



Computer-Aided Breast Cancer Classification Framework for Predictive, Preventive, and Personalized Medicine

Łukasz Jeleń and Michał Jeleń

Abbreviations

3P	Medicine 3P
BR	Bloom–Richardson
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network(s)
DPI	Dots per inch
F1	Performance measure of a classification model
FCM	Fuzzy c-means algorithm
FN	False negative
FNA	Fine needle aspiration biopsy
FP	False positive
HE	Hematoxylin and eosin
KNN	K-nearest neighbor classifier
NN	Neural network(s)
PPPM	Predictive, preventive, personalized medicine
ReLU	Rectified Linear function used in NN training
SVM	Support Vector Machines
TN	True negative
TP	True positive

Ł. Jeleń (✉)

Department of Computer Engineering, Wrocław University of Science and Technology,
Wrocław, Poland
e-mail: lukasz.jelen@pwr.edu.pl

M. Jeleń

Department of Immunopathology and Molecular Biology, Medical University of Wrocław,
Wrocław, Poland
e-mail: michal.jelen@umw.edu.pl

© The Author(s), under exclusive license to Springer Nature
Switzerland AG 2023

H. Podbielska, M. Kapalla (eds.), *Predictive, Preventive, and Personalised Medicine: From Bench to Bedside*, Advances in Predictive, Preventive and Personalised Medicine 17, https://doi.org/10.1007/978-3-031-34884-6_10

1 Introduction

World Health Organization reports that every year more than 11 million people are diagnosed with cancer [1]. In 2020 alone, ten million patients died from cancer, approximately 16.7% of deaths worldwide. These statistics place cancer as one of the leading causes of death. Additionally, breast cancer is the most frequently diagnosed cancer, with 2.26 million new cases in 2020 [2] and is one of the leading causes of death among middle-aged women.

As mentioned by Barrett et al., to achieve better patient benefits from treatment, a paradigm change is needed [3]. A key element in prevention and personalized treatment is the precise diagnosis and estimation of predictive factors [4]. Correct treatment significantly reduces the high number of deaths caused by breast cancer and increases the probability of success in treatment. This typically means high diagnosis and stage estimation accuracy. Precise and early diagnosis has a significant influence on the survival rate, indicating how many patients will live after treatment.

Cancer treatment and diagnosis are an active field among researchers in the fields of medicine and computer science. The aim of this research is to find the approach for a precise diagnosis. Here, a precise diagnosis means finding a cancer as early as possible by developing computer-aided diagnostic tools that use computer vision and neural network algorithms to estimate the predictive factors of the chosen predictive factors of breast cancer. The choice of artificial intelligence tools differs depending on the type of cancer and the type of examination used for diagnosis. This work focuses on cytological images of breast cancer that are produced during a fine needle aspiration biopsy examination. This kind of examination allows pathologists to estimate the malignancy of the cancer with very high accuracy. Malignancy estimation is very important when evaluating the survival rate of patients and the type of treatment [5]. Artificial intelligence was originally defined to describe the computer's ability to make human-like decisions. Today, with the development of other fields such as image processing, pattern recognition, or neural networks, its definition is much more complex. In this work, we describe a computer vision framework that incorporates all of these fields to build a computerized system for automated breast cancer diagnosis. This framework is capable of analyzing image information from fine needle aspirates and classifying breast cancer malignancy into three malignancy classes based on characteristics calculated according to the Bloom–Richardson grading scheme (see Sect. 2.1) [6]. Estimation of tumor grade is closely related to cancer prognosis and is considered part of breast cancer staging. During breast cancer staging, the pathomorphologist determines the extent of the cancerous tissue in the body. Both tumor grade and cancer staging, along with the mitotic activity index, are treated as prognostic factors for breast cancer. According to the literature, prognostic factors are crucial in increasing the survival rate. Furthermore, we can observe that tumors detected at an early stage and patients with identified slowly growing tumors show a much better

prognosis [4]. Cancers in their early stages are vulnerable to treatment, while cancers in their most advanced stages are usually almost impossible to treat. Therefore, early cancer detection greatly increases the probability of successful treatment. In addition to a precise diagnosis, it is necessary to foresee the course of the cancer, and being able to predict how the cancer can develop is very important for further treatment. Computer-assisted diagnosis systems based on visual interpretation of biopsy slides with neural networks and machine learning algorithms are an important step toward prevention and personalized treatment in breast cancer therapy.

2 Computer-Aided Breast Cancer Diagnosis

Computer-assisted breast cancer diagnosis has been a rapidly growing area of research for several decades. Computer-aided diagnosis (CAD) systems are used to assist in the detection and diagnosis of cancerous tissue changes. These systems are designed to operate independently of the physician. They can be used to supplement or assist in the interpretation of diagnostic images, such as mammograms, sonograms, magnetic resonance images, biopsies, or histological slides [7–9]. CAD systems are designed to identify potential lesions or other abnormalities in images that may require further evaluation or treatment. In the literature, a description of several types of CAD systems can be found that vary in the way they analyze images, the types of data they use, and the types of output they generate [10, 11]. These systems typically use image processing techniques to analyze images and extract features [8, 12], a convolutional neural network (CNN), support vector machines (SVM), or a combination of all to analyze the data.

Here, we examine the potential of CAD systems that are designed to be used as a tool to aid in determining prognostic factors in the diagnosis of breast cancer. We also discuss how applying this artificial intelligence-driven approach can help achieve PPPM therapy. Studies show that early detection and better screening have enabled earlier and better identification of breast cancer. The detection mode could be considered a prognostic factor and therefore taken into account in the management of patients, as it can affect their survival [4].

Studies have shown that CAD systems can improve the accuracy of lesion detection and diagnosis, as well as reduce the time it takes to interpret the visual image [10, 13, 14]. In addition, CAD systems can be used to reduce the number of false negative diagnoses, which can lead to a reduction in missed diagnoses and delayed treatment [15–17].

Despite the potential benefits of CAD systems, the technology is still relatively new and research is still ongoing [18–21]. There are several potential issues that may arise when using CAD systems, including false positives, false negatives, and overreliance on technology. Furthermore, more research is still needed to determine the most effective and reliable ways to use CAD systems in the prevention and personalized treatment of breast cancer.

2.1 Breast Cancer in the Context of 3P Medicine

A healthy woman's breast anatomy contains lobules connected to a nipple by ducts. These structures are supported by fat tissue. Breast cancer is abnormal cell growth that originates in the ducts and lobules.

