



Automated Detection and Classification of Breast Cancer Nuclei with Deep Convolutional Neural Network

Shanmugham Balasundaram¹, Revathi Balasundaram², Ganesan Rasuthevar³,
Christeena Joseph⁴, Annie Grace Vimala⁵, Nanmaran Rajendiran⁶ &
Baskaran Kaliyamurthy^{7,*}

¹Department of Electronics & Communication Engineering,
Sri Manakula Vinayagar Engineering College, Pondicherry, India

²Dhanalakshmi Srinivasan College of Engineering and Technology,
Mamallapuram, Tamilnadu, India

³Department of Biomedical Engineering, EGS Pillay Engineering College,
Nagapattinam, Tamil Nadu, India

⁴Department of Electronics & Communication Engineering,
SRM Institute of Science and Technology, Chennai, India

⁵Department of Electronics & Communication Engineering,
Saveetha School of Engineering-SIMATS, Thandalam, Chennai, India

⁶Department of Biomedical Engineering, Saveetha School of Engineering-SIMATS,
Thandalam, Chennai, India

⁷Department of Electronics & Communication Engineering,
Chennai Institute of Technology, Chennai, India

*E-mail: baskarank@citchennai.net

Abstract. Heterogeneous regions present in tissue with respect to cancer cells are of various types. This study aimed to analyze and classify the morphological features of the nucleus and cytoplasm regions of tumor cells. This tissue morphology study was established through invasive ductal breast cancer histopathology images accessed from the Databiox public dataset. Automatic detection and classification was carried out by means of the computer analytical tool of deep learning algorithm. Residual blocks with short skip were employed with hidden layers of preserved spatial information. A ResNet-based convolutional neural network was adapted to perform end-to-end segmentation of breast cancer nuclei. Nuclei regions were identified through color and tubular structure morphological features. Based on the segmented and extracted images, classification of benign and malignant breast cancer cells was done to identify tumors. The results indicated that the proposed method could successfully segment and classify breast tumors with an average Dice score of 90.68%, sensitivity = 98.64, specificity = 98.68, and accuracy = 98.82.

Keywords: *breast cancer; classification; deep convolutional neural network; Dice score; ResNet.*

Received March 1st, 2021, 1st Revision July 4th, 2021, 2nd Revision July 9th, 2021, Accepted for publication July 14th, 2021.

Copyright © 2021 Published by IRCS-ITB, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2021.15.2.3

1 Introduction

Breast cancer is the second most deadly form of cancer in India [1]. Out of 115,251 people affected by breast cancer, 53,592 were estimated to have died in the year 2008 [2]. The average prevalence of the disease is 22.9 for every 100,000 Indians, which is nearly one third compared to the western countries [3-4]. In 2016, the number of new cancer cases was predicted to be 1,450,000, which could cause the mortality rate to be as high as 600,000-700,000 [5-6]. The pathologist reports based on tissue investigation in clinical laboratories are also proportionate with the predicted population [7]. A report based on the Times of India Study in 2007 found that 210 out of 339 *talukas* (townships) in the state of Maharashtra do not have a pathologist [8].

Thus, the need for advanced and automatic image processing techniques to qualitatively analyze histopathological images is obvious. Suspicious tissue growth and the definite nature of suspected cancer cells can be confirmed through histological assessment. Histology, in a general sense, is the microscopic analysis of living tissues. More than 70% of diseases, most predominantly tumor cells, are first diagnosed through histopathology. Clinical diagnosis involves biopsy of tumor cells or tissue by a pathologist under a magnifying lens. The morphology of tissue patterns can be studied by the pathologist to identify the different stages of ailment in the tissue [9]. Computer-aided analysis could facilitate the pathologist's work and help to identify diseases better. Elston and Ellis in the NGS protocol [10] proposed that morphological patterns can be better diagnosed by analyzing histology tissue image processing techniques. Also, the computer-aided analysis of tissue nuclei is a great method for distinguishing between patients with a good or bad diagnosis [11].

