Development of a Cost-Effective Intelligent Clinical Decision Support System for Breast Cancer Early Diagnosis and Triage

https://doi.org/10.3991/ijoe.v18i05.29067

Marilena Tarousi[1](✉), Panagiotis Bountris[1], Panagiotis Daskalakis[2], Dimitrios D. Koutsouris[1]
[1] Biomedical Engineering Laboratory, National Technical University of Athens, Athens, Greece
[2] General hospital "Elena Venizelou", Athens, Greece
mtarousi@biomed.ntua.gr

Abstract—Women population screening using mammography has dramatically reduced breast cancer rates worldwide. Nowadays, in many countries, the prevention of breast cancer policy is based on frequent and repeated mammographies, followed by breast ultrasound and if necessary, by histological examination in the biological material of the biopsy. However, evaluating mammography findings is considered as a difficult process which can properly performed only by a highly experienced and well-trained medical staff. Subsequently, the interpretation of those findings could be easily influenced by subjective factors and therefore can be prone to diagnostic errors as evidenced in the present study. Breast MRI, is a diagnostic practice indicated in cases of high breast density and at the same time, a high-cost examination for healthcare systems. Furthermore, genetic testing that is used to diagnose hereditary breast cancer, represents a small proportion of breast cancers and at the same time it is a high-cost specialized test. Even the most widely used biopsy, Fine Needle Aspirate (FNA), in some cases involves risks and provides false negative results. This study presents a novel intelligent Clinical Decision Support System (CDSS), which uses data from common practice, non-invasive and low-cost diagnostic tests, together with medical health record, in order to provide clinicians with a viable, cost-effective and accurate diagnostic solution. After implementing several algorithms, Random Forest classifier showed the highest values of sensitivity 96,2 %, specificity 94,6%, PPV 96,2% and NPV 94,6%, being thus an effective algorithm in the development of our innovative CDSS model aiming to constitute a very useful tool in clinical practice for breast cancer early diagnosis.

Keywords—breast cancer, clinical decision support system, mammography, breast ultrasound, Random Forest classifier

1 Introduction

Breast cancer is considered "an insidious enemy" that is taking on ever-increasing proportions posing a major public health problem [1], [2]. It is a highly heterogeneous

disease and varies in its aetiology and pathological features. Regardless of the development level of the countries, breast cancer is the primary cancer that affects the female population, in terms of annual new cases. Global Cancer Observatory (GLOBOCAN) reports that Breast cancer is the most common type of cancer among women in 157 countries, in 2020 [3]. Breast cancer can occur in women of all ages after adolescence and its onset is affected by multiple risk factors [4]. According to World Health Organization (WHO) [3], in 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths all over the world. Breast cancer is considered as a prevalent cancer, as of the end of 2020, there were 7.8 million women who were diagnosed with breast cancer in the past 5 years and stayed alive [3]. According to world statistics, women who were diagnosed with breast cancer had lost more Disability-Adjusted Life Years (DALYs) than those who suffered by other types of cancer [5]. These numbers reflect the enormous impact of breast cancer on the global community as well as the need for urgent scientific collaboration for the prevention, early diagnosis, and effective treatment of the disease [5]. Regarding Greece, breast cancer is the most common type of cancer in women, and also responsible for most deaths due to cancer. In fact, it is the third leading cause of cancer death in the general population. This is translated to 7,772 women who were first diagnosed with breast cancer in 2020 and 2.333 who died of the disease [3].

Prevention and early diagnosis are crucial in dealing with the scourge of breast cancer. Early detection is the cornerstone of breast cancer control and is the right basis to achieve an improvement in breast cancer prognosis and survival rate. The earlier the disease is detected, the better the prognosis for patient survival, resulting to the avoidance of difficult treatments [6]. Each case of the disease must be treated individually as a new entity in order to achieve optimal management and treatment.

According to the European guidelines and protocols for the management and screening of women with possible breast cancer, mammography (MAMMO) is considered as the reference diagnostic tool [7]. Subsequently, breast Ultrasound (US) is used as a complementary medical imaging modality and in more rare and special cases of extremely dense breasts, Magnetic Resonance Imaging (MRI) is also performed [7]. These recommendations [7] are mainly for women with an average risk of breast cancer and not an increased risk due to genetic predisposition (mutations in the BRCA1 and BRCA2 genes or due to race/ethnicity [8]). The recommendations suggest that effective policy making to address breast cancer burden requires meaningful discussion and involvement of various stakeholders [7].

In more detail, MAMMO examination is the main method of breast imaging [9]. It is used for both diagnostic testing and symptomatic diagnosis. Breast imaging results are classified using the BI-RADS (Breast Imaging-Reporting Data System) [10] classification system. BI-RADS is a grading system that summarizes the findings of MAMMO. It facilitates the monitoring of the results and reduces the confusion in their interpretation. It is divided into seven categories, as presented in Table 1 [10]:

Table 1. BI-RADS categories

0	Insufficient examination. Additional imaging assessments and examinations or previous MAMMOs are required for comparison.
1	Negative.
2	Benign findings. There is a 0% chance of some malignancy. The evaluation is normal and the findings can be cysts, lipomas, benign calcifications, typical fibroadenomas, etc.
3	Probably benign findings. There is theoretically a non-palpable lesion, with a <2% chance of malignancy.
4	Findings suspected of malignancy. There is up to a 94% chance of cancer. This category is divided into 3 subcategories, BIRADS 4a (2-9%, low probability of malignancy), BIRADS 4b (10-49%, moderate probability of malignancy) and BIRADS 4c (50-94%, high probability of malignancy).
5	Findings with a strong indication of malignancy. The probability of malignancy is over 95% and surgery is required, as well as appropriate measures.
6	Known and biopsy-proven malignancy.

Also, breast density is classified into four categories according to ACR (American College of Radiology), specifically it is defined as A for almost entirely fatty breasts, B for breasts with scattered areas of fibroglandular density, C for heterogeneously dense breasts and D for extremely dense breasts [10]. BI -RADS can be used as a tool for risk assessment and quality assurance of results. The latest edition [10] features the findings of both US and MRI. However, the most common application of the aforementioned calibration is in MAMMO. Whereas it is the only tool for evaluating mammography findings, sometimes it can lead clinicians to wrong conclusions. Specifically, in cases where MAMMO belongs to one of the middle categories, such as 3 and 4, there is an increased concern by the surgeons side, regarding the handling of the woman's case as several studies have shown that the sensitivity of the examination is around 80% [11], [12]. In fact, the middle BI-RADS categories have been proven several times to lead the diagnosis to false negative results, thus preventing surgeons from drawing safe conclusions. In these cases, the breast surgeon is called upon to investigate, through surgery, the possibility of the existence of an underlying malignancy. BI-RADS 1, 2 and 5 provide clinicians with a clearer picture of the breast condition based on their probability. BI-RADS 6 malignancy is examined by biopsy so there is also no margin for error. As regards BI-RADS 0, this is a special category in which no conclusion can be drawn about the breast condition due to the ambiguity of the result as to the likelihood of malignancy. According to the literature, it is often observed in dense breasts, where MAMMO misses many cancers [13]. The aforementioned weaknesses of the BI-RADS classification system motivates the scientific community to seek methods to "enhance" the diagnostic tool of MAMMO.