Breast cancer is not only one of the most commonly diagnosed cancers, but also one of the most common cancers in middle-aged women. It is also one of the most deadly cancers. Screening examinations are critically necessary to reduce the high mortality rate. Regular screening can significantly reduce the death rate. Early detection and effective treatment can reduce mortality by up to 30%. A screening examination consists of mammography, ultrasound, and palpatic examination. The last can be performed at home by the patient herself or by a doctor. During a mammographic examination, doctors can detect very small lesions that cannot be distinguished during self-examination. During an ultrasound examination, we can detect the same lesions as in mammography without the risk of excessive radiation, making it safer for the patient. Unfortunately, it cannot be used for regular screening due to the fact that microcalcifications are not as clearly visible as in mammography, which can lead to misclassification of lesions [22]. Both of these methods are said to have about 25% false positive diagnoses. Furthermore, its interpretation can vary depending on the radiologist [23].

To establish a precise diagnosis, a biopsy examination is required. There are different types of biopsies, but for the purpose of this study, we will focus only on the fine needle aspiration biopsy (FNA). During this examination, a part of the abnormal tissue is collected, placed on the glass slide, and stained (see Fig. 1). When the specimen is stained, a microscopic examination is performed, during which the type of cancer is recognized, as well as its malignancy grade and prognostic and predictive factors [24]. These factors allow pathologists to foresee overall survival and disease-free survival rates, while predictive factors allow them to foresee reaction to the treatment undertaken [25].

Today, we are looking for a more holistic approach to breast cancer treatment that focuses on prevention, prediction, and personalized treatment. In the context of 3P medicine, a physician looks at the breast cancer patient as a whole and focuses on creating a customized treatment plan that meets the individual needs of the

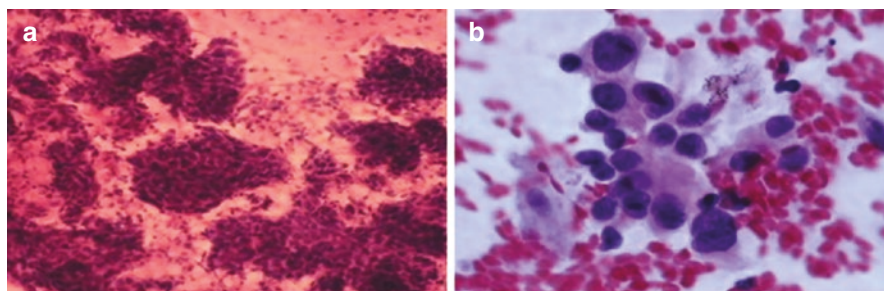


Fig. 1 Fine Needle Aspirate of a breast recorded with different magnifications. (a) 100×, (b) 400×

patient. Typically, this approach emphasizes lifestyle modifications to reduce the risk of cancer while also providing treatments to manage symptoms and improve quality of life. According to Koklesova et al., a large percentage of malignancies can be prevented with the unique properties of plant bioactive compounds [26].

When cancer is diagnosed, it is important to diagnose it quickly and accurately. As mentioned earlier, computer-aided screening and diagnosis can significantly speed up the procedure and reduce the number of false negative diagnoses, and therefore reduce missed diagnoses and delayed treatment. Here, we describe a computer-aided breast cancer classification framework that allows for the estimation of a Bloom–Richardson malignancy grade described in detail in Sect. 2.3).

Evaluation of the malignancy indicates the likelihood that the case may undergo metastasis at the time or after treatment. Personalized breast cancer treatment can include a variety of approaches depending on the type and stage of cancer and the individual’s health history and preferences [27]. Surgery, radiation therapy, chemotherapy, hormone therapy, targeted therapy, and immunotherapy are some of the options offered. Target therapy is a drug specifically designed to target cancer cells. The treatment approach is typically determined by a physician based on the medical history of the individual. This path allows for the determination of the most effective and personalized treatment plan. As cancer research advances, this type of treatment is becoming increasingly popular because it can help maximize patient success [28, 29]. In the case of breast cancer, as mentioned above, the determination of cancer malignancy influences the patient’s type of treatment and therefore it not only has a prognostic, but also a predictive value.

2.2 Machine Learning in Breast Cancer Diagnosis

Recent advances in imaging technology have made it possible to diagnose breast cancer more accurately. In particular, mammograms and ultrasound can be used to diagnose and identify the location, size, and shape of a tumor [9, 14]. Furthermore, magnetic resonance imaging and positron emission tomography scans can be used to identify the presence of hormone receptors in the tumor [30, 31]. Furthermore, to obtain a precise diagnosis of breast cancer, various machine learning techniques have been used to improve classification accuracy [11, 32]. In particular, supervised learning techniques, such as support vector machines (SVMs) and k-nearest neighbor (KNN), have been used to classify breast cancer tumors [12, 17, 31, 33]. These techniques have been shown to be effective in identifying the type of tumor, as well as its malignancy [8].

Here, we discuss a computerized breast cytology classification problem. It was first investigated by Wolberg et al. in 1990 [34]. The authors described an application of a multisurface pattern separation method to cancer diagnosis achieving an error rate as low as 4.1% on a fine needle aspiration biopsy database of 169 malignant cases and 201 benign. This data set was later publicly released as the Wisconsin Breast Cancer Database and is widely used by researchers to date [35]. This study was later expanded by Street et al. to describe ten nuclear features that were used for

the classification and prognosis of breast cancer [36]. In 2000, Street introduced a system called XCyt, the first Remote Cytological Diagnosis and Prognosis of Breast Cancer [37]. Other approaches include more recent work of Filipczuk et al., Jeleń et al. and Kowal et al. [8, 15, 38]. All of these approaches provided different ways for image segmentation. In Jeleń et al. we can find attempts to evaluate breast cancer malignancy, while in other works a discrimination between benign and malignant cases is described. Recently, with the introduction of Le Cun's convolutional neural networks [39], we can see more studies on breast cancer classification and interpretation of visual information. Deep learning algorithms have been used for the classification and identification of subtle features in mammograms, ultrasound images, and histological slides [14, 18, 19, 40]. From the above review, it is easy to see that the use of machine learning techniques has recently improved and all the authors report high accuracy in breast cancer classification. However, there is still room for improvement, particularly in the use of deep learning techniques to identify subtle features in the imaging data. There is also a great opportunity to use the potential of artificial intelligence in the design of personalized treatment.