Image pre-processing is the primary step used to remove noise from nuclei regions and the surrounding white space. This step is implemented based on identification of morphological features from the image. The digital image pre-processing tool Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to enhance the image's local contrast. CLAHE involves computing histograms for various sections of the image to distribute contrast uniformly and also defines each region of the image clearly rather than by histogram equalization. As color-based morphological features play a vital role in the distinguishing of cancer nuclei from surrounding regions, CLAHE pre-processing is required. Segmentation is another vital tool in image processing utilized to segment the nuclei region in histopathological images. This helps in extraction and further classification. It is enabled by a deep convolution neural network to detect the nucleus of tumor cells.

2 Methods

The input histopathological breast cancer images used in this study were acquired from the DatabioX public access dataset. It contains 922 images in total, with three-grade classification of invasive ductal breast carcinoma and four different levels of magnification, i.e. 4x, 10x, 20x and 40x. The dataset contains 259 grade-1, 366 grade-2 and 297 grade-3 breast cancer histologic images [12].

2.1 Network Architecture

The complete network structure of the current work is illustrated in Figure 1. The structure was stirred with a convolutional residual network [13] that contains an encoding track, i.e. a convolutional network and a decoding track network. Retrieval of the desired features from the input histopathological images and transformation to a multidimensional feature vector provides feature extraction through object segmentation of the convolutional network [14]. Various scales of feature retrieval with multiple resolution levels is contained within the convolutional and residual parts. The architecture in the present work was built using several types of architectural blocks. The block shown in Figure 2 is a residual block. It consists of triple convolutional layers (filters such as 1×1 , 3×3 , 1×1 respectively) and each convolutional level is trailed by a ReLu (rectified linear unit) as activating function. A batch normalization layer is sandwiched between every convolution layer/ReLu pair, where the batch normalization layer minimizes the internal covariance shift and increases the speed of the training process.

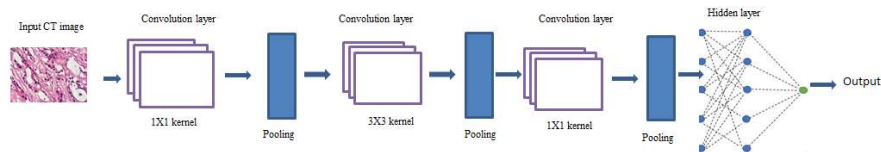


Figure 1 Deep convolutional neural network architecture.

Down sampling of the histopathological image is performed by a max pooling layer within the convolutional part for feature compression. Likewise, the input images are put together by pooling. Thus the pooling provides the pixel probability for every defined label. The ResNet-50 architecture was utilized in the present work for breast cancer segmentation. The residual block is established with a short skip connection and the convolutional part with a long skip connection in order to preserve spatial information that may be lost during the pooling operation. The previous layer is connected directly through the short skip provision rendered by the residual block. Arrangement of nested long and short connections enhances the flow of data among the various layers of the network.

The residual type of convolutional network with long and short residual layers performs end-to-end segmentation of raw histopathological image data to achieve the final segmentation result.

2.2 Deep Convolutional Neural Network

For breast cancer segmentation, transfer learning combined with a deep convolutional neural architecture is involved [15]. Arranging a greater number of stack layers allows for better distinguishing capability but is challenging during training of these networks, whereas shallow networks are comparatively easier to operate during training. Thus, increasing the number of stack layers not only leads to degradation of the network performance but also to poor convergence of the network.

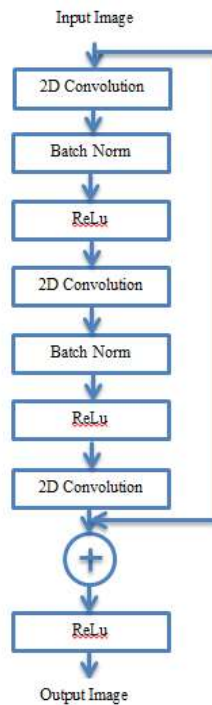


Figure 2 Residual network architecture.

This degradation problem can be overcome efficiently by incorporating hidden layer blocks, because accuracy and convergence evidently improve tasks carried out with ImageNet and MSCOCO [16]. The deep learning method involved in

the deep residual network effectively solves the degradation problem and eases network optimization through skip connection and activation after addition [17].

