JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CLASSIFICATION OF BREAST CANCER HISTOLOGY BY USING MATLAB

[1]Dr.P. Ratna Kamala, [2]U. Krupakar Reddy, [3]G. Jyothi, [4]N. Himesh, [5]D. Hemalatha, [6]M. Kamala

[1]HOD & Professor, Department of ECE,

Siddharth Institute of Engineering and Technology, Puttur, Andhra Pradesh, India.

[23456]B.tech Student, Department of ECE, Siddharth Institute of Engineering and Technology, Puttur, Andhra Pradesh, India..

Abstract: - **Among the most dangerous cancers that can affect women, breast cancer is one of the most common. For patients and their prognosis, the automatic diagnosis of breast cancer by studying histological images (HIS) is crucial. Clinical professionals can treat patients more effectively thanks to the classification of HIS, which gives them an accurate picture of disease. This paper's major goal is to use histology images to categorize the four different forms of breast cancer. Benign, in situ, Invasive, and Normal breast cancers make up the majority of cases. Using a convolutional neural network (CNN), it is possible to classify histology images accurately by the use of various layer kinds. Here, we're looking into a cutting-edge technique that can categorize breast cancer and translate the textual descriptions of the disease into auditory speech when picked up by a microphone.**

Index Terms – **Breast cancer, Convolutional Neural Network (CNN), Histological images, Breast cancer classification.**

1. INTRODUCTION

Breast tissue can grow into cancer in cases of breast cancer. A lump in the breast, a change in the form of the breast, dimpling of the skin, fluid coming from the nipple, an inverted nipple, or a red or scaly patch of skin are all indications of breast cancer. Affected individuals may experience bone discomfort, enlarged lymph nodes, shortness of breath, or yellow skin. Women are more likely to get breast cancer than males, and other risk factors include obesity, lack of exercise, alcohol use, hormone replacement therapy following menopause, ionizing radiation, early menstrual age, late or no childbearing, advanced age, a history of breast cancer, and family history. The BRCA1 and BRCA2 genes, among others, are inherited from a person's parents in about 5–10% of instances. The cells that line milk ducts and the lobules that supply those ducts with milk are where breast cancer most frequently manifests itself. Ductal carcinomas are cancers that originate from the ducts, whereas lobular carcinomas are cancers that originate from lobules. There are also more than 18 other subtypes of breast cancer. Pre-invasive lesions are the precursors to some malignancies, like ductal carcinoma in situ. By performing a biopsy on the suspicious lump, the diagnosis of breast cancer is verified. Following the diagnosis, additional tests are conducted to evaluate whether the cancer has spread outside of the breast and which treatments are most likely to be effective.

The type of breast cancer, the severity of the condition, and the patient's age all affect the outcome. The developed world has high survival rates, with 80–90% of people in England and the United States living for at least five years. There are lower survival rates in developing nations. At 25% of all occurrences, breast cancer is the most common type of cancer among women worldwide. There were 2 million new cases and 627,000 fatalities as a result of 2018. It is more prevalent in industrialized nations and affects women over 100 times more frequently than it does men.

Usually, a lump that feels different from the rest of the breast tissue is the first sign of breast cancer. When a woman feels a lump, more than 80% of breast cancer cases are identified. A mammography can identify breast cancer at its earliest stages. Breast cancer can also be detected by lumps in lymph nodes under the arms.

A nipple changing position or shape or becoming inverted, skin puckering or dimpling, a rash on or around a nipple, discharge from nipple/s, continuous pain in portion of the breast or armpit, and swelling behind the armpit or around the collarbone are all signs of breast cancer other than a lump. Although pain ("mastodynia") may be a sign of other breast health problems, it is an unreliable technique for detecting breast cancer.

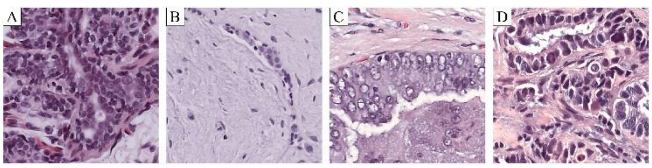


Figure 1 Examples of microscopy image patches

Paget's disease of the breast is another symptom combination associated with breast cancer. Skin changes resembling eczema, such as redness, discoloration, or slight flaking of the nipple skin, are how this syndrome manifests. Breast Paget's disease symptoms might include tingling, itching, increasing sensitivity, burning, and pain as it progresses. There might also be nipple discharge. The majority of women who are diagnosed with Paget's disease of the breast also have a breast lump.

