Machine Learning-Based A Priori Chemotherapy Response Prediction in Breast Cancer Patients using Textural CT Biomarkers*

Hadi Moghadas-Dastjerdi, *Member, IEEE*, Hira R. Sha-E-Tallat, Lakshmanan Sannachi, Laurentius O. Osapoeta, Ali Sadeghi-Naini, *Senior Member, IEEE*, Gregory J. Czarnota

***Abstract*— Early prediction of cancer response to neoadjuvant chemotherapy (NAC) could permit personalized treatment adjustments for patients, which would improve treatment outcomes and patient survival. For the first time, the efficiency of quantitative computed tomography (qCT) textural and second derivative of textural (SDT) features were investigated and compared in this study. It was demonstrated that intra-tumour heterogeneity can be probed through these biomarkers and used as chemotherapy tumour response predictors in breast cancer patients prior to the start of treatment. These features were used to develop a machine learning approach which provided promising results with cross-validated $AUC_{0.632+}$, accuracy, sensitivity and specificity of 0.86, 81%, 74% and 88%, respectively.**

***Clinical Relevance*— The results obtained in this study demonstrate the potential of textural CT biomarkers as response predictors of standard NAC before treatment initiation.**

I. Introduction

Breast cancer is the most frequently diagnosed cancer among women and accounts for ~30% of all new cancer cases diagnosed annually in Canada and United States [1], [2]. The incidence of locally advanced breast cancer (LABC) is deemed to represent up to 20% of all breast cancer cases [3]–[5]. LABC is classified as stage IIB cancer or higher with a tumour size of greater than 5 cm that may extend to the skin and/or chest wall. LABC treatment usually includes neoadjuvant chemotherapy (NAC), followed by surgery and, if required, adjuvant radiation and/or hormonal therapy [6]–[8]. However, only about 20-30% of the patients achieve a complete pathological response to standard NAC [9]–[11]. Tumour response to NAC is usually evaluated based on post-treatment anatomical imaging or histopathology on post-surgical specimens, while the window to adjust neoadjuvant treatment is closed at that time. Early evaluation of treatment efficiency based on the tumour size is quite challenging since changes in tumour dimensions in response to chemotherapy can take several months to be detectable through anatomical imaging. In some cases, no change in tumour size is traceable despite a pathological response to treatment. Prediction of response to NAC before or early after the start of the treatment can provide an opportunity to modify standard treatment plans (e.g., regimen, dose, priority of treatment options) on an individual patient basis, which is expected to enhance therapy outcome and patient survival.

Various quantitative imaging techniques have been developed to investigate breast cancer response to chemotherapy early after treatment initiation [12]–[18]. Because of its ability to provide functional information some researchers have used positron emission tomography (PET) to predict and monitor tumour response in breast cancer patients; however this imaging modality is expensive and requires using radionuclide contrast agents [12], [13]. Although dynamic contract-enhanced magnetic resonance imaging (MRI) has demonstrated promise to predict the chemotherapy response in women with breast cancer [14], its relatively high cost and limited availability restricts its use for response prediction. Diffuse optical imaging (DOI) is another modality that has been applied to monitor and predict breast cancer response to NAC [15], [16], as it can probe perfusion, oxygenation, and indirectly, tissue vascularity. However, DOI requires a long scan session to reconstruct volumetric images with an appropriate image resolution. Due to the fact that ultrasound (US) is a non-ionizing and inexpensive modality that can probe the tissue micro-structure without using any contrast agents, quantitative ultrasound (QUS) techniques have been successfully used in tissue characterization and tumour response monitoring applications [17], [18]. Lower image quality and lack of 3D information in US images have been the driving force behind research aimed at finding other efficient imaging biomarkers that can be used as NAC response predictors. Recent studies suggest that the micro-environmental characteristics of a tumour can be measured using quantitative computed tomography (qCT). Since CT images are usually acquired as part of the standard of care for LABC patients, CT could become the modality of choice to quantify spatial heterogeneity for determining tumour aggressiveness and responsiveness to treatment. Although the voxel size in CT images is considerably larger than cellular dimensions, it has been demonstrated that tissue microstructure variations could be measured through CT

*This Research was supported by the Natural Sciences and Engineering Research Council of Canada, Canadian Institutes for Health Research, Lotte & John Hecht Memorial Foundation, and the Terry Fox Foundation.