As mentioned above, breast US is the complementary examination to MAMMO during the diagnostic process [7]. It is a particularly important diagnostic test as it can provide critical information about breast findings. According to the American College of Radiology [10], utilizing breast US examination is essential for characterizing the suspicious mass using morphological features such as shape, margin and orientation. Feature categories, such as vascularisation, hardness and echogenicity characteristics, as well as techniques such as colour or dynamic Doppler and elastography, can contribute in valuable information for drawing appropriate conclusions about the mass under

investigation [10]. However, there are some cases where its results do not allow experts to draw safe conclusions. According to extensive studies [14], [15], US can lead to false positive or false negative outcomes, especially in asymptomatic women. In particular, US cannot detect the presence of calcifications. Also, both US and MAMMO cannot detect diffusely growing cancers that do not have a discrete mass, leading to false negative results. In addition, sometimes there is an overlap between benign and malignant lesions, which can lead to either false negative or false positive results if they are not thoroughly evaluated by experienced radiologists. The accuracy of US examination may vary depending on the morphology of breast cancer. For example, stellate lesions in fatty breasts and lesions accompanied by microcalcifications cannot be readily detected by US [15]. US should be used to enhance diagnosis with the information it provides on the morphology and other characteristics of the specific area of interest. While in cases where mammographic findings cannot be correlated with ultrasound findings, Mammographic results should be taken into account in order to avoid an incorrect diagnosis [14].

Breast MRI is the diagnostic test that is complementary to the aforementioned examinations and increases the diagnostic yield in cases of dense breasts. The use of MRI in the process of diagnosing breast cancer has increased significantly in recent years. MRI is a test with a high sensitivity but less specificity than other imaging methods [10]. For this reason, the use of this test is a matter of controversy for the scientific community, as it raises legitimate concern for possible overtreatment and it has been shown to cause delays in diagnosis in specific cases [9][10]. In particular, it often leads to increased rates of lateral and bilateral mastectomy without certain improvement in survival rates for patients. The inclusion of breast MRI, as a diagnostic practice, is indicated in cases of very dense breasts (ACR 3-4) and is a high-cost test for the healthcare system [16]. For that reason, MRI is not suitable for systematic use during the diagnostic procedure, however there are specific cases where its diagnostic value is unquestionable [17].

Regarding genetic testing used to detect mutations implicated in breast cancer, due to the small percentage of women with hereditary breast (5 – 10% of breast cancer patients) and the high cost of genetic testing, its universal application is an inefficient strategy for healthcare systems [18].

Finally, FNA is the most popular, non-imaging, invasive test used in the breast cancer diagnostic process. FNA is a minimally invasive biopsy method and is particularly preferred when lymph node infiltration needs to be assessed. However, according to information from the medical community, FNA test carries certain risks, such as bruising, hematoma, infection, pneumothorax (if the needle is pushed deep into the chest area) and pain. Furthermore, the need for evaluation by an experienced cytopathologist for the diagnosis is noted [19].

In conclusion, taking into account the vulnerabilities of the aforementioned diagnostic tools for the detection of breast cancer, this study presents an innovative intelligent Clinical Decision Support System (CDSS) that integrates the features of two non-invasive and low-cost diagnostic methods, MAMMO and US, combining them with the

features of patient's individual medical record, in order to provide clinicians with a viable, cost-effective and accurate diagnostic solution. The proposed system aims to untie the hands of clinicians in critical decisions concerning women's health outcomes.

2 Related work

Numerous risk assessment models have been developed in order to quantify the combined effect of various risk factors and to predict either the risk of breast cancer, the risk of having a high-risk genetic mutation, especially in the BRCA1 and BRCA2 genes, or the risk of both, using Statistical modeling. These breast cancer risk assessment models can be divided into those that use mainly hormonal and environmental factors and those that focus more on hereditary risk [20]. The most well-known hormonal/environmental models (e.g., Gail Model [21], Care Model [22], Barlow Model [23], Tice Model [24], Bodian Model [25]) include factors such as age at menarche, age at first childbirth, etc., pathological factors (e.g., number of previous breast biopsies) and hereditary factors (e.g., first-degree relatives of women with breast cancer). Another class of models for estimating genetic mutations are the genetic mutation prediction models or the inherited Models, such as Myriad I [26] and II [27], Penn and Penn II [28], Bellcross [29], etc. These models are used to determine the risk of deleterious mutations, mainly in the BRCA1 and BRCA2 genes in order to assist doctors in determining the women who are in need of genetic counselling. There are also combined models, such as Tyrer-Cuzick [30], which estimates a woman's risk of carrying a BRCA1 or BRCA2 mutation and her risk of developing breast cancer (invasive or in situ) over time, taking into account hereditary, hormonal and pathological risk factors.

Concerning breast cancer, a CDSS could assist clinicians in diagnostic decision making as it can provide information about the risk of developing the disease in specific patient cases [31]. Numerous published studies refer to CDSS for supporting diagnosis of breast cancer patients based on MAMMO, breast US and MRI image analysis (i.e. image preprocessing, segmentation, feature extraction, and classification), applying a variety of machine learning (ML) algorithms (e.g., k- Nearest Neighbor (k-NN), Support Vector Machine (SVM), Binary- Logistic regression (LR)) in order to increase diagnostic performance and identify malignant and benign tumors with high accuracy, sensitivity and specificity [32], [33], [34]. Furthermore, V. Kate et al.[35], proposed a multi-classification method based on deep learning using breast tumor histopathology images. Many research groups have extended the aforementioned risk assessment models, such as G.F Stark et al. [36], who enriched Gail Model [21] with easily accessible personal health data to predict five-year breast cancer risk with the help of ML tools.

Another large class of CDSS use the Wisconsin Breast Cancer Database (WBCD) [37], that is open to the scientific public [38], [39], [40], [41]. This database contains 32 features extracted from the digitized image of FNA breast mass examination for 699 cases. The features describe the morphology of the cell nuclei present in the image. Many studies compare different ML algorithms such as Gated Recurrent Unit (GRU)–SVM , LR, Multilayer Perceptron (MLP), k- NN, Linear Discriminant Analysis (LDA), Softmax Regression, and others have chosen the combination of Ensemble Learning

algorithms in order to evaluate the effectiveness of algorithms in terms of accuracy, validity, sensitivity, and specificity based on WBCD [42], [43], [44], [45]. Furthermore, A. Osareh et al. [46], in their study used two publicly available datasets. The first dataset contained fine needle aspirate samples from breast lumps (FNAB) and the second dataset consisted of gene microarrays. The best overall diagnosis accuracy achieved by SVM classifier was 98.80% and 96.33% respectively for the two datasets. Another research team, by M.M.Rahman et al. [47], applied well known ML algorithms on anthropometric and clinical characteristics, specifically body mass index (BMI), age and levels of glucose, MCP-1 chymokine, resistin and insulin. The classification accuracy achieved was 93.9%. Additionally, Y. Chang et al.[48], recently presented their hybrid ML approach based on WBCD. Similarly, M.F Aslan et al. [49], processed the results of women's blood tests using various ML methods in order to achieve early diagnosis of breast cancer. The dataset they used originated from the UCI [50] library and included characteristics such as age, BMI, glucose and insulin levels, etc.

Various CDSS would be useful in critical decisions such as choosing which screening tests a woman should undergo, interpreting test results and deciding whether a woman should be referred for an invasive or expensive diagnostic test. These decisions are currently guided by clinical practice guidelines (CPGs), which are unlikely to work effectively in the general population given its heterogeneity. The research team of A.M. Alaa et al. [51], proposed a personalised DSS tailored to the characteristics of individuals. Their system, ConfidentCare, creates groups of patients with similar characteristics (e.g., imaging characteristics, age, breast density, family history, etc.), combines the data from their personal medical records with the results of MAMMO, US and MRI, and using ML algorithms, suggests the appropriate screening policies and practices that would be most effective in managing each group of patients individually. A similar method, of grouping patients based on their characteristics, was used by M.Alamelumangai et al. [52], that proposed a CDSS that combines the data of an individual's history (age, breast density, family history, previous biopsies) with the results of MAMMO, US and MRI (using BI-RADS scores).

Regarding the aforementioned models that have been developed for risk assessment and early diagnosis of breast cancer, several of them use data from invasive tests including genetic information, information from haematological tests, biopsy data (core biopsy and FNA). Another large category uses data from MAMMO, US and MRI image analysis, regardless of the cost, and features obtained from women's medical records. In conclusion, to the best of our knowledge, no equivalent intelligent system to the one proposed in this study has been published to date in the literature, meaning a system that combines non-invasive and cost-effective diagnostic tools along with the characteristics of the medical history form, as presented below, to increase diagnostic performance.

