International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Digital Mammography Data Transformation Model to Improve the Accuracy of Benign and Malignant Detection in Breast Cancer

Anak Agung Ngurah Gunawan[1]*, I. Wayan Supardi[1], Nyoman Wendri[1], Anak Agung Ngurah Frady Cakra Negara[2], Anak Agung Ngurah Bagaskara[1], Putu Patriawan[3], I. Bagus Gede Dharmawan[4]

Submitted: 30/08/2023 **Revised**: 21/10/2023 **Accepted**: 01/11/2023

Abstract: Digital mammography is a tool for early detection of breast cancer, but its sensitivity is low in dense breasts, and the false negatives are high. Therefore, this research aims to create a digital mammography data transformation model to improve the accuracy of benign and malignant detection in breast and compare it with other advanced methods. This research took data from Dokter Sutomo Hospital Surabaya and Sanglah Hospital Denpasar in the form of mammogram images taken from the hospital database with complete pathology results from 2010 to 2023. A total of 442 mammograms consisting of 114 benign and 328 malignant, 50% taken for training as many as 57 benign and 164 malignant, and 50% for trials as many as 57 benign and 164 malignant. This research used three methods: K-Nearest Neighbor (KNN), Naïve Bayes, and Support Vector Machine (SVM). Then, this research compared the three methods. Inclusion criteria: complete pathology results, no radiotherapy, and chemotherapy. The local hospital ethics committee approved this research. All patients were informed and obtained their verbal consent.The proposed KNN method with binary transformation had the best value with a sensitivity of 96.66% and False Negative 0%, compared to the Naïve Bayes method, which had a sensitivity value of 94.44% and False Negative 4%. In comparison, the SVM method had a sensitivity value of 84.85% and False Negative 4.16 %. Our meta-analysis showed that transformed physical parameters could increase sensitivity and decrease False Negatives. However, these findings must be proven on larger and multiple datasets with different mammography scanners

***Keywords*:** *Benign, Breast cancer, KNN, Malignant, Mammography, Naïve Baye*

1. Introduction

Breast cancer is the disease most feared by women. According to the World Health Organization In 2020, 7.8 million women suffered from breast cancer in the last 5 years[1]. Breast cancer is a significant public health problem due to the increasing death rate due to breast cancer[2]. The second main cause of death in women is breast cancer[3]. The main cause of death in women and the most commonly diagnosed is breast cancer[4], [5]. Breast cancer is the most prevalent in women throughout the world and is also the leading cause of death among women[6], [7]. Breast cancer is the main cause of death in Nigeria[8]. The main cause of death in women worldwide is breast cancer[9].

Breast cancer is the leading contributing factor to cancer-related deaths for women in the age group 20 to 59 years worldwide[4], [10]. Early detection is needed to reduce mortality rate. In many countries, breast cancer is one of the most frequently diagnosed[11]. Early detection remains the mainstay of breast cancer control and better treatment[12]. Early detection with mammography screening can reduce breast cancer morbidity and mortality[11]. Early detection and appropriate treatment can reduce breast cancer mortality[13]. Early detection can be improved by combining biomarkers with screening mammograms[14]. Microcalcifications were found to be a sign of breast cancer. Signs of breast cancer are if microcalcification is found[15]. Microcalcification has different absorption and scattering properties between benign and malignant[16]. The appearance of isolated microcalcifications on postoperative mammography is characteristic of breast cancer recurrence[17]. Microcalcification identification is an early diagnosis of breast cancer and can reduce breast cancer morbidity and mortality[18]. Mammography is able to detect microcalsification that is very small in size. Mammography plays an essential role in breast cancer screening as it can be used to diagnose breast masses and areas of breast calcification early[19]. The main modality used for breast cancer screening is mammography[18]. The most widely used tool for breast screening is mammography[20]–[22]. Breast soft tissue can be seen by mammography[23]. In developed countries, women over

1Department of Physics, Faculty of Mathematics and Natural Sciences, Udayana University, Bali, Indonesia.
Orcid id: 0000-0002-0576-7599
2Department of Informatics, Faculty of Mathematics and Natural Sciences, Udayana University, Bali Indonesia.
Orcid id: 0009-0002-5265-309x
3Radiology Department, Prof. Ngurah Hospital, Bali, Indonesia
Orcid id: 0009-0005-0251-7737
4Radiology Department, Prima Medika Hospital, Bali, Indonesia
** Corresponding Author Email: a.a.ngurahgunawan@unud.ac.id*

40 are encouraged to undergo mammography screening yearly. The mortality rate of breast cancer at the age of 50-60 years can be reduced by 20%-35% by mammography examination[24]. In Spain, women aged 50-69 years are advised to undergo mammography screening every 2 years[25]. The American College of Radiology since 2017 recommended mammography screening starting at age 40, increasing life expectancy by 5-7 years[26]. Before the 2000s, screen-film mammography was used to screen for breast cancer. Since the invention of digital selectors in 2000, digital mammography has been developed for clinical use[27]. In clinical practice, digital mammography is recommended to detect breast tumors[28]. Clinical practice follow-up guidelines updated by the American society starting in 2013. Annual mammographic evaluations should be performed post-treatment and after completion of radiation therapy[29], [30]. The mortality rate of breast cancer can decrease by about 23% if routine mammography examination is carried out among women invited to attend screening[31], [32]. However, in reality the mortality rate from breast cancer is still high. The number of deaths from breast cancer in 2018 was 626,679 people and new cases occurred around 2.09 million people[33]. Cancer is the second leading cause of death in the world after heart disease (8.97 million deaths) And the death toll is likely to occur around 18.63 million people by 2060[3], [34]. In Europe, it is estimated that the incidence of female breast cancer between 2020 and 2040 will increase by 7%, and the mortality rate will increase by 20.6%[35]. In 2021, the American Cancer Society reported that the number of new cases was around 281,550 people and the mortality rate was 43,600 people [36]. In 2022, the American Cancer Society estimates there will be about 1,918,030 new cases and 609,360 deaths in the United States[37], [38]. Mammography has disadvantages, such as: Breast density can affect screening value[39]. Mammography cannot screen for dense breast tissue[40]–[42]. Sensitivity will decrease if there is fibroglandular tissue covering the cancer[43]. Dense breast tissue can increase false negatives and decrease sensitivity[44]–[48]. Mammography sensitivity depends on breast density, the denser the breast, the smaller the sensitivity[19]. Breast cancer detection still uses original data thus far, previous research showed the original data results have particularly low accuracy. Khamis et al in 2017[49] used original data and obtained an accuracy of 83.3%, whereas Ma et al in 2017 [50] used original data and obtained an accuracy of 87.5%. Moreover, Fusco et al in 2020 [51] used original data and obtained an accuracy of 83%. Meanwhile, in this research used original data and obtained an accuracy of 62.89%. There is still limited research that performs data transformation for the purpose of increasing the accuracy of breast cancer detection. Many still do not know how to increase sensitivity and reduce false negatives. In addition, many still do not know that there are ten physical parameters in medical images that can be used to increase sensitivity and reduce false negatives. Besides, many also do not know that data transformation and variable selection input, plays a very important role in increasing sensitivity and reducing false negatives. Therefore, this research aims to create a digital mammography data transformation model to improve the accuracy of benign and malignant detection in breast cancer and compare it with other advanced methods. This research is very important, considering that the number of breast cancers is increasing every year[35]. This research is important because 10%-30% of cases of dense breasts are undetected using mammography[14], [52].

2. Materials and Methods

This research is quantitative research, with a population in the form of secondary data on benign and malignant mammograms that have never been biopsied. Samples were taken randomly of 442 mammograms from the entire population which is divided into 50% mammograms for training and 50% mammograms for trial. The research instrument was a document of the results of an anatomical pathology examination. This research procedure carried out is as in Figure 4.

2.1 Datasets

Before obtaining secondary mammography data from the radiology database at Dokter Sutomo Surabaya Hospital, the researchers took care of ethical clearance first. After obtaining ethical clearance with number 07/Panke.KKE/I/2012 dated January 26, 2012, and No: 1204/UN.14.2/KEP/2017 dated Denpasar, May 17, 2017, the next day starting January 27, 2012 until now, this research started taking secondary mammography data. The stages to obtain secondary mammography data were as follows. First, the research took the patient record data in the anatomical pathology room. Then, the research recorded the patient record number and anatomical pathology results of the operated patients. Further, the research looked for the patient's mammography data based on the patient's record data in the radiology room. This research only took the mammography data with a data record and a definite disease from anatomical pathology examination. The researchers did not have access to information that could identify individual participants during or after data collection. This research took data from Dokter Sutomo Hospital Surabaya and Sanglah Hospital Denpasar in the form of mammogram images taken from the hospital database with complete pathology results from 2010 to 2023. A total of 442 mammograms, consisting of 114 benign and 328 malignant, were taken. 50% for training with 57 benign

and 164 malignant, and 50% for testing with 57 benign and 164 malignant, respectively. Inclusion criteria: complete pathology results, no radiotherapy, and chemotherapy. The local hospital ethics committee approved this research. All patients were informed and obtained their verbal consent. The obtained mammography image was cropped with a suspicious mass of 2 cm x 2 cm. Then, this research calculated ten physical parameters with a distance between pixels of 1 to 10 pixels. Thus, this research obtained 100 variables for each mammography image. Then, the obtained data was made into three models. The first model was untransformed (original) data, the second was binary transformed, and the third was bipolar transformed. Further, this research conducted an ANOVA test to select significant variables to classify benign and malignant. The results can be seen in Table 2 and Table 3. Furthermore, this research used the best data model between the three models as input variables for the Naïve Bayes and SVM methods.