The deep layer residual element in Figure 2 is used to build the deep neural network. A hidden layer in the deep neural unit transmits the signal between the blocks through skip connection. Every single neural layer contains triple convolutional layers with filters such as $(1 \times 1, 3 \times 3, 1 \times 1)$ and each convolutional layer followed by a batch normalization layer. The neural network is trained by learning the following function:

$$x_{l+1} = x_l + R(x_l, w_l) \quad (1)$$

where R denotes the residual function, i.e. a double convolutional layer stack accompanied with batch normalized layers as shown in Figure 2; x_l and x_{l+1} are the input and output features of the l -th ResNet unit; and w_l is a set of weights and biases associated with the l -th ResNet unit.

3 Results

Segmentation of cancer cell regions in standard hematoxylin and eosin (H&E) stained histopathological images of breast cancer tissue (and a number of other tissues) is a key finding of computer-aided review and assessment of histopathological slides. Segmentation of cancer cell regions is a basic way of automatically categorizing stained histochemical slides of tumor-holding cells. This is a safeguard against mistakenly analyzing the supporting tissue of an organ rather than tumor stromal zones. The evaluation system was created according to the manner in which pathologists analyze tumor slides, as shown in Figure 3.

Estimating the density of tubules is among the most imperative predictors to determine the grade of breast cancer. The occurrence rate of tubules formed by tumors indicates the grade of tumor distinction and also to what extent the breast cancer cells affect its surrounding regions or other normal cells. Identification of tubules in histopathology images is done by observing white or pale pink colored lumen bounded by the nuclei of cells, as depicted in Figure 4.

This color morphology and nearby lumen region with nuclei boundary helps to predict the incidence of a tubular region. However, tubular region identification with these data could lead to erroneous detection of fat or supporting tissue regions instead of tubules. This kind of tubular detection is observed manually with the interpretation of surface knowledge. Various tissue regions from the inputted breast cancer histopathological image are identified by color variations due to H&E staining (cytoplasm and connectivity = pink, lumen and fat = white, and cell nuclei = purple). Extraction of lumen nuclei along with their surrounding region is done through spatial and proximity features.

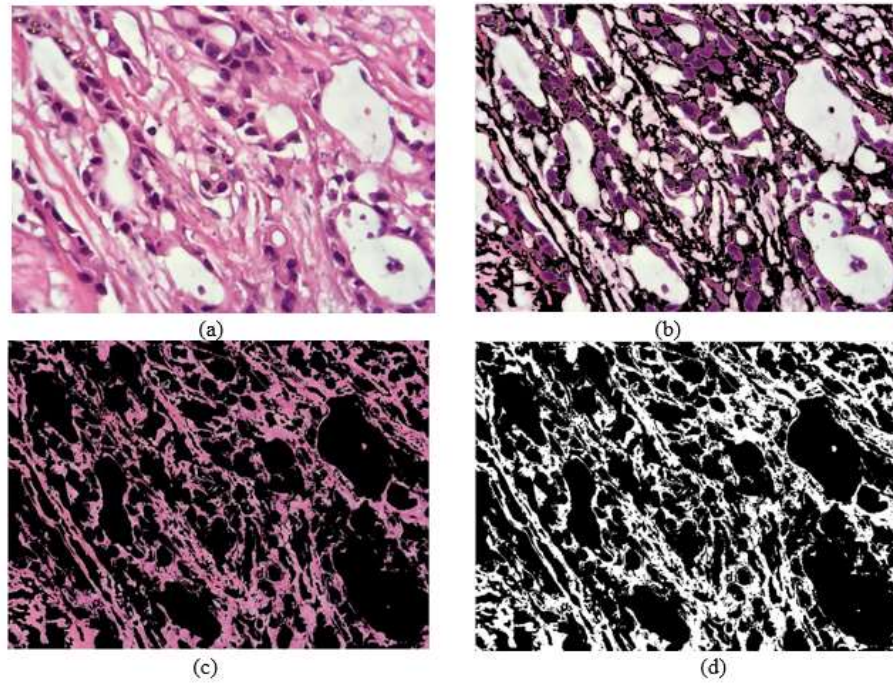


Figure 3 Clustered biopsy images: (a) normal image, (b) nuclei region, (c) stoma cells and connective region, and (d) white region.