3 Materials and methods

3.1 Clinical data

Current study's dataset was extracted from a population of women who had referred to Breast Cancer Center of General Hospital of Athens (Greece) "Elena Venizelou" over a period of 15 months, in order to undergo a clinical examination and diagnostic tests to investigate the possibility of developing breast cancer. All medical data (diagnostic test results, pathological and anatomical findings, detailed personal and family history) were registered and stored anonymously in order to be processed with respect to personal data protection and GDPR (General Data Protection Regulation) rules [53], and after the approvance of Scientific and Administrative Council of "Elena Venizelou" Hospital.

The study is based on 489 women (mean age = 54.8, sd = 13.5) who underwent breast surgery, either conservative breast surgery (ingectomy) or mastectomy, due to suspicious findings or medical history. According to the histological analysis reports of surgical excision specimens, 287 malignant cases and 202 benign cases were found. Histological result was used as the "Gold Standard" of the present study so the cases were classified in two categories: Negative for breast cancer (202/489 - Class1) and Positive for breast cancer (287/489 - Class 2). The data included in the study for each woman were a medical history form, a MAMMO report classified by BI-RADS scale and a breast US report.

As mentioned earlier, the BI-RADS classification system [10] is used as a benchmark to evaluate MAMMO and acts as a diagnostic tool for assessing breast cancer risk based on Mammographic findings. Women's mammographies were divided into the 6 BI-RADS categories as shown in Table 2:

Table 2. Number of MAMMO cases by BI-RADS category

BI-RADS 0	120
BI-RADS 1	6
BI-RADS 2	50
BI-RADS 3	105
BI-RADS 4a	34
BI-RADS 4b	36
BI-RADS 4c	80
BI-RADS 5	58

Having the 489 mammographies classified by BI-RADS, we investigated the diagnostic performance of MAMMO. Specifically, based on probabilities assigned to each BI-RADS category, representing the likelihood of malignant lesion (see Table 1), we grouped mammographic results in two clusters. One cluster represents the negative prediction (in terms of malignancy) and the other cluster represents the positive prediction.

Different positivity thresholds were used to group MAMMO results into the aforementioned clusters (see Table 3, Table 4 & Table 5). This process showed that MAMMO yields a large number of mostly false negative and fewer false positive results in all cases, in whichever direction we "shift" the positivity threshold. It is important to note that, despite the fact that mammographies belonging to the BI-RADS 0 category constitute a significant proportion of all cases (120/489), these cases were excluded from the grouping due to their characterization as insufficient examination [54], meaning that they do not provide any information about the possibility of breast malignancy. Specifically, this category does not allow any conclusion to be drawn in relation to the likelihood of malignancy so the cases cannot be included in either of the 2 aforementioned clusters.

To sum up, we used the results of the histological analysis in comparison to mammographies assessment (via BI-RADS) in order to create three Confusion Matrices (CM), as presented in Tables 3, 4 and 5:

Table 3. Confusion matrix 1 - MAMMO results with a positivity threshold of 4a and above (according to BI-RADS)

CM 1	Positive histology result	Negative histology result
BI-RADS 4a, 4b, 4c, 5	177	31
BI-RADS 1,2,3	50	111

Table 4. Confusion matrix 2 - MAMMO results with a positivity threshold of 4b and above (according to BI-RADS)

CM 2	Positive histology result	Negative histology result
BI-RADS 4b, 4c, 5	166	8
BI-RADS 1,2,3, 4a	61	134

Table 5. Confusion matrix 3 - MAMMO results with a positivity threshold of 4c and above (according to BI-RADS)

CM 3	Positive histology result	Negative histology result
BI-RADS 4c, 5	136	2
BI-RADS 1,2,3, 4a, 4b	91	140

There are many measures and methods to evaluate the performance of a diagnostic test for classifying cases in a binary situation. The most widely used measures are the accuracy, the sensitivity and specificity. These measures assess the probability of test result correctness, in relation to the actual condition of the individual. Other important evaluation measures include positive predictive value (PPV), negative prognostic value (NPV), the Youden's index, that is a measure to evaluate the diagnostic test's ability to balance sensitivity and specificity and the F1 score that constitutes the harmonic mean of PPV and sensitivity of the model [55], [56]. Table 6 presents the pooled results for the diagnostic measures of MAMMO for all 3 positivity thresholds explained previously.

Table 6. Diagnostic efficiency of MAMMO for breast cancer detection in the study population

	MAMMO results		
	threshold BI-RADS 4a	***threshold BI-RADS 4b***	***threshold BI-RADS 4c***
Accuracy (%)	78	81	74.8
Sensitivity (%)	77.9	73.1	59.9
Specificity (%)	78.1	94.4	98.6
PPV (%)	88.5	95.4	98.5
NPV (%)	68.9	68.7	59
Youden's index	0.56	0.58	0.59
F1 score	0.83	0.83	0.74

After identifying the diagnostic weaknesses of MAMMO, demonstrated in Table 6, we excluded the BI-RADS score from our feature input vector, and instead we used descriptive mammography features to assist the decision support system in increasing diagnostic performance as presented below.

3.2 Data preprocessing and feature vector

For each case of the study population, a feature vector was created, which consists of 35 variables. These variables were derived from women's medical history forms and of both diagnostic tests (MAMMO and breast US). The histopathological result (Gold Standard) was used as the target variable (output).

The selection of the input vector features of the models was based on an extensive review of the international literature according to the most important and determinant risk factors for breast cancer. To begin with, woman's age seems to have a significant influence on the level of risk of developing breast cancer according to the international scientific and medical community [57], [58], [59]. Additionally, reproductive factors, including age of onset of menstruation, number of childbirths, number of abortions or miscarriages [60], age of first childbirth [61], age of last childbirth [62], breastfeeding, oral contraceptive pill use, age of onset of menopause and possible menopausal hormone therapy, have been shown, in a large number of studies, to play an important role in assessing the level of breast cancer risk [58], [63], [64]. Another risk factor of developing breast cancer seems to be the Body Mass Index (BMI) and obesity, since several studies have been published on the subject but with contradictory results [65], [66], [67], [68]. Subsequently, particular characteristics of the breast, such as its density [69], [70], [71] and size [72], [73], have been extensively studied for their significance on the risk of breast cancer. According to the medical community and relevant studies, some of the most important risk factors are woman's personal history disorders (previous female disorders, previous breast diseases, thyroid disorders [74], [75], hypertension [76], high cholesterol levels [77]) and her family history of breast and other cancers [78], [79], [80]. Finally, another risk factor relates to the woman's habits and specifically periods of smoking as well as the amount of cigarettes smoked during this period [81].

Regarding imaging examinations, features extracted from MAMMO and breast US were included in the feature space and their selection was based on their importance according to the international literature. In particular, one of the most important points that breast surgeons pay attention, during the diagnostic process through MAMMO, is whether the detected tumor or lesion is a new finding compared to a previous relevant examination. Another feature that is emphasized in the explanation of MAMMO findings is whether the breasts have micronodular morphology and calcifications [82]. The architectural disorder in the breast and the breast density based on the ACR scale, detected on MAMMO examination, is a very important risk factor for mammographic findings, as highlighted in numerous published studies [82], [83]. Complementary to MAMMO findings, US can provide critical information regarding the shape of the tumor detected in the examination (e.g., oval shape, radial shape, etc). The shape of the detected tumor is a strong indicator of the US finding, as the "smoother" the shape of the tumor is (e.g., oval or round) the more likely it is to be a benign lesion, as opposed to radial or spindle-shaped lesions which are "incriminated" as more likely to be malignant. Another important feature that US report contains, is the characterization of the boundaries of the localized lesion in terms of their smoothness. Additionally, the vascularization of the tumor as well as the condition of the lymph nodes, are frequently described in the US report, information of particular concern to the clinicians as they can be the tell-tale signs of a malignant lesion [82].

