Nederveen, E. Boven, and S. B. Schagen, "Cerebral hyporesponsiveness and cognitive impairment 10 years after chemotherapy for breast cancer," *Hum. Brain Mapping*, vol. 32, no. 8, pp. 1206–1219, Aug. 2011. doi: 10.1002/hbm.21102.

[35] M. Iwase, M. Hattori, M. Sawaki, A. Yoshimura, H. Kotani, N. Gondo, Y. Adachi, A. Kataoka, S. Onishi, K. Sugino, N. Horisawa, M. Mori, M. Terada, and H. Iwata, "Presence of small residual malignant lesions in pathologic complete response after neo-adjuvant chemotherapy in patients with breast cancer," *Breast J.*, 2019. doi: 10.1111/tbj.13473.

[36] T. Konishi, E. Shinozaki, K. Murofushi, S. Taguchi, Y. Fukunaga, S. Nagayama, Y. Fujimoto, T. Akiyoshi, T. Nagasaki, M. Suenaga, A. Chino, H. Kawachi, N. Yamamoto, Y. Ishikawa, M. Oguchi, N. Ishizuka, M. Ueno, and K. Yamaguchi, "Phase II trial of neoadjuvant chemotherapy, chemoradiotherapy, and laparoscopic surgery with selective lateral node dissection for poor-risk low rectal cancer," *Ann. Surgical Oncol.*, vol. 26, no. 8, pp. 2507–2513, Aug. 2019. doi: 10.1245/s10434-019-07342-7.

[37] M. S. van Ramshorst, A. van der Voort, E. D. van Werkhoven, I. A. Mandjes, I. K. MANP, V. O. Dezentjé, I. M. Oving, A. H. Honkoop, L. W. Tick, A. J. van de Wouw, C. M. Mandigers, L. J. van Warmerdam, P. J. Wesseling, M.-J. T. V. Peeters, P. S. C. Linn, and G. S. Sonke, "Neoadjuvant chemotherapy with or without anthracyclines in the presence of dual HER2 blockade for HER2-positive breast cancer (TRAIN-2): A multicentre, open-label, randomised, phase 3 trial," *Lancet Oncol*, vol. 19, no. 12, pp. 1630–1640, Dec. 2018. doi: 10.1016/S1470-2045(18)30570-9.

[38] A. D. Altman, J. McGee, T. May, K. Lane, L. Lu, W. Xu, P. Ghatage, and B. Rosen, "Neoadjuvant chemotherapy and chemotherapy cycle number: A national multicentre study," *Gynecol. Oncol.*, vol. 147, no. 2, pp. 257–261, Nov. 2017. doi: 10.1016/j.ygyno.2017.08.006.

[39] M. Jebbink, E. van Werkhoven, I. A. M. Mandjes, J. Wesseling, E. H. Lips, M.-J. T. D. F. V. Peeters, C. E. Loo, G. S. Sonke, S. C. Linn, C. F. Zamora, and S. Rodenhuis, "The prognostic value of the neoadjuvant response index in triple-negative breast cancer: Validation and comparison with pathological complete response as outcome measure," *Breast Cancer Res. Treat.*, vol. 153, no. 1, pp. 145–152, Aug. 2015. doi: 10.1007/s10549-015-3510-4.

[40] S. Rodenhuis, I. A. M. Mandjes, J. Wesseling, M. J. van de Vijver, M.-J. T. D. F. V. Peeters, G. S. Sonke, and S. C. Linn, "A simple system for grading the response of breast cancer to neoadjuvant chemotherapy," *Ann. Oncol.*, vol. 21, no. 3, pp. 481–487, Mar. 2010. doi: 10.1093/annonc/mdp348.

[41] D. Blum, I. Liepelt-Scarfone, D. Berg, T. Gasser, C. la Fougère, and M. Reimold, "Controls-based denoising, a new approach for medical image analysis, improves prediction of conversion to Alzheimer's disease with FDG-PET," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 46, no. 11, pp. 2370–2379, 2019. doi: 10.1007/s00259-019-04400-w.

[42] H. J. Al-Barakati, H. Saigo, R. H. Newman, and D. B. Kc, "RF-GlutarySite: A random forest based predictor for glutarylation sites," *Mol. Omics*, vol. 15, no. 3, pp. 189–204, Jan. 2019. doi: 10.1039/c9mo00028c.