H. Moghadas-Dastjerdi, L. Sannachi, L. O. Osapoeta and G. J. Czarnota are with the Departments of Medical Biophysics, and Radiation Oncology, University of Toronto, Toronto, ON, M5G 2M9 Canada, also with the Departments of Radiation Oncology, and Physical Sciences Platform, Sunnybrook Health Sciences Center, Toronto, ON, M4N 3M5 Canada (corresponding author, phone: +1-416-480-6100 x83744; e-mail: hadi.moghadas@sunnybrook.ca; lakshmanan.sannachi@sunnybrook.ca; laurentiusoscar.osapoetra@sunnybrook.ca; gregory.czarnota@sunnybrook.ca).

H. R. Sha-E-Tallat, was with Physical Sciences Platform, Sunnybrook Health Sciences Center. She is now with the Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada (e-mail: shan-e-tallathira.rahman@sunnybrook.ca).

A. Sadeghi-Naini is with the Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada; also with the Department of Radiation Oncology and Physical Sciences Platform, Odette Cancer Center and Sunnybrook Research Institute, Sunnybrook Health Sciences Center, Toronto, ON, Canada; also with the Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada (e-mail: asn@yorku.ca).

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

voxel intensity alteration as a result of the partial volume effect [19], [20].

In this study, the efficiency of qCT biomarkers has been investigated for the first time for use as predictors of tumour response to NAC for LABC patients prior to initiation of treatment. Textural and second derivative of textural (SDT) features were extracted from pre-treatment CT images acquired from 39 LABC patients. The best feature subsets were selected through a multi-step feature ranking and selection process and then used to train a machine learning algorithm. Results indicated that SDT features in conjunction with an adaptive-boosting decision tree classifier can be used to predict pathological response of LABC tumours to NAC before the start of the treatment with a sensitivity and specificity of 74% and 88%, respectively.

II. Methods

A. Study Protocol and Data Acquisition

This study was conducted in accordance with institutional research ethics board approval. Thirty nine women aged 27 to 83 years who were diagnosed with LABC were enrolled in this study after receiving an informed consent. The patients were planned to be treated with a full course of NAC followed by surgery. A core needle biopsy was performed for all patients to confirm the cancer diagnosis, and to determine the histological subtype, the hormone receptor status, and the initial cellularity of the tumour. As part of the institutional standard of care, all patients underwent contrast-enhanced CT imaging of the breast prior to treatment initiation. CT Scans were performed with a multi-slice CT scanner (LightSpeed, GE Medical Systems, Chicago, United States) using a helical acquisition mode. The scan parameters included x-ray tube current: 10-367mA, tube voltage: 120 kV, slice size: 512×512 pixels, slice thickness: 2.5 mm, and pixel spacing: 0.8×0.8 mm. Additionally, MRI scans were acquired for all patients before and after the treatment to measure tumour size and assess chest wall involvement.

B. Tumour Response and Histopathological Analysis

Breast surgery was performed for all patients after the completion of NAC. The surgical specimens were used to assess tumour response to NAC based on standard clinical histopathology. For each patient, the clinical/pathological tumour response was determined at the end of their treatment using a modified response (MR) grading system. The MR score was based on both response evaluation criteria in solid tumours (RECIST) and histopathological criteria and defined as follows: MR1: no diminishment in tumour size; MR2: <30% diminishment in tumour size; MR3: an estimated 30-90% reduction in tumour size (or a very low residual tumour cellularity determined histopathologically); MR4: a diminishment of >90% in tumour size; MR5: no evident tumour and no malignant cells identifiable in sections from the site of the tumour; only vascular fibroelastic stroma remains, often containing macrophages; however, ductal carcinoma in situ may be present. A binary classification was investigated where the patients with MR scores of $\geq$3 and $\leq$2 were deemed to be responders (R) and non-responders (NR), respectively. In keeping with this, 27 and 12 patients were identified as responders and non-responders, respectively.