Table 7 describes in detail the variables resulting from women's medical history forms and the diagnostic tests (MAMMO and US), as well as the values of these variables (feature vector). The resulting feature vector is used to develop the following CDSS model.

Table 7. Variables of feature vector

	Variable' Name	**Description**	**Range of values**
1	Age	Woman's age	Numeric Variable
2	Acr	Grading of breast density based on the ACR scale [54]	1= The breasts are almost completely fatty 2= There are scattered areas of fibroglandular tissue 3= Breasts are heterogeneously dense 4= Breasts are extremely dense
3	Bmi	Body Mass Index = (Weight/Height^2)	1= (< 18.5 kg/m^2) Underweight 2= (18.5-24.9 kg/m^2) Healthy weight 3= (25-29.9 kg/m^2) Overweight 4= (>30 kg/m^2) Obesity
4	Breast Size	Woman's Breast size (Categories were based on woman's bra number)	0= Small size 1= Middle size 2= Big size 3= Very big size
5	Smoking	Duration of Smoking in years	Numeric Variable
6	Number of Cigarettes	Number of Cigarettes per Day	Numeric Variable
7	Gynaecological diseases	Previous or current Gynaecological diseases/conditions	0= There are no gynaecological conditions in woman's history, 1= Uterine polyps/ Ovarian cysts/ Polycystic ovary syndrome/ Ectopic pregnancy,

			2= Fibroids, 3= Cervical cancer/Endometrial cancer
8	Previous Breast Diseases	Previous Breast Diseases	0= There are no breast diseases in woman's history, 1= Breast cysts, 2= Fibroadenoma, 3= Mastitis 4= Calcifications 5= Breast cancer
9	Hypertension	Woman suffers from Hypertension	0= No 1= Yes
10	Thyroid Problems	Woman has problems with the thyroid gland	0= No 1= Yes
11	High Cholesterol	Woman has high cholesterol levels	0= No 1= Yes
12	Cancer History (1rst Degree- Same Type)	Mother, Father, Children or Sibling with Breast cancer	0= None >=1 Number of relatives with Breast Cancer
13	Cancer History (2nd Degree - Same Type)	Uncle/Aunt, Cousin, Grandmother/Grandfather, Nephew/Niece with Breast cancer	0= None >=1 Number of relatives with Breast cancer
14	Cancer History (1rst Degree - Other Type)	Mother, Father, Children or Sibling with another type of cancer (other than breast cancer)	0= None >=1 Number of relatives with other cancer
15	Cancer History (2nd Degree - Other Type)	Uncle/Aunt, Cousin, Grandmother/Grandfather, Nephew /Niece with another type of Cancer (other than breast cancer)	0= None >=1 Number of relatives with other cancer
16	Onset of menstruation	Age of Menstruation onset	1= up to 11 years old 2= 12 - 14 years old 3= 15 years and over
17	Age of Menopause	Age of Menopause onset	0= Woman is not in menopause 1= 26 - 42 years old 2= 43 - 50 years old 3= 51 years and over
18	Menopause	Woman is menopausal	0= No 1= Yes
19	Pregnancies	Number of Pregnancies	0= None >=1 Number of Pregnancies
20	Miscarriages	Number of Miscarriages	0= None >=1 Number of Miscarriages
21	Abortions	Number of Abortions	0= None >= 1 Number of Abortions
22	Age of 1st Childbirth	Age of 1st Childbirth	1= No pregnancy 2= up to 22 years old 3= 23 - 26 years old 4= 27 - 32 years old 5= 33 years and over
23	Age of Last Childbirth	Age of Last Childbirth	1= One or no pregnancy 2= up to 25 years old 3= 26 - 32 years old

			4= 33 years and over
24	Breastfeeding	Total breastfeeding duration of children in months	0= Not at all 1= Up to 12 months 2= More than 12 months
25	Contracting Pills	Duration of taking contraceptive pills in months	0= Not at all 1= Up to 12 months 2= More than 12 months
26	Hormones	Duration of hormone intake (excluding thyroxine) in years	0= Not at all 1= Up to 5 years 2= More than 5 years
27	Thyroid Related Hormones	Duration of thyroxine intake in years	0= Not at all 1= Up to 1 year 2= More than 1 year
28	New Finding	New finding in MAMMO report compared to the previous one	0= No 1= Yes
29	Micronodular Morphology	Breasts with micronodular morphology (in MAMMO report)	0= No 1= Yes
30	Architectural Distortion	Distortion of the architecture of the breasts (in MAMMO report)	0= No 1= Yes
31	Calcifications	Calcifications' morphology (in MAMMO report)	0= No existence 1= Benign 2=Macrocalcifications 3= Non clustered 4= Scattered 5= Clustered 6= Suspicious morphology
32	Tumor Shape	Nodular shading Shape (in US report)	0= No shading of a breast 1= Round 2= Oval 3= Longitudinal 4= Lobed 5= Spindle-shaped 6= Micro-lobed 7= Irregular 8= Radial
33	Tumor Margins	Tumor margins (in US report)	0= Indistinct 1= Smooth
34	Vascularity	Hematoma - Vascularisation of tumor (in US report)	0= Poor/ Not increased/ Normal 1= Increased
35	Lymph Nodes	Lymph node morphology (in US report)	0= No abnormal swelling/ no presence 1= Benign morphology 2= Intramammary 3= Gastric lymph nodes 4= Suspicious morphology 5= Pathologically swollen

Python (3.6.13 version) was used to preprocess our clinical data, build ML algorithms and perform data analysis. The availability of numerous open-source libraries and tools make Python an ideal choice for ML models development. Particularly, in

order to develop our ML models, *pandas*, *numpy*, *scikit-learn* and *statistics* libraries were used.

Before training our models, all variables were investigated for the existence of missing values. As all the data used in this study were obtained from the Breast Cancer Center's records of "Elena Venizelou" hospital, we were expecting for the presence of missing values. Thus, when more than 60% of a variable's values were missing, this variable was excluded from the study. Following the aforementioned rule, we excluded bloud group (74%), alcohol consumption (70%) and tumor's compactness (64%). For the remaining features, we replaced missing values by applying an iterative imputation (IterativeImputer class – sklearn library, Python 3.0) strategy, a sophisticated approach that involves defining a model to predict each missing feature as a function of all other features.

4 Results

4.1 Classification/prediction models

Two prominent ML classifiers were used and tested in order to find the model with the highest diagnostic performance. Classifiers' role was to categorize the cases in two classes (Positive result and Negative result) corresponding to the breast histology. Each classifier was fed by the feature vector presented in Table 7 and extracted the corresponding classification group, providing thereby a prediction about the actual condition of each woman's breasts. It is worth mentioning that although the feature vector consists of boolean, ordinal and numerical variables (see Table 7), the models selected do not require feature scaling. In particular, neither Naive Bayes, which involves multiple application of the Bayes rule, nor the tree-based Random Forest algorithm make use of any distance related optimization that would be affected by significant differences of features' magnitudes. Their principal characteristics, that led to their selection, are detailed below:

Naive Bayes classifier. The Bayesian classifier that operates on the assumption that all features are conditionally independent of each other is called naïve. The Naive Bayes (NB) [84] classifier is the simplest probabilistic classifier. Despite the naive design and the oversimplified assumptions of feature independence, it has been shown that performs particularly well on many really complex classification problems, while at the same time it is robust to violations of the independence assumption features. NB works directly and requires no special design; in order to build it, it is sufficient only to train it with the training dataset, a process in which the distributions of the features and the prior probabilities of the classes are estimated [84]. In our study, we used GaussianNB method that implements the Gaussian NB algorithm for classification, meaning that the likelihood of the continuous features is assumed to be Gaussian. In order to test NB's diagnostic performance, we used the resampling method of 10 – fold cross-validation, that splits the dataset into training and test set iteratively so as to avoid over-fitting problems. The confusion matrix resulting from testing NB classifier was used to evaluate the diagnostic performance of the model (see Table 8).