2.3 Bloom–Richardson Grading Scheme

In Sect. 2.1 we mentioned that prognostic and predictive factors allow pathologists to foresee the course of cancer. In the case of breast cancer, the most important prognostic factors are histological grade and mitotic count. These factors are described by the Bloom–Richardson (BR) grading scheme grading system, which is the most common malignancy grading scale used by pathologists. The system was originally introduced by Bloom and Richardson in 1957 for grading histological slides [6]. In 1989, the originally proposed scheme was modified by Scarff and is now recognized as a modified Scarff–Bloom–Richardson scheme. For the diagnosis of breast cancer, this scheme is one of the best-known prognostic factors [41]. In our studies, we use this scale to assess malignancy for cytological smears. Based on this system, three factors are considered in grading cancerous tissue, each evaluated on a three-point scale according to the following description:

1. *Degree of structural differentiation*—describes the degree of tubule formation in histological slides. In cytology, tubules are not retained, the scoring is based on the determination of cell groupings (see Fig. 2—Grade 1). For this factor, we will grant one point if cells are grouped regularly, two points are given when grouped and single cells are visible, and three points are awarded when cells are irregularly spread.
2. *Pleomorphism (P)*—This factor takes into account the differences in the size, shape, and staining of the nuclei. This scoring is fairly straightforward, since as the nuclei irregularities increase, the prognosis worsens. Here, nuclei with uniform size, shape, and staining will receive one point, while those with moderate variations will receive two points. Three points are given to the nuclei with sig-

nificant variations. These deviations are depicted in Fig. 2—Grade 2. It can be seen that G2 ductal carcinoma has more uniform nuclei and less staining variations than G3 ductal carcinoma (Fig. 2—Grade 3).

3. *Frequency of Hyperchromatic and Mitotic Figures*—this factor is used to assess the number of mitoses one can find in the image. The more mitoses are present in the image, the worse the prognosis. If occasional mitotic figures are found, then one point is awarded. Two points are given for the slides with two or three figures, and for more than three figures, we will grant three points.

The final BR grade is assigned depending on the sum of the quantitative values of the above factors. According to Bloom and Richardson, the grade distribution is as described below [6]:

Points		
345	67	89
Grade I	Grade II	Grade III

where:

- Grade I—Low malignancy,
- Grade II—Intermediate malignancy,
- Grade III—High malignancy.

In the literature, other malignancy classification schemes can be found that are used for other types of cancer. The basis for all of them is similar to that of the Bloom–Richardson scheme. All assess cell pleomorphism, tubules, and mitosis. Variations usually consist of additional features that are taken into account. For histological slide grading, the most common variation of the Scraff–Bloom–Richardson scheme is the Elston and Ellis modification known as the Nottingham–Bloom–Richardson scheme. In this scoring system, the amount of gland formation or cell differentiation, nuclear features, and mitotic activity are assessed.

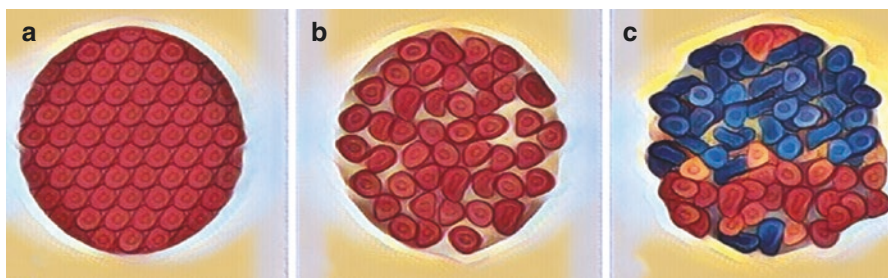


Fig. 2 Illustration of Malignancy Grades; (a) Grade 1, (b) Grade 2, (c) Grade 3

3 Material and Methods

In this section, a database of fine needle aspirates from breast cancer will be described, as well as the methodology of feature extraction. We will show classical features that can be calculated to resemble the Bloom–Richardson features. In addition, a set of features derived with convolutional neural networks is depicted. Based on these features, we show the classification methodology that allows for the estimation of predictive factors for breast cancer, namely the malignancy grade.

3.1 Database of Breast Cancer Slides

In our study, we have used a collection of images recorded during breast cytological examinations (see Fig. 3). Images were collected in the Department of Pathology and Clinical Cytology of the Medical University of Wrocław, Poland. The preparation of the slides includes staining with hematoxylin and eosin, known as the HE technique. The choice of staining agents allows visualization of nuclei with purple and black dyes and cytoplasm with shades of pink and red blood cells with orange and red dyes. The slides were digitized with an Olympus BX 50 microscope with CCD-IRIS camera mounted on the head of the microscope. Using MultiScan Base 08.98 software, we were able to record images with a resolution of 96 dots per inch (dpi) and a size of 764×572 pixels.

All images were recorded at two different magnifications of the same tissue region for each patient. On the low magnification images, we can see whether the cells are forming groups or are loosely spread in the image. These images were recorded with 100 \times magnification and comprise 50% of all images in the database. The low magnification images are used for the estimation of features based on the cells' tendency to form groups. Healthy and low malignant cases tend to form one or two large groups in the image, while those cases with large malignancy are loosely spread and the groups usually consist of only a few cells.

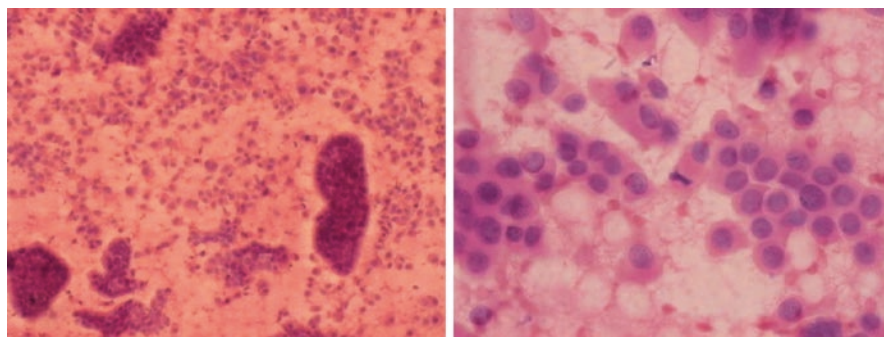


Fig. 3 Breast Cancer Fine needle aspirates. Sample Database Images: left—Low magnification, right—high magnification

The second subset of images was recorded with 400× magnification. This type of image allowed for the determination of features describing the pleomorphic cells' features. These shape-based features provide important information about cell nuclei. Here, low malignancy cases have uniform size and staining, while in more malignant cases this tendency is disturbed, and the nuclei in the image will assume nonuniform sizes and will have stronger staining variations. Currently, the database consists of 480 FNA biopsy images (both 100× and 400×) that were graded by an expert pathologist and will later be treated as the “gold standard” for malignancy classification. All images represent the three classes of cancer malignancy, namely low (G1), intermediate (G2), and high (G3) malignancy grades. There are 22 cases of low malignancy, 134 of intermediate and 84 of high malignancy. For all cases, a follow-up examination was performed. When a tissue was surgically removed, a histopathological examination was graded using the Bloom–Richardson [6] grading scale that confirmed the classification of FNA. Therefore, all the cases in our database were histopathologically confirmed.

