Breast Cancer Diagnosis based on Joint Variable Selection and Constructive Deep Neural Network

R. Zemouri*, N. Omri†, C. Devalland‡, L. Arnould§, B. Morello†, N. Zerhouni† and F. Fnaiech¶
*Cedric-Lab, CNAM, Paris (France), ryad.zemouri@cnam.fr
†FEMTO-ST, Univ. of Bourgogne-Franche-Comté, Besançon (France), nabil.omri@femto-st.fr
‡Pathological Anatomy and Cytology, Hôpital Nord Franche-Comté, Belfort, (France), christine.devalland@hnfc.fr
§Dept. of Biologie and Tumor Pathology, Centre G-F Leclerc, Dijon (France), LArnould@cgfl.fr
¶Université de Tunis, ENSIT, LR13ES03 SIME, Tunis (Tunisia), fnaiech@ieee.org

***Abstract*—Breast cancer is the second most common cancer (after lung cancer) that affect women both in the developed and less developed countries. Nowadays, using the Computer Aided Diagnosis (CAD) techniques becomes a necessity for several reasons: assisting and improving physicians, speed in data processing, harmonization and aid of diagnosis, better access to advanced online-medicine. Recently, several works about Breast Cancer Computer Aided Diagnosis (BC-CAD) have been published, and Neural Networks techniques, particularly deep architectures represent a significant part of these works. In this paper, we prpose a BC-CAD based on joint variable selection and a Constructive Deep Neural Network "ConstDeepNet". A feature variable selection method is applied to decrease the number of inputs used to train a Deep Learning Neural Network. Experiments have been conducted on two datasets, the Wisconsin Breast Cancer Dataset (WBCD) and real data from the north hospital of Belfort (France) to predict the recurrence score of the Oncotype DX. Consequently, the use of joint variable algorithm with ConstDeepNet outperforms the use of the Deep Learning arechitecture alone.**

***Index Terms*—Tumor detection, clinical data, breast cancer, deep learning neural networks, classifier, feature selection.**

I. Introduction

According to World Health Organization (WHO), cancers remain the leading cause of death in France: they are responsible for nearly 150,000 deaths each year. Breast cancer is the first cancer of women in France and represents a public health problem, 54,000 new cases were diagnosed in 2015 in France. A 5-year net survival improved to 87% for patients diagnosed between 2005 and 2010 [1]. This is partly due to improved treatment and screening by mammography. Confirmation of the diagnosis is made by examining a fragment of the tissue by biopsy and microscopic analysis by a pathologist. Tumors are classified on the one hand following histological criteria and also thanks to immune-histochemical biomarkers which are prognostic factors of the tumor. The World Health Organization (WHO) classification of breast tumors defines 21 entities. Indeed, the pathologist provides a phenotypic immune-histochemical and morphological characterization. Thus, the different types of cancer are classified by their signatures are determined several types (luminal A, luminal B, BASAL LIKE, HER 2) which are attached to the classification WHO. For several years, the use of the molecular characteristics of tumors makes it possible to improve this classification thus conditioning the tumor prognosis. Expression profile production requires sophisticated technology reserved for specialized centers. Among the sophisticated technologies, bioinformatics, more precisely neural networks and Deep Learning has been so far applied to Biological Data [2], [3], [4], [5].

BREAST cancer diagnosis is usually performed by doctors based on Digital Mammography (DM) or on Medical Images (MI). In order to assist doctors to process big amount of images for different patients, Breast Cancer Computer Aided Diagnosis (BC-CAD) is becoming, nowadays, an appealing area of research. The techniques used for the BC-CAD depends whether the input data are medical images or not:

- BC-CAD using digital medical images: the diagnosis is based on numerical medical images processing such as histopathology or ultrasound images. In this case, image techniques are used for the pre-processing, segmentation and feature extraction step. More details about these techniques are given by three recently published review papers [6], [7], [8], [9]. One of the main problems of these breast cancer analyses is that the dataset used to train and to test the classifiers are too small and very hard to obtain. Three public datasets used by most of the scientists are available in [10], [11], [12].
- BC-CAD without digital medical images: other information can also be used for the BC-CAD. These data are obtained from the medical analysis as biopsy or biomarkers, or from more general information such as the age of the patient. Most of the literature methods are tested on the Wisconsin Breast Cancer Dataset (WBCD) taken from the University of California at Irvine (UCI) machine learning repository [13],

Feature selection is an important step in the BC-CAD. The first interest of this technique is to select the most important and significant features to improve the diagnosis. In the two techniques of medical diagnosis (BC-CAD with and without medical images), several features can be considered and often the doctors do not know which one is more significant than others. The second goal for the feature selection is to reduce the size of the input data when using machine learning algorithms.

In this paper, we intend to propose a BC-CAD perform-

ing jointly the optimization variable selection step with the application of a Constructive Deep learning Neural Network. Therefore, in section 2, a short review on related works on BC-CAD is given. While in sections 3 we present the constructive deep learning neural network methodologies and the variable selection procedure used to breast cancer diagnosis. Datasets and knowledge Extraction Procedure will be presented and discussed in section 4. Finally, in section 5, results and discussions are presented. The last section gives conclusions and further works.

II. Related works

A. BC-CAD based on image processing

For the BC-CAD based on image processing, Deep Convolutional Neural Network (DCNN) is the mostly used neural network architecture in recent publications [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. Deep Convolutional Neural Networks are used in several tasks specially in image classification and face recognition [32]. DCNNs are composed by several layers trained with supervised procedure to learn a complex model. They are able to automatically extract visual features from the pixel-level content such as edges and to identify complex visual concepts. Other NN architectures are used in some comparative studies such as [33], [34]. In [33], four types of NN (a multilayer perceptron (MLP), a 10-layer feedforward neural network and two types of SVM classifiers) were compared for malignancy grading classification from fine needle aspiration biopsy (FNA) images. Three classes were considered based on Bloom–Richardson grade distribution: Grade 1, 2 and 3 according to the risk low, intermediate and high malignancy. The best performances were obtained by the MLP while the 10-layer feedforward neural network is supposed having better classification performances. In [34] four classification models were compared namely, the multilayer perceptron (MLP) using back-propagation algorithm, probabilistic neural network (PNN), learning vector quantization (LVQ), and support vector machine (SVM). The results for malignancy diagnosis of cytological images shows that the best classifier is PNN, followed in order by SVM, LVQ, and MLP. SVM classifiers were used in [35] for malignancy diagnosis for breast cancer histopathology images and in [36] for breast ultrasound images. In [37], an adaptive gradient descent (AGD) algorithm based on variable learning rate is used for malignancy diagnosis classification of breast tumors ultrasound images.

B. BC-CAD based on other techniques

The second way to diagnosis the breast cancer is to study various pathological/biological aspects of the tumor that provide prognostic and predictive information. These include, for example, some surgical pathology (biomarkers) as the tumor type, tumor size, tumor grade and lymph node status or more general aspects as post-surgical measurements, personal data, and type of treatment [38]. A literature review showed that many types of Neural Networks are used as pattern recognition. Three comparison studies on the WBCD dataset were made by [39], [40] and [41]. In [40], the Support Vector Machine (SVM) gives the highest accuracy comparing to Decision Tree algorithm, Naive Bayes (NB) and k-Nearest Neighbours (k-NN). Learning Vector Quantization performs better results in comparison with the Back-Propagation Algorithm in [42]. Support Vector Machines (SVM) was used in [43], [40], [44]. The obtained results by [43] indicates that SVM can be effectively used for breast cancer diagnosis to help oncologists in order to decide which women should participate in a screening program and which should not. In [40], a BC-CAD based on a SVM classifier with F-score technique for feature selection was tested in the WBCD dataset. The best diagnosis model was obtained with five input features instead of nine. The F-score method is a simple technique which measures the discrimination of two sets of real numbers [45]. It was also used by [44] to evaluate the importance of five DNA viruses affecting the breast tumor. The diagnosis was performed by a SVM and the results shows that the accuracy of the SVM is slightly superior to the accuracy of a linear discriminate analysis. In [41], an Association Rules Neural Network (AR + NN) method is proposed for WBCD breast cancer diagnosis problem.