2.2 Mammography

The mammography used in the Dokter Sutomo Hospital dataset was the Kodak brand type Dryview 6800 laser image for printing mammography films and direct view cr 975 for placing mammography films. The mammography tool settings used KV = 30, brightness = - 7, MAS = 25, contrast = -4, latitude = 11, and film size = 18 x 24 cm.

2.3 Image analysis

To calculate the physical parameters obtained from mammographic images from the dataset of Dokter Sutomo Hospital in Surabaya, first, the researchers cropped the suspicious mass with a size of 2x2 cm. Then, it was saved in BMP format and improved image quality with histogram equalization. Furthermore, the research analyzed the image by counting the number of gray level pairs appearing at a distance of 1 pixel, as shown in Figure 1, and gray level pairs at a distance of 2 pixels, as shown in Figure 2.

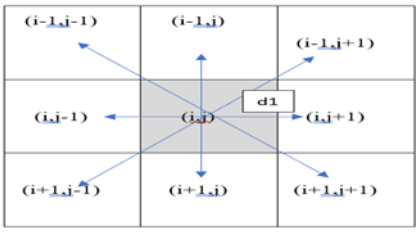

Fig. 1. Gray Level Pairs at a Distance of 1 Pixel

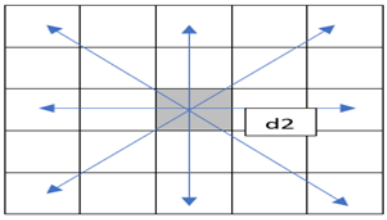

Fig. 2. Gray Level Pairs at a Distance of 2 Pixels

And so on until the gray level pair at distance 10, as shown in Figure 3

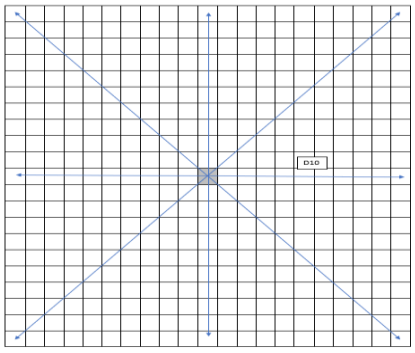

Fig. 3. Gray Level Pairs at a Distance of 10 Pixels

The number of gray level pairs was obtained from H(y1,y1,d1) to H (y255,y255,d10) from the ten distances between pixels. Then, continuing with calculating the entropy, contrast, angular second moment, moment difference inverse, correlation, mean, deviation, entropy of hdiff, angular second moment of hdiff, and mean hdiff using equations (1 to 11): [53], [54]

$$Entrophy = -\sum_{y_q=y_1}^{y_1}\sum_{y_r=y_1}^{y_t} H(y_q, y_r, d)\, log\, H\,(y_q, y_r, d) \quad (1)$$

$$Contrast = \sum_{y_q=y_1}^{y_1}\sum_{y_r=y_1}^{y_t}(y_q - y_r)^2\left[H(y_q, y_r, d)\right]^2 \quad (2)$$

$$Anguler\ Second\ Moment = \sum_{y_q=y_1}^{y_1} \sum_{y_r=y_1}^{y_t} \left[H(y_q, y_r, d)\right]^2 \quad (3)$$

$$\text{Moment Differential Inverse} = \sum_{y_q=y_1}^{y_1} \sum_{y_r=y_1}^{y_t} \left[\frac{H(y_q,y_r,d)}{1+(y_q-y_r)^2}\right] \quad y_r \neq y_q \quad (4)$$

$$Correlation = \frac{\sum_{y_q=y_1}^{y_t} \sum_{y_r=y_1}^{y_t} y_q y_r H(y_q,y_r,d) - \mu H_m(y_q,d) H_m(y_r,d)}{\sigma\left(H_m(y_q,d)\right)\sigma(H_m(y_r,d))} \quad (5)$$

With

$$H_m(y_q, d) = \sum_{y_r=y_1}^{y_t} H(y_q, y_r, d), \qquad H_m(y_r, d) = \sum_{y_q=y_1}^{y_t} H(y_q, y_r, d)$$

$$Mean = \sum_{y_q=y_1}^{y_t} y_q H(y_q, y_r, d) \quad (6)$$

$$Deviation = \sqrt{\sum_{y_q=y_1}^{y_t} \left[y_q - \sum_{y_q=y_1}^{y_t} y_p H_m(y_p, d)\right]^2 H_m(y_q, d)} \quad (7)$$

$$\text{Second} - \text{order histogram} = H_{diff}(i, d) = \sum_{y_q=|y_q-y_r|=i}^{y_t} \sum_{y_r=y_1}^{y_t} H(y_q, y_r, d) \quad (8)$$

$$Entropy\ of\ the\ second - order\ histogram = -\sum_{i=i_1}^{i_t} H_{diff}(i, d)\ log\ H_{diff}\ (i, d) \quad (9)$$

Where EHD is Entropy of the second-order histogram

$$Anguler\ Second\ Moment\ of\ the\ second - order\ histogram = \sum_{i=i_1}^{i_t} \left[H_{diff}(i, d)\right]^2 \quad (10)$$

$$Mean\ of\ the\ second - order\ histogram = \sum_{i=i_1}^{i_t} iH_{diff}(i, d) \quad (11)$$

With yq, yr, and d, respectively, the gray-level value of the first pixel, the gray-level value of the second pixel, and the distance between the first pixel and the second pixel. H(yq, yr, d) is the number of gray-level pairs at a distance d.

2.4 Related Work

2.4.1 Firmness Difference Based Method

Several works in the literature focused on the problem of benign malignant classification using Elastography. This technique utilized the theory that benign and malignant have differences in firmness. The image results obtained would be compared to the suspicious mass with its surroundings. If a suspicious mass has a stiffness that is uneasily deformed compared to its surroundings, it is described as a dark color. If a suspicious mass has a stiffness that is easily deformed compared to its surroundings, it is described as a light color. Malignant masses usually appeared dark and highly contrasted with the surrounding breast tissue during deformation. Benign masses usually appeared lighter and had lower contrast to the surrounding breast tissue during deformation. Benign tumors were loosely attached to the surrounding tissue, whereas malignant tumors were usually characterized by firm desmoplasticity. Research conducted by[55] to differentiate tumor firmness using the Role of supersonic shear wave imaging quantitative Elastography (SSI) by measuring shear wave velocity using a color scale calibrated in kilopascals (kPa). The results obtained for malignant were 134 kPa, while for benign was 50 kPa.

2.4.2 K-Nearest Neighbor-Based Method

Previous studies on sampling classification models or dataset collection applied the K-NN method. Research conducted by [56] optimized K-Nearest Neighbor model was used to find the best hyperparameters for the best K. From the Kaggle dataset, the best k value was 9 with an accuracy of 94.35%.

2.5. The Proposed Method

2.5.1 Data Normalization

This research proposed the input data be normalized first before using the KNN method to improve Accuracy, Sensitivity, Specificity, and Precision. There were three data models used in this research, namely non-normalized data, binary normalized data, and bipolar normalized data. The equation for normalizing data was proposed as follows:

$$\text{Binary Normalization} = 0.8 * (X - Min) / (Max - Min) + 0.1 \quad (12)$$

$$\text{Bipolar Normalization} = 0.8*(X - Min)/(Max - Min) + 0.1 + (X - Min)/(Max - Min) - 1 \quad (13)$$

2.5.2 Classification Algorithm

The proposed model involved using a classification algorithm to classify benign-malignant and comparing its performance with similar data sets. This classification algorithm consisted of K-Nearest Neighbors, Naïve Bayes, and Support Vector Machines.