3.2 Breast Cancer Malignancy Classification Framework

The framework described in this study uses the bright-field light microscope with an additional mirror mounted behind the objective to split the image into two visible images. The first image is visible through the eyepiece, and the second image is projected to the camera and then recorded by the camera. The data obtained in this way is then processed by machine learning algorithms, and one of the three BR grades is assigned. In Fig. 4, we show a machine learning pipeline to estimate the grade of malignancy. This scheme can be divided into several stages and will depend on the approach taken. For the classical approach, where SVM and KNN classification is performed, we need to introduce image preprocessing and segmentation (see Sect. 3.3). When deep learning is applied, we use convolutional neural networks to construct the feature vector; see Sect. 3.4 for more details.

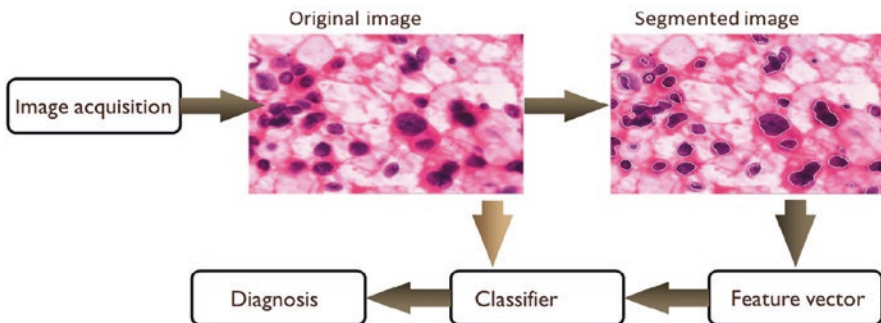


Fig. 4 Machine learning framework for malignancy classification

3.3 Preprocessing and Nuclei Segmentation

Preprocessing is a stage in which the image recorded during the acquisition step is usually modified in a way to remove unwanted noise that could have been introduced by the imaging setup. Typically, these operations would include filtering to remove noise introduced during analog-to-digital conversion. In some applications, we also try to deconvolve the image with a so-called point spread function, which should remove the aberrations of the optics of the acquisition setup. In our study, we applied a median filter that allows for the removal of random noise. When this filter is applied, the value of the pixels is replaced by a median value of the 3×3 neighborhood of the pixel. The main advantage of median filters is their ability to preserve edges. This allowed us to prepare the images for image segmentation by smoothing the homogeneous regions, which in fact reduces the number of colors representing these areas. For this purpose we applied a well-known segmentation technique that uses a fuzzy version of a k-means clustering, called a Fuzzy c-means segmentation.

Fuzzy c-means is an approach proposed by Klir and Yuan [42] that can be used to divide image information and extract nuclei. Generally, a set of data of $X = x_1, x_2, \dots, x_n$ is divided into c clusters. It is assumed that $P = A_1, A_2, \dots, A_c$ is a known pseudopartition and A_i is a vector that fits all x_k members to the i cluster. Now, using the Eq. (1), one can calculate the center of the c cluster [43].

$$v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m}, \quad i = 1, 2, \dots, c \quad (1)$$

where $m > 1$ is a weight that controls the fuzzy membership. Memberships are defined by Eq. 2 if $\|x_k - v_i\|^2 > 0$ for all $i \in \{1, 2, \dots, c\}$ and if $\|x_k - v_i\|^2 = 0$ for some $i \in I \subseteq \{1, 2, \dots, c\}$ the memberships are defined as a non-negative real number that satisfies the Eq. (3) for $i \in I$.

$$A_{-i}(x_{-k}) = \left[\sum_{j=i}^c \left(\frac{x_k - v_j^2}{x_k - v_j^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \quad (2)$$

$$\sum_{i \in I} A_i(x_k) = 1. \quad (3)$$

The clustering algorithm looks for a set P that minimizes the performance index $J_m(P)$ defined by Eq. (4).

$$J_m(P) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m x_k - v_i^2. \quad (4)$$

In contrast to all other segmentation techniques, the fuzzy c-means algorithm (FCM) requires no additional processing and as such was applied to segment the color information in the image. In Fig. 5 an example of FCM segmentation is presented.

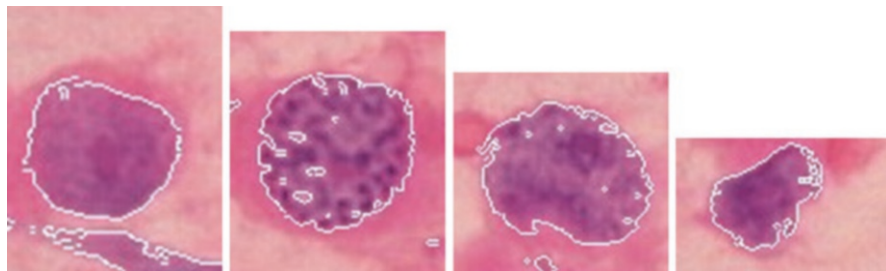


Fig. 5 Nuclei segmentation with fuzzy c-means algorithm

3.4 Feature Extraction

Feature extraction in classification is the process of selecting and extracting meaningful features from a data set to identify patterns, classify data, and make predictions. It is a crucial step in machine learning and data mining. Feature extraction involves selecting a subset of the most important, meaningful, and useful features from a larger set of raw data, which can then be used to train a machine learning model. The goal is to reduce the dimensionality of the data set while preserving important information and patterns. This helps to reduce the complexity of the model and improve its accuracy. Feature extraction can be performed manually, using domain knowledge and prior experience, or automated, using algorithms.

For the classification of the breast cancer malignancy data, two types of feature vectors were created. One used classical image processing methods and the second created with a convolutional neural network.

3.4.1 Classical Feature Extraction

The calculation of these features depended on the magnification of the image. For 100× magnification, we calculated the average area of the visible groups in the image, the number of groups, and the dispersion defined with Eq. (5).

$$\frac{1}{D} = \frac{1}{n-1} \sum_{i=1}^n (A_c - A_{100})^2, \quad (5)$$

For the 400× magnification images, we calculated features that resembled nuclear features. All features were described in detail by Jeleń et al. [8] and include:

- Area that was calculated as the sum of all nuclei pixels of the nucleus [37].
- Perimeter defined as the length of the nuclear boundary of a nucleus that is approximated by a length of the polygonal approximation of the boundary.
- Convexity determined as the ratio of the nucleus area and the area of the minimal convex polygon that contains the nucleus, called a convex hull.
- Eccentricity described the circularity of the nucleus taking into account the fact that healthy nuclei will assume circular shapes, while cancerous nuclei can have arbitrary shapes. Eccentricity is calculated as a ratio of the distance between

focal points of an ellipse matched with a nucleus having the same second moments as the segmented nuclei, and its major axis length. It will assume values between 0 and 1, where 0 corresponds to a circle and $\underline{1}$ to a line segment.