C. Quantitative CT Analysis

The regions of interest (ROI) in CT image slices were outlined manually to include the entire tumour within the breast. In order to derive textural and STD features within each ROI, textural analysis was performed using a grey-level co-occurrence matrix (GLCM) method. For each patient, the parametric maps of textural features were generated using texture analysis on grey-scale CT images using a sliding window approach. A three-pixel by three-pixel window was used to sweep the whole ROI with a step size of one pixel. The GLCM method statistically quantifies the angular relationship between neighboring pixels with different intensities and the distance/direction between them [21]. All CT voxel values were quantized into 128 grey levels and, at four different angels of 0º, 45 º, 90 º and 135 º, the GLCMs were computed symmetrically within one pixel distance from each reference pixel. The average of each textural feature was computed over all four GLCMs to represent the textural characteristic of a given pixel (at the center of the sliding window). The textural features included entropy (ENT), contrast (CON), correlation (COR), maximum GLCM probability (MAX), GLCM mean (MEA), homogeneity (HOM), GLCM standard deviation (STD) and energy (ENE). To calculate the average of each feature over the entire tumour volume, a weighted averaging scheme was used based on the ROI size to address the differences in tumour cross-section area over different slices. The eight textural features which were computed for each pixel were used to form eight textural parametric maps. These parametric maps were used to extract the GLCM-based SDT features.

D. Feature Selection and Response Prediction

For each patient, eight textural features (and eight corresponding parametric maps) were derived from a CT image and eight textural (SDT) features were extracted from each textural parametric map. Consequently, 8 textural and 64 SDT features were extracted for each CT image. Two separate sets of textural and SDT features were constructed to compare their effectiveness in response prediction. In each set, features were ranked based on minimal-redundancy-maximal-relevance (mRMR) criterion [22] prior to the start of the feature selection. To find the best feature subset, a sequential forward feature selection (SFS) scheme was applied. In this method, starting from a null subset, the efficacy of adding each feature to the subset was investigated iteratively based on their mRMR ranking. The $AUC_{0.632+}$ [23]–[25] was used to assess the performance of different feature subsets in response prediction. The best feature subsets of the textural and SDT features were used separately to train two independent classifiers to predict the treatment response. The best feature subsets consisted of three features since adding more features did not enhance the results, and given the small size of the dataset, using more features could increase the risk of overfitting. When data samples are not equally distributed across different classes (imbalanced dataset) classifiers tend frequently to be trained toward the majority class. To address this issue with the imbalanced dataset in this study, the minority class was oversampled to 24 samples using the SMOTE method [26], while the majority class was undersampled by taking 200 bootstrapped samples with the same size of the oversampled minority group. These bootstrapped samples of majorities then

combined with the oversampled set of minorities to form 200 balanced training sets. Each of the training sets was used to train an adaptive boosting (AdaBoost) classifier [27], using a decision tree (DT) as the weak learner classifier (AdaBoost-DT) with the following parameters: maximum depth = 3, estimators = 100 and learning rate = 0.1. A majority vote over all 200 classifiers yielded the predicted label of each test sample. The response prediction framework was evaluated using a leave-one-patient-out (LOPO) cross-validation scheme. For this purpose, one of the patients was left out iteratively before oversampling and bootstrapping and used as the test sample. This process was repeated until all samples were tested.

III. Results

The majority of women who have participated in this study were diagnosed with ductal carcinoma (IDC) (93%), with 5% and 2% of the patients diagnosed with invasive lobular carcinoma (ILC) and invasive metaplastic carcinoma (IMPC), respectively. Sixty-five percent of tumours had positive estrogen (ER+) and progesterone (PR+) receptors, while only 30% of the patients had tumours with a positive Her2/Neu receptor (HER2+), and 25% of the patients were diagnosed with a triple negative tumour. Half of the patients received Adriamycin, Cytoxan, followed by paclitaxel (AC-T) for NAC, 41% received 5-fluorouracil, epirubicin, cyclophosphamide followed by docetaxel (FEC-T), 6% received doxorubicin, cyclophosphamide followed by docetaxel (AC-D) and 3% received doxorubicin and cyclophosphamide (TC). Additionally, all patients with HER2+ tumours received monoclonal antibody trastuzumab (TRA). On average, tumour size was reduced by 40% after the treatment, compared to the initial size. The treatment regimen was not modified based on the imaging findings over the course of study.

