Prediction of malignancy upgrade rate in high-risk breast lesions using an artificial intelligence model: a retrospective study

Özge Aslan [ID]
Ayşenur Oktay [ID]
Başak Katuk [ID]
Rıza Cenk Erdur [ID]
Oğuz Dikenelli [ID]
Levent Yeniay [ID]
Osman Zekioğlu [ID]
Süha Süreyya Özbek [ID]

PURPOSE

High-risk breast lesions (HRLs) are associated with future risk of breast cancer. Considering the pathological subtypes, malignancy upgrade rate differs according to each subtype and depends on various factors such as clinical and radiological features and biopsy method. Using artificial intelligence and machine learning models in breast imaging, evaluations can be made in terms of risk estimation in different research areas. This study aimed to develop a machine learning model to distinguish HRL cases requiring surgical excision from lesions with a low risk of accompanying malignancy.

METHODS

A total of 94 patients who were diagnosed with HRL by image-guided biopsy between January 2008 and March 2020 were included in the study. A structured database was created with clinical and radiological characteristics and histopathological results. A machine learning prediction model was created to make binary classifications of lesions as malignant or benign. Random forest, decision tree, K-nearest neighbors, logistic regression, support vector machine (SVM), and multilayer perceptron machine learning algorithms were used. Among these algorithms, SVM was the most successful. The estimations of malignancy for each case detected by artificial intelligence were combined and statistical analyses were performed.

RESULTS

Considering all cases, the malignancy upgrade rate was 24.5%. A significant association was observed between malignancy upgrade rate and lesion size ($P = 0.004$), presence of mammography findings ($P = 0.022$), and breast imaging-reporting and data system category ($P = 0.001$). A statistically significant association was also found between the artificial intelligence prediction model and malignancy upgrade rate ($P < 0.001$). With the SVM model, an 84% accuracy and 0.786 area-under-the-curve score were obtained in classifying the data as benign or malignant.

CONCLUSION

Our artificial intelligence model (SVM) can predict HRLs that can be followed up with a lower risk of accompanying malignancy. Unnecessary surgeries can be reduced, or second line vacuum excisions can be performed in HRLs, which are mostly benign, by evaluating on a case-by-case basis, in line with radiology–pathology compatibility and by using an artificial intelligence model.

KEYWORDS

Artificial intelligence, breast, cancer, high risk lesion of breast, image-guided biopsy

From the Department of Radiology (Ö.A. [envelope] dr.ozgeaslan@gmail.com, A.O., S.S.Ö.), Ege University Faculty of Medicine, İzmir, Turkey; Department of Computer Engineering (B.K., R.C.E., O.D.), Ege University Engineering Faculty, İzmir, Turkey; Department of General Surgery (L.Y.), Ege University Faculty of Medicine, İzmir, Turkey; Department of Medical Pathology (O.Z.), Ege University Faculty of Medicine, İzmir, Turkey.

The increase in breast cancer screening with mammography increases the rate of non-palpable lesions detected in the breast.[1,2] In the diagnosis of these lesions, percutaneous biopsy methods are increasingly applied under the guidance of imaging methods. Percutaneous needle biopsy is a fast, easy-to-apply, inexpensive, and well-tolerated biopsy alternative to open surgical biopsies.[1,3] The prevalence of high-risk breast lesion (HRL) detection with core needle biopsy (CNB) is 5–9% in all breast biopsies.[2,4,5] HRLs are defined as lesions with a high risk of malignant transformation and the possibility of synchronous or ad-

jacent breast malignancy.[6] They are detected together with breast malignancy by imaging-guided percutaneous breast biopsies as well as by excisional biopsies and during surgical procedures.[5,7,8] Additionally, HRLs are a heterogeneous group of proliferative lesions with variable malignant potential and can be considered as falling in a "gray zone" between benign and malignant lesions (Figure 1).