Like other malignancies, breast cancer develops when an external (environmental) component interacts with a genetically predisposed host. Normal cells divide as often as required before stopping. They cling to other cells and maintain their position in tissues. When cells can no longer stop dividing, join with other cells, stay put, and pass away at the appropriate time, they are considered malignant.

When they are no longer required, normal cells will commit cell suicide (programmed cell death). They are guarded against cell suicide up until that point by a number of protein clusters and pathways. The PI3K/AKT pathway and the RAS/MEK/ERK pathway are two of the protective pathways. These defensive pathways' genes can occasionally have mutations that turn them permanently "active," preventing the cell from committing suicide when it is no longer required. This is one of the processes that, when combined with other mutations, results in cancer. Typically, when a cell is prepared for programmed cell death, the PTEN protein shuts off the PI3K/AKT pathway.

2. LITERATURE SURVEY

According to Siegel R. L., Miller K. D. [1], and Jemal A., breast cancer is one of the most lethal types of cancer that can affect women. Histology pictures are used to determine the presence of breast cancer. This study aims to categorize various forms of breast cancer using histology photos. Image processing techniques can be used to classify histology pictures successfully. Deep learning outperforms other image processing methods when used for image classification applications. Convolutional neural network topologies like Alex Net, Inception-Net, and ResNet, among others, are used for categorization. Capsule networks, which capture spatial and orientation information, were employed for the current study's architecture because traditional convolutional neural networks have numerous shortcomings.

In another study Ashikigmin M., Gooch B., Shirley P., Reinhard E [2]. Automating the diagnosis of any form of disease is crucial, and it's progressing quickly as more and more deep learning techniques are used to solve various issues. One of these computer-aided systems might be a decision-support tool that can correctly distinguish between several histological breast cancer picture kinds, such as normal tissue or carcinoma (benign, in situ or invasive). The authors of this research describe a convolutional capsule network-based deep learning approach for the categorization of four different breast tissue biopsy image types when hematoxylin and eosin staining is used.

In yet another paper [3], Fevens T., Krzyzak A., Monczak R., Filipczuk P stated that breast cancer must be discovered early for treatment to be effective. The cytological analysis of breast tissue taken straight from the tumour is a preliminary stage in the diagnosis. This study describes improvements in fine needle biopsies' ability to distinguish between benign and malignant tissue through the examination of cytological pictures. The circular Hough transform is used to estimate the cell nuclei rather than depending on precise segmentation of the nuclei. A support vector machine classifies detected circles as correct or incorrect based on texture features and the percentage of nuclei pixels according to a nuclei mask obtained using Otsu's thresholding method. The resulting circles are then filtered to keep only high-quality estimates for further analysis. Four separate classifiers each classify the biopsies using a set of 25 nuclear characteristics.

In alternative study [4], the authors El-Yaniv R., Issler E., Kimmel R., Meir R., and Peleg D. Brook A, On the basis of microscopic biopsy images, a fully automatic method for diagnosing breast cancer is given. The technique uses general feature vectors that are based on level-set statistics of the images to apply multi-class support vector machines to, which results in high recognition rates. Weal thus takes into account the issue of categorization with rejection and display early findings that suggest potential advantages.

In another study [5], the authors Ko J., Swetter S., Novoa R. A., Esteva A., and Kuprel B proposed the majority of skin cancer diagnoses are made visually, starting with an initial clinical screening and perhaps include dermoscopic study, a biopsy, and histological investigation. Skin cancer is the most prevalent human malignancy. The fine-grained variety in how skin lesions form makes it difficult to classify skin lesions automatically using photographs. Deep convolutional neural networks (CNNs) exhibit promises in a wide range of fine-grained object categorization for broad and highly variable applications. Here, we show how to classify skin lesions using a single CNN that was trained end-to-end from raw images utilizing only the inputs of pixels and disease labels. A dataset of 129,450 clinical images, which is two orders of magnitude greater than previous datasets and contains images of 2,032 different disorders, is used to train a CNN.