Fig. 1 illustrates representative CT image slices overlaid by textural parametric maps of all eight GLCM textural features. SDT features are presented in bar plots next to the corresponding maps for a responding and a non-responding patient, acquired prior to NAC initiation. A 3×3 pixel sliding window analysis with a shifting step size of one pixel was used to generate the parametric maps, which visualize the spatial variation of the features within the tumour. These parametric maps subsequently underwent GLCM textural analysis to calculate the SDT features.

The best feature subset selected from the textural and SDT features are tabulated in Table I. According to the feature selection results, these feature combinations yield the most promising prediction outcome. In this table, the SDT features were named based on the textural characteristic extracted from the parametric maps. For instance, COR_MAX is the MAX feature that was extracted from the COR parametric map.

Table II summarizes the results of cross-validated response prediction using the best feature subsets of Table I and the Adaboost-DT classifier. The higher scores demonstrated in the bottom row indicate that the SDT features outperformed the standard textural features for NAC response prediction, and may potentially have more prognostic power in this application.

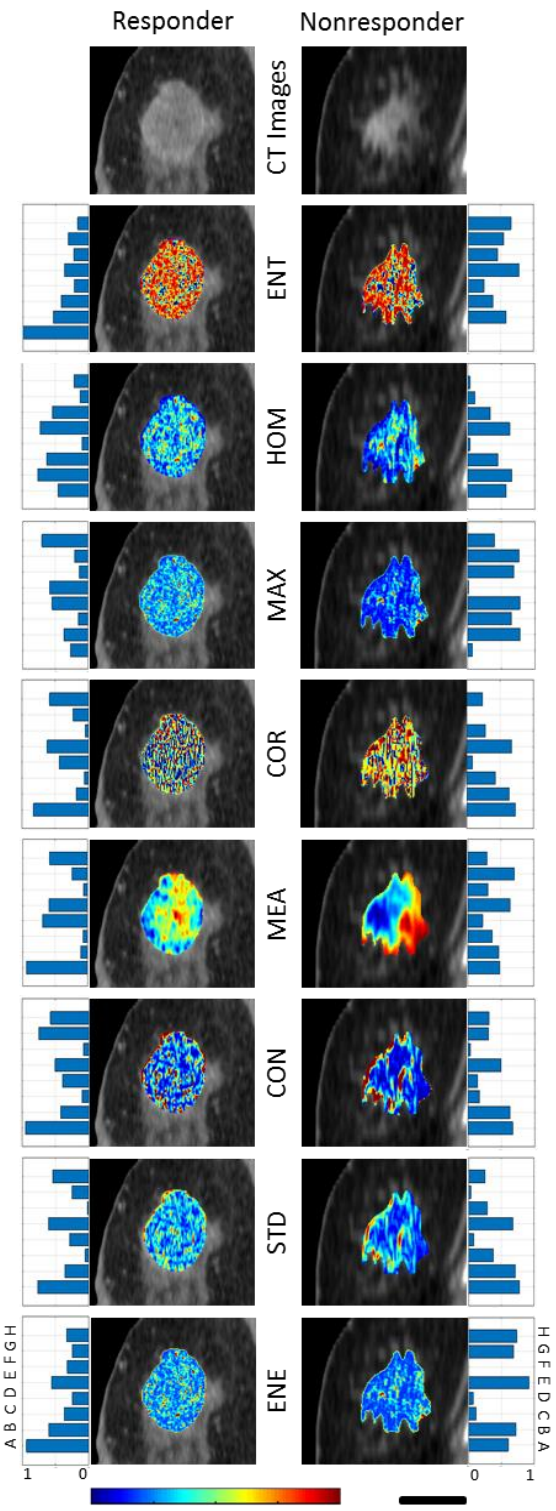

Figure 1. Representative CT images with parametric map overlays acquired for a responding (left) and a non-responding (right) patient. From top to bottom, the parametric maps demonstrate entropy, homogeneity, maximum GLCM probability, correlation, GLCM mean, contrast, GLCM standard deviation and energy. The bar plots next to the parametric maps show SDT features' values. These values normalized to a range of 0-1 for a better demonstration. The titles of the bars are as follows: A: ENT, B: HOM, C: MAX, D: COR, E: MEA, F: CON, G: STD, H: ENE. The color bar represents a scale of the range [1.5, 2.5] for ENT, [0, 0.8] for HOM, [-0.1, 0.8] for MAX, [-0.2, 0.7] for COR, [0, 50] for MEA, [0, 100] for CON, [-0.9, 8] for STD and [0.1, 0.8] for ENE. The scale bar represents 4 cm.

IV. Discussion

