 doi:10.20944/preprints202206.0186.v1

Article

Developing An Over-the-Counter Screening Model for Breast Cancer among the Asian Women Population

Tengku Muhammad Hanis[1], Nur Intan Raihana Ruhaiyem[2], Wan Nor Arifin[3], Juhara Haron[4,5], Wan Faiziah Wan Abdul Rahman[5,6], Rosni Abdullah[2] and Kamarul Imran Musa[1,*]

[1] Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; tengkuhanismokhtar@gmail.com (T.M.H.); drkamarul@usm.my (K.I.M.)
[2] School of Computer Sciences, Universiti Sains Malaysia, USM 11800, Penang, Malaysia; intanraihana@usm.my (N.I.R.R.); rosni@usm.my (R.A.)
[3] Biostatistics and Research Methodology Unit, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; wnarifin@usm.my
[4] Department of Radiology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; drjuhara@usm.my
[5] Breast Cancer Awareness and Research Unit, Hospital Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; drjuhara@usm.my (J.H.); wfaiziah@usm.my (W.F.W.A.R.)
[6] Department of Pathology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; wfaiziah@usm.my
* Correspondence: drkamarul@usm.my

Abstract: This study aimed to determine the feasibility of the development of an over-the-counter (OTC) screening model using machine learning for breast cancer screening in the Asian women population. Data were retrospectively collected from women who came to the Hospital Universiti Sains Malaysia, Malaysia. Five screening models were developed based on machine learning methods; random forest, artificial neural network (ANN), support vector machine (SVM), elastic-net logistic regression and extreme gradient boosting (XGBoost). Features used for the development of the screening models were limited to information from the patients' registration form. The model performance was assessed across the dense and non-dense groups. SVM had the best sensitivity while elastic-net logistic regression had the best specificity. In terms of precision, both random forest elastic-net logistic regression had the best performance, while, in terms of PR-AUC, XGBoost had the best performance. Additionally, SVM had a more balanced performance in terms of sensitivity and specificity across the mammographic density groups. The three most important features were age at examination, weight and number of children. In conclusion, OTC models developed from machine learning methods can improve the prognostic process of breast cancer in Asian women.

Keywords: screening model; breast cancer; explainable model; machine learning; Asian women

1. Introduction

Artificial Intelligence (AI) is a subfield of computer science that aims to develop a system capable to perform a task that usually requires human intelligence. The rise of AI is expected to improve many areas including the field of healthcare and medicine. Machine learning and deep learning, a subfield of AI had been studied to be used as health data analytic tools including for drug discovery, genomic medicine, disease prognosis and diagnosis and personalised healthcare [1,2]. However, the adoption of machine learning in healthcare and medicine is slower than in the other fields [3]. The three pillars for the successful application of machine learning or AI in healthcare include data and security, analytics and insight and shared expertise [4]. Thus, the emergence of explainable or interpretable machine learning research is in line with the three pillars and is expected to further help machine learning to be fully incorporated in this field. Explainable machine learning provides insight into the model and increases confidence and trust in the decision of the model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

Breast cancer is the commonest cancer among women in at least 140 countries [5]. WHO aims to reduce global breast cancer mortality by 25% annually between 2020 and 2040, which is equivalent to 2.5 million breast cancer death worldwide [6]. The three key points in achieving this goal outlined by the WHO were early detection, timely diagnosis and comprehensive management of the disease. Thus, ensuring an efficient health delivery system is essential in supporting the three approaches outlined by the WHO. Generally, delay in the management of the cancer is divided into two; patient delay and provider delay [7]. Patient delay is the delay during the period between the first discovery of the symptom and medical consultation. Provider delay is the delay between medical consultation and the beginning of the cancer treatment. Additionally, the combination of both types of delay is known as a total delay. However, a more complicated model of the total delay had also been proposed. For example, the total patient delay model detailed the total delay into five stages [8] while the total breast cancer delay model detailed it into eight stages [9]. Nonetheless, the total delay of more than 1 to 3 months has been observed to be associated with advanced stages of cancer and reduced survival of the patients [7,10,11]. In Malaysia, about 66.8% of breast cancer patients had a consultation delay [12]. Thus, this study planned to improve the efficiency of the process for breast cancer patients in getting a medical consultation.