We evaluate its effectiveness in two crucial binary classification use cases—keratinocyte carcinomas against benign seborrheic keratoses and malignant melanomas vs benign nevi—against 21 board-certified dermatologists using biopsy-proven clinical pictures. The first case illustrates the most prevalent cancers, whereas the second case illustrates the deadliest skin cancer. The CNN achieves performance on par with all tested experts across both tests, indicating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists. Mobile devices with deep neural networks may enable dermatologists to treat patients outside of the clinic. By 2021, 6.3 billion smartphone subscriptions are anticipated, which could potentially offer low-cost universal access to essential diagnostic services.

According to Swapna G, Soman, K. P., and Vinayakumar R, diabetes, also known as diabetes mellitus, affects the great majority of people worldwide. Diabetes can only be managed; it cannot be cured. Diabetes is identified in this study by analyzing Heart Rate Variability (HRV) signals derived from ECG readings. To automatically detect the abnormality, we used deep learning networks that included CNN and CNN-LSTM (LSTM = Long Short-Term Memory) convolutional neural networks. Deep learning approaches do not require feature extraction, in contrast to the standard analysis methods that have been used up to this point. We divided the database into separate training and testing sets before doing classification.

Sowmya V., Deepika J., and Soman K. P urged that, one of the most serious types of cancer for women is breast cancer. Histology pictures are used to determine the presence of breast cancer. This study aims to categories various forms of breast cancer using histology photos. Image processing techniques can be used to classify histology pictures successfully. Deep learning outperforms other image processing methods in applications involving picture categorization. Several Convolutional neural network architectures, including Alex Net, Inception-Net, ResNet, etc., are utilized for classification purposes. Capsule networks, which capture spatial and orientation information, were employed for the current study's architecture because traditional convolutional neural networks have numerous shortcomings. The suggested work demonstrates that the pre-processing of the histology pictures increases the Capsule Network model's accuracy.

3. EXISTING SYSTEM

The statistical distribution of the observed combinations of intensities at specific points relative to one another in the image is used to compute the texture characteristics in statistical texture analysis. Statistics are divided into first-order, second-order, and higher-order statistics based on the quantity of intensity points (pixels) in each combination. Second order statistical texture features can be extracted using the Gray Level Concurrence Matrix (GLCM) approach. The method has been applied in a variety of contexts. Third and higher order textures take the relationships between three or more pixels into consideration. Although theoretically feasible, these are rarely used because of the lengthy calculating process and the complexity of the interpretation. The number of rows and columns in a GLCM matrix equals the number of grey levels, G, in an image. The frequency with which two pixels, separated by a pixel distance (x, y), one with intensity I and the other with intensity "j," appear in a particular neighborhood is represented by the matrix element P (i, j |Δx, Δy). The second order statistical probability values for changes between grey levels i and 'j' at a specific displacement distance d (Δx, Δy) and at a specific angle (ɵ) are contained in the matrix element P (i, j | d, ɵ). When using many intensities levels G, it is necessary to store a lot of temporary data, such as a G x G matrix for each combination of (Δx, Δy) or (d, ɵ). The amount of the texture samples used to estimate the GLCMs is crucial because of their high dimensionality. Consequently, there are frequently fewer grey levels. The example given for four different grey levels can be used to explain GLCM matrix formulation. The offset utilized here is one pixel (a reference pixel and its immediate neighbor). It is feasible to use a larger offset if the window is big enough. The top left cell will be filled with the number of times the combination 0,0 happens, or how often a neighboring pixel with grey level 0 falls to the right of another neighboring pixel with grey level 0 within the image area (reference pixel).