Table 8. Naive Bayes classifier's confusion matrix

Naive Bayes	Positive histology result	Negative histology result
Classified as Positive	268	13
Classified as Negative	19	189

Random Forest classifier. Random Forest (RF) [85] is one of the most efficient and stable classifiers. Fernández-Delgado et al. [86], conducted a huge comparative study on the diagnostic performance of different classifiers and RFs demonstrated the best performance compared to 170 other classifiers, using more than 100 different datasets. An RF is a collective classification tree classifier. The generalisation error is quite limited since a very large number of trees are developed. The random selection of features reduces the ratio of large and non-pruned trees, which makes the whole method quite unbiased. RF has many advantages as it can automatically handle the missing values of variables, it combines categorical and continuous data, and there is no need to apply feature scaling methods [85], [87]. Compared to decision trees, RF achieves improved prediction accuracy without increasing the computational cost [88]. The algorithm's hyperparameters that affect its predictive performance were tuned after performing grid-search. The number of trees was set to 100, the maximum number of features considered at each split was log base 2 of the total number of features and the minimum number of samples in each leaf, which helps the model avoid overfitting, was set to 2% of the number of samples.

Given the stochastic nature of RF algorithm, we ran the model several times, compared the outcomes and came up with the best results presented in the following section. We, also, used the resampling method of 10 – fold cross-validation for RF's diagnostic performance evaluation. The confusion matrix resulting from testing RF classifier is presented in Table 9. Given the stochastic nature of RF algorithm, we ran the model several times, compared the outcomes and came up with the best results.

Table 9. Random Forest classifier's confusion matrix

Random Forest	Positive histology result	Negative histology result
Classified as Positive	276	11
Classified as Negative	11	191

4.2 Diagnostic performance evaluation

Table10 presents the results of basic measures that quantify the diagnostic accuracy of the two classifiers in order to compare and select the one with the highest diagnostic performance for our CDSS model.

Table 10. Diagnostic efficiency of selected ML algorithms

	Naive Bayes	**Random Forest**
Accuracy (%)	93.5	95.5
Sensitivity (%)	93.4	96.2
Specificity (%)	93.6	94.6
PPV (%)	95.4	96.2
NPV (%)	90.9	94.6
Youden's index	0.87	0.91
F1 score	0.94	0.96

The above results (Table 10) show that RF algorithm achieved the highest accuracy 95.5 %, demonstrating thus its effectiveness in combining binary, categorical and numerical variables. RF classifier showed the highest values in the measures of sensitivity (96.2 %), specificity (94.6%), PPV (96.2%) and NPV (94.6%), indicating its suitability in creating our innovative CDSS model as it gives extremely reduced false negative and false positive results compared to the diagnostic tool of MAMMO. Specifically, based on RF's Confusion Matrix (Table 9), we found 276 true positive and 191 true negative results in the total of 489 cases examined by using cross-validation technique. Important statistical indicators are also Youden's index (0.91) and the F1 score (0.96). Youden's index incorporates the information of sensitivity and specificity. Its value ranges from 0 to 1, with the highest value indicating a significant diagnostic performance. Regarding F1 score, it is a way of combining model's PPV and sensitivity and it is defined as the harmonic mean of these two values. Its value also ranges from 0 to 1, with the highest value indicating the model's excellent ability to identify true positive results.

5 Discussion and conclusions

In order to evaluate the proposed CDSS as an assistive tool in the diagnosis of breast cancer we referred to the 3 different cases of MAMMO's diagnostic performance evaluation explained above (see Tables 3, 4 and 5) and also, we compared the values in Table 6 with the results of the RF algorithm (Table 10). Considering diagnostic performance measures in Table 9 & 10, it is evident that our CDSS model, based on RF classifier, shows obviously higher performance compared to the diagnostic tool of MAMMO based on BI-RADS scoring.

Noting the high diagnostic performance of the proposed CDSS, it is important to highlight the fact that we used the data set of 489 women, including those with a BI-RADS 0 score on their mmammography. As mentioned above, the 120 cases with BI-RADS 0 were excluded from the MAMMO Confusion Matrices and so they were not considered in the calculation of its diagnostic measures. However, these cases were not excluded from our model, as we chose to investigate their contribution to the diagnostic process despite the fact that they represent a difficult and "insufficient category". It is considered critical that our model can assist clinicians in drawing conclusions based on features of simple daily clinical practice. Namely the results from non-invasive and

cost-effective diagnostic procedures, diagnosing all cases of women without excluding any category. Proposed CDSS is therefore a tool with universal application and an essential added value in the breast cancer diagnosis process.

Our results led us to conclude that the proposed model predicts in accuracy the vast majority of cases. In particular the cases of women with a BI-RADS score of 3, 4a or 4b on their MAMMO, which, according to Table 2, constitute a significant proportion of the total dataset. The significant added value of the proposed CDSS, concerns the cases of the middle categories (according to the BI-RADS scale) where diagnostic errors and mismatches with the final histological evaluation often occur (see Table 3, 4 and 5). The complacency caused by a false negative MAMMO result (according to its BI-RADS category) is crucial for the management of a breast cancer patient as it may cause a remarkable delay in the investigation of the case or lead to mismanagement.

The fact that the system can distinguish highly accurately the cases of patients from healthy women means that it can play a doubly beneficial role for healthcare systems. On the one hand, it has the ability to identify promptly the cases of women at increased risk of breast cancer, resulting in early intervention by clinicians, preparing both the patient and the healthcare system for the personalized management of each case. On the other hand, by providing accurate information on healthy women's cases, it leads to the reduction of unnecessary costs by avoiding overdiagnosis and overtreatment. Being aware that breast cancer is a scourge for women worldwide and identifying the weaknesses of the diagnostic procedure followed so far, in the current study, we present a CDSS that is able to support life-critical decisions of clinicians. The proposed CDSS combines individual medical history data with the results of basic, non-invasive and low-cost diagnostic tests, providing clinicians with patient-specific predictions of the diagnostic outcome that leads to appropriate and timely decisions.

To sum up, this research presents a viable and cost-effective solution for national healthcare systems that may contribute to a significant cost reduction regarding the management of women at risk of developing breast cancer while achieving a remarkable increase in the validity of diagnosis.

6 Future work

The ultimate goal, of this study, is to create a comprehensive, ML based, application that includes all the input variables of our feature vector and will be designed to support clinicians' decisions in daily clinical practice. This application could be extended to several hospitals, in national level, so that the CDSS could be trained and re-evaluated with a larger amount of data. As the proposed CDSS maintains its high diagnostic performance it could be extended to an international level, incorporating the factor of race [8] that appears to influence breast cancer risk.

7 Acknowledgment

This work was funded by the UNITI project financed by the European Union’s Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261.

8 References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA. Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, Jan. 2019. https://doi.org/10.3322/caac.21551

[2] L. Iacoviello, M. Bonaccio, G. de Gaetano, and M. B. Donati, "Epidemiology of breast cancer, a paradigm of the 'common soil' hypothesis," *Seminars in Cancer Biology*, vol. 72. Academic Press, pp. 4–10, Jul. 01, 2020. https://doi.org/10.1016/j.semcancer.2020.02.010

[3] "Cancer Today." https://gco.iarc.fr/today/home (accessed May 28, 2021).