One of the strongest indicators of breast cancer is the mammographic density which reflects the amount of dense and fatty tissue in the breast [13,14]. Women with denser breasts had four to six times higher chances of developing breast cancer than those with less dense breasts [15]. Asian women or women with Asian ancestry had denser breasts compared to other populations [16]. Thus, it is imperative to take mammographic density into account in developing a machine learning model for Asian breast cancer women in Malaysia. This study aimed to evaluate the performance of machine learning methods as an over-the-counter (OTC) screening model to prioritise women with a high probability of getting breast cancer for medical consultation. Further, this study intended to assess the performance of the OTC model across dense and non-dense women. Lastly, we aimed to determine the top five most influential features for each of the developed models.

2. Materials and Methods

2.1. Sample population

Hospital Universiti Sains Malaysia (HUSM) is located in the northeast coast region of Malaysia in the state of Kelantan. HUSM is a university- and research-based hospital associated with Universiti Sains Malaysia. Breast Cancer Awareness and Research Unit (BestARi) is a breast cancer resource centre in HUSM. BestARi receives women with breast-related problems from the northeast coast region of Malaysia, especially from the state of Kelantan. The unit acts as a one-stop centre to facilitate a multidisciplinary approach to breast cancer management of the incoming women.

2.2. Breast cancer data

Breast cancer data were collected retrospectively from the BestARi, Department of Radiology and Department of Pathology in HUSM. All the data were limited to 1st January 2014 and 30th June 2021. Twenty-seven related pieces of information were collected in this study. There were 24 pieces of information collected from the BestARi including hospital identification number, date of examination, age at examination, race, marital status, number of children, age at menarche, weight, height, handedness, information regarding the symptoms; lump, nipple discharge, nipple retraction, axillary mass, pain and skin changes and information regarding the medical history; history of breast surgery or implant, history of breast trauma, history of birth control or hormone replacement therapy, history of the previous mammography, history of breast self-examination, breastfeeding history, history of total abdominal hysterectomy bilateral salpingo-oophorectomy (TAHBSO), family history of breast cancer and menopausal status. This information was used in the machine learning model development except for hospital identification number and date

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

of examination. Breast imaging-reporting and data system (BIRADS) classification information and BIRADS density (or mammographic density) were collected from the Department of Radiology, HUSM. This information was used as a reference and was not included in the model development. The breast cancer diagnosis was collected from histopathological examination (HPE) from the Department of Pathology, HUSM.

The data from the Department of Radiology and Department of Pathology were combined with the BestARi's data if both data were dated within a year after the date of BestARi's data for each patient. The latest medical record was taken if patients had several records in the BestARi and a single record from the Department of Radiology or Department of Pathology. Afterwards, a body mass index was further calculated from the individual weight and height and was added to the existing list of variables. Each patient was classified into normal and abnormal classification. The normal class was patients with a BIRADS classification of 1. The abnormal class was patients with a BIRADS classification of 2, 3, 4, 5 and 6 or who had a diagnosis of benign or malignant subtype of breast cancer from HPE result. Patients with a BIRADS classification of 0 and missing BIRADS classification or mammographic density were excluded from the study. Additionally, non-dense breast women were those with BIRADS density of A and B, while dense breast women were those with BIRADS density of C and D.