Sl.No	GLCM Feature	Formula
1.	*Contrast*	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
2.	*Correlation*	$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$
3.	*Dissimilarity*	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
4.	*Energy*	$\sum_{i,j=0}^{N-1} P_{i,j}^2$
5.	*Entropy*	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$
6.	*Homogeneity*	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
7.	*Mean*	$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})\ ,\ \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$
8.	*Variance*	$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2\ ,\ \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2$
9.	*Standard Deviation*	$\sigma_i = \sqrt{\sigma_i^2}\ ,\ \sigma_j = \sqrt{\sigma_j^2}$

Figure 2. Features of GLCM

4. PROPOSED SYSTEM

A. Convolutional Neural Network

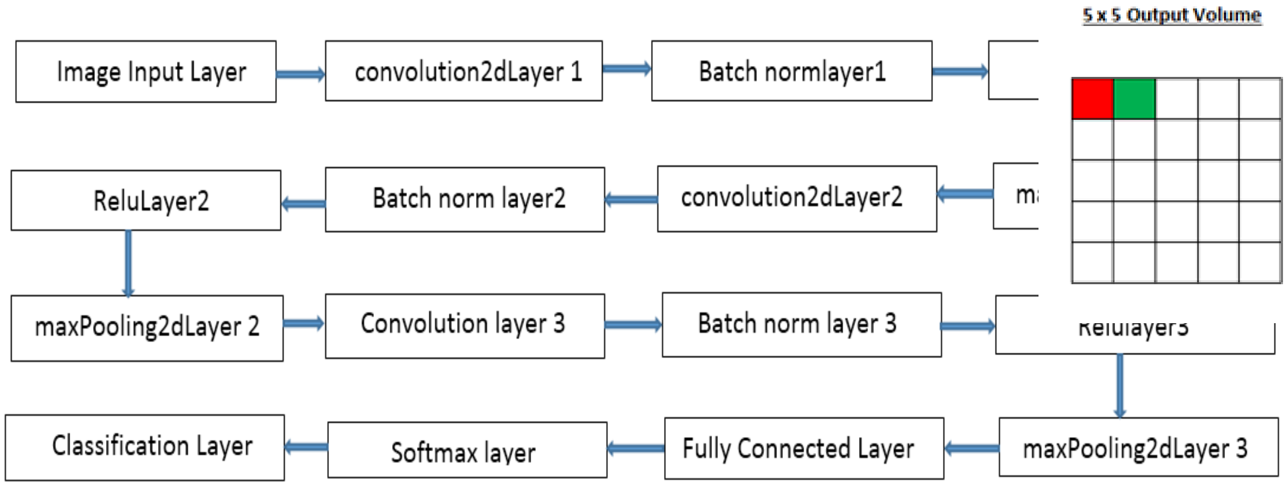
Convolutional Neural Networks (CNNs) have a basic design made up of a few stages that can repeatedly emerge in a network in an orderly fashion. Convolution layer, pooling layer, and fully linked layer are the three primary layers of a CNN. These layers are organized so that the input layer comes first and the output layer comes last. The subsections that follow provide a brief description of these strata.

1. Input Layer:

In every CNN, this is the bottom layer. Here, typically input are the pixel values obtained from any image. Raw pixel values are occasionally fed to the network, but pre-processing can also be used for better outcomes. For grayscale images, the input would have the following dimensions: (image height) x (image width) x (number of channels = 1). The input shape for RGB or true color images will be (image height) x (image width) x (Number of Channels = 3). Yet, CNN is also capable of processing 1D or 3D data volumes with little adjustment.

2. Convolutional Layer

The key component of D-CNN is this layer. This layer's foundation is the convolution operation. We need some filters or kernels that can extract specified structures from given image data in order to do convolution operations. Stride is a phrase used in D-CNN that describes the quantity of cells that need to be shifted. The filter is positioned in the top-right corner of the input image, and after performing an element-wise production between the filter's values and corresponding values of the image pixels, the produced results are added up to obtain a value. The obtained total is then positioned in one output cell. After that, the filter is moved to the right one stride and same calculations are repeated.



3. Pooling Layer

The input's height and width are decreased and its basic features are extracted by the pooling layer. Usually, this layer is applied following one or more convolution layers. This layer's computing performance is improved by lowering its dimension and its number of parameters.
reduces the likelihood of overfitting. There are numerous ways to pool data, including max pooling (where the maximum value is picked from a specific type of filter) and average pooling. The max pooling process is represented by (the average value from a certain shape of filter).

4. Fully Connected Layers

The output from the final layer is flattened (i.e., molded into a linear array) and provided as input to the nodes of a fully connected layer after the input data has been passed through some convolution and pooling layers. The outputs of all preceding layers are combined into a single linear array, and then a matrix multiplication with weights and addition with a bias value are performed.