HRLs are associated with future breast cancer risk and are precursors of breast carcinogenesis.[5,9] Lesions defined as high-risk in thick-needle biopsies [CNB and vacuum assisted biopsy (VAB)] may upgrade to malignancy when a surgical excision is performed. The overall positive predictive value (PPV) for malignancy is approximately 10–30%. After detecting an HRL in a thick-needle biopsy, a clinical decision is required between surgical excision or follow-up of the lesion to avoid unnecessary surgery due to the possibility of concomitant malignancy based on radiology–pathology compatibility. As a general approach, surgical excision is often recommended for most of these lesions because of the risk of malignancy.[10] However, the malignancy upgrade rate of HRLs reported in the literature is varied and depends on various factors such as pathological subtype, clinical and radiological features, and biopsy method.[2] In recent studies, a case-by-case approach was recommended. Upgrade rates are higher in lesions with atypia compared to other HRLs.[7] The Second International B3 Lesions Consensus Conference recommends excision with vacuum biopsy as an alternative to open surgery in HRLs except atypical ductal hyperplasia (ADH) and phyllodes tumors.[11]

The success of image-guided needle biopsies depends on the evaluation after the biopsy as well as the biopsy procedure. When evaluating biopsy results, radiopathological compatibility is considered. Patholo-gy results can be expected to adequately explain imaging findings.[2] A multidisciplinary case-based approach is key to optimal patient care.[2]

Previous research suggests that artificial intelligence (AI) algorithms can support breast radiologists in diagnosis, treatment, and case follow-up management by using large quantities of high-quality imaging data, however more studies are needed into this.[12] Using AI and machine learning models in breast imaging, evaluations can be made in terms of risk estimation in different research areas.[13] In the literature, there are many examples of successful computer-aided diagnosis systems that have used traditional machine learning and deep learning algorithms to classify breast cancer.[14] However, there is insufficient research into risk determination in HRLs.

In this study, we aimed to develop a machine learning model to distinguish HRLs with a low risk of accompanying malignancy from cases requiring surgical excision. For this purpose, a structured dataset consisting of HRL patients with known surgical outcomes was created. Then, a machine learning model was trained with this dataset to develop a model for classifying patients whose surgical outcomes were unknown.

Methods

Approval for this study was obtained from the ethics committee of our institution (approval no: 20-11.1T/42, date: 25.11.2020). Before the biopsies were performed, the procedure was explained to all patients, and they signed a consent form. The pathology results of 2.249 patients who underwent image-guided thick-needle biopsy between January 2008 and March 2020 in our breast radiology clinic were retrospectively evaluated, and 120 patients diagnosed with high-risk lesion were identified from these cases. The pathology results of those who underwent surgical excision and the radiological follow-up results of those who were followed-up without surgery were evaluated. A total of 26 patients, who were followed up for less than one year after having a thick-needle biopsy or whose pathology results were unknown were excluded from the study. A structured database was created with the following information: age at the time of diagnosis, breast cancer history and family history, age of menarche, hormonal therapy history, other cancer history, smoking status, lesion size, radiological imaging features, breast imaging reporting and data system (BI-RADS) category, biopsy type, needle thickness, sampling number (<4 or ≥4), biopsy histopathology result, excision histopathology results, and follow-up findings (Figure 2).

Mammographic images were obtained with full-field digital mammography and digital breast tomosynthesis mammograms (Lorad Selenia and Selenia Dimensions, Hologic). Stereotactic VABs were performed on a prone table unit (Multicare Platinum; Hologic), with a 9-G needle (Encore biopsy probe; Bard). Magnetic resonance imagings (MRI) were performed with 1.5T (Magnetom Amira, Smphony Siemens) and 3T (Magnetom Verio Siemens) MRI devices using conventional and dynamic contrast sequences. Ultrasonography (US) and US-guided biopsy procedures were performed with Hitachi and Siemens devices using a high-frequency linear probe. A 14-G needle was used in US-guided thick-needle biopsy. The BI-RADS category was determined according to the American College of Radiology BI-RADS Atlas

High risk breast lesion subtypes	
Ductal proliferative lesions	Atypical ductal hyperplasia (ADH), Columnar cell lesions
Lobular proliferative lesions (Lobular neoplasia)	Atypical lobular hyperplasia (ALH), Lobular carcinoma in situ (LCIS)
Radial sclerosing lesions	Radial scar, complex sclerosing lesion
Papillary lesions	Intraductal papilloma, Atypical intraductal papilloma (AIP)
Cellular fibroepithelial tumor	Flat epithelial atypia (FEA)

Figure 1. Classification of high-risk breast lesion subtypes.