2.5.3 K-Nearest Neighbor Algorithm

After transforming the data, this research conducted an ANOVA test to select significant variables that could differentiate malignant and benign. This research selected variables whose p-value was smaller than 0.05, as shown in Table 2. Then, the selected variables were used as input variables. The KNN method used the following equation:

$$D = \sqrt{\sum_{i=1}^{i=n} \left(T_i - U_i\right)^2} \quad (14)$$

2.5.4 Naïve Bayes Method

This research also used the Naïve Bayes method to classify benign and malignant. The results are shown in Figure 5, Table 5, Table 6 and Table 7.

2.5.5 Support Vector Machine Method

This research also used the SVM method to classify benign and malignant. The results are shown in Table 5, Table 6, and Table 7.

2.6 Evaluation Method

This research used an evaluation method to obtain the performance of each stage of the recommended model, using the confusion matrix Table 1 and equations (15 to 20):

Table 1. Confusion Matrix

		Actually	
		Malignant	Benign
Predictions	Malignant	T.P	F.P
	Benign	FN	TN

Accuracy = (TP+TN) / (TP+TN+FP+FN) (15)

Sensitivity (Recall) = TP/(TP+FN) (16)

Specificity = TN /(FP+TN) (17)

Precision = TP / (TP+FP) (18)

TPR = TP / (TP+FN) (19)

FPR = FP / (FP+TN) (20)

The results are shown in Table 4 to Table 6, and Figure 6.

2.7 Research Flowchart

The sequence of research steps carried out is shown in Figure 4.

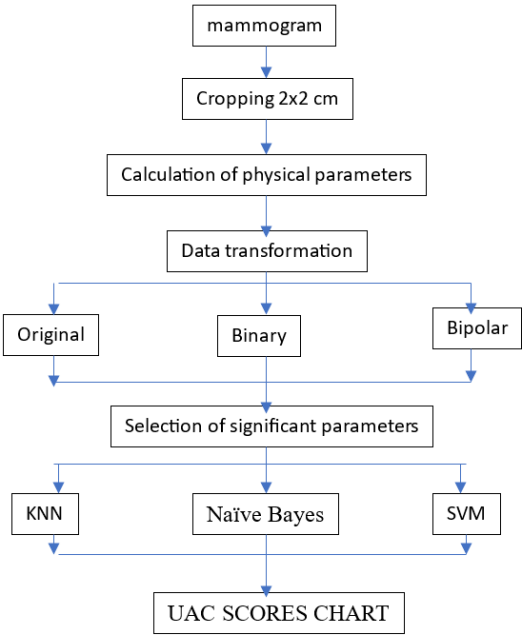

Fig. 4. Research Flowchart

3. Results and Discussions

3.1 Results

This research first conducted an ANOVA test to select genuinely significant variables for classifying benign and malignant to increase the sensitivity value and reduce False Negative. The results are shown in Table 2 and Table 3. Then, this research entered the significant variables using various methods, such as KNN, Naïve Bayes, and SVM. Further, the research calculated TP, FP, TN, FN, accuracy, sensitivity, specificity, precision, and percent positive phase, as shown in Table 4, Table 5, Table 6, and Table 7. This research calculated True Positive Rate (TPR) and False Positive Rate (FPR) values by making an ROC curve to evaluate the best model. The best result is a graph close to 1, as seen in Figure 6, Figure 7, and Figure 8.

Table 2. ANOVA Test Results

No	parameter	Type	Transformation											
			Original				Binary				Bipolar			
			Means	Median	Variances	P	Means	Median	Variances	P	Means	Median	Variances	P
1	Entry2	B	3.623918	3.672390	0.027	0.33	0.681884	0.708737	0.017	0.004	0.409240	0.469659	0.084	0.01
		m	3.648060	3.681460	0.026		0.620683	0.676851	0.043		0.271538	0.397315	0.217	
2	Entry10	B	3.655704	3.688220	0.023	0.21	0.749778	7813758	0.015	0,00	0.430735	0.514450	0.106	0,00
		m	3.684946	3.723875	0.023		0.328108	2927412	0.029		-0.38675	-0.46633	0.149	
3	Contr1	B	316.332798	279.885670	31241	0.18	0.205000	0.169125	0.017	0,00	-0.66374	-0.74446	0.085	0,00
		m	384.860324	282.776475	136639		0.313783	0.266059	0.034		-0.41898	-0.52636	0.172	
4	Contr10	B	1381.88089	1091.65350	813465	0.11	0.647764	0.549663	0.161	0,00	-0.43138	-0.51961	0.130	0.03
		m	1644.74618	1365.60101	130434		0.137083	0.112864	0.011		-0.81656	-0.87105	0.058	
5	MA7	B	0.0004844	0.0002600	0,000	0.69	0.108358	0.101552	0.004	0.00	-0.88119	-0.89650	0.020	0.01
		m	0.0007224	0.0002400	0,000		0.137039	0.115286	0.011		-0.81666	-0.86560	0.056	
6	MA10	B	0.0003630	0.0002600	0,000	0.58	0.104367	0.100677	0.001	0,00	-0.88057	-0.8962	0.020	0,00
		m	0.0006111	0.0002300	0,000		0.314685	0.297516	0.025		-0.41695	-0.45558	0.126	
7	MD1	B	0.0550107	0.0533000	0,000	0.64	0.409749	0.384755	0.030	0,00	-0.2030	-0.25929	0.150	0,00
		m	0.0540364	0.0521550	0,000		0.320979	0.282600	0.028		-0.40279	-0.48914	0.140	

8	MD10	B	0.0288426	0.0269400	0,000	0.24	0.052386	0.026469	0.015	0,00	-0.30410	-0.37642	0.114	0.23
		m	0.0271359	0.0251850	0,000		0.484554	0.480614	0.029		-0.03475	-0.04361	0.149	
9	Dev1	B	29.2661230	27.9802200	103,00	0.02	0.400318	0.381038	0.026	0.52	-0.22428	-0.26766	0.134	0.50
		m	33.1841234	31.7979850	136,94		0.414223	0.397494	0.029		-0.19299	-0.23063	0.145	
10	Dev2	B	28.5529330	27.6270500	87,491	0.01	0.397482	0.379023	0.026	0.50	-0.23066	-0.27219	0.129	0.24
		m	32.9303829	31.5994350	132,76		0.424049	3970499	0.033		-0.17088	-0.23163	0.165	
11	Dev3	B	28.9365904	27.4532600	98,362	0.03	0.394299	0.379741	0.025	0.24	-0.23782	-0.27058	0.129	0.26
		m	32.6469437	31.5952250	132,68		0.424048	0.398218	0.033		-0.17089	-0.22900	0.165	
12	Dev4	B	28.8757035	27.4605400	97,705	0.02	0.395873	0.381646	0.025	0.26	-0.23428	-0.26629	0.129	0.29
		m	32.7302653	31.7013000	132,99		0.421530	0.395159	0.032		-0.17655	-0.23589	0.164	
13	Dev5	B	28.7876132	27.3374300	98063	0.02	0.395095	0.381000	0.025	0.29	-0.23603	-0.26774	0.129	0.30
		m	32.6556002	31.6354850	132,95		0.419446	0.392636	0.032		-0.18124	-0.24156	0.163	
14	Dev6	B	28.7310381	27.2506300	98,074	0.02	0.393566	0.380157	0.025	0.30	-0.23947	-0.26964	0.128	0,00
		m	32.5526713	31.5815950	132,59		0.416203	0.389225	0.032		-0.18854	-0.24924	0.161	
15	Dev7	B	28.6428042	27.1467300	98.102	0.01	0.311570	0.299941	0.015	0,00	-0.42396	-0.45013	0.075	0.38
		m	32.9683803	31.7754500	155,04		0.414723	0.386952	0.032		-0.19187	-0.25435	0.161	

16	Dev8	B	28.6340682	27.0875000	98,433	0.02	0.392578	0.377886	0.025	0.38	-0.24169	-0.27475	0.128	0.40
		m	32.4748835	31.4074550	133,24		0.413016	0.385135	0.032		-0.19571	-0.25844	160	
17	Dev9	B	28.6236523	27.0638900	98,930	0.02	0.392246	0.379386	0.025	0.40	-0.24244	-0.27138	0.127	0.44
		m	32.4376877	31.5001500	133,00		0.411298	0.383720	0.031		-0.19957	-0.26162	0.159	
18	Dev10	B	28.6039161	27.0547300	99,289	0.02	0.390189	0.379838	0.026	0.40	-0.24239	-0.26527	0.126	0.00
		m	32.4558556	31.7116850	132,83		0.560291	0.589023	0.034		0.135656	0.200302	0.173	
19	entryHd1	B	1.5199742	1.5367100	0.012	0.65	0.478615	0.484046	0.030	0.00	-0.04811	-0.03589	0.150	0.28
		m	1.5274054	1.5307500	0.011		0.532586	0.552104	0.043		0.073320	0.1172	0.215	
20	entryHd10	B	1.8235537	1.8287500	0.022	0.13	1.015179	1.025284	0.059	0,00	0.173836	0.189415	0.139	0,00
		m	1.8578421	1.8640650	0.022		0.159448	0.106797	0.036		-0.76624	-0.88470	0.184	
21	MaHd1	B	0.1416254	0.0362900	0.146	0.00	0.251599	0.232538	0.009	0,00	-0.55890	-0.60178	0.046	0,00
		m	0.0378843	0.0357250	0,000		0.155258	0.105064	0.033		-0.77566	-0.88860	0.168	
22	MaHd2	B	0.1385170	0.0290100	0.158	0.00	0.246222	0.233054	0.008	0,00	-0.57099	-0.60062	0.041	0,00
		m	0.0304798	0.0290650	0,000		0.153280	0.104490	0.032		-0.78011	-0.88989	0.160	
23	MaHd3	B	0.1371625	0.0255800	0.166	0.00	0.242939	0.224892	0.008	0,00	-0.57838	-0.61899	0.040	0,00
		m	0.0267680	0.0249400	0,000		0.152229	1042226	0.031		-0.78248	-0.89049	0.156	