[43] Z.-G. Wei and S.-W. Zhang, "DMSC: A dynamic multi-seeds method for clustering 16S rRNA sequences into OTUs," *Frontiers Microbiol*, vol. 10, p. 428, Mar. 2019. doi: 10.3389/fmicb.2019.00428.

[44] R. Ge, M. Zhou, Y. Luo, Q. Meng, G. Mai, D. Ma, G. Wang, and F. Zhou, "McTwo: A two-step feature selection algorithm based on maximal information coefficient," *BMC Bioinf.*, vol. 17, p. 142, Mar. 2016. doi: 10.1186/s12859-016-0990-0.

[45] M. Zhou, Y. Luo, G. Sun, G. Mai, and F. Zhou, "Constraint programming based biomarker optimization," *BioMed Res. Int.*, vol. 2015, Dec. 2014, Art. no. 910515. doi: 10.1155/2015/910515.

[46] A. Allam, M. Nagy, G. Thoma, and M. Krauthammer, "Neural networks versus Logistic regression for 30?days all-cause readmission prediction," *Sci. Rep.*, vol. 9, no. 1, p. 9277, Jun. 2019. doi: 10.1038/s41598-019-45685-z.

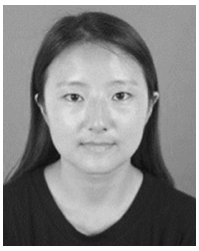
[47] J. Xie, R. Liu, J. T. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers Genet.*, vol. 10, p. 80, Feb. 2019. doi: 10.3389/fgene.2019.00080.

[48] J. Lin and K.-C. Wong, "Off-target predictions in CRISPR-Cas9 gene editing using deep learning," *Bioinformatics*, vol. 34, no. 17, pp. i656–i663, Sep. 2018. doi: 10.1093/bioinformatics/bty554.

[49] X. Xia, H. Weng, R. Men, R. Sun, B. C. Y. Zee, K. C. Chong, and M. H. Wang, "Incorporating methylation genome information improves prediction accuracy for drug treatment responses," *BMC Genet.*, vol. 19, p. 78, Sep. 2018. doi: 10.1186/s12863-018-0644-5.