5th edition classification, based on mammography, US, and MRI findings.

Morphology and distribution features of microcalcification, structural distortion, asymmetry, and mass opacity were evaluated in the mammography. Lesions were classified as mass and non-mass (abnormal echogenicity and structural distortion) findings on the US and recorded. The presence of mass and non-mass enhancement in the MRI was evaluated. In cases diagnosed with more than one HRL by biopsy, diagnoses that included atypia and had a higher risk of malignancy were accepted as the main lesion. Those who had a malignant diagnosis (invasive ductal carcinoma, ductal carcinoma in situ, or invasive lobular carcinoma) with surgical excision were accepted as upgraded to malignancy.

Patients with benign histopathology results and those who were stable in the long-term follow-up were included in the benign group, and those with a malignant excision diagnosis were included in the malignant group. The upgrade rate of existing HRLs t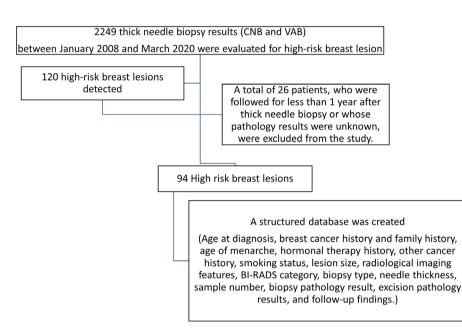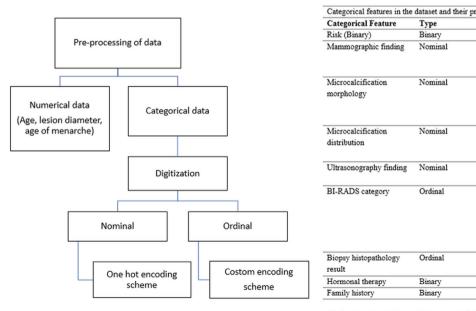o malignancy in the AI prediction model was defined from the highest to the lowest, considering the ranges specified in the literature [ADH > atypical intraductal papilloma (AIP) > lobular neoplasia > radial scar > intraductal papilloma without atypia].[2,7,15,16]

Artificial intelligence model technique

Libraries and technologies used

Python programming language and related libraries (Numpy, Pandas, and Scikit-learn) were used in data preprocessing and training the machine learning algorithms.

Pre-processing of data

Data were preprocessed prior to the creation of the AI prediction model. The data set contained columns with numerical data and categorical data. Pre-processing steps were carried out on these columns. In the preprocessing stage, categorical data were digitized, and all data were normalized. For digitization, a one-hot encoding scheme or a custom encoding scheme was used depending on the type of categorical data (nominal or ordinal) (Figure 3).

For example, in the mammography findings column, which contains nominal categorical data, one-hot vectors were created for each of the column values of asymmetry, mass opacity, microcalcification, and structural distortion (Figure 3). These vectors were added to the data set as a new feature, and the original column was removed from the dataset.

Figure 2. Flow chart of the steps followed while creating the database. CNB, core needle biopsy; VAB, vacuum assisted biopsy; BI-RADS, breast imaging reporting and data system.