This study investigated the capability of qCT biomarkers to predict response to NAC in LABC patients prior to start of treatment. Patients were assessed prior to the initiation of treatment and were subsequently followed up during and after the course of chemotherapy. Treatment response was assessed based on standard clinical and histopathological methods. A GLCM approach was used to extract the textural and SDT features. The best features subsets were selected through a sequential forward feature selection technique in conjunction with the mRMR method and $AUC_{0.632+}$ criterion. The selected features were used to train an Adaboost-DT classifier. The performance of the response prediction framework was evaluated by a LOPO cross-validation approach. Results indicated that SDT features provided better performance with an accuracy, F1-score and $AUC_{0.632+}$ of 80.8%, 80.0% and 86.1%, respectively. The higher discriminatory power of SDT features might be due to the fact that they could capture more local distribution variation of tumour texture characteristics. Although a large-cohort study is required to further investigate the efficiency of the proposed biomarkers, results obtained in this study are promising and motivate future studies towards adapting quantitative CT imaging in conjunction with machine learning techniques for chemotherapy response prediction in cancer patients.

TABLE I. LIST OF THE FEATURES SELECTED THROUGH THE FETURE SELECTION METHOD FROM TEXTURAL AND SDT FEATURE SETS.

Features Type	**Selected Features**		
Textural	CON	MAX	ENE
SDT	CON_MEA	STD_HOM	ENT_STD

TABLE II. SUMMERY OF ADABOOST-DT CLASSIFIER CROSS-VALIDATED PERFORMANCE USING THE SELCETED FEATURE SUBSETS.

Features Type	**Scores**					
	$AUC_{0.632+}$ (%)	**ACC (%)**	**AUC (%)**	**SPC (%)**	**SNS (%)**	**F1 (%)**
Textural	82.8	76.9	77.3	88.0	66. 7	75.0
SDT	86.1	80.8	81.0	88.0	74.1	80.0

References

[1] 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.

[2] Canadian Cancer Statistics Advisory Committee, *Canadian Cancer Statistics 2018*. Toronto, ON: Canadian Cancer Society, 2018.

[3] V. Valero, A. U. Buzdar, and G. N. Hortobagyi, "Locally advanced breast cancer," *Oncologist*, vol. 1, no. 1 & 2, pp. 8–17, 1996.

[4] C. Allemani *et al.*, "Breast cancer survival in the US and Europe: A CONCORD high-resolution study," *Int. J. Cancer*, vol. 132, no. 5, pp. 1170–1181, Mar. 2013.

[5] K. Tryfonidis, E. Senkus, M. J. Cardoso, and F. Cardoso, "Management of locally advanced breast cancer—perspectives and future directions," *Nat. Rev. Clin. Oncol.*, vol. 12, p. 147, Feb. 2015.

[6] B. Fisher *et al.*, "Effect of preoperative chemotherapy on the outcome of women with operable breast cancer.," *J. Clin. Oncol.*, vol. 16, no. 8, pp. 2672–2685, Aug. 1998.