2.3. Machine learning models

Five screening models were developed from machine learning methods including random forest, artificial neural network (ANN), support vector machine (SVM), elastic-net logistic regression and extreme gradient boosting (XGBoost). All five machine learning methods had been used in various breast cancer-related problems. The random forest algorithms had been implemented previously to predict an early diagnosis of breast cancer and had achieved 98% accuracy [17]. Additionally, the ANN algorithm had been proved excellent in breast cancer diagnosis using a mammographic input and performed 3-5% better than the accuracy of the radiologist [18]. Also, this algorithm had been used to predict breast cancer recurrence [19]. The SVM algorithm had been used in breast cancer risk estimation in previous research [20]. Elastic regression is a regularised method combining a least absolute shrinkage and selection operator (LASSO) and ridge penalties methods. The elastic regression and XGBoost, both had been used to predict breast cancer recurrence in previous studies [19,21]. All machine learning methods were run using the parsnip package [22] with the kernlab package [23] as a backend for SVM, ranger package [24] for random forest, nnet package [25] for ANN, glmnet package [26] for elastic-net logistic regression and xgboost package [27] for XGBoost. R version 4.1.3 was used to develop all the screening models in the Windows 10 OS machine [28].

2.3.1. Pre-processing steps

Initially, all 25 predictor features (Table 1) were included in the model development. Next, missing values in the data were imputed using a bagged tree model. Subsequently, variables with large absolute correlations with other variables were removed. Then, the training dataset was balanced using a random over-sampling examples (ROSE) algorithm [29]. All numerical features were normalised and transformed using a Yeo-Johnson transformation [30]. A dummy coding variable was created for all categorical features for all machine learning models except for the random forest model. The random forest model had been shown to have at least similar performance if not better when categorical features were used as factor variables as opposed to when the dummy variables were used in the model [31].

The data was split into a 75% training dataset and a 25% testing dataset. The training dataset was further split into 10-folds cross-validation groups for hyperparameter tuning. The ROSE algorithm was implemented using a **themis** package [32]. The remaining pre-processing steps were implemented using a **recipes** package [33]. The testing dataset was split into a dense breast dataset and a non-dense breast dataset. Thus, there were three

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

testing datasets available; 1) the whole testing dataset, 2) the dense breast testing dataset and 3) the non-dense breast testing dataset.

2.3.2. Hyperparameter tuning

Two stages of hyperparameter tuning were applied to each machine learning model. The first stage of tuning was using a random grid search with a Latin hypercube grid design of 1000 combinations of hyperparameters. The second stage was using a simulated annealing approach. The best result based on the sensitivity metrics from the first stage was used as an initial parameter for the simulated annealing approach. The simulated annealing algorithm was run for 500 iterations for each machine learning model. The hyperparameter tuning process was done using a finetune package [34].

2.3.3. Performance metrics

A true positive (TP) case was defined as an abnormal case and predicted abnormal by the model, while a true negative (TN) case was a normal case and predicted normal by the model. A false negative (FN) case was an abnormal case but predicted normal by the model, while a false positive (FP) case was a normal case but predicted abnormal by the model. The four performance metrics used to assess the performance of the machine learning models were sensitivity, specificity, precision and precision recall-area under the curve (PR-AUC).

Sensitivity or recall was defined as:

$$\frac{TP}{TP + FN}$$

Specificity was defined as:

$$\frac{TN}{TN + FP}$$

Precision was defined as:

$$\frac{TP}{TP + FP}$$

The four metrics were applied to the whole testing dataset, dense breast testing dataset and non-dense breast testing dataset to assess the final performance of each ML model.

2.4. Explainable approach

Variable importance was assessed as a change in loss function after variable permutations. The loss function was defined as 1 – PR-AUC. The number of permutations was set to 50. Only the top five important variables were displayed in the variable importance plots. The explainable models were implemented using DALEX and DALEXtra packages [35,36].