- **Centroid**—For each nucleus, the centroid is a point (x_i, y_i) called a center of mass of the extracted nucleus along each row (X) and column (Y). It is calculated as follows:

$$\bar{x}_i = \frac{1}{A_i} \sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} j N_{i(j,k)}; \quad (6)$$

$$\bar{y}_i = \frac{1}{A_i} \sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} k N_{i(j,k)}; \quad (7)$$

where $N_i(j, k)$ equals 1 if the pixel j, k is in the nucleus N_i , and 0 otherwise.

- **Orientation** is also called an axis of the least second moment and contains information on the orientation of the nucleus. For the coordinate system placed at the centroid (x_i, y_i) of the nucleus, we can define the orientation as follows:

$$Or_i = \tan(2\theta_i), \quad (8)$$

where the angle θ_i is measured counterclockwise from the x -axis.

- **Projections** are calculated along rows and columns.
- **Moment-based features**—here we use seven normalized central moments to calculate the rotation, scaling, and translation invariant features.
- **Histogram-based features** are statistical features calculated based on the image histogram and included five features: mean, standard deviation, skew, energy, and width.
- **Textural features**—These features require a determination of the gray-level co-occurrence matrix that describes the relationships between a pair of pixels and their gray levels. Assuming that the distance between the pixels and the directions are given, we can extract energy, homogeneity, inertia, and correlation features.
- **Color-based features**—for these features, we treated each color component as a separate intensity image. To calculate color features, we can use the same logic as for the Textural features and applied them to each color band.

3.4.2 Convolutional Neural Networks

Convolution neural networks (CNNs) are a type of deep neural networks that are the most commonly used computer vision applications. They became a popular artificial intelligence tool because they are capable of automatic extraction of valuable information from real-world images or video streams. Unlike fully connected layers in classical neural networks, the CNN model extracts simple features from input through a single or multiple convolution layer and executes convolution operations.

Each layer is a set of nonlinear functions that combine weights in different coordinates and allow the weight to be reused from the spatial subsets of the previous layer output.

The first CNN model described by Yann LeCun in 1989 was a neural network that processed data with a known grid-like topology using at least one convolution operation instead of a matrix multiplication [39]. The convolution of two real functions can be determined with Eq. (9).

$$F_v(t) = (x - \omega)(t) \quad (9)$$

where x is an input, ω is the convolution kernel, and F_v is a feature vector. Since the images are two dimensional, the convolution kernel (Ω) should also be two dimensional. This leads to a definition of convolution with Eq. (10).

$$F_M(i,j) = (\Omega \odot \text{Img})(I,j) \sum_m \sum_n \text{Img}(j+m, i+n) \Omega(m,n). \quad (10)$$

To build a network capable of performing feature extraction, we build a network consisting of several convolution layers, where each layer performs several parallel convolutions that yield a set of linear activation values. Each value passes through the rectified linear function called ReLU that maintains the non-linearity of the resulting feature vector. Finally, we modify the output of the layer to reduce the output dimensionality. This is performed using a maximum output pool function (Eq. 11) and is called max-pooling.

$$O(x) = \max(O,x) \quad (11)$$

In this study, we used one of the most popular CNN architectures, called VGG-16, which was introduced by Simonyan and Zisserman in 2014 [44]. This is a CNN model that is built from 16 convolutional layers and was pretrained on the ImageNet dataset. This data set consists of more than 14 million images categorized into almost 1000 classes.

Based on the output of the VGG-16 network, we constructed a feature vector that was then used in the classification of the malignancy.

3.5 Classification Scheme

Pattern classification deals with the issue of assigning a specific class to the given pattern. There are many methods that can be used to classify data. Here, we apply the K-nearest neighbor rule, support vector machines, random forests, and neural networks to classify the data sets prepared according to the description in Sect. 3.4.

3.5.1 K-Nearest Neighbor

K-nearest neighbor (KNN) is one of the simplest classification algorithms. It is based on the distance calculation between the pattern in question and its k neighbors. The decision is made based on the closest association between the pattern and

the neighbors. The pattern is classified into the closest class in terms of distance between its k neighbors.

The training procedure is very simple and is based on recording the entire training set. Testing usually uses a Euclidean distance to calculate the distances between the training samples and the tested sample. The class assigned to the sample is the one for which the distance is the smallest. To be able to calculate the Euclidean norm, it is usually necessary to normalize the data to avoid any data inconsistency. To classify our data, a KNN was calculated for five neighbors.

3.5.2 Support Vector Machines

Support vector machines are used to separate two or more classes of patterns or data points by constructing a boundary between them. An unknown point will be classified according to its orientation with respect to that boundary. To estimate the boundary between classes, we use boundary points from each class. These points are called support vectors. This procedure is an iterative approach that minimizes some error function (Eq. 12),

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \varepsilon_i \quad (12)$$

with the following restrictions:

$$y_i (w^T \phi(x_i) + b) \geq 1 - \varepsilon_i \quad \text{and} \quad \varepsilon_i \geq 0, i = 1, \dots, N \quad (13)$$

where C and b are constants, w is the weight vector, ε_i is a bias value that deals with overlapping cases, and ϕ is a kernel function that transforms the input data into the feature space. The constant C has a major influence on the error rate and has to be carefully estimated during the training process.

Depending on the error function, we can distinguish between different SVMs and different kernels. Here, we make use of the radial base function (RBF) kernel (Eq. 14).

$$\phi = \exp(-\gamma x_i - x_j^2) \quad (14)$$

The learning process uses the Adatron algorithm [45] which guarantees the convergence to the solution assuming that the solution exists.

3.5.3 Random Forests

Random forests, as their name suggests, are made up of a large number of individual decision trees that operate as a whole. Each tree in the random forest produces a class prediction and the class with the highest votes becomes our model prediction. Based on the training data, the trees are created from bootstrap data samples. During training, a random subset of attributes is drawn from which the best decision tree split is selected. The final decision is based on the majority of trees that have developed in the forest. In our case, ten trees were chosen.

3.5.4 Neural Networks

The idea of neural networks is based on the real interactions of human nerve system.

The basic element of the neural network is the neuron, sometimes also called a perceptron. It is a mathematical model of a biological neuron. Combining a few neurons together in such a way that neurons can interact with each other makes a neural network that is capable of processing input data and providing us with a certain decision.