[4] A. R. Brentnall, J. Cuzick, D. S. M. Buist, and E. J. A. Bowles, "Long-Term accuracy of breast cancer risk assessment combining classic risk factors and breast density," *JAMA Oncol.*, vol. 4, no. 9, p. 180174, Sep. 2018. https://doi.org/10.1001/jamaoncol.2018.0174

[5] Z. Q. Tao, A. Shi, C. Lu, T. Song, Z. Zhang, and J. Zhao, "Breast Cancer: Epidemiology and Etiology," *Cell Biochem. Biophys.*, vol. 72, no. 2, pp. 333–338, Jun. 2015. https://doi.org/10.1007/s12013-014-0459-6

[6] O. Ginsburg *et al.*, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, May 2020. https://doi.org/10.1002/cncr.32887

[7] H. J. Schünemann *et al.*, "Breast cancer screening and diagnosis: A synopsis of the european breast guidelines," *Ann. Intern. Med.*, vol. 172, no. 1, pp. 46–56, Jan. 2020. https://doi.org/10.7326/M19-2125

[8] J. Krammer *et al.*, "Breast cancer detection and tumor characteristics in BRCA1 and BRCA2 mutation carriers," *Breast Cancer Res. Treat.*, vol. 163, no. 3, pp. 565–571, Jun. 2017. https://doi.org/10.1007/s10549-017-4198-4

[9] P. Steyerova, "Imaging of the Breast," in *Breast Cancer Management for Surgeons*, Cham: Springer International Publishing, 2018, pp. 127–145. https://doi.org/10.1007/978-3-319-56673-3_12

[10] E. A. Sickles, C. M. Appleton, E. S. Burnside, and S. C. Gavenonis, "Breast imaging reporting and data system (BI-RADS) Atlas- Ultrasound 5th edn.," *Am. Coll. Radiol. BI-RADS-Atlas*, pp. 121–132, 2013, [Online]. Available: https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/US-Reporting.pdf

[11] L. Domingo *et al.*, "Cross-national comparison of screening mammography accuracy measures in U.S., Norway, and Spain," *Eur. Radiol.*, vol. 26, no. 8, pp. 2520–2528, Aug. 2016. https://doi.org/10.1007/s00330-015-4074-8

[12] C. D. Lehman, R. D. Wellman, D. S. M. Buist, K. Kerlikowske, A. N. A. Tosteson, and D. L. Miglioretti, "Diagnostic accuracy of digital screening mammography with and without computer-aided detection," *JAMA Intern. Med.*, vol. 175, no. 11, pp. 1828–1837, Nov. 2015. https://doi.org/10.1001/jamainternmed.2015.5231

[13] K. M. Kelly, J. Dean, W. S. Comulada, and S.-J. Lee, "Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts," *Eur. Radiol. 2009 203*, vol. 20, no. 3, pp. 734–742, Sep. 2009. https://doi.org/10.1007/s00330-009-1588-y

[14] H. M. Zonderland, "The role of ultrasound in the diagnosis of breast cancer," *Semin. Ultrasound, CT MRI*, vol. 21, no. 4, pp. 317–324, Aug. 2000. https://doi.org/10.1016/S0887-2171(00)90026-X

[15] S. V. Sree, "Breast imaging: A survey," *World J. Clin. Oncol.*, vol. 2, no. 4, p. 171, 2011. https://doi.org/10.5306/wjco.v2.i4.171

[16] I.-W. Pan, K. C. Oeffinger, and Y.-C. T. Shih, "Out-of-pocket cost of screening with breast MRI for women at high risk for breast cancer.," *J. Clin. Oncol.*, vol. 38, no. 15_suppl, pp. 2017–2017, May 2020. https://doi.org/10.1200/JCO.2020.38.15_suppl.2017

[17] N. H. G. M. Peters, I. H. M. Borel Rinkes, N. P. A. Zuithoff, W. P. T. M. Mali, K. G. M. Moons, and P. H. M. Peeters, "Meta-analysis of MR imaging in the diagnosis of breast lesions," *Radiology*, vol. 246, no. 1, pp. 116–124, Jan. 2008. https://doi.org/10.1148/radiol.2461061298

[18] E. F. Long and P. A. Ganz, "Cost-effectiveness of Universal BRCA1/2 Screening: Evidence-Based Decision Making," *JAMA Oncol.*, vol. 1, no. 9, pp. 1217–1218, Dec. 2015. https://doi.org/10.1001/jamaoncol.2015.2340

[19] J. T. Casaubon, S. Tomlinson-Hansen, and J.-P. Regan, "Fine Needle Aspiration Of Breast Masses," *StatPearls*, Aug. 2021, Accessed: Oct. 07, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK470268/

[20] J. A. Cintolo-Gonzalez *et al.*, "Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications," *Breast Cancer Res. Treat.*, vol. 164. https://doi.org/10.1007/s10549-017-4247-z

[21] G. MH *et al.*, "Projecting individualized probabilities of developing breast cancer for white females who are being examined annually," *J. Natl. Cancer Inst.*, vol. 81, no. 24, pp. 1879–1886, Dec. 1989. https://doi.org/10.1093/jnci/81.24.1879

[22] R. Lindsey *et al.*, "Deep neural network improves fracture detection by clinicians," *Proc. Natl. Acad. Sci.*, vol. 115, no. 45, pp. 11591–11596, Nov. 2018. https://doi.org/10.1073/pnas.1806905115

[23] B. WE *et al.*, "Prospective breast cancer risk prediction model for women undergoing screening mammography," *J. Natl. Cancer Inst.*, vol. 98, no. 17, pp. 1204–1214, Sep. 2006. https://doi.org/10.1093/jnci/djj331

[24] N. F. Boyd, L. J. Martin, M. J. Yaffe, and S. Minkin, "Mammographic density: A harmonally responsive risk factor for breast cancer," *J. Br. Menopause Soc.*, vol. 12, no. 4, pp. 186–193, Dec. 2006. https://doi.org/10.1258/136218006779160436

[25] S. E. Singletary, "Rating the Risk Factors for Breast Cancer," *Ann. Surg.*, vol. 237, no. 4, p. 474, 2003. https://doi.org/10.1097/01.SLA.0000059969.64262.87

[26] D. Shattuck-Eidens *et al.*, "BRCA1 sequence analysis in women at high risk for susceptibility mutations: Risk factor analysis and implications for genetic testing," *J. Am. Med. Assoc.*, vol. 278, no. 15, pp. 1242–1250, Oct. 1997. https://doi.org/10.1001/jama.278.15.1242

[27] T. S. Frank *et al.*, "Sequence analysis of BRCA1 and BRCA2: Correlation of mutations with family history and ovarian cancer risk," *J. Clin. Oncol.*, vol. 16, no. 7, pp. 2417–2425, 1998. https://doi.org/10.1200/JCO.1998.16.7.2417

[28] D. D. Couch *et al.*, "BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer," *N. Engl. J. Med.*, vol. 336, no. 20, pp. 1409–1415, May 1997. https://doi.org/10.1056/NEJM199705153362002

[29] C. Bellcross, A. Lemke, L. Pape, L. Tess, and L. Meisner, "Evaluation of a breast/ovarian cancer genetics referral screening tool in a mammography population," *Genet. Med.*, vol. 11, no. 11, pp. 783–789, Nov. 2009. https://doi.org/10.1097/GIM.0b013e3181b9b04a

[30] J. Tyrer, S. Duffy, and J. Cuzick, "A breast cancer prediction model incorporating familial and personal risk factors," *Stat. Med.*, vol. 23, no. 7, pp. 1111–1130, Apr. 2004. https://doi.org/10.1002/sim.1668

[31] S. M. Carter, W. Rogers, K. T. Win, H. Frazer, B. Richards, and N. Houssami, "The ethical, legal and social implications of using artificial intelligence systems in breast cancer care," *Breast*, vol. 49, pp. 25–32, Feb. 2020. https://doi.org/10.1016/j.breast.2019.10.001

[32] N. I. R. Yassin, S. Omran, E. M. F. El Houby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," *Comput. Methods Programs Biomed.*, vol. 156, pp. 25–45, 2018. https://doi.org/10.1016/j.cmpb.2017.12.012

[33] H. Alsalman and N. Almutairi, "IDSS: An Intelligent Decision Support System for Breast Cancer Diagnosis," *2nd Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2019*, May 2019. https://doi.org/10.1109/CAIS.2019.8769579