III. The Constructive Deep Neural Network

A Deep Multi Layer Perceptron with m hidden layers is formally described with two parameters (Λ, Φ). Λ is a vector that gives the number of neurons n_l for each layer l : $\Lambda = \big(n_0,\ n_1,\ ...,\ n_l,\ ...,\ n_m,\ n_{m+1}\big)$, where $l = 0$ and $l = m+1$ are respectively the input and the output layers, $l = 1\ to\ m$ are the m-hidden layers. Φ is the weight connections vector: $\Phi = \big(W^1,\ W^2,\ ...,\ W^l,\ ...,\ W^{m+1}\big)$. Each component W^l of this vector is the weight connection matrix, and each element w^l_{ij} of W^l is the weight connection between the i^{th} neuron of the layer l and the j^{th} neuron of the layer $l-1$. The proposed Constructive Deep Neural Network (ConstDeepNet) is a Deep Neural Network architecture that evolves gradually during the training process [46]. Let the size of the hidden layer l at the step t be denoted by n^t_l. To avoid infinite loop in the constructive process, two parameters must be defined: Max_{HL} and Max_n used to set respectively the maximum number of hidden layers and the maximum number of neurons per layer allowed by the user considered as thresholds.

The convergence condition ($ConvCond$) of the ConstDeepNet algorithm is based on quantitative criteria taken from the confusion matrix. This matrix contains information about the actual and the predicted classifications. True Positive (TP) values are the number of Positive classification correctly classified as Positive, True Negative (TN) values represent the number of Negative classifications correctly classified as Negative, False Positive (FP) values are the number of Negative classifications incorrectly classified as Positive, False Negative (FN) values represent the number of Positive classification incorrectly classified as Negative. Based on these values, three metrics are calculated:

- Accuracy : $Acc = \frac{TP+TN}{TP+TN+FP+FN}$
- Precision or Positive Predictive Value (PPV) : $PPV = \frac{TP}{FP+TP}$
- Sensitivity or True Positive Rate (TPR) : $TPR = \frac{TP}{TP+FN}$

As usually known in the community of learning machine, the reduction of the input variables is a useful operation. It has a great impact on the computing time and on the accuracy. In this paper, a variable STRASS algorithm [47] and the mRMR algorithms [48] are used. The Whole STRASS feature selection process is presented by the algorithm 2.

IV. The dataset

A. The Wisconsin-Madison breast cancer data set

The data set consist of 699 samples that were collected by Dr. W. H. Wolberg at the University of Wisconsin-Madison Hospitals taken from needle aspirates from human breast cancer tissue [49]. The Wisconsin-Madison breast cancer data set consists of nine features obtained from fine needle aspirates. The features are Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, Single epithelial cell size, Bare Nuclei, Bland Chromatin, Normal nucleoli, Mitoses, each of which is ultimately represented as an integer value between 1 to 10.

B. Recurrence Score (RS) prediction of the Oncotype DX (ODX)

The specific prognosis and treatment of the more and less aggressive breast tumors, is usually performed with Oncotype DX (ODX) gene expression assay [50]. The ODX is a commercial assay developed by the Genomic Health Inc (GHI) to determine the expression of a panel of 21 genes in tumor tissue [51], [52]. The result is reported as a Recurrence Score (RS) ranging from 0 to 100, divided into three categories: class 1 for low-risk ($RS < 18$), class 2 for the intermediate-risk ($18 \leq RS < 31$), and class 3 for the high-risk ($RS \geq 31$). In general, the more aggressive cancers (high-risk) require adjuvant chemotherapy while the more benign (low-risk) respond well to hormonal therapy alone. However, the