24	MaHd4	B	0.1366188	0.0236100	0.171	0.00	0.243453	0.224677	0.008	0,00	-0.57722	-0.61947	0.042	0,00
		m	0.0243793	0.0226150	0,000		0.151415	0.103972	0.030		-0.78431	-0.89106	0.153	
25	MaHd5	B	0.1361481	0.0220900	0.175	0.00	0.245800	0.226554	0.009	0,00	-0.57194	-0.61525	0.046	0,00
		m	0.0226401	0.0209650	0,000		0.150846	0.103647	0.030		-0.78559	-0.89179	0.152	
26	MaHd6	B	0.1357433	0.0207200	0.178	0.00	0.248402	0.228608	0.010	0,00	-0.56609	-0.61063	0.049	0,00
		m	0.0213110	0.0197050			0.150532	0.103348	0.030		-0.78630	-0.89246	0.151	
27	MaHd7	B	0.1355581	0.0195700	0.181	0.00	0.252410	0.227456	0.011	0,00	-0.55707	-0.61322	0.054	0,00
		m	0.0203074	0.0184250	0,000		0.150214	0.103274	0.030		-0.78701	-0.89263	0.151	
28	MaHd8	B	0.1353756	0.0191000	0.183	0.00	0.258499	0.232491	0.012	0,00	-0.54337	-0.60189	0.060	0,00
		m	0.0194421	0.0176400	0,000		0.149871	0.103068	0.030		-0.78779	-0.89309	0.150	
29	MaHd9	B	0.1351939	0.0183500	0.184	0.00	0.266989	0.239381	0.014	0,00	-0.52427	-0.58639	0.068	0,00
		m	0.0187470	0.0170050	0,000		0.149574	0.102885	0.029		-0.78845	-0.89350	0.149	
30	MaHd10	B	0.1349186	0.0177500	0.186	0.00	0.076938	0.063764	0.004	0.05	-0.50473	-0.56350	0.079	0.05
		m	0.0180974	0.0166050	0,000		0.467607	0.467859	0.035		-0.07288	-0.07231	0.175	

31	mean Hd4	B	20.4479156	19.7990100	37,055	0.35	0.407286	0.391188	0.023	0.03	-0.20860	-0.24482	0.118	0.03
		m	21.2548643	20.5856350	40,282		0.463994	0.434334	0.046		-0.08101	-0.14774	0.232	
32	mean Hd10	B	27.0714639	25.7139600	76,320	0.13	1.250394	1.217572	0.253	0.00	-0.15003	-0.17470	0.143	0.00
		m	29.2466852	28.6197800	92,262		0.616221	0.648187	0.040		0.261498	0.333421	0.200	

Table 3. Significant Data to Differentiate Benign and Malignant

Data type	Significant Physical Parameters
Original	Dev1, Dev2, Dev3, Dev4, Dev5, Dev6, Dev7, Dev8, Dev9, Dev10, MaHd1, MaHd2, MaHd3, MaHd4, MaHd5, MaHd6, MaHd7, MaHd8, MaHd9, and MaHd10
Binary Transformation	Entr10, Contr10, MA10, MD10, EntrHd10, MaHd10, and MeanHd10
Bipolar Transformation	Entr2, Contr1, MA7, MD1, EntrHd1, MaHd1 and MeanHd4

Table 4. TP, FP, TN and FN Calculation Results Using the KNN Method

Data Type	T.P	F.P	FN	TN
Original	22	47	35	117
Binary Transformation	57	2	0	162
Bipolar Transformation	48	21	9	143

Table 5. Performance comparison with KNN, Naïve Bayes, and SVM methods with Original transformation data

Metode	accuracy	Sensitivity	Specificity	Precision	False Negatif
KNN	62.90 %	38.60 %	71.34%	31.88 %	62.90 %
Naïve Bayes	72.00 %	72.00 %	0.00 %	100 %	28.00 %
SVM	97.76 %	96.49 %	91.46 %	79.71 %	0.91%

Table 6. Performance comparison with KNN, Naïve Bayes, and SVM methods with binary transformation data

Metode	accuracy	Sensitivity	Specificity	Precision	False Negatif
KNN	99.09 %	96.66 %	98.78 %	96.61 %	0.00 %
Naïve Bayes	92.00 %	94.44 %	85.71 %	94.44 %	4.00 %
SVM	95.83 %	84.85 %	100 %	100 %	4.16%

Table 7. Performance comparison with KNN, Naïve Bayes, and SVM methods with bipolar transformation data

Metode	accuracy	Sensitivity	Specificity	Precision	False Negatif
KNN	86.43 %	84.21%	87.19%	69.57 %	86.43 %
Naïve Bayes	84.00 %	81.82 %	100 %	100 %	16.00 %
SVM	91.85 %	91.23 %	92.07 %	80 %	2.26%

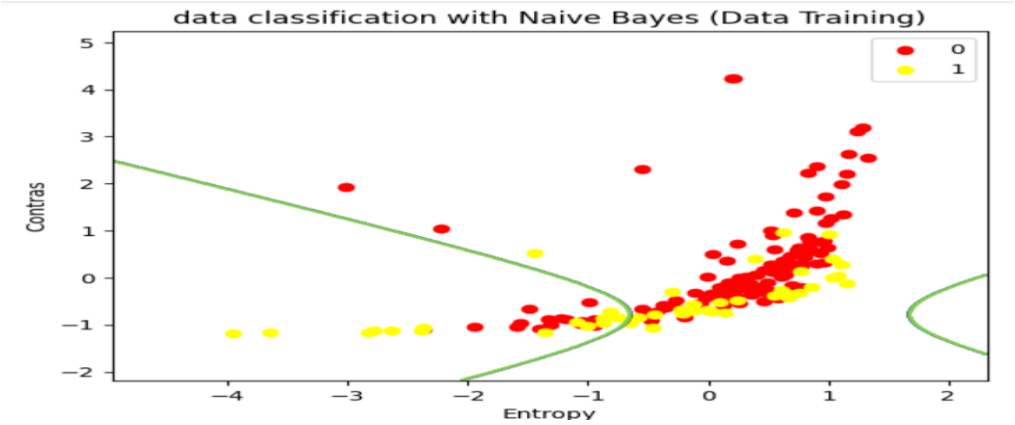

Fig. 5. Classification Graph Using Naïve Bayes.