Categorical features in the dataset and their preprocessing steps		
Categorical Feature	**Type**	**Preprocessing**
Risk (Binary)	Binary	Positive:1, Negative: 0
Mammographic finding	Nominal	One-hot encoding scheme was used for asymmetry, mass, microcalcification, and structural distortion values.
Microcalcification morphology	Nominal	One-hot encoding scheme was used for amorphous, coarse heterogeneous, suspicious on follow-up, fine linear branching, and fine pleomorphic values.
Microcalcification distribution	Nominal	One-hot encoding scheme was used for microcalcification values in cluster, segmental, regional, and multigroups.
Ultrasonography finding	Nominal	One-hot encoding scheme was used for mass and non-mass finding values.
BI-RADS category	Ordinal	A custom encoding scheme was used for BI-RADS3:1, BI-RADS4a:2, BI-RADS4b:3, BI-RADS4c:4, BI-RADS5:5.
Biopsy histopathology result	Ordinal	A custom encoding scheme was used: IP:1, RS: 2, LN:3, AIP:4, ADH:5.
Hormonal therapy	Binary	Positive:1, Negative:0
Family history	Binary	Positive:1, Negative:0

IP: Intraductal papilloma without atypia; RS: Radial scar; LN: Lobular neoplasia; AIP: Atypical intraductal papilloma; ADH: Atypical ductal hyperplasia.

Figure 3. Categorical features in the dataset and their preprocessing steps.

In the BI-RADS category column, which contains ordinal categorical data, a custom encoding scheme was used to match BI-RADS3:1, BI-RADS4a:2, BI-RADS4b:3, BI-RADS4c:4, and BI-RADS5:5. Minimum-maximum normalization was used for normalization of the data.

Machine learning model development

The data passed through the preprocessing stage were divided into training and test datasets. The test data set comprised 20% of the entire data set (19 samples). In splitting the dataset, the proportions of samples in each class observed in the original dataset were preserved, and a stratified train–test split was applied.

The prepared data sets were used to create a machine learning prediction model to make binary classification as "malignant" or "benign." Random forest, decision tree, K-nearest neighbor, logistic regression, support vector machine (SVM), and multilayer perceptron machine learning algorithms were run with the training data set, and their performances were measured with the test data set (Figure 4). In the specified machine learning algorithms trained by hyperparameter optimization and using cross-validation, the models were compared by looking at the accuracy and area under the curve (AUC) score. Although the AUC scores of the logistic regression (0.743) and SVM (0.786) models are relatively close, SVM made a more accurate prediction for the "malign" samples. In addition, the accuracy of SVM (0.84) was 0.05 points higher than the logistic regression (0.79). The AUC score and accuracy of the K-nearest neighbor model was lower than the SVM model (Figures 5-7).

The SVM, which gave the most successful results, was selected. For the hyperparameters of the SVM algorithm, the C, gamma, and kernel parameters were optimized for various values (Figure 4). In the fine-tuning of the hyperparameters, five-fold cross-validation was performed with the grid search algorithm.

The performance of the SVM classification model was measured by using the metrics accuracy, sensitivity, specificity, and F1 Score (Figure 5), and a confusion matrix was obtained (Figure 6). The AUC score of the model was then calculated (Figure 7).

The estimation of malignancy of each case detected by AI and clinical and radiological case features were combined and statistical analyses were performed with the IBM SPSS 25.0 program.
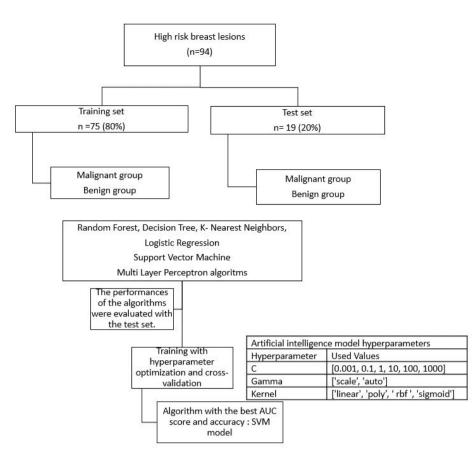

Figure 4. Machine learning model development stages. AUC, area-under-the-curve; SVM, support vector machine.

Figure 5. Logistic regression, support vector machine, and K-nearest neighbors model detailed classification reports.

The estimation of malignancy upgrade was evaluated in all cases according to each HRL pathological subtype.