3. Results

After the removal of duplicates, there were 1091 women who came to the BestARi for breast cancer-related problems. Most of the women who attended to BestARi were Malay, married, did breast self-examination and had no family history of breast cancer. Most women had no symptoms or presentation of lump, nipple discharge, nipple retraction, axillary mass, breast pain and skin changes. Additionally, most women had no history of breast trauma and total abdominal hysterectomy bilateral salpingo-oophorectomy (TABHSO). Details of the characteristics of the data were presented in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

Table 1. Characteristics of women who attended BestARi in Hospital Universiti Sains Malaysia (n = 1091).

Characteristic	Abnormal, n = 861 n (%)	Normal, n = 230 n (%)	Missing values n (%)	Overall, n = 1,0911
Age at examination[1, 2]	53.7 (9.6)	50.0 (8.1)	3 (0.3%)	52.9 (9.4)
Age at menarche[1, 2]	13.1 (1.5)	13.0 (1.5)	97 (8.9%)	13.1 (1.5)
No of children[1, 2]	3.6 (2.4)	3.8 (2.7)	85 (7.8%)	3.7 (2.5)
Weight (kg)[1, 2]	63.5 (12.8)	64.2 (12.9)	263 (24.0%)	63.7 (12.8)
Height (cm)[1]	155.0 (6.4)	156.4 (5.5)	692 (63.0%)	155.2 (6.3)
BMI[1, 2]	26.7 (5.6)	27.1 (5.7)	696 (64.0%)	26.8 (5.6)
Race[2]			34 (3.1%)	
Chinese	112 (13.4%)	21 (9.4%)		133 (12.6%)
Indian	4 (0.5%)	0 (0.0%)		4 (0.4%)
Malay	706 (84.8%)	201 (89.7%)		907 (85.8%)
Others	3 (0.4%)	0 (0.0%)		3 (0.3%)
Siamese	8 (1.0%)	2 (0.9%)		10 (0.9%)
Marriage status[2]			59 (5.4%)	
Divorced	4 (0.5%)	0 (0.0%)		4 (0.4%)
Married	759 (93.1%)	208 (95.9%)		967 (93.7%)
Single	46 (5.6%)	8 (3.7%)		54 (5.2%)
Widowed	6 (0.7%)	1 (0.5%)		7 (0.7%)
Breastfeeding[2]			541 (50.0%)	
No	131 (30.7%)	30 (24.4%)		161 (29.3%)
Yes	296 (69.3%)	93 (75.6%)		389 (70.7%)
Lump[2]			41 (3.8%)	
No	588 (70.9%)	148 (67.0%)		736 (70.1%)
Yes	241 (29.1%)	73 (33.0%)		314 (29.9%)
Nipple discharge[2]			52 (4.8%)	
No	793 (96.4%)	205 (94.9%)		998 (96.1%)
Yes	30 (3.6%)	11 (5.1%)		41 (3.9%)
Nipple retraction[2]			45 (4.1%)	
No	784 (94.8%)	213 (97.3%)		997 (95.3%)
Yes	43 (5.2%)	6 (2.7%)		49 (4.7%)
Axillary mass[2]			55 (5.0%)	
No	764 (93.2%)	203 (94.0%)		967 (93.3%)
Yes	56 (6.8%)	13 (6.0%)		69 (6.7%)
Pain[2]			54 (4.9%)	
No	691 (84.1%)	172 (80.0%)		863 (83.2%)
Yes	131 (15.9%)	43 (20.0%)		174 (16.8%)
Skin changes[2]			55 (5.0%)	
No	772 (94.3%)	204 (94.0%)		976 (94.2%)
Yes	47 (5.7%)	13 (6.0%)		60 (5.8%)
Surgery/implant[2]			76 (7.0%)	
No	531 (65.7%)	143 (69.1%)		674 (66.4%)
Yes	277 (34.3%)	64 (30.9%)		341 (33.6%)
Trauma[2]			108 (9.9%)	
No	754 (96.5%)	191 (94.6%)		945 (96.1%)
Yes	27 (3.5%)	11 (5.4%)		38 (3.9%)
BC-HR[2]			51 (4.7%)	
No	554 (67.6%)	130 (59.1%)		684 (65.8%)
Yes	266 (32.4%)	90 (40.9%)		356 (34.2%)
Previous mammogram[2]			40 (3.7%)	
No	348 (41.9%)	116 (52.5%)		464 (44.1%)
Yes	482 (58.1%)	105 (47.5%)		587 (55.9%)
Breast self-examination[2]			106 (9.7%)	
No	149 (19.3%)	44 (20.9%)		193 (19.6%)
Yes	625 (80.7%)	167 (79.1%)		792 (80.4%)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