Algorithm 1 : The ConstDeepNet Algorithm

Step t = 0,
initialize the NN (one HL with one neuron),
let $l = 1$, the index of the 1^{st} hidden layer $HL(l)$,
//Begin the constructive procedure:
While (Constructive Process)
  // Evaluate the performances of the Deep Neural Network
  // on the training Set
  If ($Acc > \theta$ and $PPV > \theta$ and $TPR > \theta$) then
    $ConvCond = True$
  Else
    $ConvCond = False$
  End
  if $l > Max_{HL}$ and $ConvCond = False$ then
    The Deep Neural Network is NOT successfully built
    End of the Constructive Process
  End
  if $ConvCond = True$:
    The Deep Neural Network is successfully built
    End of the Constructive Process
  End
  If $n_l^t \leq Max_n$ and $ConvCond = False$ then
    add a new neuron ($n_l^{t+1} = n_l^t + 1$) for $HL(l)$,
    initialize randomly all the new weights,
  End
  If $n_l^t > Max_n$ and $l \leq Max_{HL}$ and $ConvCond = False$ then
    add a new hidden layer ($l = l + 1$) with one neuron,
    initialize randomly all the new weights,
  End
  Update the weights W^l of the $HL(l)$
  if Deep Neural Network is succesfuly built
    fine tuned of the last layer W^{l+1}
  End
END

Algorithm 2 : The STRASS Algorithm

Input:
  E : the whole set of pairs,
  $S_o = x_1, \ldots, x_r$: the set of features
  $DC_{tot} = DC(S_o, E)$:DC of S_o
  ρ : threshold for loss of DC
Output:
  S_f : selected features,
  Initialize S_f to ϕ
1. Selection of strongly predominant features
  For each $x_k \in S_o$do
    if $DCG(x_k, S_o - \{x_k\}, E) > 0$ then
      $S_f = S_f + \{x_k\}$
      $S_o = S_o \setminus \{x_k\}$
    end
  end
  Update E, as $E = E \setminus \{discriminated\ pairs\}$
2. Selection of the remaining weak relevant features
  While $DC(S_f, E) < \rho * DC_{tot}$ do
    $DC_{max} = 0$
    For each $x_k \in S_o$ do
      if $DC(x_k, E) > DC_{max}$ then
        $DC_{max} = DC(x_k, E)$
        $x_{max} = x_k$
      end
    end $S_f = S_f + \{x_{max}\}$
    $S_o = S_o \setminus \{x_{max}\}$
  end
  Update E, as $E = E \setminus \{discriminated\ pairs\}$
3. Elimination of redundant features
  For each $x_k \in S_o$ do
    if $DC(S_f \setminus \{x_k\}, E) = DC(S_f, E)$ then
      $S_f = S_f \setminus x_k$
    end
  end

TABLE I
INPUT FEATURE SELECTION RESULTS OBTAINED BY THE STRASS AND THE MRMR ALGORITHM FOR THE WISCONSIN-MADISON DATA SET

All the features	SRASS	mRMR
Clump thickness	**Yes**	**Yes**
Uniformity of cell size	No	**Yes**
Uniformity of cell shape	No	No
Marginal adhesion	**Yes**	No
Single epithelial cell size	**Yes**	**Yes**
Bare Nuclei	**Yes**	**Yes**
Bland Chromatin	No	**Yes**
Normal nucleoli	**Yes**	**Yes**
Mitoses	**Yes**	No

Oncotype DX tends to be expensive, the laboratory facilities with specialized equipment are limited and the time between biopsy and prognostic prediction can be very high.

The study data set contains 90 cases carcinoma mammary luminal B with available Oncotype DX test results from 2012 to 2017 taken from the Georges Francois Leclerc cancer centre and the North Trévenans County Hospital located respectively in Dijon and Belfort in France. The Recurrence Score is calculated from a set of ten input features: the age, the tumor size, the ganglionic status, four different tumor grading information, the Estrogen Receptor (RE), the Progesterone Receptor (RP) and ki67. According to the RS, three classes were formed: class 1 for the low risk with 40 cases, class 2 for the intermediate risk with 38 cases and class 3 for the high risk with 12 cases. The first half of the data set were used for the training process, and the second half for the test.