Statistical analysis

The distribution of cases across the age groups was expressed as the mean ± stan-
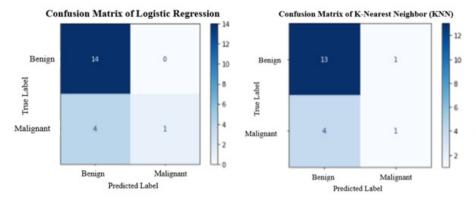
Figure 6. Confusion matrices of the logistic regression, support vector machine, and K-nearest neighbors.
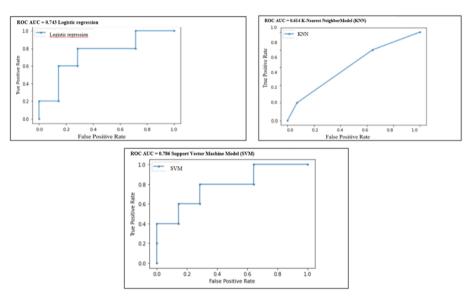

Figure 7. Logistic regression, support vector machine, and K-nearest neighbors algorithm area-under-the-curve scores. AUC, area-under-the-curve; ROC, receiver operating characteristic curve; SVM, support vector machine.

dard deviation, and categorical data were expressed as frequencies (n) and percentages (%). All statistical analyses were performed with SPSS software version 25.0 (IBM). Kolmogorov–Smirnov and Shapiro–Wilk tests were used to assess the normal distribution of data. Pearson's chi-square and Fisher's exact tests were employed to compare the malignancy upgrade rate and AI SVM model assessment. Student's t-tests were used to compare differences in continuous variables. Pearson's chi-square test was used to evaluate the relationship between the AI SVM model assessment and radiological–clinical features of cases.
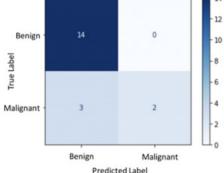
Results

In the 94 female patients, the mean age was 47.22 ± 10.7; range: 17–73 years, the mean lesion size was 1.8 ± 4.9 cm; range: 5–100 cm, and the mean age of menarche was 13.26 ± 1.4; range: 10–18 years.

A rate of 25.5% (n = 24) of the patients had a positive family history of breast cancer. Hormonal therapy was applied in 26.6% (n = 25). When evaluated in terms of family history, hormonal therapy, previous breast cancer and HRL history, risk factor status of breast cancer was positive in 35% (n = 33) of the patients. There was a history of smoking in 36% (n = 34). Considering the imaging characteristics, 53% (n = 50) had positive mammography findings (microcalcification, asymmetry, structural distortion, and mass opacity). Suspicious microcalcification was present in 31% (n = 29). The most common microcalcification morphology was amorphous (14%; n = 13), and the most common distribution pattern was clustered type (17%; n = 16). The most common BI-RADS category was 4A (55.3%; n = 52). US findings [53% (n = 50) with a mass and 22% (n = 21) without a mass] were observed in 75.5%. MRI findings [20.2% (n = 19) mass enhancement or 21.3% (n = 20) non-mass enhancement] were present in 41.5% of the cases (Table 1). Of the mass-shaped lesions (n = 19), 47.4% (n = 9)

had smooth contours and 52.6% (n = 10) had irregular contours in MRI. In the pharmacokinetic evaluation of lesions, 92.3% (n = 36) type-1 and type-2 curves and 7.7% (n = 3) type-3 curve patterns were observed.

Mammography-guided (stereotactic) VAB was performed on 25.5% (n = 24) of the patients, and US-guided CNB was performed on 75.5% (n = 70). Vacuum biopsies were performed using 9-G needles, and 14-G needles were used in CNBs. The number of samples was below four in 20% of the patients and four or more in 80% of the patients (Table 1).

According to the thick-needle biopsy histopathology results, the pathological subtypes of the cases were ADH (44.7%; n = 42), intraductal papilloma (37.2%; n = 35), AIP (10.6%; n = 10), radial scar (5.3%; n = 5), and lobular neoplasia (2.1%; n = 2). Of the cases, 84% were removed by surgical excision, and 16% were followed up. Of the 79 excised cases, 41% were diagnosed as benign, 30% with atypia, and 29% as malignant. Fifteen patients who were followed up without surgery were stable in clinical and radiological follow-up, and these cases were placed in the benign group.