[34] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, B. Alami, and M. Maaroufi, "Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms," *Int. J. Online Biomed. Eng.*, vol. 17, no. 11, pp. 60–75, Nov. 2021. https://doi.org/10.3991/ijoe.v17i11.24781

[35] V. Kate and P. Shukla, "Breast Cancer Image Multi-Classification Using Random Patch Aggregation and Depth-Wise Convolution based Deep-Net Model," *Int. J. Online Biomed. Eng.*, vol. 17, no. 01, pp. 83–100, Jan. 2021. https://doi.org/10.3991/ijoe.v17i01.18513
[36] G. F. Stark, G. R. Hart, B. J. Nartowt, and J. Deng, "Predicting breast cancer risk using personal health data and machine learning models," *PLoS One*, vol. 14, no. 12, p. e0226765, Dec. 2019. https://doi.org/10.1371/journal.pone.0226765
[37] "UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set." https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic) (accessed Dec. 16, 2021).
[38] M. Nilashi, O. Ibrahim, H. Ahmadi, and L. Shahmoradi, "A knowledge-based system for breast cancer classification using fuzzy logic method," *Telemat. Informatics*, vol. 34, no. 4, pp. 133–144, Jul. 2017. https://doi.org/10.1016/j.tele.2017.01.007
[39] A. F. M. Agarap, "On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset," in *ACM International Conference Proceeding Series*, Feb. 2018, pp. 5–9. https://doi.org/10.1145/3184066.3184080
[40] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis," *Des. 2018, Vol. 2, Page 13*, vol. 2, no. 2, p. 13, May 2018. https://doi.org/10.3390/designs2020013
[41] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," *J. Healthc. Eng.*, vol. 2019, 2019. https://doi.org/10.1155/2019/4253641
[42] E. A. Bayrak, P. Kirci, and T. Ensari, "Comparison of machine learning methods for breast cancer diagnosis," Apr. 2019. https://doi.org/10.1109/EBBT.2019.8741990
[43] C. Baneriee, S. Paul, and M. Ghoshal, "A comparative study of different ensemble learning techniques using Wisconsin breast cancer dataset," Nov. 2018. https://doi.org/10.1109/ICCECE.2017.8526215
[44] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "ScienceDirect The 6th International Symposium on Frontiers in Ambient and Mobile Systems Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," *Procedia Comput. Sci.*, vol. 83, pp. 1064–1069, 2016. https://doi.org/10.1016/j.procs.2016.04.224
[45] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine Learning Classification Techniques for Breast Cancer Diagnosis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 495, no. 1, p. 012033, Apr. 2019. https://doi.org/10.1088/1757-899X/495/1/012033
[46] A. Osareh and B. Shadgar, "Machine Learning Techniques to Diagnose Breast Cancer," *2010 5th Int. Symp. Heal. Informatics Bioinforma.*, 2020. https://doi.org/10.1109/HIBIT.2010.5478895
[47] M. M. Rahman, Y. Ghasemi, E. Suley, Y. Zhou, S. Wang, and J. Rogers, "Machine Learning Based Computer Aided Diagnosis of Breast Cancer Utilizing Anthropometric and Clinical Features," *IRBM*, vol. 42, no. 4, pp. 215–226, Aug. 2021. https://doi.org/10.1016/j.irbm.2020.05.005
[48] Y. Chang and F. Abu-Amara, "An Efficient Hybrid Classifier for Cancer Detection," *Int. J. Online Biomed. Eng.*, vol. 17, no. 03, pp. 76–97, Mar. 2021. https://doi.org/10.3991/ijoe.v17i03.19683
[49] M. F. Aslan, Y. Çelik, K. Sabancı, and A. Durdu, "Breast cancer diagnosis by different machine learning methods using blood analysis data," *Orig. Res. Pap. Int. J. Intell. Syst. Appl. Eng. IJISAE*, vol. 6, no. 4, pp. 289–293, 2018. https://doi.org/10.18201/ijisae.20186484 55
[50] "UCI Machine Learning Repository." https://archive.ics.uci.edu/ml/index.php (accessed Dec. 16, 2021).
[51] A. M. Alaa *et al.*, "ConfidentCare: A Clinical Decision Support System for Personalized Breast Cancer Screening," *IEEE Trans. Multimed.*, vol. 18, no. 10, pp. 1942–1955, Feb. 2016. https://doi.org/10.1109/TMM.2016.2589160

[52] M. Alamelumangai and B. Sathiyabhama, "Personalized Care: A Clinical Decision Support System for Breast Cancer Screening Using Clustering and Classification," *SSRN Electron. J.*, Nov. 2017. https://doi.org/10.2139/ssrn.3134277
[53] "Data protection in the EU | European Commission." https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en (accessed May 31, 2021).
[54] "Breast Imaging Reporting & Data System | American College of Radiology." https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads (accessed Jun. 02, 2021).
[55] S. Safari, A. Baratloo, M. Elfil, and A. Negida, "Evidence Based Emergency Medicine Part 2: Positive and negative predictive values of diagnostic tests," *Emergency*, vol. 3, no. 3, p. 87, 2015. https://doi.org/10.22037/aaem.v3i3.274
[56] G. Hughes, "Youden's index and the weight of evidence," *Methods Inf. Med.*, vol. 54, no. 2, pp. 198–199, 2015. https://doi.org/10.3414/ME14-04-0003
[57] C. Gajdos, P. I. Tartter, I. J. Bleiweiss, C. Bodian, and S. T. Brower, "Stage 0 to stage III breast cancer in young women," *J. Am. Coll. Surg.*, vol. 190, no. 5, pp. 523–529, May 2000. https://doi.org/10.1016/S1072-7515(00)00257-X
[58] A. T. Toriola and G. A. Colditz, "Trends in breast cancer incidence and mortality in the United States: Implications for prevention," *Breast Cancer Research and Treatment*, vol. 138, no. 3. Breast Cancer Res Treat, pp. 665–673, Apr. 2013. https://doi.org/10.1007/s10549-013-2500-7
[59] B. C. Yankaskas *et al.*, "Performance of First Mammography Examination in Women Younger Than 40 Years," *JNCI J. Natl. Cancer Inst.*, vol. 102, no. 10, pp. 692–701, May 2010. https://doi.org/10.1093/jnci/djq090
[60] G. Erlandsson, S. M. Montgomery, S. Cnattingius, and A. Ekbom, "Abortions and breast cancer: Record-based case-control study," *Int. J. Cancer*, vol. 103, no. 5, pp. 676–679, Feb. 2003. https://doi.org/10.1002/ijc.10877
[61] M. Lambe, C. Hsieh, H. Chan, A. Ekbom, D. Trichopoulos, and H.-O. Adami, "Parity, age at first and last birth, and risk of breast cancer: A population-based study in Sweden," *Breast Cancer Res. Treat. 1996 383*, vol. 38, no. 3, pp. 305–311, 1996. https://doi.org/10.1007/BF01806150
[62] W.-C. Chie *et al.*, "Age at Any Full-term Pregnancy and Breast Cancer Risk," *Am. J. Epidemiol.*, vol. 151, no. 7, pp. 715–722, Apr. 2000. https://doi.org/10.1093/oxfordjournals.aje.a010266
[63] H. D. Nelson *et al.*, "Risk factors for breast cancer for women aged 40 to 49 years: A systematic review and meta-analysis a systematic review and meta-analysis," *Ann. Intern. Med.*, vol. 156, no. 9, pp. 635–648, May 2012. https://doi.org/10.7326/0003-4819-156-9-201205010-00006
[64] R. Verma, R. L. Bowen, S. E. Slater, F. Mihaimeed, and J. L. Jones, "Pathological and epidemiological factors associated with advanced stage at diagnosis of breast cancer," *British Medical Bulletin*, vol. 103, no. 1. Br Med Bull, pp. 129–145, Sep. 2012. https://doi.org/10.1093/bmb/lds018
[65] M. S. R. Shawon, M. Eriksson, and J. Li, "Body size in early life and risk of breast cancer," *Breast Cancer Res.*, vol. 19, no. 1, Jul. 2017. https://doi.org/10.1186/s13058-017-0875-9
[66] H. J. Baer *et al.*, "Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: A prospective cohort study," *Breast Cancer Res.*, vol. 7, no. 3, Feb. 2005. https://doi.org/10.1186/bcr998
[67] E. V. Bandera *et al.*, "Body size in early life and breast cancer risk in African American and European American women," *Cancer Causes Control*, vol. 24, no. 12, p. 2231, 2013. https://doi.org/10.1007/s10552-013-0302-1
[68] G. Fagherazzi, G. Guillas, M. C. Boutron-Ruault, F. Clavel-Chapelon, and S. Mesrine, "Body shape throughout life and the risk for breast cancer at adulthood in the French E3N cohort," *Eur. J. Cancer Prev.*, vol. 22, no. 1, pp. 29–37, Jan. 2013. https://doi.org/10.1097/CEJ.0b013e328355ec04