V. RESULTS AND DISCUSSION

A. The Wisconsin-Madison breast cancer data set

The constructive deep learning neural network has been trained on the Wisconsin-Madison breast cancer data set in two steps. In a first attempt, the ConstDeepNet, has been applied without the STRASS variable structure. The obtained results are given in Table II. The accuracy is about 96%, a cross validation has been performed with k=5. In a second attempt the ConstDeepNet has been applied with the STRASS and mRMR variable structure (table I), and the obtained results are given in table III. The obtained accuracy is about 96% with the same cross validation procedure with k=5. At each new run of the algorithm, and since all the synaptic coefficients are randomly fixed, a new architecture in terms of number of hidden layers and number of neurons in each hidden layer is obtained.

Here in all the experiments, we have only one output which takes 1 or 0 respectively when the patient has cancer the output is 1, if not the output is zero. By these experiments, one should retain that the use of a variable structure algorithm in order to decrease the number of inputs is very useful. From table II and table III, we have almost the same results in terms of accuracy. This input reduction of 33% is of great importance

TABLE II
RESULTS ON THE WISCONSIN-MADISON DATA SET WITHOUT INPUT REDUCTION

N.N Architecture	Accuracy
9 / 27 / 27 / 24 / 16 / 1	0.950 ± 0.012
9 / 34 / 23 / 17 / 1	0.963 ± 0.016
9 / 50 / 19 / 1	0.969 ± 0.008
9 / 37 / 39 / 24	0.964 ± 0.011

TABLE III
RESULTS ON THE WISCONSIN-MADISON DATA SET WITH THE STRASS AND MRMR INPUT REDUCTION

	Accuracy	
N.N Architecture	STRASS	mRMR
6 / 20 / 36 / 30 / 1	0.959 ± 0.007	0.960 ± 0.007
6 / 39 / 26 / 16 / 1	0.962 ± 0.010	0.963 ± 0.008
6 / 37 / 24 / 1	0.959 ± 0.012	0.966 ± 0.014
6 / 50 / 30 / 1	0.962 ± 0.014	0.963 ± 0.021

TABLE IV
INPUT FEATURE SELECTION RESULTS OBTAINED BY THE STRASS AND THE MRMR ALGORITHM FOR THE ONCOTYPE DX PREDICTION

All the features	SRASS	mRMR
Age	**Yes**	**Yes**
Tumor Size	**Yes**	**Yes**
Ganglionic Status	**Yes**	No
SBR Grade	**Yes**	No
Glande Grade	No	**Yes**
Nuclei Grade	No	No
Mitosies Grade	**Yes**	**Yes**
Estrogen Receptor (RE)	**Yes**	**Yes**
Progesterone Receptor (PR)	**Yes**	**Yes**
ki67	No	**Yes**

in computation time and in cost, because the doctors have to record only six measures instead of nine on each patient.

B. Recurrence Score (RS) prediction of the Oncotype DX (ODX)

Table IV gives the features selected by the STRASS and the mRMR. The figure 1 presents the comparison results obtained for each class and for each metric (Acc, PPV, TPR). Different convergence conditions of the ConstDeepNet algorithm were tested by varying the values of the threshold θ from 0,1 to 0,9. To evaluate the repeatability of the algorithm, each test point θ of the figure 1 gives the mean value obtained by running the ConstDeepNet algorithm 100 times. The red plots present the results obtained with all the input features, the blue and the green plots the results with the input features selection obtained by respectively the STRASS and the mRMR algorithm. The objective to reach for a good classifier is to reduce the size of the False Positive and False Negative population. This means that the metric value must be close to one. For example, if $PPV = 0.5$ means that for each one