[7] D. Mauri, N. Pavlidis, and J. P. A. Ioannidis, "Neoadjuvant Versus Adjuvant Systemic Treatment in Breast Cancer: A Meta-Analysis," *JNCI J. Natl. Cancer Inst.*, vol. 97, no. 3, pp. 188–194, Feb. 2005.

[8] N. M. Rueth *et al.*, "Underuse of trimodality treatment affects survival for patients with inflammatory breast cancer: an analysis of treatment and survival trends from the National Cancer Database," *J. Clin. Oncol.*, vol. 32, no. 19, p. 2018, 2014.

[9] T. Byrski *et al.*, "Pathologic Complete Response Rates in Young Women With BRCA1 -Positive Breast Cancers After Neoadjuvant Chemotherapy," *J. Clin. Oncol.*, vol. 28, no. 3, pp. 375–379, Jan. 2010.

[10] J. R. Broadwater, M. J. Edwards, C. Kuglen, G. N. Hortobagyi, F. C. Ames, and C. M. Balch, "Mastectomy following preoperative chemotherapy. Strict operative criteria control operative morbidity.," *Ann. Surg.*, vol. 213, no. 2, p. 126, 1991.

[11] W. Haque, V. Verma, S. Hatch, V. Suzanne Klimberg, E. Brian Butler, and B. S. Teh, "Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy," *Breast Cancer Res. Treat.*, vol. 170, no. 3, pp. 559–567, Aug. 2018.

[12] J. S. Lewis and K. R. Keshari, *Imaging and Metabolism*. Cham: Springer International Publishing, 2018.

[13] M. E. Juweid and B. D. Cheson, "Positron-Emission Tomography and Assessment of Cancer Therapy," *N. Engl. J. Med.*, vol. 354, no. 5, pp. 496–507, Feb. 2006.

[14] A. Tudorica *et al.*, "Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI," *Transl. Oncol.*, vol. 9, no. 1, pp. 8–17, Feb. 2016.

[15] W. T. Tran *et al.*, "Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis," *Br. J. Cancer*, vol. 116, no. 10, p. 1329, 2017.

[16] A. Cerussi *et al.*, "Predicting response to breast cancer neoadjuvant chemotherapy using diffuse optical spectroscopy," *Proc. Natl. Acad. Sci.*, vol. 104, no. 10, pp. 4014–4019, Mar. 2007.

[17] A. Sadeghi-Naini *et al.*, "Low-frequency quantitative ultrasound imaging of cell death in vivo," *Med. Phys.*, vol. 40, no. 8, p. 082901, Jul. 2013.

[18] L. Sannachi *et al.*, "Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters," *Med. Image Anal.*, vol. 20, no. 1, pp. 224–236, Feb. 2015.

[19] H. Moghadas-Dastjerdi, M. Ahmadzadeh, and A. Samani, "Towards computer based lung disease diagnosis using accurate lung air segmentation of CT images in exhalation and inhalation phases," *Expert Syst. Appl.*, vol. 71, pp. 396–403, 2017.

[20] H. Moghadas-Dastjerdi, M. Ahmadzadeh, E. Karami, M. Karami, and A. Samani, "Lung CT image based automatic technique for COPD GOLD stage assessment," *Expert Syst. Appl.*, vol. 85, pp. 194–203, Nov. 2017.

[21] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[22] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[23] B. Efron and R. Tibshirani, "Improvements on Cross-Validation: The 632+ Bootstrap Method," *J. Am. Stat. Assoc.*, vol. 92, no. 438, pp. 548–560, Jun. 1997.

[24] B. Sahiner, H.-P. Chan, and L. Hadjiiski, "Classifier performance prediction for computer-aided diagnosis using a limited dataset," *Med. Phys.*, vol. 35, no. 4, pp. 1559–1570, Mar. 2008.

[25] W. Gómez Flores, W. C. de A. Pereira, and A. F. C. Infantosi, "Improving classification performance of breast lesions on ultrasonography," *Pattern Recognit.*, vol. 48, no. 4, pp. 1125–1136, Apr. 2015.

[26] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[27] Y. Freund, "Boosting a Weak Learning Algorithm by Majority," *Inf. Comput.*, vol. 121, no. 2, pp. 256–285, Sep. 1995.