Considering all cases, the malignancy upgrade rate was 24.5% (n = 23). According to the pathological subtypes, the malignancy upgrade rates were 50% (n = 1) for lobular neoplasia, 40% (n = 2) for radial scar, 31% (n = 13) for ADH, 30% (n = 3) for AIP, and 11.4% (n = 4) for intraductal papilloma (Table 2).

When evaluated with Pearson's chi-square test for the upgrade rate to malignancy with all variables, a statistically significant association was found with the variables of BI-RADS category, lesion diameter, and presence of mammographic findings ($P < 0.05$; Table 3). No statistically significant relationship was found between family history and smoking and upgrade to malignancy ($P = 0.631$, $P = 0.247$, respectively).

Table 1. Benign and malignant lesion distribution characteristics

| | | Benign group | Malignant group | SVM model prediction result (n) | | Total |
		n (%)	n (%)	Wrong	Right	n (%)
Biopsy type	CNB	55 (59)	15 (16)	8	62	70 (74)
	VAB	16 (17)	8 (9)	1	23	24 (26)
Risk	Positive	25 (27)	8 (9)	1	32	33 (35)
	Negative	46 (49)	15 (16)	8	53	61 (65)
Needle thickness	9 G	16 (17)	8 (9)	1	23	24 (26)
	14 G	55 (59)	15 (16)	8	62	70 (74)
Sampling number	<4	15 (16)	4 (4)	3	16	19 (20)
	≥4	56 (60)	19 (20)	6	69	75 (80)
Lesion diameter	<1.5 cm	46 (49)	7 (7)	4	49	53 (56)
	≥1.5 cm	25 (27)	16 (17)	5	36	41 (44)
Mammography finding	Microcalcification	19 (40)	10 (21)	1	28	29 (62)
	Mass	6 (13)	3 (6)	1	8	9 (19)
	Non-mass	6 (13)	3 (6)	2	7	9 (19)
US finding	Mass	40 (56)	10 (14)	4	46	50 (70)
	Non-mass	14 (20)	7 (10)	4	17	21 (30)
MRI finding	Mass	16 (41)	3 (8)	1	18	19 (49)
	Non-mass	13 (33)	7 (18)	5	15	20 (51)
Microcalcification morphology	Amorphous, course heterogeneous	11 (38)	4 (14)	1	14	15 (52)
	Fine linear branching, fine pleomorphic	8 (28)	6 (21)	-	14	14 (48)

CNB, core needle biopsy; VAB, vacuum assisted biopsy; SVM, support vector machine; US, ultrasonography; MRI, magnetic resonance imaging.

Table 2. Distribution characteristics of pathological subgroups and malignancy upgrade rates

Pathologic subgroups	Benign group n (%)	Malignant group n (%)	Total n (%)	Malignancy upgrade rate (%)
IP	31 (33)	4 (4)	35 (37)	11.4
ADH	29 (31)	13 (14)	42 (45)	31
AIP	7 (7)	3 (3)	10 (11)	30
Radial scar	3 (3)	2 (2)	5 (5)	40
Lobular neoplasia	1 (1)	1 (1)	2 (2)	50
Total	71 (76)	23 (24)	94 (100)	24.5

IP, intraductal papilloma without atypia; ADH, atypical ductal hyperplasia; AIP, atypical intraductal papilloma.

Table 3. Variables that have a statistically significant relationship with malignancy upgrade rate

		Benign group n (%)	Malignant group n (%)	Total n (%)	P value
BI-RADS category	3-4A-4B	64 (68)	14 (15)	78 (83)	0.003
	4C-5	7 (7)	9 (10)	16 (17)	
Lesion diameter	<1.5 cm	46 (49)	7 (7)	53 (56)	0.004
	≥1.5 cm	25 (27)	16 (17)	41 (44)	
Mammographic finding	Positive	33 (35)	17 (18)	50 (53)	0.022
	Negative	38 (40)	6 (6)	44 (47)	

BI-RADS, breast imaging reporting and data system.