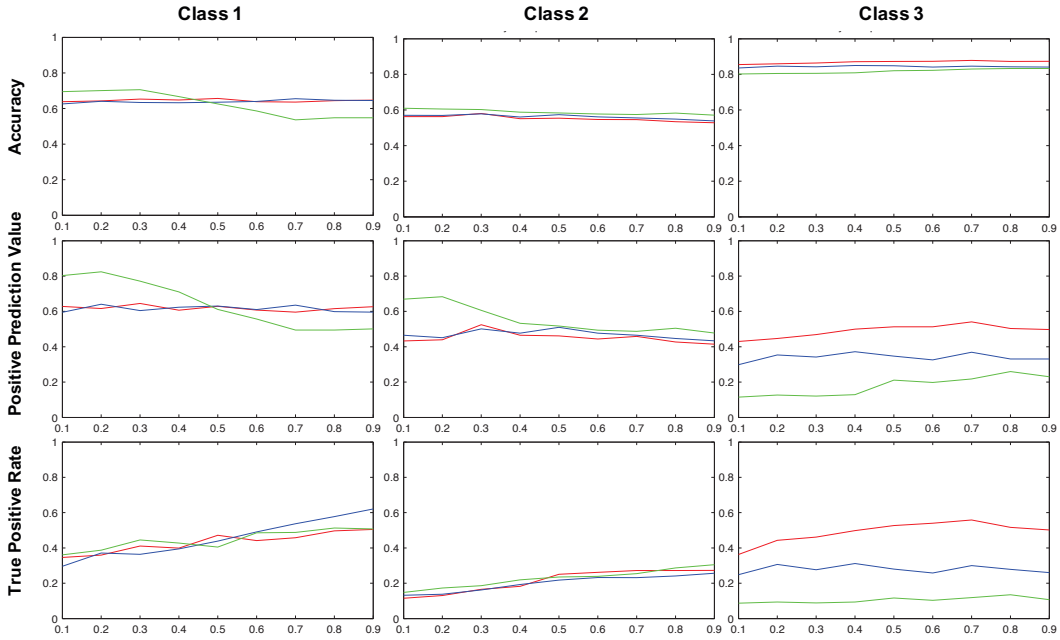

Fig. 1. The obtained results for the Recurrence Score (RS) prediction of the Oncotype DX (ODX). The three metrics (Accuracy, Positive Predictive Value (PPV) and True Positive Rate (TPR)) are calculated for each class (grade) according to the variation of the convergence threshold θ. Red plot: results obtained with all the features. Blue plot: results obtained with the features selected by the STRASS algorithm. Green plot: results obtained with the features selected by the mRMR algorithm.

True Positive there is one False Positive prediction, and if $PPV > 0.5$ means that there is more True than False Positive prediction. As for the previous data set, the accuracy of the Recurrence Score prediction obtained with less input features is almost the same with all the features. The input reduction of 30% is a great importance in the reduction of the clinical investigations costs.

VI. Conclusion

This paper deals with a new joint variable selection and constructive deep learning neural network algorithm used to the diagnosis of breast cancer disease. The reduction of clinical variables by STRASS and mRMR selection variable algorithm leads a drastically computing time and medical cost investigations. As a future works, we are working on using more data sets of patients to test the proposed joint algorithm. Moreover, we intend to compare this joint algorithm with another joint neural network linked with a pruning and growing standard algorithm rather than deep learning algorithm.

References

[1] L. Schoutteten, M. Colonna, H. Curé, P. Delafosse, N. Mitha, N. Zerhouni, G. Gavazzi, and A. Seigneurin, "Breast cancer incidence and survival in elderly women during the 1989–2012 period: A population-based study in a french area," *Cancer Treatment and Research Communications*, vol. 11, pp. 6 – 9, 2017.

[2] M. Mahmud, M. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Trans. Neural Netw. Learn. Syst*, 2018.