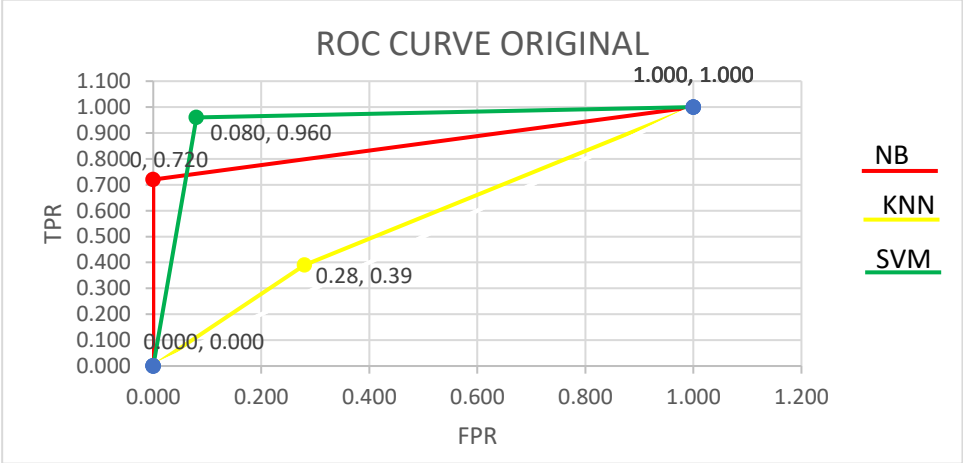

Fig. 6. Roc curve original

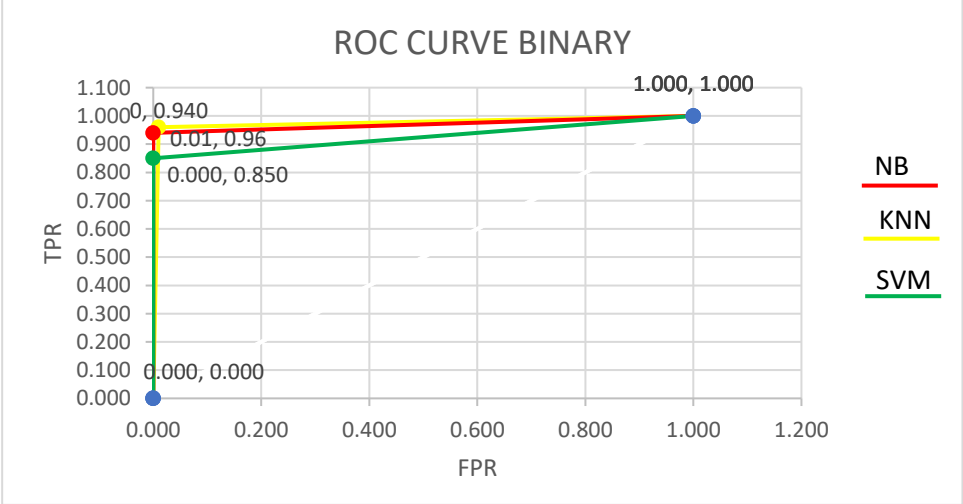

Fig. 7. Roc curve binary

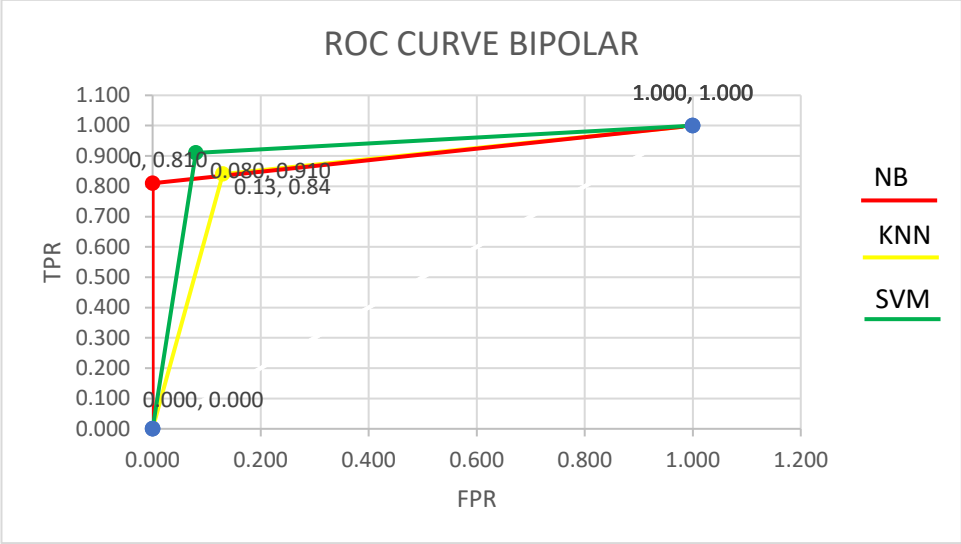

Fig. 8. Roc curve bipolar

3.2 Discussions

Not all physical parameters or features on a mammogram could differentiate between benign and malignant. Thus, an ANOVA test was needed to find significant parameters to differentiate between benign and malignant. Using original measurement data could not increase sensitivity and reduce false negatives, so data transformations such as binary and bipolar transformations were needed. Binary transformation was able to increase sensitivity and reduce false negatives. By using the same data and the same data transformation (binary transformation), the KNN method performance had the best performance, followed by the Naïve Bayes and SVM methods. The results of the literature study showed that data from the American Cancer Society Statistics in 2021 revealed 281,550 new cases of invasive breast cancer diagnosed among American women, ultimately resulting to the death of about 43,600[36]. According to the American Cancer Society, there will be an estimated 1,918,030 new cancer cases and 609,360 deaths in the United States by 2022[38]. In Europe, it is estimated that the incidence of female breast cancer between 2020 and 2040 will increase by 7%, and the mortality rate will increase by 20.6%[35]. To overcome this case, the government should recommend routine examinations for women aged 40 years and above and conduct breast cancer screening. In addition, medical professionals must intervene in the community to socialize how to detect breast cancer early.

The results of previous literature studies showed that 10% -30% of cases of dense breasts were undetected using mammography[14], [52], and dense breast tissue could reduce the sensitivity of mammography[40]–[42] Besides, mammography had False Negative ranging from 10-15% of cases[57]. Literature studies showed that the comparison of accuracy, sensitivity, and specificity values with other sophisticated methods can be seen in Table 8.

Table 8. Summary of Performance Comparison Results of the Method Classification with the Proposed Data Normalization and Other Advanced Methods.

Study	years	method	accuracy	Sensitivity	Specificity
Awad[55]	2013	(SSI)	Not mentioned	Not mentioned	Not mentioned
Khamis et al.[49]	2017	sonoelastographic strain ratio	83.3 %	100%	88 %
Aleem et al.[58]	2018	diffusion-weighted magnetic resonance imaging	Not mentioned	89 %	94.7 %
Abraham et al.[59]	2018	contrast enhanced MRI, diffusion-weighted MRI and proton-MRS	Not mentioned	87.5 %	93.3 %

Mus et al.[60]	2017	Time to enhancement derived from ultrafast breast MRI	Not mentioned	Not mentioned	Not mentioned
Li et al.[61]	2018	contrast-enhanced sonography	Not mentioned	Not mentioned	Not mentioned
Liu et al.[62]	2019	Histogram analysis of diffusion kurtosis imaging	Not mentioned	85.71 %	93.33 %
Forte et al.[16]	2020	grating interferometry-based mammography	Not mentioned	Not mentioned	Not mentioned
Ma et al.[7]	2021	diffusion weighted imaging and intravoxel incoherent motion	Not mentioned	87 %	80%
Ma et al.[63]	2021	deep learning neural architecture	Not mentioned	70 %	90%
Wang et al.[64]	2022	contrast-enhanced ultrasound	Not mentioned	87 %	79 %
Ma et al.[50]	2017	contrast-enhanced MRI	87.5 %	88.8 %	85.5 %
Kanao et al.[65]	2018	Value of T2 weighted and diffusion weighted MR images	Not mentioned	Not mentioned	Not mentioned
Fusco et al.[51]	2020	magnetic resonance imaging and diffusion weighted MRI imaging	83 %	88 %	77 %
Zhang et al.[66]	2022	Quantitative transport mapping	Not mentioned	Not mentioned	Not mentioned
Han et al.[67]	2021	Genome-wide nucleosome profiles of plasma cell-free DNA	Not mentioned	70.8 %	76.5 %
Choi et al.[68]	2017	Comparison of 3D and 2D shear-wave elastography	Not mentioned	81.9 %	84 %
Wubulihasimu et al.[69]	2018	contrast-enhanced ultrasound	Not mentioned	87 %	80%
Wasan et al.[70]	2019	digital breast tomosynthesis	97 %	Not mentioned	Not mentioned
Chen et al.[71]	2020	Multiple b-valued diffusion-weighted imaging	Not mentioned	97.7 %	55 %
Yu et al.[72]	2021	shear-wave elastography	Not mentioned	98.7 %	62.1 %
Eroglu et al.[23]	2021	Convolutional Neural Networks	95.6 %	95.6 %	97.78
Mosapour et al.[73]	2020	Expression levels of VLDL receptor and VLDL-c levels	Not mentioned	Not mentioned	Not mentioned
Shia[74]	2021	support vector machine	Not mentioned	94.34 %	93.22 %

Davoudi et al.[75]	2014	Doppler Sonography	Not mentioned	Not mentioned	Not mentioned
Handayani et al.[76]	2021	K-Nearest Neighbor	97 %	Not mentioned	Not mentioned
		Naïve Bayes	94 %	Not mentioned	Not mentioned
Assegie T.[56]	2021	Optimized K-Nearest Neighbor	94.35 %	Not mentioned	Not mentioned
Gunawan et al.	2023	K-Nearest Neighbor with Binary Transformed physics parameter data	99.095 %	96.661 %	98.781 %
Gunawan et al.	2023	K-Nearest Neighbor with Bipolar Transformed physics parameter data	86.425 %	84.210%	87.195%
Gunawan et al.	2023	K-Nearest Neighbor with non- Transformed physical parameter data	62.896 %	38.596 %	71.341%
Gunawan El al.	2023	Naïve Bayes	92,000 %	94.444 %	85.714 %
Gunawan et al	2023	Support Vector Machine	95.833 %	84.848 %	100%

Our proposed method is not much different from other state-of-the-art methods.