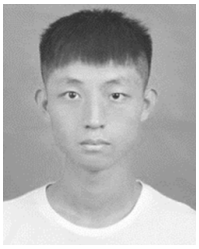
[50] C. Jia, Y. Zuo, and Q. Zou, "O-GlcNAcPRED-II: An integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique," *Bioinformatics*, vol. 34, no. 12, pp. 2029–2036, Jun. 2018. doi: 10.1093/bioinformatics/bty039.
[51] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks," *Briefings Bioinf.*, vol. 17, no. 2, pp. 193–203, Mar. 2016. doi: 10.1093/bib/bbv033.
[52] D. Ma, S. Zhong, X. Liu, H. Mai, G. Mai, C. Xu, and F. Zhou, "CD3D and PRKCQ work together to discriminate between B-cell and T-cell acute lymphoblastic leukemia," *Comput. Biol. Med.*, vol. 77, pp. 16–22, Oct. 2016. doi: 10.1016/j.compbiomed.2016.07.004.
[53] C. Xu, J. Liu, Y. Shu, Z. Wei, W. Zheng, X. Feng, and F. Zhou, "An OMIC biomarker detection algorithm TriVote and its application in methylomic biomarker detection," *Epigenomics*, vol. 10, no. 4, pp. 335–347, Jan. 2018. doi: 10.2217/epi-2017-0097.
[54] Y. Ye, R. Zhang, W. Zheng, S. Liu, and F. Zhou, "RIFS: A randomly restarted incremental feature selection algorithm," *Sci. Rep.*, vol. 7, no. 1, p. 13013, Oct. 2017. doi: 10.1038/s41598-017-13259-6.
[55] K. E. McCann, S. A. Hurvitz, and N. McAndrew, "Advances in targeted therapies for triple-negative breast cancer," *Drugs*, vol. 79, no. 11, pp. 1217–1230, Jul. 2019. doi: 10.1007/s40265-019-01155-4.
[56] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," in *Handbook of Pattern Recognition and Computer Vision World Scientific*. 1993, pp. 88–107.
[57] J. Behr, M. Kolb, J. W. Song, F. Luppi, B. Schinzel, S. Stowasser, M. Quaresma, and F. J. Martinez, "Nintedanib and sildenafil in patients with idiopathic pulmonary fibrosis and right heart dysfunction (INSTAGE): A pre-specified sub-group analysis of a double-blind, randomized clinical trial," *Amer. J. Respiratory Crit. Care Med.*, vol. 2019. doi: 10.1164/rccm.201903-0488OC.
[58] E. J. Lenze, E. Lenard, M. Bland, P. Barco, J. P. Miller, M. Yingling, C. E. Lang, "Effect of enhanced medical rehabilitation on functional recovery in older adults receiving skilled nursing care after acute rehabilitationa randomized clinical trial," *JAMA Netw. Open*, vol. 2, no. 7, Jul. 2019, Art. no. e198199. doi: 10.1001/jamanetworkopen.2019.8199.
[59] M. Weiser, L. Levi, D. Zamora, A. Biegon, J. P. SanGiovanni, M. Davidson, and S. Burshtein, "Effect of adjunctive estradiol on schizophrenia among women of childbearing AgeA randomized clinical trial," *JAMA Psychiatry*, Jul. 2019. doi: 10.1001/jamapsychiatry.2019.1842.
[60] G. Isola, G. Matarese, A. Alibrandi, D. Dalessandri, M. Migliorati, E. Pedullà, and E. Rapisarda, "Comparison of effectiveness of etoricoxib and diclofenac on pain and perioperative sequelae after surgical avulsion of mandibular third molars: A randomized, controlled, clinical trial," *Clin. J. Pain*, Jul. 2019. doi: 10.1097/AJP.0000000000000748.
[61] C. Rodrigues *et al.*, "The effect of a rapid molecular blood test on the use of antibiotics for nosocomial sepsis: A randomized clinical trial," *J. Intensive Care*, vol. 7, p. 37, Jul. 2019. doi: 10.1186/s40560-019-0391-3.
[62] S. Jeong, J. H. Lee, G. Lee, Y.-I. Kwon, and J.-B. Oh, "Effect of steamed onion (ONIRO) consumption on body fat and metabolic profiles in overweight subjects: A 12-week randomized, double-blind, placebo-controlled clinical trial," *J. Amer. College Nutrition*, pp. 1–10, Aug. 2019. doi: 10.1080/07315724.2019.1635052.
[63] M. M. Islam, S. Huq, M. I. Hossain, A. M. S. Ahmed, A. Ashworth, A. H. Mollah, and T. Ahmed "Efficacy of F-100, diluted F-100, and infant formula as rehabilitation diet for infants aged?<?6 months with severe acute malnutrition: A randomized clinical trial," *Eur. J. Nutrition*, Jul. 2019. doi: 10.1007/s00394-019-02067-5.

XIN FENG received the master's degree from the School of Software, Jilin University, in 2013, where she is currently pursuing the Ph.D. degree with the School of Computer Science. She is mainly devoted to the research of biomedical big data during the reading period.



LELIAN SONG received the master's degree in surgery from Jilin University, in 2018. Her research interests include early diagnosis and individualized comprehensive treatment of breast cancer.

SHAOFEI WANG received the B.Sc. degree from the School of Computer Science and Technology, Jilin University, Changchun, China, in 2014. He is currently pursuing the master's degree in data science with the Engineering School, Columbia University, NY, USA. His research interests include bioinformatics, data mining, and natural language processing.

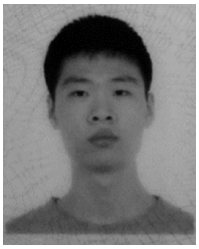


HAOQIU SONG is currently pursuing the master's degree with the School of Computer Science, Hubei University of Technology. His research interests include bioinformatics, data mining, and natural language processing.

HANG CHEN received the B.Sc. degree from the College of Computer Science and Technology, Jilin University, China, in 2018. He is an Algorithm Researcher with Megvii Inc., Beijing, China. His research interests include computer vision, especially text detection and recognition, data science, and dataset establishment to machine learning applications.