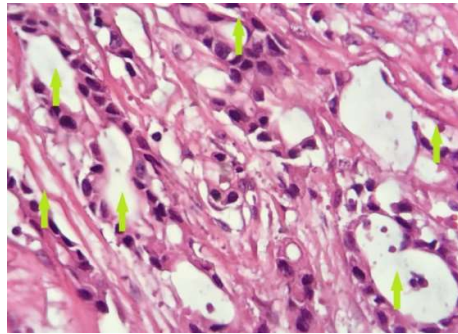


Figure 4 Tubular region identified in a histopathological image.

3.1 Evaluation Parameters

The accuracy of the current technique can be quantitatively evaluated by using a performance measure such as the Dice coefficient (DC). This parameter helps to compute the variance between automated segmented output and manually

segmented reference standard output. DC in Eq. (2) is predicted by computing the ratio between two times the intersection of binary masks by the sum of total the number of elements in each set, i.e.

$$DC = \frac{2(A_s \cap B_m)}{|A_s + B_m|} \quad (2)$$

where A_s is a binary mask created in the current method and B_m is the binary mask of the manual reference standard. DC always lies in the range of $[0, 1]$, where 1 indicates 100% overlap and 0 indicates no overlap between output and reference standard [18,19]. Sensitivity is defined by:

$$\text{Sensitivity} = \frac{TP}{FN+TP} \quad (3)$$

The equation for specificity is defined by:

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (4)$$

The equation of accuracy is defined by:

$$\text{Accuracy} = \frac{TP+TN}{FP+FN+TP+TN} \quad (5)$$

where, TP is the number of true positives (actual affected nuclei identified correctly), TN is the number of true negatives (non-affected nuclei predicted as affected nuclei), FP is the number of false positives (affected nuclei not predicted as affected nuclei), and FN is the number of false negatives (non-affected nuclei predicted correctly).

The detection module is trained with a mix of histopathological images of all three grades, consisting of 554 and 368 test images. The accuracy level between the training and the test dataset is given in Figure 5.

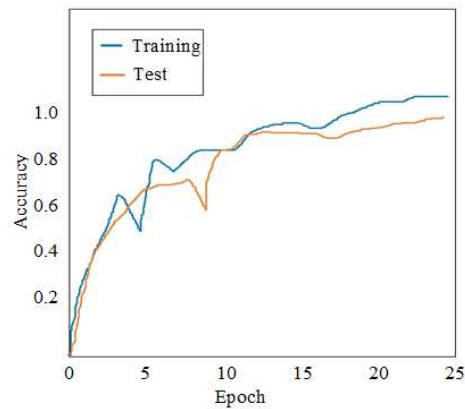


Figure 5 Validation of DCNN for the given dataset.

4 Discussion

The current work focused exclusively on the detection, segmentation and classification of regions of interest in breast tumor histopathological images. Breast tumor identification and classification was done based on heterogeneous surroundings of cancer nuclei. This synchronizes well with a previous single-cell pathology study that analyzed the stromal surroundings of the tumor [20]. Feature extraction of tumor cells using CNN with a UNET architecture that establishes long skip connections has been done in a previous study. It produced a Dice coefficient for tumor identification of 0.88 [21], while the current work achieved about 0.90. Also, the long skip connection was avoided in the current work through step-by-step adding of a deep residual CNN architecture. Based on their shape, size and appearance, the nuclei regions can be identified in H&E stained breast histopathology images using the marked point process technique [22,23]. This kind of marking process may result in misidentification of fat tissues instead of tumor regions. Comparison of classification of benign and malignant tumor images as in the current study has not been done previously. One study has previously reported the detection of mitosis cells in histopathology breast cancer images by morphological features such as color of the nuclei and cytoplasm in the tumor location. The classification of mitosis and non-mitosis was carried out to help pathologists. A detection rate of 71% was achieved based on statistical features [24].