[3] A. Pouliakis, E. Karakitsou, N. Margari, P. Bountris, M. Haritou, J. Panayiotides, D. Koutsouris, and P. Karakitsos, "Artificial neural networks as decision support tools in cytopathology: Past, present, and future," *Biomedical Engineering and Computational Biology*, vol. 7, pp. 1–18, 02 2016.

[4] A. Sheikhtaheri, F. Sadoughi, and Z. Hashemi Dehaghi, "Developing and using expert systems and neural networks in medicine: A review on benefits and challenges," *Journal of Medical Systems*, vol. 38, no. 9, p. 110, Jul 2014.

[5] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[6] Y. L. Kergosien and D. Racoceanu, "Semantic knowledge for histopathological image analysis: from ontologies to processing portals and deep learning," in *Proc. SPIE 10572, 13th International Conference on Medical Information Processing and Analysis, 105721F*, Nov. 2017.

[7] M. Saha, R. Mukherjee, and C. Chakraborty, "Computer-aided diagnosis of breast cancer using cytological images: A systematic review," *Tissue and Cell*, vol. 48, no. 5, pp. 461 – 474, 2016.

[8] M. Aswathy and M. Jagannath, "Detection of breast cancer on digital histopathology images: Present status and future possibilities," *Informatics in Medicine Unlocked*, 2016.

[9] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[10] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, July 2016.

[11] A. Pêgo and P. Aguiar, "Bioimaging 2015," 2015.

[12] D. C. Moura, M. A. G. López, P. Cunha, N. G. de Posada, R. R. Pollan, I. Ramos, J. P. Loureiro, I. C. Moreira, B. M. F. de Araújo, and T. C. Fernandes, *Benchmarking Datasets for Breast Cancer Computer-Aided Diagnosis (CADx)*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 326–333.

[13] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Systems with Applications*, vol. 46, pp. 139 – 144, 2016.

[14] M. Saha, C. Chakraborty, and D. Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29 – 40, 2018.

[15] N. Wahab, A. Khan, and Y. S. Lee, "Two-phase deep convolutional neural network for reducing class skewness in histopathological images

based breast cancer detection," *Computers in Biology and Medicine*, vol. 85, pp. 86 – 97, 2017.

[16] B. E. Bejnordi, J. Lin, B. Glass, M. Mullooly, G. L. Gierach, M. E. Sherman, N. Karssemeijer, J. van der Laak, and A. H. Beck, "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 929–932.

[17] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PLOS ONE*, vol. 12, no. 6, pp. 1–14, 06 2017.

[18] M. Saha, C. Chakraborty, I. Arun, R. Ahmed, and S. Chatterjee, "An advanced deep learning approach for ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer," *Scientific Reports*, vol. 7, no. 1, 2017.

[19] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific Reports*, vol. 7, no. 1, 2017.

[20] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Scientific Reports*, vol. 7, 2017.

[21] W. Sun, T.-L. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 4 – 9, 2017, recent Developments in Machine Learning for Medical Imaging Applications.

[22] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer, "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303 – 312, 2017.

[23] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Medical Image Analysis*, vol. 37, pp. 114 – 128, 2017.

[24] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214 – 223, 2016.

[25] M. Veta, P. J. van Diest, and J. P. W. Pluim, *Cutting Out the Middleman: Measuring Nuclear Area in Histopathology Slides Without Segmentation*. Cham: Springer International Publishing, 2016, pp. 632–639.

[26] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, July 2016, pp. 2560–2567.

[27] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken, and J. Van Der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific reports*, vol. 6, p. 26286, 2016.

[28] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, May 2016.

[29] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248 – 257, 2016.

[30] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221 – 231, 2016.

[31] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, *Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 411–418.

[32] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, and S. Tubaro, "Deep convolutional neural networks for pedestrian detection," *Signal Processing: Image Communication*, vol. 47, pp. 482 – 489, 2016.