YUXUAN LIU received the B.Sc. degree in computer science from the College of Computer Science and Technology, Jilin University, in 2019. His research interest includes machine learning in biomedical big data.

CHENWEI LOU is currently pursuing the master's degree with the College of Computer and Technology, Jilin University. His research interests include machine learning and data mining.



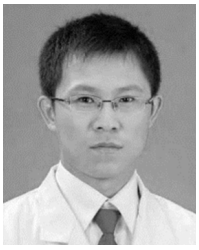
JIAN ZHAO is currently pursuing the bachelor's degree with the Department of Computer Science, Jilin University, China. His current research interests include machine learning and neural networks.

QUEWANG LIU is currently pursuing the master's degree with the School of Computer Science and Technology, Jilin University (JLU), Changchun, China, where he has been a Freshman with the School of Computer of Science and Technology, since 2014. He has been a Member Fellow of the HILab, JLU, since 2018.

YANG LIU received the B.Sc. degree from the School of Computer of Science and Technology, Changchun Institute of Technology (CCIT), in 2017. He is currently pursuing the master's degree with the College of Software, Jilin University. His research interests include developing efficient feature selection and prediction algorithms for the high-dimensional biomensional data.

RUIXUE ZHAO received the B.Sc. degree from the College of Computer Science and Technology, Jilin University, in 2018. She is currently pursuing the master's degree in information system from The Hong Kong University of Science and Technology. Her research interests include developing data mining algorithms for financial and biomedical big data.

KAI XING received the B.Sc. degree from the College of Electronic Science and Engineering, Jilin University, Changchun, China, in 2018. He is currently working with ICarbonX.



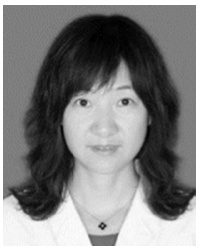
SIJIE LI is the Deputy Chief Physician and a Medical Doctor specializes in early diagnosis of breast cancer, breast-conserving surgery, minimally invasive surgery for breast surgery, and breast reconstruction after breast cancer surgery.

YUNHE YU received the master's degree in surgery from Jilin University, in 2018. Her research interests include early diagnosis, treatment, and prognosis analysis of breast cancer.

ZHENYU LIU received the Ph.D. degree in surgery from Jilin University, in 2017. He is currently a Resident of breast surgery with The First Hospital of Jilin University. His main research areas include early diagnosis, treatment, and survival analysis of breast cancer.

CHENGYANG YIN is currently pursuing the master's degree in breast surgery with The First Hospital of Jilin University. His research interests include early diagnosis and standardized treatment of breast cancer.

BING HAN is a Chief Physician, a Professor, a Doctor of medicine, and the master's degree Tutor specializes in mammotomy of breast benign tumors, breast cancer surgery, neoadjuvant therapy, standardization, and personalized treatment.



YE DU received the Ph.D. degree in surgery. She is the Deputy Chief Physician, the master's degree Instructor specializes in early diagnosis of breast cancer, surgery, and postoperative individualized comprehensive treatment, benign breast tumor minimally invasive surgery, and intravenous infusion port implantation.

RUIHAO XIN was born in Jilin, China, in 1989. He received the Ph.D. degree from the Changchun University of Science and Technology, in 2018. His research interests include artificial intelligence and control technology.

LAN HUANG received the Ph.D. degree from the College of Computer Science and Technology, Jilin University, Changchun, China, in 2003, where she is currently a Professor with the College of Computer Science and Technology. Her research interests include data mining and business intelligence.

ZHIMIN FAN received the bachelor's degree in medicine from Bethune Medical University, in 1985, the master's degree in thoracic surgery from Bethune Medical University, in 1990, and the Ph.D. degree in surgery from Jilin University, in 2004. He is currently the Director of breast surgery with The First Hospital of Jilin University, where he is also a Professor, the Chief Physician, and the master's and Ph.D. degree Tutor. His main research areas include differentially expressed protein analysis of breast cancer and early diagnosis, treatment, survival analysis, and prognostic factors of breast cancer.

FENGFENG ZHOU received the master's and Ph.D. degrees in computer science from the University of Science and Technology of China, in 2000 and 2005, respectively. He is currently with his laboratory at Jilin University focuses on the development of feature selection algorithms for biomedical big data.