This color morphology identification coincides well with the present study, but by incorporating deep residual convolutional training, the classification achieved better accuracy. A review study found that a complete digital workflow of histopathological breast cancer images is lacking in many places, so incorporation of artificial intelligence through deep learning algorithms could produce better results in classification and detection [25]. According to the recommendation, a deep convolutional neural network was employed for classification of tumor regions present in breast tissue. The comparison of performance metrics for sensitivity, specificity and accuracy are shown in Figures 6, 7 and 8 respectively. Specificity indicates the likelihood of a classifier that produces a negative result when disease is absent. This is also known as the true negative (TN) rate. The sensitivity analysis in Figure 6 provides the proportion of positive cases that are classified correctly and reveals the ability of the classifier to correctly predict positive cases, as shown in Figure 7. Accuracy indicates the general prediction capability of the projected deep learning model. True positives and true negatives indicate the capacity of the classifier to correctly predict the presence and absence of breast cancer. Fig. 8 compares the accuracy of several existing method and the proposed method, where the X axis represents the number of images taken for analysis and the Y axis represents the accuracy values obtained in percentage.

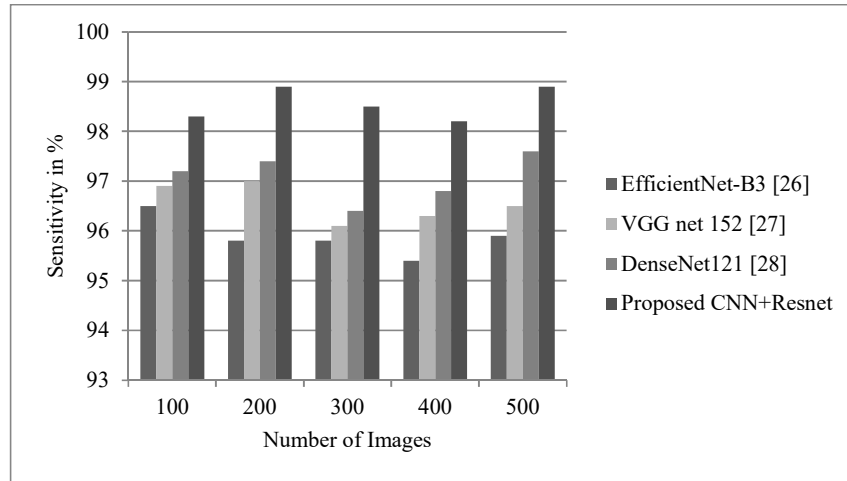


Figure 6 Comparison of the sensitivity parameter between EfficientNet, VGGNet, Densenet121 and ResNet.

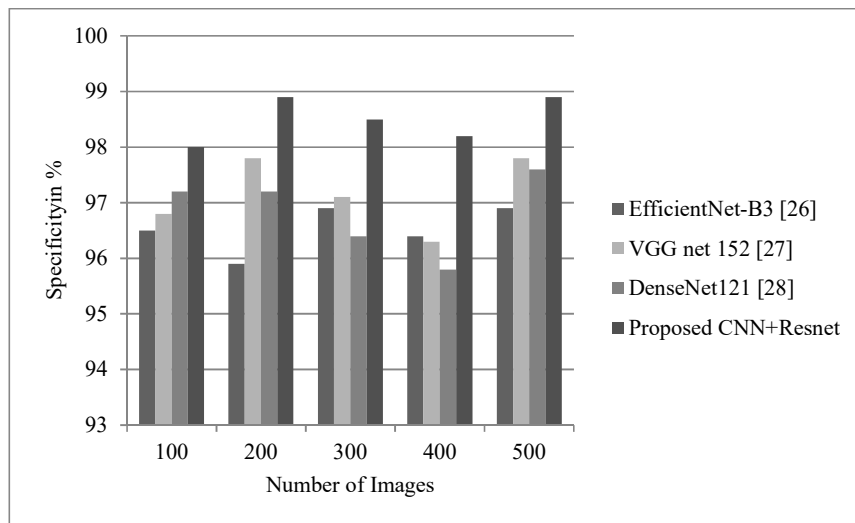


Figure 7 Comparison of the specificity parameter between EfficientNet, VGGNet, Densenet121 and ResNet.



Figure 8 Comparison of the accuracy parameter between EfficientNet, VGGNet, Densenet121 and ResNet.