The AI analysis identified 85 cases correctly and 9 cases incorrectly (Tables 1 and 4). The SVM AI model, which was trained using certain hyperparameters, had 84% accuracy (Figure 5) and an AUC score of 0.786 (Figure 7) in classifying the data as benign or malignant.

No statistically significant difference was found between needle thickness/biopsy type and erroneous AI estimation (*P* = 0.297).

A statistically significant difference was found between the AI prediction and the malignancy upgrade rate of the patients (*P* < 0.001). The sensitivity of the malignant case prediction set of the AI model was 60.87%, the specificity was 100%, PPV was 100%, and negative predictive value was 88.75%.

Discussion

The most significant problem in the management of HRLs is upgrading to malignancy. The upgrade rate to malignancy in this study was 24.5%, which is similar to the rates reported in the literature.[10]

Considering pathological subtypes, the rate of upgrade to malignancy differs according to each subtype. The malignancy upgrade rate of ADH, which was the most common lesion subtype among our cases, was similar to the literature. A wide range of malignancy upgrade rates for ADH and AIP has been reported in the literature.[2,7] In this study, for AIP, as in ADH, there were erroneous AI predictions in three of the patients,

Table 4. Distribution characteristics of pathological subgroups malignancy upgrade rates and the accuracy of the SVM model in subtypes

Pathologic subgroup	True malignancy upgrade rates n (%)	SVM model malignancy upgrade rates n (%)	SVM model accuracy (%)
ADH (n = 42)	13 (31)	10 (23.8)	92.8
AIP (n = 10)	3 (30)	-	70
LN (n = 2)	1 (50)	1 (50)	100
RS (n = 5)	2 (40)	-	60
IP (n = 35)	4 (11.4)	3 (8.6)	97.1
Total	23 (24.5)	14 (14.9)	90.4

ADH, atypical ductal hyperplasia; AIP, atypical intraductal papilloma; LN, lobular neoplasia; RS, radial scar; IP, intraductal papilloma without atypia; SVM, support vector machine.

and biopsies were performed with tru-cut in both groups.

The malignancy upgrade rates for radial scar and lobular neoplasia were in the upper limit of the rates stated in the literature.[7] This may be due to the low number of cases in these subgroups.

The SVM model made an incorrect prediction in nine malignant cases in total (Tables 1 and 4). One of these cases was diagnosed by VAB with a 9-G needle, and all others were diagnosed by CNB with a 14-G needle. No statistically significant correlation was found between needle thickness/biopsy type and erroneous AI estimation, but the low number of cases is a limitation in the evaluation of this variable. In the study of Bahl et al.[13], which included 1,006 HRLs, the AI prediction model had a prediction accuracy of 97.4% in malignant cases and 69.4% in benign cases, and they reported that unnecessary surgeries could be reduced in benign cases. In the present study, the AI model made a correct prediction in all cases that were diagnosed as benign by surgical excision and considered as stable in long-term follow-up. More than half of the patients who underwent surgical excision were diagnosed as benign. Considering the radiopathological fit and AI model estimation, if these cases had been followed up radiologically and clinically, the rate of unnecessary surgery could have been reduced by 71%.

The majority of HRLs are benign but most are surgically excised because of the associated risk of malignancy. Post-biopsy evaluation and biopsy procedure are important in the management of these lesions.[1,15]

Comparable to the literature, there was a statistically significant relationship between VAB as a biopsy guide method, 9-G needle thickness, and sufficient number of samples with the malignancy upgrade rate. In cases that have these features, a more appropriate decision can be made in terms of follow-up and excision. In order to increase the correct prediction rates with the AI model, studies containing more cases and data sets are needed.

There are some further limitations to this study. Firstly, it is a retrospective study. The differences in the number of pathological subtypes and the low number of patients were our biggest limitations. Significant results could not be obtained in many statistical analyses due to the differences in the number of pathological subtypes such as lobular neoplasia and radial scar, and the low number of cases. In addition, due to the small number of cases and the limited number of histopathological features in terms of the degree of atypia, a clear analysis of the variables that may be effective in the erroneous predictions of the SVM model could not be made. For this reason, better statistical results can be obtained by adding features such as the degree of pathological atypia, which will further strengthen the data set, and by including more patients.