4. Conclusion

Breast cancer detection still uses original data thus far, previous research showed the original data results have particularly low accuracy. Khamis et al in 2017[49] used original data and obtained an accuracy of 83.3%, whereas Ma et al in 2017 [50] used original data and obtained an accuracy of 87.5%. Moreover, Fusco et al in 2020 [51] used original data and obtained an accuracy of 83%. Meanwhile, in this research used original data and obtained an accuracy of 62.89%. This research proposed using transformed data to increase accuracy. In this research, binary and bipolar transformation data were used. The result obtained in this research were able to increase accuracy to 99.09% for binary transformed data and 86.43% for bipolar transformed data, which previously was 62.9% for original data. The weakness of using original data is the low accuracy value and high false negatives. Therefore, this research used data transformation techniques. The data transformation is carried out using 3 data models, namely original, bipolar, and binary model. From the result of this research analysis, it was found that the original data was not significant for detecting benign and malignant breast cancer. Meanwhile, binary data showed significant result. Besides that. Bipolar data showed significant result, although not as good as binary data, however more advanced than the original data. Based on data transformation in this analysis, it was found that binary data was more significant in increasing the accuracy of cancer detection. The contribution of this research is that with the data transformation method, benign and malignant cancer types are more accurately detected.

Author Contributions

Gunawan: project design; data analysis; script writing. supardi: Manuscript editing. Wendri: Manuscript editing. Frady: program maker. Baskara: manuscript editor. Patriawan: data collection; data analysis; Dharmawan: data collection; data analysis.

Acknowledgments

The authors would like to thank Suarsana for the financial resources needed to carry out this research. We also thank Triyono for their help in preparing the testing protocol and obtaining the data.

Conflict Of Interest Statement

For all authors, there are no conflicts of interest.

Orcid

Anak Agung Ngurah Gunawan

https://orcid.org/0000-0002-0576-7599

List of Abbreviations

KKN : K-Nearest Neighbor

NB : Naïve Bayes

SVM : Support Vector Machine.

X : Original data pixel intensity value

Min : Minimum value of all data`

Max : Maximum value of all data.

D : Euclidean Distance

Ti : Training Data

Ui : Data to be Tested

TP : Predicted to be malignant, actually malignant

FP : Predicted to be malignant but actually benign

FN : Predicted to be benign actually malignant

TN : Predicted to be benign actually benign

Dev1 : Deviation at a distance of 1 pixel

Dev10 : Deviation at a distance of 10 pixels

MaHd1 : Angular second moment of hdiff at a distance of 1 pixel

MaHd10 : Angular second moment of hdiff at a distance of 10 pixels

Entr10 : Entropy at a distance of 10 pixels

Contr10 : Contrast at a distance of 10 pixels

MA10 : Angular second moment at a distance of 10 pixels

MD10 : Inverse difference moment at a distance of 10 pixels

EntrHd10 : Entropy of hdiff at a distance of 10 pixels

MeanHd10 : Mean hdiff at a distance of 10 pixels

Entry2 : Entropy at a distance of 2 pixels

Contr1 : Contrast at a distance of 1 pixel

MA7 : Angular second moment at a distance of 7 pixels

MD1 : Inverse difference moment at a distance of 1 pixel

EntrHd1 : Entropy of hdiff at a distance of 1 pixel

MeanHd4 : Mean hdiff at a distance of 4 pixels

ROC : Receiver Operating Characteristic

TPR : True Positive Rate

FPR : False Positive Rate

References

[1] S. Di Maria, S. Vedantham, and P. Vaz, 'X-ray dosimetry in breast cancer screening: 2D and 3D mammography', *Eur. J. Radiol.*, vol. 151, Jun. 2022, doi: 10.1016/j.ejrad.2022.110278.

[2] E. Mirbagheri, M. Shafiee, M. Shanbezadeh, and H. Kazemi-Arpanahi, 'Developing the required data set for the integration of breast cancer registry systems in Iran', *Informatics Med. Unlocked*, vol. 32, Jan. 2022, doi: 10.1016/j.imu.2022.101011.

[3] M. R. Mogal, M. R. Mahmod, S. A. Sompa, A. Junayed, M. Z. Abedin, and M. A. Sikder, 'Association between ankyrin 2 gene and breast cancer progression: A preliminary computational

assessment using the database approach', *Informatics Med. Unlocked*, vol. 25, Jan. 2021, doi: 10.1016/j.imu.2021.100663.

[4] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[5] J. J. Gommers *et al.*, 'Breast magnetic resonance imaging as a problem solving tool in women recalled at biennial screening mammography: A population-based study in the Netherlands', *Breast*, vol. 60, pp. 279–286, Dec. 2021, doi: 10.1016/j.breast.2021.11.014.

[6] A. Jemal, F. Bray, and J. Ferlay, 'Global Cancer Statistics: 2011', *CA Cancer J Clin*, vol. 49, no. 2, pp. 1,33-64, 1999, doi: 10.3322/caac.20107.Available.

[7] W. Ma, J. Mao, T. Wang, Y. Huang, and Z. H. Zhao, 'Distinguishing between benign and malignant breast lesions using diffusion weighted imaging and intravoxel incoherent motion: A systematic review and meta-analysis', *Eur. J. Radiol.*, vol. 141, Aug. 2021, doi: 10.1016/j.ejrad.2021.109809.

[8] A. D. Omisore *et al.*, 'Practice, Perceptions, and Prospects of Mammography Screening in Nigeria: Insights From a National Survey of Female Health Workers', *Clin. Breast Cancer*, vol. 22, no. 5, pp. 462–472, 2022, doi: 10.1016/j.clbc.2022.02.009.

[9] M. M. Rahman, M. S. I. Khan, and H. M. H. Babu, 'BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer', *Array*, vol. 16, Dec. 2022, doi: 10.1016/j.array.2022.100256.

[10] S. I. Khan, A. Shahrior, R. Karim, M. Hasan, and A. Rahman, 'MultiNet: A deep neural network approach for detecting breast cancer through multi-scale feature fusion', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6217–6228, Sep. 2022, doi: 10.1016/j.jksuci.2021.08.004.

[11] J. R. C. Lameijer *et al.*, 'Trends in delayed breast cancer diagnosis after recall at screening mammography', *Eur. J. Radiol.*, vol. 136, Mar. 2021, doi: 10.1016/j.ejrad.2020.109517.

[12] E. M. F. El Houby and N. I. R. Yassin, 'Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks', *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, doi: 10.1016/j.bspc.2021.102954.

[13] A. D. Nelson and S. Krishna, 'An effective approach for the nuclei segmentation from breast histopathological images using star-convex polygon', *Procedia Comput. Sci.*, vol. 218, pp. 1778–1790, 2023, doi: 10.1016/j.procs.2023.01.156.

[14] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, 'Comparative analysis of breast cancer detection using machine learning and biosensors', *Intelligent Medicine*, vol. 2, no. 2. Elsevier B.V., pp. 69–81, May 01, 2022. doi: 10.1016/j.imed.2021.08.004.

[15] Y. J. Yeow, C. C. Yu, Y. C. Cheung, and S. C. Chen, 'A cohort study of mammography-guided vacuum-assisted breast biopsy in patients with compressed thin breasts (≦ 3 cm)', *Asian J. Surg.*, 2023, doi: 10.1016/j.asjsur.2023.04.074.

[16] S. Forte *et al.*, 'Can grating interferometry-based mammography discriminate benign from malignant microcalcifications in fresh biopsy samples?', *Eur. J. Radiol.*, vol. 129, Aug. 2020, doi: 10.1016/j.ejrad.2020.109077.

[17] R. M. Heaney, L. Sweeney, F. Flanagan, A. O'Brien, and C. Smith, 'Ipsilateral microcalcifications after breast-conserving surgery: is it possible to differentiate benign from malignant calcifications?', *Clin. Radiol.*, vol. 77, no. 3, pp. 216–223, Mar. 2022, doi: 10.1016/j.crad.2021.12.005.

[18] Q. Lin *et al.*, 'Artificial intelligence-based diagnosis of breast cancer by mammography microcalcification', *Fundam. Res.*, Jun. 2023, doi: 10.1016/j.fmre.2023.04.018.

[19] M. L. Huang and T. Y. Lin, 'Considering breast density for the classification of benign and malignant mammograms', *Biomed. Signal Process. Control*, vol. 67, May 2021, doi: 10.1016/j.bspc.2021.102564.