An early study of breast cancer classification using EfficientNet-B3 reported an accuracy level of 97%. The method was found to be image-specific, i.e. the classifier was abstractive to multi-resolution images except for its input [26]. Meanwhile, the proposed method consisting of CNN + ResNet achieved an accuracy of 98.82% for input images with four different resolutions.

The classifier with VGGNet was observed doing classification between normal and malignant breast cancer cells. It achieved an accuracy level of 96.19% and a specificity level of 93.33% [27]. The current work involved more efficient detection of multiple classes of input images with better accuracy and specificity (98.68%). A study on a DenseNet deep classifier with a pooling layer has previously been reported, using more parameters to achieve better classification with histopathological images. It had a complex sub-network with inflated model size and over-fitting network layers [28]. The present study with ResNet utilized optimum parameters in the hidden layer model to overcome over-fitting limitations.

5 Conclusion

The present work developed an enhanced histopathological image analysis technique for segmentation and classification of objects of interests related to breast cancer, taking advantage of digital image processing to detect tumors in histopathological images. It can detect various graded breast cancer images correctly through proper segmentation while overcoming obstacles due to uneven

stacking of residual layers. Morphological features of tubular structures are detected through robust segmentation algorithms. This segmentation preceded automatic detection and extraction of nuclei and lumen regions from H&E-stained breast cancer histopathological images. The classification of breast cancer nuclei as benign or malignant is achieved through a ResNet-based deep convolutional neural network. The detection and classification of tumor-affected breast nuclei regions were qualitatively analyzed in terms of Dice coefficient, specificity and accuracy.

References

- [1] Xiaomei, M. & Yu, H., *Global Burden of Cancer*, The Yale Journal of Biology and Medicine, **79**(3-4) pp. 85-94, 2006.
- [2] *IARC Fact sheet*, Available from: <http://www.globocan.iarc.fr/factsheet.asp> (10 June 2013).
- [3] Manoharan, N., Nair, O., Shukla, N.K. & Rath, G.K., *Descriptive Epidemiology of Female Breast Cancer in Delhi, India*, Asian Pacific journal of cancer prevention: APJCP, **18**(4), pp. 1015-1018, 2017.
- [4] Stewart, B. & Wild, C., *World Cancer Report 2014*, Lyon: International Agency for Research on Cancer, 2014.
- [5] India State-level Disease Burden Initiative Cancer Collaborators, *The Burden of Cancers and Their Variations Across the States of India: The Global Burden of Disease Study 1990-2016*, The Lancet. Oncology, **19**(10), pp. 1289-1306, 2018.
- [6] D'Souza, N.D., Murthy, N.S. & Aras, R.Y., *Projection of Burden of Cancer Mortality for India 2011-2026*, Asian Pacific Journal of Cancer Prevention: APJCP, **14**(7), pp. 4387-4392, 2013.
- [7] Agarwal, G. & Ramakant, P., *Breast Cancer Care in India: The Current Scenario and the Challenges for the Future*, Breast Care (Basel, Switzerland), **3**(1), pp. 21-27, 2008.
- [8] Timesofindia.indiatimes.com/city/pune (10 November 2007).
- [9] Hanby, A.M. & Walker, C., *Tavassoli FA. Devilee P: Pathology and Genetics: Tumours of the Breast and Female Genital Organs. WHO Classification of Tumours Series – Volume IV*, Lyon, France: IARC Press, Breast Cancer Res., **6**(133), 2004.
- [10] Elston, C.W. & Ellis, I.O., *Pathological Prognostic Factors in Breast Cancer. I. The Value of Histological Grade in Breast Cancer: Experience from a Large Study with Long-term Follow-up*, Histopathology, **19**(5), pp. 403-410, 1991.
- [11] Milosevic, M., Jankovic, D., Milenkovic, A. & Stojanov, D., *Early Diagnosis and Detection of Breast Cancer. Technol Health Care*, **26**(4), pp. 729-759, 2018.