In neural networks, each neuron accepts an input signal of the form $X = [x_1, x_2, \dots, x_n]$ and each of the subsignals is assigned a weight. $F(s_i)$ is called an activation function of the neuron and, depending on the type of neuron, activates its output. In our case, the activation function used is the rectifier linear unit function—ReLU. The network architecture is based on 200 hidden layers and one input layer and one output layer. Before we can use our neural network, it is necessary to train it so that it can recognize the desired patterns. Training is based on weight adjustment depending on the output value. Our network uses the Adam optimization algorithm that iteratively updates the weights of the network [46]. Training is performed on known patterns for which the output is known. Such a set of known patterns is called a training set. Analogously, a set of unknown patterns is called a testing set.

Training of the network was carried out in a maximum of 320 iterations and for validation we used a leave-one-out technique. This technique is often used for small data sets and involves leaving one data sample for validation and the remaining samples are used for training. The procedure is repeated for all the samples in the data.

3.6 Classification Performance

To be able to say how the proposed classifiers behave and to be able to evaluate their performance, we introduce quantitative criteria. The most popular and reliable evaluation method is based on the confusion matrix that contains information on actual and predicted classifications. The fields of the matrix are filled depending on the classification result of the tested samples. Based on these responses, we can determine the number of positive classifications correctly classified as positive, called true positives (TP), the number of negative classifications correctly classified as negative, called true negatives (TN), the number of negative classifications incorrectly classified as positive, called false positives (FP), and the number of positive classifications incorrectly classified as negative, called false negatives (FN). Based on these values, we can now define additional measures such as F1, precision, and recall. Precision is the model's ability to avoid identifying irrelevant items as relevant and is calculated as a ratio of TP and the total number of predicted positives. Recall, on the other hand, measures the model's ability to identify all relevant items and is calculated as a ratio of TP to positives. F1 measure evaluates the overall performance of a classification model. It is calculated as a ratio of the sum of precision and recall and their doubled multiplication. The F1 score ranges from 0 to 1, where 1 is a perfect performance.

4 Results and Discussion

The objective of this study was to investigate the achievability of the methods described in Sect. 3.5 to computerized breast cancer malignancy and therefore for the problem of automated predictive factor estimation. The results described in this section were obtained for experimental investigations that included a comparative study of classification performance in the feature vector extracted with conventional image processing techniques called FVC23 and the feature vector created with the convolutional neural network called FVDL23 (as described in Sect. 3.4). During these tests, we made decisions only between the G2 and G3 classes, as this was the setup commonly described in the literature. Additionally, we checked how the introduction of the G1 class will influence the accuracy of the classification. These tests were performed only for the feature vector (FVDL123) extracted with the convolutional neural networks.

In Table 1 classification results for the two-class classification problem are gathered in the form of area under curve, classification accuracy, F1, precision, and recall measures. The corresponding confusion matrix is presented in Table 2. From the classification results, we can see that neural networks achieved the best classification accuracy (84.6%), which is confirmed by the high F1 measure. It can also be noticed that the Precision and Recall measures are also highest for decisions made with a neural network. Of all, kNN was the classifier with the worst performance, for which the lowest accuracy of 59.4% was observed for the feature vector created with classical image processing methods. It should be noted that accuracy and quality measures increased significantly when convolutional neural networks were used as feature extractor. These rates show that convolutional neural networks allow for the determination of more meaningful features. The accuracy obtained in this study could possibly be much higher if the fine aspirate database contained more images and the balance between classes were preserved.

The second part of the experiments was to perform a classification of breast cancer malignancies in all three malignancy classes to resemble the Bloom–Richardson grading. The results of these analyzes are presented in Table 3 as a three-class problem, and the corresponding confusion matrix is presented in Table 4. For comparison purposes, we also put the results of a two-class problem discussed earlier. Here, we can see again that the neural network is the best performing classifier achieving 78.7% accuracy with a similar 78.6% for F1 and precision measures. This means

Table 1 Classification results for the two-class problem

Model	FVC23					FVDL23				
	AUC	Acc	F1	Precision	Recall	AUC	Acc	F1	Precision	Recall
kNN	0.651	0.594	0.594	0.595	0.594	0.866	0.782	0.779	0.779	0.782
SVM	0.785	0.752	0.752	0.751	0.752	0.884	0.796	0.797	0.799	0.795
Rand. Forest	0.847	0.792	0.790	0.792	0.792	0.829	0.756	0.748	0.754	0.756
NN	0.821	0.792	0.791	0.791	0.792	0.919	0.846	0.846	0.846	0.846

Table 2 Confusion matrices for the two-class problem

Actual	Predicted			
	FVC23		FVD23	
	G2	G3	G2	G3
G2	81.0%	23.3%	87.9%	20.5%
G3	19.0%	76.7%	12.1%	79.5%

Table 3 Classification results for the three-class vs. two-class problem

Model	3-Class					2-Class				
	AUC	Acc	F1	Precision	Recall	AUC	Acc	F1	Precision	Recall
kNN	0.833	0.717	0.706	0.709	0.717	0.866	0.782	0.779	0.779	0.782
SVM	0.866	0.723	0.688	0.656	0.723	0.884	0.796	0.797	0.799	0.795
Rand. Forest	0.837	0.721	0.705	0.711	0.720	0.829	0.756	0.748	0.754	0.756
NN	0.897	0.787	0.786	0.786	0.787	0.919	0.846	0.846	0.846	0.846

Table 4 Confusion matrix for the three-class problem

Actual	Predicted		
	G1	G2	G3
G1	81.0%	18.2%	40.5%
G2	11.7%	81.2%	4.8%
G3	7.3%	0.6%	54.8%

that the introduction of an additional class did not significantly alter the results. As expected, the accuracy of the classification dropped, but for the size of the database, we can still treat it as a very good result. These additional experiments confirm that convolutional neural networks are a good choice for feature extraction. They provide accurate features that lead to good classification performance. The disadvantage of the method is the time it takes to perform the calculations. For our setup, it took roughly 1.5 times longer to evaluate one image. CNNs also require databases at least ten times larger than those used in our study, and therefore the training process tends to be time consuming. The results of this study suggest that the use of neural networks increases the precision of the classification of malignancy grade and therefore a better estimation of prognostic factors. Furthermore, the study showed that classical feature extraction methods based solely on image processing methods, such as filtering, segmentation, and geometric moments, resulted in less accurate classification.

5 Conclusions

In this chapter, a computerized framework for the estimation of breast cancer predictive factors was described. As mentioned in Sect. 2.1 Bloom–Richardson grade is one of the best prognostic and predictive factors for breast cancer [24, 25]. In this

context, our research on computer-aided malignancy grading shows that machine learning algorithms have a very large potential to estimate these factors. The results presented in Sect. 4 clearly show that it is possible to create a computer vision algorithm that is capable of classifying fine needle aspiration biopsy slides and assigning one of the three malignancy grades with high accuracy and precision. Additionally, we can also conclude that.