[20] S. W. Duffy *et al.*, 'The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish Counties: A collaborative evaluation', *Cancer*, vol. 95, no. 3, pp. 458–469, 2002, doi: 10.1002/cncr.10765.

[21] L. Nyström, I. Andersson, N. Bjurstam, J. Frisell, B. Nordenskjöld, and L. E. Rutqvist, 'Long-term effects of mammography screening: Updated overview of the Swedish randomised trials', *Lancet*, vol. 359, no. 9310, pp. 909–919, 2002, doi: 10.1016/S0140-6736(02)08020-0.

[22] J. H. Yoon, K. Han, H. J. Suh, J. H. Youk, S. E. Lee, and E. K. Kim, 'Artificial intelligence-based computer-assisted detection/diagnosis (AI-CAD) for screening mammography: Outcomes of AI-CAD in the mammographic interpretation workflow', *Eur. J. Radiol. Open*, vol. 11, Dec. 2023, doi: 10.1016/j.ejro.2023.100509.

[23] Y. Eroğlu, M. Yildirim, and A. Çinar, 'Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR', *Comput. Biol. Med.*, vol. 133, Jun. 2021, doi: 10.1016/j.compbiomed.2021.104407.

[24] R. Verma, G. K. Jain, and A. Chougule, 'Evaluation and analysis of entrance surface air kerma of the thyroid gland during breast cancer screening mammography', *Radiat. Med. Prot.*, vol. 2, no. 2, pp. 67–71, Jun. 2021, doi: 10.1016/j.radmp.2021.03.001.

[25] J. Louro *et al.*, 'Differences in breast cancer risk after benign breast disease by type of screening diagnosis', *Breast*, vol. 54, pp. 343–348, Dec. 2020, doi: 10.1016/j.breast.2020.09.005.

[26] S. M. Oshima, S. D. Tait, L. Fish, R. A. Greenup, and L. J. Grimm, 'Primary care provider perspectives on screening mammography in older women: A qualitative study', *Prev. Med. Reports*, vol. 22, Jun. 2021, doi: 10.1016/j.pmedr.2021.101380.

[27] I. Sechopoulos, J. Teuwen, and R. Mann, 'Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art', *Seminars in Cancer Biology*, vol. 72. Academic Press, pp. 214–225, Jul. 01, 2021. doi: 10.1016/j.semcancer.2020.06.002.

[28] Z. Li *et al.*, 'Diagnostic Performance of Mammographic Texture Analysis in the Differential Diagnosis of Benign and Malignant Breast Tumors', *Clin. Breast Cancer*, vol. 18, no. 4, pp. e621–e627, Aug. 2018, doi: 10.1016/j.clbc.2017.11.004.

[29] J. L. Khatcheressian *et al.*, 'Breast cancer follow-up and management after primary treatment: American society of clinical oncology clinical practice guideline update', *J. Clin. Oncol.*, vol. 31, no. 7, pp. 961–965, 2013, doi: 10.1200/JCO.2012.45.9859.

[30] W. Taher, J. Benson, and S. Leinster, 'The effectiveness of mammography surveillance after treatment of primary breast cancer: A single centre retrospective cohort study', *Ann. Med. Surg.*, vol. 65, May 2021, doi: 10.1016/j.amsu.2021.102272.

[31] B. L. Secretan, D. Ph, C. Scoccianti, D. Ph, D. Loomis, and D. Ph, 'Spe ci a l R e p or t Breast-Cancer Screening — Viewpoint of the IARC Working Group', pp. 2353–2358, 2015.

[32] J. Wang *et al.*, 'Cost-effectiveness of abbreviated-protocol MRI screening for women with mammographically dense breasts in a national breast cancer screening program', *Breast*, vol. 61, pp. 58–65, Feb. 2022, doi: 10.1016/j.breast.2021.12.004.

[33] F. Rahman *et al.*, 'A multi-omics approach to reveal the key evidence of GDF10 as a novel therapeutic biomarker for breast cancer', *Informatics Med. Unlocked*, vol. 21, Jan. 2020, doi: 10.1016/j.imu.2020.100463.

[34] C. Mattiuzzi and G. Lippi, 'Current cancer epidemiology', *J. Epidemiol. Glob. Health*, vol. 9, no. 4, pp. 217–222, 2019, doi: 10.2991/jegh.k.191008.001.

[35] T. Sterlingova, E. Nylander, L. Almqvist, and B. Møller Christensen, 'Factors affecting women's participation in mammography screening in Nordic countries: A systematic review', *Radiography*, vol. 29, no. 5. W.B. Saunders Ltd, pp. 878–885, Aug. 01, 2023. doi: 10.1016/j.radi.2023.06.010.

[36] T. Zheng *et al.*, 'Deep learning-enabled fully automated pipeline system for segmentation and classification of single-mass breast lesions using contrast-enhanced mammography: a prospective, multicentre study', 2023. [Online]. Available: www.itksnap.org

[37] R. L. Siegel, K. D. Miller, and A. Jemal, 'Cancer statistics, 2020', *CA. Cancer J. Clin.*, vol. 70, no. 1, 2020, doi: 10.3322/caac.21590.

[38] K. Loizidou, R. Elia, and C. Pitris, 'Computer-aided breast cancer detection and classification in mammography: A comprehensive review', *Computers in Biology and Medicine*, vol. 153. Elsevier Ltd, Feb. 01, 2023. doi: 10.1016/j.compbiomed.2023.106554.

[39] R. W. Woods, L. R. Salkowski, M. Elezaby, E. S. Burnside, R. M. Strigel, and A. M. Fowler, 'Image-based screening for men at high risk for breast cancer: Benefits and drawbacks', *Clin. Imaging*, vol. 60, no. 1, pp. 84–89, 2020, doi: 10.1016/j.clinimag.2019.11.005.

[40] M. Le Boulc'h *et al.*, 'Comparison of breast density assessment between human eye and automated software on digital and synthetic mammography: Impact on breast cancer risk', *Diagn. Interv. Imaging*, vol. 101, no. 12, pp. 811–819, Dec. 2020, doi: 10.1016/j.diii.2020.07.004.

[41] I. Saarenmaa *et al.*, 'The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultasonography', *Breast Cancer Res. Treat.*, vol. 67, no. 2, pp. 117–123, 2001, doi: 10.1023/A:1010627527026.

[42] V. Tesic, B. Kolaric, A. Znaor, S. K. Kuna, and B. Brkljacic, 'Mammographic density and estimation of breast cancer risk in intermediate risk population', *Breast J.*, vol. 19, no. 1, pp. 71–78, 2013, doi: 10.1111/tbj.12051.

[43] B. Malik, E. Iuanow, and J. Klock, 'An Exploratory Multi-reader, Multi-case Study Comparing Transmission Ultrasound to Mammography on Recall Rates and Detection Rates for Breast Cancer Lesions', *Acad. Radiol.*, vol. 29, pp. S10–S18, Jan. 2022, doi: 10.1016/j.acra.2020.11.011.

[44] V. P. Jackson, R. E. Hendrick, S. A. Feig, and D. B. Kopans, 'Imaging of the radiographically dense breast', *Radiology*, vol. 188, no. 2, pp. 297–301, 1993, doi: 10.1148/radiology.188.2.8327668.

[45] C. H. Van Gils, J. D. M. Otten, A. L. M. Verbeek, J. H. C. L. Hendriks, and R. Holland, 'Effect of

mammographic breast density on breast cancer screening performance: A study in Nijmegen, the Netherlands', *J. Epidemiol. Community Health*, vol. 52, no. 4, pp. 267–271, 1998, doi: 10.1136/jech.52.4.267.

[46] E. D. Pisano *et al.*, 'Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in DMIST', *Radiology*, vol. 246, no. 2, pp. 376–383, 2008, doi: 10.1148/radiol.2461070200.

[47] H. D. Nelson, E. S. O'Meara, K. Kerlikowske, S. Balch, and D. Miglioretti, 'Factors associated with rates of false-positive and false-negative results from digital mammography screening: An analysis of registry data', *Ann. Intern. Med.*, vol. 164, no. 4, pp. 226–235, 2016, doi: 10.7326/M15-0971.

[48] S. M. Albeshan, S. Z. Hossain, M. G. Mackey, J. K. Peat, F. M. Al Tahan, and P. C. Brennan, 'Preliminary investigation of mammographic density among women in Riyadh: association with breast cancer risk factors and implications for screening practices', *Clin. Imaging*, vol. 54, no. January, pp. 138–147, 2019, doi: 10.1016/j.clinimag.2019.01.002.

[49] M. E. M. Khamis, A. M. Alaa El-deen, and A. A. Azim Ismail, 'The diagnostic value of sonoelastographic strain ratio in discriminating malignant from benign solid breast masses', *Egypt. J. Radiol. Nucl. Med.*, vol. 48, no. 4, pp. 1149–1157, Dec. 2017, doi: 10.1016/j.ejrnm.2017.05.005.