- [12] Hamidreza, B., Elham, A., Maryam, T. & Somayyeh, J.J., *A Histopathological Image Dataset for Grading Breast Invasive Ductal Carcinomas*, Informatics in Medicine Unlocked, **19**, 2020, DOI:10.1016/j.imu.2020.100341.
- [13] Wickstrøm, K., Kampffmeyer, M. & Jenssen, R., *Uncertainty and Interpretability in Convolutional Neural Networks for Semantic Segmentation of Colorectal Polyps*, Medical Image Analysis, **60**, 2019.
- [14] Ronneberger, O., Fischer, P. & Brox, T., *U-net: Convolutional Networks for Biomedical Image Segmentation*, Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds), *Medical Image Computing and Computer-assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, pp. 234-241, 2015.
- [15] Ponzio, F., Macii, E., Ficarra, E. & Cataldo, S.D., *Colorectal Cancer Classification using Deep Convolutional Networks – An Experimental Study*, 5th International Conference on Bioimaging, pp. 58-66, 2018.
- [16] Liu, X., Guo, S., Zhang, H., He, K., Mu, S., Guo, Y. & Li, X., *Accurate Colorectal Tumor Segmentation for CT scans Based on the Label Assignment Generative Adversarial Network*, Medical Physics, **46**(8), pp. 3532-3542, 2019.
- [17] Krizhevsky, A., Sutskever, I. & Hinton, G.E., *Imagenet Classification with Deep Convolutional Neural Networks*, Communications of the ACM, **60**(6), pp. 84-90, 2017.
- [18] Crum, W. R., Camara, O. & Hill, D.L.G., *Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis*, IEEE Transactions on Medical Imaging, **25**(11), pp. 1451-1461, 2006.
- [19] Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells III, W.M., Jolesz, F.A. & Kikinis, R., *Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index*, Academic Radiology, **11**(2), pp. 178-189, 2004.
- [20] Jackson, H.W., Fischer, J.R., Zanotelli, V., Ali, H.R., Mechera, R., Soysal, S.D., Moch, H., Muenst, S., Varga, Z., Weber, W.P. & Bodenmiller, B., *The Single-cell Pathology Landscape of Breast Cancer*, Nature, **578**(7796), pp. 615-620, 2020.
- [21] Ujjwal, B., Sanjay, T., Swapnil, R., Sudeep, G., Meenakshi, T., Aliasgar, M., Nilesh, S., Mayuresh, A. & Abhishek, M., *A Novel Approach for Fully Automatic Intra-Tumor Segmentation With 3D U-Net Architecture for Gliomas*, Frontiers in Computational Neuroscience, **14**(10), 2020.
- [22] Avenel, C. & Kulikova, M.S., *Marked Point Processes with Simple and Complex Shape Objects for Cell Nuclei Extraction from Breast Cancer H&E Images*, Proceedings of SPIE Medical Imaging: Digital pathology, 2013.
- [23] Zhan, X., Cheng, J., Huang, Z., Han, Z., Helm, B., Liu, X., Zhang, J., Wang, T. F., Ni, D. & Huang, K., *Correlation Analysis of Histopathology*

- and Proteogenomics Data for Breast Cancer*, Molecular & Cellular Proteomics, **18**(8 suppl 1), pp. S37-S51, 2019.
- [24] Irshad, H., *Automated Mitosis Detection in Histopathology Using Morphological and Multi-channel Statistics Features*, Journal of Pathology Informatics, **4**(1), 10, 2013.
- [25] Robertson, S., Azizpour, H., Smith, K. & Hartman, J., *Digital Image Analysis in Breast Pathology-From Image Processing Techniques to Artificial Intelligence*, Translational Research: The Journal of Laboratory and Clinical Medicine, **194**, pp. 19-35, 2018.
- [26] Wang, J., Liu, Q., Xie, H., Yang, Z. & Zhou, H., *Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks*, Cancers, **13**, 661, 2021. DOI: 10.3390/cancers13040661.
- [27] Wang, P., Hu, X., Li, Y., Liu, Q. & Zhu, X., *Automatic Cell Nuclei Segmentation and Classification of Breast Cancer Histopathology Images*, Signal Process., **122**, pp. 1-13, 2016.
- [28] Li, X., Shen, X., Zhou, Y., Wang, X. & Li, T.Q., *Classification of Breast Cancer Histopathological Images Using Interleaved DenseNet with SENet (IDSNet)*, PLoS ONE **15**(5), e0232127, 2020. DOI: 10.1371/journal.pone.0232127.