In conclusion, this study's AI model (SVM) can predict HRLs that can be followed up with a lower risk of accompanying malignancy. Both ADH and AIP cases should be surgically excised because of the high risk of malignancy associated with them. Apart from these subtypes, HRLs, which are mostly benign, can be evaluated on a case-by-case basis, in line with radiology–pathology compatibility and using an AI prediction model, to reduce unnecessary surgeries, or excision can be performed with second-line VAB.

Conflict of interest disclosure

The authors declare no conflicts of interest.

References

1. Oktay A. Meme hastalıklarında görüntüleme. İçinde: Tunacı M, ed. Radyoloji-patoloji sonuçlarının korelasyonu, bölüm 7. 2nd ed. Ankara: *Dünya Tıp Kitabevi*; 2020:249-255. **[Crossref]**

2. Neal L, Sandhu NP, Hieken TJ, et al. Diagnosis and management of benign, atypical, and indeterminate breast lesions detected on core needle biopsy. *Mayo Clin Proc*. 2014;89(4):536-547. **[Crossref]**

3. Liberman L. Clinical management issues in percutaneous core breast biopsy. *Radiol Clin North Am*. 2000;38(4):791-807. **[Crossref]**

4. Oktay A. Radyolojik-patolojik korelasyon: yüksek risk lezyonlarda ne yapmalıyız? *Türk Radyoloji Seminerleri*. 2014;2:217-229. **[Crossref]**

5. Gulla S, Lancaster R, De Los Santos J. High-risk breast lesions and current management. *Semin Roentgenol*. 2018;53(4):252-260. **[Crossref]**

6. Bianchi S, Caini S, Renne G, et al. Positive predictive value for malignancy on surgical excision of breast lesions of uncertain malignant potential (B3) diagnosed by stereotactic vacuum-assisted needle core biopsy (VANCB): a large multi-institutional study in Italy. *Breast*. 2011;20(3):264-270. **[Crossref]**

7. Nakhlis F. How do we approach benign proliferative lesions? *Curr Oncol Rep*. 2018;20(4):34. **[Crossref]**

8. Georgian-Smith D, Lawton TJ. Controversies on the management of high-risk lesions at core biopsy from a radiology/pathology perspective. *Radiol Clin North Am*. 2010;48(5):999-1012. **[Crossref]**

9. Thomas PS. Diagnosis and management of high-risk breast lesions. *J Natl Compr Canc Netw*. 2018;16(11):1391-1396. **[Crossref]**

10. Boateng S, Tirada N, Khorjekar G, Richards S, Ioffe O. Excision or observation: the dilemma of managing high-risk breast lesions. *Curr Probl Diagn Radiol*. 2020;49:124-132. **[Crossref]**

11. Rageth CJ, O'Flynn EAM, Pinker K, et al. Second international consensus conference on lesions of uncertain malignant potential in the breast (B3 lesions). *Breast Cancer Res Treat*. 2019;174(2):279-296. **[Crossref]**

12. Mendelson EB. Artificial intelligence in breast imaging: potentials and limitations. *AJR Am J Roentgenol*. 2019;212(2):293-299. **[Crossref]**

13. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology*. 2018;286(3):810-818. **[Crossref]**

14. Murtaza G, Shuib L, Wahab AWA, et al. Deep learning-based breast cancer classification through medical imaging modalities: state of

the art and research challenges. *Artif Intell Rev.* 2019;53:1655-1720. **[Crossref]**

15. Krishnamurthy S, Bevers T, Kuerer H, Yang WT. Multidisciplinary considerations in the management of high-risk breast lesions.

16. Armes JE, Galbraith C, Gray J, Taylor K. The outcome of papillary lesions of the breast diagnosed by standard core needle biopsy *AJR Am J Roentgenol.* 2012;198(2):132-140. **[Crossref]**

within a BreastScreen Australia service. *Pathology.* 2017;49(3):267-270. **[Crossref]**