[69] E. Couto *et al.*, "Hormone therapy use and mammographic density in postmenopausal Norwegian women," *Breast Cancer Res. Treat.*, vol. 132, no. 1, pp. 297–305, Feb. 2012. https://doi.org/10.1007/s10549-011-1810-x
[70] N. Boyd, G. Lockwood, J. Byng, … D. T.-C. E., and U. 1998, "Mammographic densities and breast cancer risk.," *AACR*, 2007, Accessed: Sep. 10, 2021. [Online]. Available: https://cebp.aacrjournals.org/content/7/12/1133.short
[71] K. Rojas and A. Stuckey, "Breast Cancer Epidemiology and Risk Factors," *Clin. Obstet. Gynecol.*, vol. 59, no. 4, pp. 651–672, 2016. https://doi.org/10.1097/GRF.0000000000000239
[72] L. Jansen, R. Backstein, and M. Brown, "Breast size and breast cancer: a systematic review," *J. Plast. Reconstr. Aesthet. Surg.*, vol. 67, no. 12, pp. 1615–1623, 2014. https://doi.org/10.1016/j.bjps.2014.10.001
[73] A. Markkula *et al.*, "Given breast cancer, does breast size matter? Data from a prospective breast cancer cohort," *Cancer Causes Control*, vol. 23, no. 8, pp. 1307–1316, Aug. 2012. https://doi.org/10.1007/s10552-012-0008-9
[74] E. Y. Kim *et al.*, "Serum concentration of thyroid hormones in abnormal and euthyroid ranges and breast cancer risk: A cohort study," *Int. J. Cancer*, vol. 145, no. 12, pp. 3257–3266, Dec. 2019. https://doi.org/10.1002/ijc.32283
[75] M. Simon, M. Tang, and L. Bernstein, "Do thyroid disorders increase the risk of breast cancer?," *AACR*, 2002, Accessed: Oct. 21, 2021. [Online]. Available: https://cebp.aacrjournals.org/content/11/12/1574.short
[76] H. Han *et al.*, "Hypertension and breast cancer risk: a systematic review and meta-analysis," *Sci. Reports 2017 71*, vol. 7, no. 1, pp. 1–9, Mar. 2017. https://doi.org/10.1038/srep44877
[77] M. Touvier *et al.*, "Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies," *Br. J. Nutr.*, vol. 114, no. 3, pp. 347–357, Aug. 2015. https://doi.org/10.1017/S000711451500183X
[78] R. J. MacInnis, D. R. English, A. M. Haydon, J. L. Hopper, D. M. Gertig, and G. G. Giles, "Body size and composition and risk of rectal cancer (Australia)," *Cancer Causes Control*, vol. 17, no. 10, pp. 1291–1297, 2006. https://doi.org/10.1007/s10552-006-0074-y
[79] J. L. Hopper *et al.*, "Age-specific breast cancer risk by body mass index and familial risk: prospective family study cohort (ProF-SC)," *Breast Cancer Res. 2018 201*, vol. 20, no. 1, pp. 1–11, Nov. 2018. https://doi.org/10.1186/s13058-018-1056-1
[80] Y. Guo *et al.*, "Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent," *PLOS Med.*, vol. 13, no. 8, p. e1002105, Aug. 2016. https://doi.org/10.1371/journal.pmed.1002105
[81] C. Catsburg, V. A. Kirsh, C. L. Soskolne, N. Kreiger, and T. E. Rohan, "Active cigarette smoking and the risk of breast cancer: a cohort study," *Cancer Epidemiol.*, vol. 38, no. 4, pp. 376–381, Aug. 2014. https://doi.org/10.1016/j.canep.2014.05.007
[82] M. A. Benito and J. C. Herrero, "Breast Imaging," *Learn. Diagnostic Imaging 100 Essent. Cases*, pp. 1–25, 2008. https://doi.org/10.1007/978-3-540-71207-7_1
[83] M. Bahl, J. A. Baker, E. N. Kinsey, and S. V. Ghate, "Architectural Distortion on Mammography: Correlation With Pathologic Outcomes and Predictors of Malignancy," vol. 205, no. 6, pp. 1339–1345, Nov. 2015. https://doi.org/10.2214/AJR.15.14628
[84] D. Lowd and P. Domingos, "Naive Bayes Models for Probability Estimation," 2005. https://doi.org/10.1145/1102351.1102418
[85] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees," 2012, Accessed: Oct. 19, 2021. [Online]. Available: www.IJCSI.org
[86] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, and A. Fernández-Delgado, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?," *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, 2014, Accessed: Dec. 20, 2021. [Online]. Available: http://www.mathworks.es/products/neural-network

[87] P. Bountris, M. Haritou, A. Pouliakis, P. Karakitsos, and D. Koutsouris, "A decision support system based on an ensemble of random forests for improving the management of women with abnormal findings at cervical cancer screening," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-November, pp. 8151–8156, Nov. 2015. https://doi.org/10.1109/EMBC.2015.7320286

[88] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, B. Alami, and M. Maaroufi, "Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms," *Int. J. Online Biomed. Eng.*, vol. 17, no. 11, pp. 60–75, Nov. 2021. https://doi.org/10.3991/ijoe.v17i11.24781

9 Authors

Marilena Tarousi is a Ph.D. Candidate in the Biomedical Engineering Laboratory at the National Technical University of Athens**.** She holds a diploma in Electrical and Computer Engineering (National Technical University of Athens) and a Master's degree in Biomedical Engineering. She has a firm background in Clinical Decision Support Systems, Biomedical Signal Analysis and Mathematical Modeling (email: mtarousi@biomed.ntua.gr).

Dr. Panagiotis Bountris, received his degree in Electrical & Computer Engineering and his PhD degree in Medical Informatics and Biomedical Engineering from the National Technical University of Athens (NTUA), Greece, in 2008 and 2017, respectively. His main areas of expertise are artificial intelligence, machine learning, decision support systems and medical informatics (email: p.bountris@biomed.ntua.gr).

Dr. Panagiotis Daskalakis is the chief Director of Breast Cancer Center of General Hospital "Elena Venizelou" – Athens. He holds a diploma of Medicine in Medical School of Athens, four Masters, including Medical Informatics Technology and one PhD thesis (email: pddask67@gmail.com).

Prof. Dimitris Koutsouris, received his Diploma in Electrical Engineering in 1978 (Greece), DEA in Biomechanics in 1979 (France), Doctorat in Genie Biologie Medicale (France), Doctorat d'Etat in Biomedical Engineering 1984 (France). He is currently Professor at the Dept. of Electrical & Computers Engineering of National Technical University of Athens and head of the Biomedical Engineering Laboratory. He has published over 100 research articles and book chapters and more than 150 conference communications. Recently he has also been elected as president of the Greek Society of Biomedical Technology (email: dkoutsou@biomed.ntua.gr).

Article submitted 2021-12-22. Resubmitted 2022-02-23. Final acceptance 2022-02-24. Final version published as submitted by the authors.