5. Output Layer

The output layer comes after the completely linked layer. The final two layers of conventional neural networks can be compared to these two layers. However, the output is typically obtained using a SoftMax output device.

B. Convolution 2D-layer:

By far the most crucial component of a CNN, the convolutional 2D layer is where the majority of the processing actually takes place. Among many other things, it needs data input, a kernel filter, and a feature map. A feature detector examines the picture's receptive fields for something like the presence of the feature. It is also known as a kernel or filter. This method is in fact called convolution. A section of the image is represented by the two-dimensional (2-D) weighted array that serves as the feature detector.

When the neural network can truly be trained, three hyperparameters that decide the larger size must always be given. These are a few instances:
1. The number of filters used and the intensity of the outcome are related. For instance, three different maps from three different filters would result in a depth of 3.

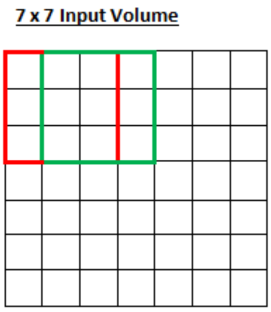


Figure 3 Stride

2. The kernel's stride may be the amount of input matrix pixels it traverses. A longer stride results in less output, despite the fact that two or more stride values are uncommon.
3. Zero-padding is frequently used when the filters can't accept the image as input. The outcome is higher or equal because all entries outside of the input matrix are set to zero.

C. Rectified Linear Unit Layer (ReLU):

An activation function that looks and behaves like a linear function but is actually a nonlinear function is necessary to train deep neural networks using stochastic gradient with back propagation of defects. The aim must be more
responsive to the active aggregate input in order to minimize overloading. This activation function has been created using a rectified linear activation unit, or ReLU. Rectified networks are a general term for neural networks with buried convolutional layers that employ the rectifier function.
The Rectified Linear Unit Layer will incorporate both the rectification layer and the non-linearity function. These combinations will function as a rectified linear unit layer in the space between convolutional neural network layers. Briefly stated, the ReLU layer is a rectified linear unit.

$$Y_i^{(l)} = \max\left(0, Y_i^{(l-1)}\right) \qquad \dots (1)$$

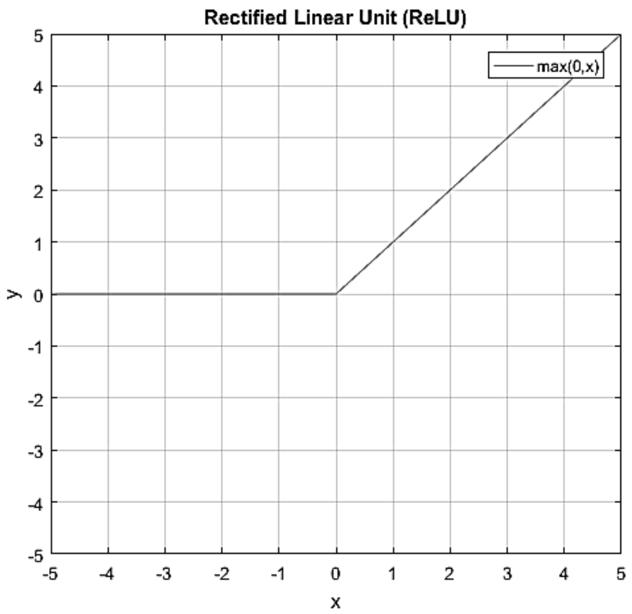


Figure 4 ReLU curve

Batch Normalization Layer:

A batch normalization layer separately normalizes a mini-batch of data for every observation made for every channel. Use batch normalization layers between convolutional layers and nonlinearities, such as ReLU layers, to quicken the convolutional neural network's training process and lessen its susceptibility to network initialization.

D. Maximum pooling Layer:

Down-sampling, also known as max pooling, is a technique for decreasing the number of input pieces to the absolute minimum. Similar to the convolutional 2d layer, the pooling technique applies a filter to the full input image, however this filter lacks weights. The values of the receptive field are instead added to the result array by the kernel using a summing function. The primary objective of the pooling layer should be to give feature lessening because feature detection is much more important for the precise position of such features, especially in picture identification projects. Max pooling outperforms average pooling in terms of results. We shall therefore carry out our research utilizing the highest pooling layer.