- Computer-aided diagnosis (CAD) is an important tool for the analysis of breast imaging data. They can not only detect breast cancer by analyzing digital cytological slides, but also provide prognostic and predictive factors. It could also be used to help assess personalized treatment plans. This needs further evaluation and research. It can be pointed out as an open problem.
- Artificial intelligence can be used to analyze fine needle aspiration biopsy slides, which can then be used to detect and classify breast cancer malignancy grade.
- Deep learning (DL) algorithms have been used to extract features from biopsy images and estimate predictive factors for breast cancer. These algorithms can be used to analyze this type of data with high precision.
- Image processing techniques can also be used to analyze biopsy images to classify breast cancer. These algorithms did not show as good accuracy as the DL algorithms, but can still be treated as a good alternative if the database is small.
- The fine needle aspirate database was collected and prepared for malignancy grade classification and is maintained continuously. To the best of our knowledge, no other similar database is publicly available. To obtain better classification rates, the database needs to be enlarged, which can also be pointed out as an open problem.
- 3P medicine can benefit from artificial intelligence that provides machine learning tools for personalized predictive and preventive treatment of breast cancer.
- CAD PPM system can be created based on machine learning algorithms that are not only capable of estimating prognostic factors but also of analyzing medical history and other personalized profiles to create an individualized risk profile.

In breast cancer, 3P medicine can be used to identify and treat patients at risk of developing breast cancer, predict the likelihood of developing certain types of breast cancer, and prevent recurrence. In this study, we focus on predictive and prognostic factors for breast cancer. Other examples of 3P approaches to breast cancer would include genetic testing to identify high-risk patients, mammogram screening to detect early stage tumors, and personalized treatments such as targeted drugs and immunotherapies. Furthermore, lifestyle modifications, such as exercise, diet, and stress management, can be incorporated into 3P approaches to reduce the risk of developing breast cancer or reduce the risk of recurrence. All these areas can be addressed for artificial intelligence-driven systems. AI can be used to detect and diagnose the disease in its early stages, allowing for the estimation of fast personalized therapy. In addition, it could also be used to analyze biopsy results to determine the best treatment options for each patient. Machine learning algorithms could also be used to analyze medical history, genetic profile, lifestyle, and environmental

factors to create an individualized risk profile. This profile can be used to identify individuals who are at an increased risk of developing breast cancer and to recommend preventive measures to reduce their risk. In general, the results of this study demonstrate that the computerized malignancy grading framework will allow repeatability in the decision-making process, which is of great concern among the pathological community. The computerized scheme described in this investigation complies with this requirement, and the classification results obtained are very good. In our opinion, such a proposed solution could significantly affect physicians' day-to-day work and help in the estimation of a personalized therapy and prediction of the disease depending on the treatment undertaken. Preventing breast cancer and estimating individual risk factors could also be possible when additional patient information would be available to train machine learning algorithms. In such a case, we would be able to build a CAD PPM system that can fulfill the requirements of the 3P paradigm change of the future.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al (2021) Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249
2. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Pineros M, Znaor A et al (2021) Cancer statistics for the year 2020: an overview. *Int J Cancer* 149(4):778–789
3. Barrett M, Boyne J, Brandts J, Rocca HPB-L, Maesschalck LD, Wit KD et al (2019) Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care. *EPMA J* 10:445–464. <https://doi.org/10.1007/S13167-019-00188-9>
4. Soerjomataram I, Louwman MWJ, Ribot JG, Roukema JA, Coebergh JWW (2008) An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat* 107:309–330
5. Jeleń Ł (2009) Computerized cancer malignancy grading of fine needle aspirates. PhD thesis, Concordia University
6. Bloom HJ, Richardson WW (1957) Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* 11:359. <https://doi.org/10.1038/bjc.1957.43>
7. Eltrass AS, Salama MS (2020) Fully automated scheme for computer-aided detection and breast cancer diagnosis using digitised mammograms. *IET Image Process* 14:495–505. <https://doi.org/10.1049/IET-IPR.2018.5953>
8. Jeleń Ł, Krzyżak A, Fevens T, Jeleń M (2016) Influence of feature set reduction on breast cancer malignancy classification of fine needle aspiration biopsies. *Comput Biol Med* 79:80
9. Drukker K, Pesce L, Giger M (2010) Repeatability in computer-aided diagnosis: application to breast cancer diagnosis on sonography. *Med Phys* 37:2659–2669. <https://doi.org/10.1118/1.3427409>
10. Yassin NIR, Omran S, Houbay EMFE, Allam H (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput Methods Programs Biomed* 156:25–45. <https://doi.org/10.1016/J.CMPB.2017.12.012>
11. ElOuassif B, Idri A, Hosni M, Abran A (2021) Classification techniques in breast cancer diagnosis: a systematic literature review. *Comput Methods Biomech Biomed Engin* 9:50–77. <https://doi.org/10.1080/21681163.2020.1811159>