[33] Łukasz Jeleń, A. Krzyżak, T. Fevens, and M. Jeleń, "Influence of feature set reduction on breast cancer malignancy classification of fine needle aspiration biopsies," *Computers in Biology and Medicine*, vol. 79, pp. 80 – 91, 2016.

[34] Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury, "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Systems Journal*, vol. 8, no. 3, pp. 949–964, Sept 2014.

[35] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Processing*, vol. 122, pp. 1 – 13, 2016.

[36] X. Xi, H. Shi, L. Han, T. Wang, H. Y. Ding, G. Zhang, Y. Tang, and Y. Yin, "Breast tumor segmentation with prior knowledge learning," *Neurocomputing*, vol. 237, pp. 145 – 157, 2017.

[37] B. K. Singh, K. Verma, and A. Thoke, "Adaptive gradient descent backpropagation for classification of breast tumors in ultrasound imaging," *Procedia Computer Science*, vol. 46, pp. 1601 – 1609, 2015, proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014 at Bolgatty Palace & Island Resort, Kochi, India.

[38] J. M. Jerez-Aragonés, J. A. Gómez-Ruiz, G. Ramos-Jiménez, J. Muñoz-Pérez, and E. Alba-Conejo, "A combined neural network and decision trees model for prognosis of breast cancer relapse," *Artificial Intelligence in Medicine*, vol. 27, no. 1, pp. 45 – 63, 2003.

[39] R. D. H. Devi and P. Deepika, "Performance comparison of various clustering techniques for diagnosis of breast cancer," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec 2015, pp. 1–5.

[40] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3240 – 3247, 2009.

[41] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3465 – 3469, 2009.

[42] D. Furundzic, M. Djordjevic, and A. J. Bekic, "Neural networks approach to early breast cancer detection," *Journal of Systems Architecture*, vol. 44, no. 8, pp. 617 – 633, 1998.

[43] L. Álvarez Menéndez, F. de Cos Juez, F. S. Lasheras, and J. Álvarez Riesgo, "Artificial neural networks applied to cancer detection in a breast screening programme," *Mathematical and Computer Modelling*, vol. 52, no. 7, pp. 983 – 991, 2010, mathematical Models in Medicine, Business & Engineering 2009.

[44] C.-L. Huang, H.-C. Liao, and M.-C. Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," *Expert Systems with Applications*, vol. 34, no. 1, pp. 578 – 587, 2008.

[45] Y.-W. Chen and C.-J. Lin, *Combining SVMs with Various Feature Selection Strategies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 315–324.

[46] R. Zemouri, "An evolutionary building algorithm for deep neural networks," in *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, June 2017, pp. 1–7.

[47] B. Chebel-Morello, S. Malinowski, and H. Senoussi, "Feature selection for fault detection systems: application to the tennessee eastman process," *Applied Intelligence*, vol. 44, no. 1, pp. 111–122, Jan 2016.

[48] X. V. Nguyen, J. Chan, S. Romano, and J. Bailey, "Effective global approaches for mutual information based feature selection," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: ACM, 2014, pp. 512–521.

[49] M. Lichman, "UCI machine learning repository," 2013.

[50] e. a. M.B. Flanagan, D.J. Dabbs, "Histopathologic variables predict oncotype dx recurrence score." *Mod Pathol,*, vol. 21, no. 10, pp. 1255–1261, Oct. 2008.

[51] M. E. Klein, D. J. Dabbs, Y. Shuai, A. M. Brufsky, R. Jankowitz, S. L. Puhalla, and R. Bhargava, "Prediction of the oncotype dx recurrence score: use of pathology-generated equations derived by linear regression analysis," *Modern Pathology, Nature Publishing Group*, vol. 26, p. 658–664, 2013.

[52] F. O. Ademuyiwa, A. Miller, T. O'Connor, S. B. Edge, M. A. Thorat, G. W. Sledge, E. Levine, and S. Badve, "The effects of oncotype dx recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort," *Breast Cancer Research and Treatment*, vol. 126, no. 3, pp. 797–802, Apr 2011.