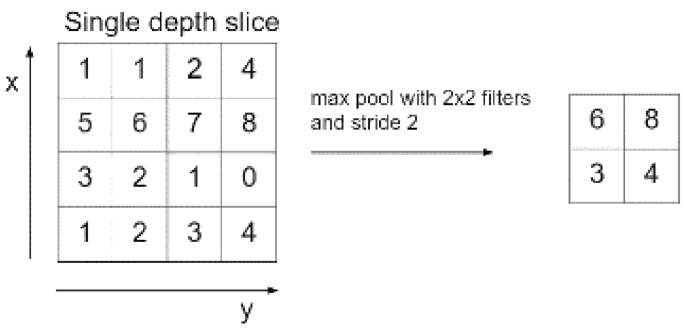


Figure 5 Maximum pooling

E. Output Layers:

SoftMax and Classification Layers

A SoftMax function is applied to the input using a SoftMax layer. With SoftMax Layer, produce a SoftMax layer. For classification and weighted classification problems with classes that are mutually exclusive, a classification layer calculates the cross-entropy loss. Use classification-Layer to create a classification layer.

The primary benefit of CNN over its forerunners is that it does it without the need for human intervention. The classification layer, or output layer, then provides the categorized output, which is shown as text that has been transformed to speech and is audible through a microphone.

5. EXPERIMENTAL RESULTS

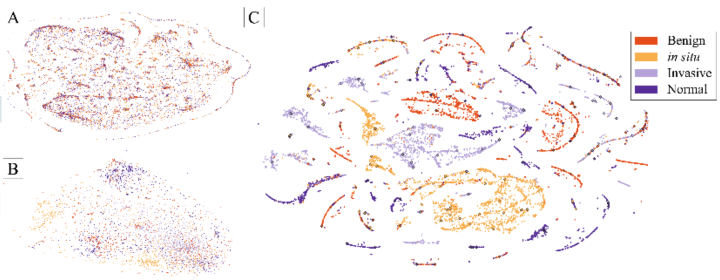


Figure 6 classification

6. CONCLUSION

The classification of breast cancer utilizing capsule net architecture is the foundation of the current work. It is evident from our work that data pre-processing and parameter adjustment can boost the performance of traditional designs. The findings indicate that this approach can be utilized as an automated tool to help physicians diagnose diseases, which may encourage them to focus more on early treatment rather than diagnosis and raise the survival rate for cancer patients.

REFERENCES

[1] Siegel R. L., Miller K. D., Jemal A., "Cancer statistics", CA Cancer JClin. 68:7- 30. W.-K. Chen, Linear Networks and Systems, Belmont, CA: Wadsworth, pp. 123–135, 2018.

[2] Sabour Sara, Frosst Nicholas, E. Hinton Geoffrey, "Dynamic Routin Between Capsules", 31st Conference on Neural Information Processing Systems NIPS, 2017.

[3] Iesmantas T., Alzbutas R., "Convolutional Capsule Network for Classification of Breast Cancer Histology Images", In: Campilho A.,Karray F., ter Haar Romeny B. (eds), Image Analysis and Recognition, Lecture Notes in Computer Science, vol 10882. Springer, Cham, ICIAR2018.

[4] Ferreira C. A. et al. Classification of Breast Cancer Histology Images through Transfer Learning Using a Pre-trained Inception Resnet V2. In: Campilho A., Karray F., ter Haar Romeny B. (eds), Lecture Notes in Computer Science, vol 10882. Springer, Cham. Image Analysis and Recognition, ICIAR 2018.

[5] Teresa Arajo, Guilherme Aresta, Eduardo Castro, Jos Rouco, PauloAguiar, Catarina Eloy, Antnio Polnia, Aurlio

Campilho, “Classification of breast cancer histology images using Convolutional Neural Networks”, vol. 12, 2017.
[6] Rakhlin A., Shvets A., Iglovikov V., Kalinin A. A., “Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis”, In: Campilho A., Karray F., ter Haar Romeny B. (eds), Lecture Notes in Computer Science, vol. 10882. Springer, Cham, Image Analysis and Recognition, ICIAR 2018.