12. Jeleń Ł (2020) Texture description for classification of fine needle aspirates. In: Korbicz J, Maniewski R, Patan K, Kowal M (eds) *Current trends in biomedical engineering and bioimages analysis*. Springer, Cham, pp 107–116
13. Liu Q, Liu Z, Yong S, Jia K, Razmjoooy N (2020) Computer-aided breast cancer diagnosis based on image segmentation and interval analysis. *Automatika* 61:496–506. <https://doi.org/10.1080/00051144.2020.1785784>
14. Rybiatek A, Jeleń Ł (2020) Application of densenets for classification of breast cancer mammograms. In: Saeed K, Dvorsky J (eds) *Computer information systems and industrial management*. Springer, Cham, pp 266–277
15. Filipczuk P, Fevens T, Krzyżak A, Monczak R (2013) Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE Trans Med Imaging* 32:2169–2178. <https://doi.org/10.1109/TMI.2013.2275151>
16. Krawczyk B, Jeleń Ł, Krzyżak A, Fevens T (2012) Oversampling methods for classification of imbalanced breast cancer malignancy data. In: *Computer vision and graphics*, vol 7594 LNCS, pp 483–490. <https://doi.org/10.1007/978-3-642-33564-8-58>
17. Krawczyk B, Jeleń Ł, Krzyżak A, Fevens T (2014) One-class classification decomposition for imbalanced classification of breast cancer malignancy data. In: *Artificial intelligence and soft computing*, vol 8467LNAI, pp 539–550. <https://doi.org/10.1007/978-3-319-07173-2-46>
18. Hameed Z, Garcia-Zapirain B, Aguirre JJ, Isaza-Ruget MA (2022) Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. *Sci Rep* 12:1–21. <https://doi.org/10.1038/s41598-022-19278-2>
19. Eroglu Y, Yildirim M, Cinar A (2021) Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using MR-MR. *Comput Biol Med* 133:104407. <https://doi.org/10.1016/j.compbiomed.2021.104407>
20. Ranjbarzadeh R, Dorosti S, Ghoushchi SJ, Caputo A, Tirkolae EB, Ali SS et al (2023) Breast tumor localization and segmentation using machine learning techniques: overview of datasets, findings, and methods. *Comput Biol Med* 152:106443. <https://doi.org/10.1016/J.COMPBIOMED.2022.106443>
21. Raaj RS (2023) Breast cancer detection and diagnosis using hybrid deep learning architecture. *Biomed Signal Process Control* 82:104558. <https://doi.org/10.1016/J.BSPC.2022.104558>
22. Kolb TM, Lichy J, Newhouse JH (1998) Occult cancer in women with dense breasts: detection with screening US—diagnostic yield and tumor characteristics. *Radiology* 207:191–199
23. Beam CA, Layder PM, Sullivan DC (1996) Variability in the interpretation of screening mammograms by US radiologists. *Arch Intern Med* 156:209–213
24. Agrawal AK, Jeleń M, Grzebieniak Z, Zukrowski P, Rudnicki J, Nienartowicz E (2008) Androgen receptors as a prognostic and predictive factor in breast cancer. *Folia Histochem Cytobiol* 46(3):269–276
25. Stankov A, Bargallo-Rocha JE, Silvio N-S, Ramirez MT, StankovaNinova K, Meneses-Garcia A (2012) Clinical study prognostic factors and recurrence in breast cancer: experience at the National Cancer Institute of Mexico. *ISRN Oncol* 2012:825258. <https://doi.org/10.5402/2012/825258>
26. Koklesova L, Liskova A, Samec M, Qaradakhi T, Zulli A, Smejkal K et al (2020) Genoprotective activities of plant natural substances in cancer and chemopreventive strategies in the context of 3p medicine. *EPMA J* 11:261–287. <https://doi.org/10.1007/S13167-020-00210-5>
27. Harbeck N, Salem M, Nitz U, Gluz O, Liedtke C (2010) Personalized treatment of early-stage breast cancer: present concepts and future directions. *Cancer Treat Rev* 36:584–594. <https://doi.org/10.1016/j.ctrv.2010.04.007>
28. Marcom PK, Anders CK, Ginsburg GS, Potti A, Nevins JR (2020) Utilization of genomic signatures for personalized treatment of breast cancer. In: *Pharmacogenetics of breast cancer*, pp 107–122. <https://doi.org/10.1201/9780429137723-8>
29. Kern R, Correa SC, Scandolara TB, da Silva JC, Pires BR, Panis C (2020) Current advances in the diagnosis and personalized treatment of breast cancer: lessons from tumor biology. *Pers Med* 17:399–420. <https://doi.org/10.2217/pme-2020-0070>

30. Satoh Y, Imokawa T, Fujioka T, Mori M, Yamaga E, Takahashi K et al (2022) Deep learning for image classification in dedicated breast positron emission tomography (dbPET). *Ann Nucl Med* 36:401–410. <https://doi.org/10.1007/s12149-022-01719-7>
31. Vidic I, Egnell L, Jerome NP, Teruel JR, Sjøbakk TE, Østlie A et al (2018) Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: preliminary study. *J Magn Reson Imaging* 47:1205–1216. <https://doi.org/10.1002/jmri.25873>
32. Fondon I, Sarmiento A, Garcia AI, Silvestre M, Eloy C, Polonia A et al (2018) Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Comput Biol Med* 96:41–51. <https://doi.org/10.1016/j.combiomed.2018.03.003>
33. Rajyalakshmi U, Rao SK, Prasad KS (2017) Supervised classification of breast cancer malignancy using integrated modified marker controlled watershed approach. In: *Proceedings - 7th IEEE international advanced computing conference, IACC 2017*, pp 584–589. <https://doi.org/10.1109/IACC.2017.0125>
34. Wolberg WH, Mangasarian OL (1990) Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proc Natl Acad Sci U S A* 87:9193–9196
35. Mangasarian OL, Setiono R, Wolberg WH (1990) Pattern recognition via linear programming: theory and application to medical diagnosis. *Large-Scale Num. Opt.* SIAM, Philadelphia, pp 22–31
36. Street WN, Wolberg WH, Mangasarian OL (1993) Nuclear feature extraction for breast tumor diagnosis. In: *Imaging Science and Technology/Society of Photographic Instrumentation Engineers. International Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, vol 1905, pp 861–870
37. Street NW (2000) Xcyt: a system for remote cytological diagnosis and prognosis of breast cancer. In: Jain LC (ed) *Soft computing techniques in breast cancer prognosis and diagnosis*. World Scientific Publishing, Singapore, pp 297–322
38. Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R (2013) Computeraided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Comput Biol Med* 43:1563–1572. <https://doi.org/10.1016/j.combiomed.2013.08.003>
39. Cun YL, Jackel LD, Boser B, Denker JS, Graf HP, Guyon I et al (1989) Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Commun Mag* 27:41–46. <https://doi.org/10.1109/35.41400>
40. Kowal M, Skobel M, Gramacki A, Korbicz J (2021) Breast cancer nuclei segmentation and classification based on a deep learning approach. *Int J Appl Math Comput Sci* 31:85–106. <https://doi.org/10.34768/amcs-2021-0007>
41. Doussal VL, Tubiana-Hulin M, Friedman S, Hacene K, Spyrtos F, Brunet M (1989) Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (SBR). An improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer* 64:1914–1921. [https://doi.org/10.1002/1097-0142\(19891101\)64:9<1914::aid-cnrcr2820640926>3.0.co;2-g](https://doi.org/10.1002/1097-0142(19891101)64:9<1914::aid-cnrcr2820640926>3.0.co;2-g)
42. Klir GJ, Yuan B (1995) *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, Upper Saddle River, NJ
43. Theera-Umporn N (2005) Patch-based white blood cell nucleus segmentation using fuzzy clustering. *ECTI EEC* 3(1):15–19
44. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*. <https://doi.org/10.48550/arxiv.1409.1556>
45. Friess TT, Cristianini N, Campbell C (1998) The kernel adatron algorithm: a fast and simple learning procedure for support vector machines. In: *15th International conference on machine learning*. Morgan Kaufman Publishers, Burlington, MA
46. Kingma DP, Ba JL (2014) Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*