[50] [50] D. Ma *et al.*, 'Intravoxel incoherent motion diffusion-weighted imaging as an adjunct to dynamic contrast-enhanced MRI to improve accuracy of the differential diagnosis of benign and malignant breast lesions', *Magn. Reson. Imaging*, vol. 36, pp. 175–179, Feb. 2017, doi: 10.1016/j.mri.2016.10.005.

[51] R. Fusco *et al.*, 'Blood oxygenation level dependent magnetic resonance imaging and diffusion weighted MRI imaging for benign and malignant breast cancer discrimination', *Magn. Reson. Imaging*, vol. 75, pp. 51–59, Jan. 2021, doi: 10.1016/j.mri.2020.10.008.

[52] R. L. Siegel, K. D. Miller, and A. Jemal, 'Cancer statistics, 2016', *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, 2016, doi: 10.3322/caac.21332.

[53] A. Agung, N. Gunawan, I. Suhariningsih, K. S. P. Triyono, and B. Widodo, 'Determination of Physical Parameter Model for the Photo Film Mammographic X-Ray Results on the Breast Cancer Histology Classification', 2012.

[54] A. A. N. Gunawan, S. Poniman, and I. W. Supardi, 'Classification of breast cancer grades using physical parameters and K-nearest neighbor method', *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 2, pp. 637–644, Apr. 2019, doi: 10.12928/TELKOMNIKA.V17I2.9797.

[55] F. M. Awad, 'Role of supersonic shear wave imaging quantitative elastography (SSI) in differentiating benign and malignant solid breast masses', *Egypt. J. Radiol. Nucl. Med.*, vol. 44, no. 3, pp. 681–685, Sep. 2013, doi: 10.1016/j.ejrnm.2013.04.001.

[56] T. A. Assegie, 'An optimized K-Nearest neighbor based breast cancer detection', *J. Robot. Control*, vol. 2, no. 3, pp. 115–118, May 2021, doi: 10.18196/jrc.2363.

[57] D. S. Grewal *et al.*, 'The most widely used is mammography, which has false- negative results in 10e15% cases.', *Med. J. Armed Forces India*, no. xxxx, pp. 1–7, 2022, doi: 10.1016/j.mjafi.2022.01.015.

[58] R. A. Abd El-Aleem, E. Abo El-Hamd, H. A. Yousef, M. E. M. Radwan, and R. A. A. Mohammed, 'The added value of qualitative and quantitative diffusion-weighted magnetic resonance imaging (DW-MRI) in differentiating benign from malignant breast lesions', *Egypt. J. Radiol. Nucl. Med.*, vol. 49, no. 1, pp. 272–280, Mar. 2018, doi: 10.1016/j.ejrnm.2017.10.015.

[59] Y. G. S. Ebrahim, M. R. Louis, and E. A. Ali, 'Multi-parametric dynamic contrast enhanced MRI, diffusion-weighted MRI and proton-MRS in differentiation of benign and malignant breast lesions: Imaging interpretation and radiology-pathology correlation', *Egypt. J. Radiol. Nucl. Med.*, vol. 49, no. 4, pp. 1175–1181, Dec. 2018, doi: 10.1016/j.ejrnm.2018.07.018.

[60] R. D. Mus *et al.*, 'Time to enhancement derived from ultrafast breast MRI as a novel parameter to discriminate benign from malignant breast lesions', *Eur. J. Radiol.*, vol. 89, pp. 90–96, Apr. 2017, doi: 10.1016/j.ejrad.2017.01.020.

[61] J. Li *et al.*, 'Can different regions of interest influence the diagnosis of benign and malignant breast lesions using quantitative parameters of contrast-enhanced sonography?', *Eur. J. Radiol.*, vol. 108, pp. 1–6, Nov. 2018, doi: 10.1016/j.ejrad.2018.09.005.

[62] W. Liu, C. Wei, J. Bai, X. Gao, and L. Zhou, 'Histogram analysis of diffusion kurtosis imaging in the differentiation of malignant from benign breast lesions', *Eur. J. Radiol.*, vol. 117, pp. 156–163, Aug. 2019, doi: 10.1016/j.ejrad.2019.06.008.

[63] J. Ma *et al.*, 'Distinguishing benign and malignant lesions on contrast-enhanced breast cone-beam CT with deep learning neural architecture search', *Eur. J. Radiol.*, vol. 142, Sep. 2021, doi: 10.1016/j.ejrad.2021.109878.

[64] J. Wang, R. Zhao, and J. Cheng, 'Diagnostic accuracy of contrast-enhanced ultrasound to

differentiate benign and malignant breast lesions: A systematic review and meta-analysis', *Eur. J. Radiol.*, vol. 149, Apr. 2022, doi: 10.1016/j.ejrad.2022.110219.

[65] S. Kanao, M. Kataoka, M. Iima, D. M. Ikeda, M. Toi, and K. Togashi, 'Differentiating benign and malignant inflammatory breast lesions: Value of T2 weighted and diffusion weighted MR images', *Magn. Reson. Imaging*, vol. 50, pp. 38–44, Jul. 2018, doi: 10.1016/j.mri.2018.03.012.

[66] Q. Zhang *et al.*, 'Quantitative transport mapping (QTM) for differentiating benign and malignant breast lesion: Comparison with traditional kinetics modeling and semi-quantitative enhancement curve characteristics.', *Magn. Reson. Imaging*, vol. 86, pp. 86–93, Feb. 2022, doi: 10.1016/j.mri.2021.10.039.

[67] B. W. Han *et al.*, 'Noninvasive discrimination of benign and malignant breast lesions using genome-wide nucleosome profiles of plasma cell-free DNA', *Clin. Chim. Acta*, vol. 520, pp. 95–100, Sep. 2021, doi: 10.1016/j.cca.2021.06.008.

[68] H. Y. Choi, Y. M. Sohn, and M. Seo, 'Comparison of 3D and 2D shear-wave elastography for differentiating benign and malignant breast masses: focus on the diagnostic performance', *Clin. Radiol.*, vol. 72, no. 10, pp. 878–886, Oct. 2017, doi: 10.1016/j.crad.2017.04.009.

[69] M. Wubulihasimu, M. Maimaitusun, X. L. Xu, X. D. Liu, and B. M. Luo, 'The added value of contrast-enhanced ultrasound to conventional ultrasound in differentiating benign and malignant solid breast lesions: a systematic review and meta-analysis', *Clin. Radiol.*, vol. 73, no. 11, pp. 936–943, Nov. 2018, doi: 10.1016/j.crad.2018.06.004.

[70] R. K. Wasan *et al.*, 'Can digital breast tomosynthesis accurately predict whether circumscribed masses are benign or malignant in a screening population?', *Clin. Radiol.*, vol. 74, no. 4, pp. 327.e1-327.e5, Apr. 2019, doi: 10.1016/j.crad.2018.12.020.

[71] B. Y. Chen *et al.*, 'Multiple b-value diffusion-weighted imaging in differentiating benign from malignant breast lesions: comparison of conventional mono-, bi- and stretched exponential models', *Clin. Radiol.*, vol. 75, no. 8, pp. 642.e1-642.e8, Aug. 2020, doi: 10.1016/j.crad.2020.03.039.

[72] Y. Yu *et al.*, 'Application of a shear-wave elastography prediction model to distinguish between benign and malignant breast lesions and the adjustment of ultrasound Breast Imaging Reporting and Data System classifications', *Clin. Radiol.*, vol. 77, no. 2, pp. e147–e153, Feb. 2022, doi: 10.1016/j.crad.2021.10.016.

[73] A. Mosapour, F. S. Karami Tehrani, and M. Atri, 'Expression level of VLDL receptor and VLDL-c levels in the malignant and benign breast tumors: The correlation with miRNA-4465 and miRNA-1297', *Mol. Cell. Probes*, vol. 53, Oct. 2020, doi: 10.1016/j.mcp.2020.101624.

[74] W. C. Shia and D. R. Chen, 'Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine', *Comput. Med. Imaging Graph.*, vol. 87, Jan. 2021, doi: 10.1016/j.compmedimag.2020.101829.

[75] Y. Davoudi, B. Borhani, M. P. Rad, and N. Matin, 'The role of doppler sonography in distinguishing malignant from benign breast lesions', *J. Med. Ultrasound*, vol. 22, no. 2, pp. 92–95, 2014, doi: 10.1016/j.jmu.2013.12.001.

[76] I. Handayani, 'IJISCS | 1 COMPARISON OF K-NEAREST NEIGHBOR AND NAÏVE BAYES FOR BREAST CANCER CLASSIFICATION USING PYTHON'.