Handedness[2]			667 (61.0%)	
Left	20 (5.8%)	6 (7.4%)		26 (6.1%)
Right	323 (94.2%)	75 (92.6%)		398 (93.9%)
Height (cm)[1, 2]	155.0 (6.4)	156.4 (5.5)	692 (63.0%)	155.2 (6.3)
TABHSO[2]			70 (6.4%)	
No	720 (89.4%)	187 (86.6%)		907 (88.8%)
Yes	85 (10.6%)	29 (13.4%)		114 (11.2%)
Mammographic density			0 (0.0%)	
Non-dense	468 (54.4%)	124 (53.9%)		592 (54.3%)
Dense	393 (45.6%)	106 (46.1%)		499 (45.7%)
Family history[2]			520 (48.0%)	
No	352 (79.1%)	101 (80.2%)		453 (79.3%)
Yes	93 (20.9%)	25 (19.8%)		118 (20.7%)
Menopause status[2]			0 (0.0%)	
No	385 (44.7%)	139 (60.4%)		524 (48.0%)
Yes	476 (55.3%)	91 (39.6%)		567 (52.0%)

Notes: BestARi = breast cancer awareness and research unit; Family history = family history of breast cancer; BC-HR = history of birth control or hormone replacement; TAHBSO = history of total abdominal hysterectomy bilateral salpingo-oophorectomy; [1] mean (SD); [2] Features included in the model development.

Table 2 presents the final performance of the screening models after simulated annealing hyperparameter tuning using the training datasets and Table 3 presents the performance of the models on the testing dataset. SVM had the highest sensitivity and the lowest specificity, while the remaining four models had a more balanced performance between sensitivity and specificity. In terms of PR-AUC and precision, elastic-net logistic regression had the highest performance at 0.83 and 0.80, respectively.

Table 2. The final performance of the screening models and the corresponding ML methods after simulated annealing hyperparameter tuning on the training dataset.

ML methods for the screening model	PR-AUC	Precision	Sensitivity	Specificity
Random forest	0.82	0.82	0.64	0.45
Artificial neural network	0.81	0.82	0.64	0.43
Support vector machine	0.82	0.80	0.74	0.29
Elastic-net logistic regression	0.83	0.84	0.53	0.60
XGBoost	0.82	0.81	0.64	0.44

XGBoost=extreme gradient boosting, PR-AUC=precision recall-area under the curve

Table 3. Performance of the screening models and the corresponding ML methods on the testing dataset.

ML methods for the screening models	PR-AUC	Precision	Sensitivity	Specificity
Random forest	0.83	0.80	0.70	0.42
Artificial neural network	0.77	0.76	0.67	0.30
Support vector machine	0.78	0.76	0.83	0.14
Elastic-net logistic regression	0.83	0.80	0.67	0.45
XGBoost	0.84	0.79	0.68	0.41

XGBoost=extreme gradient boosting, PR-AUC=precision recall-area under the curve

Figure 1 illustrates the confusion matrices of each machine learning model on the testing dataset. SVM had the best performance based on Figure 1, despite a lower proportion of true negative cases and a higher proportion of false positive cases compared to other models. As an over-the-counter screening model, the model with the highest proportion of true positive cases and the lowest proportion of false negative cases was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

preferred. Misclassification of an abnormal case as a normal case was undesired as it would delay the consultation of the patients with a high probability of getting breast cancer.

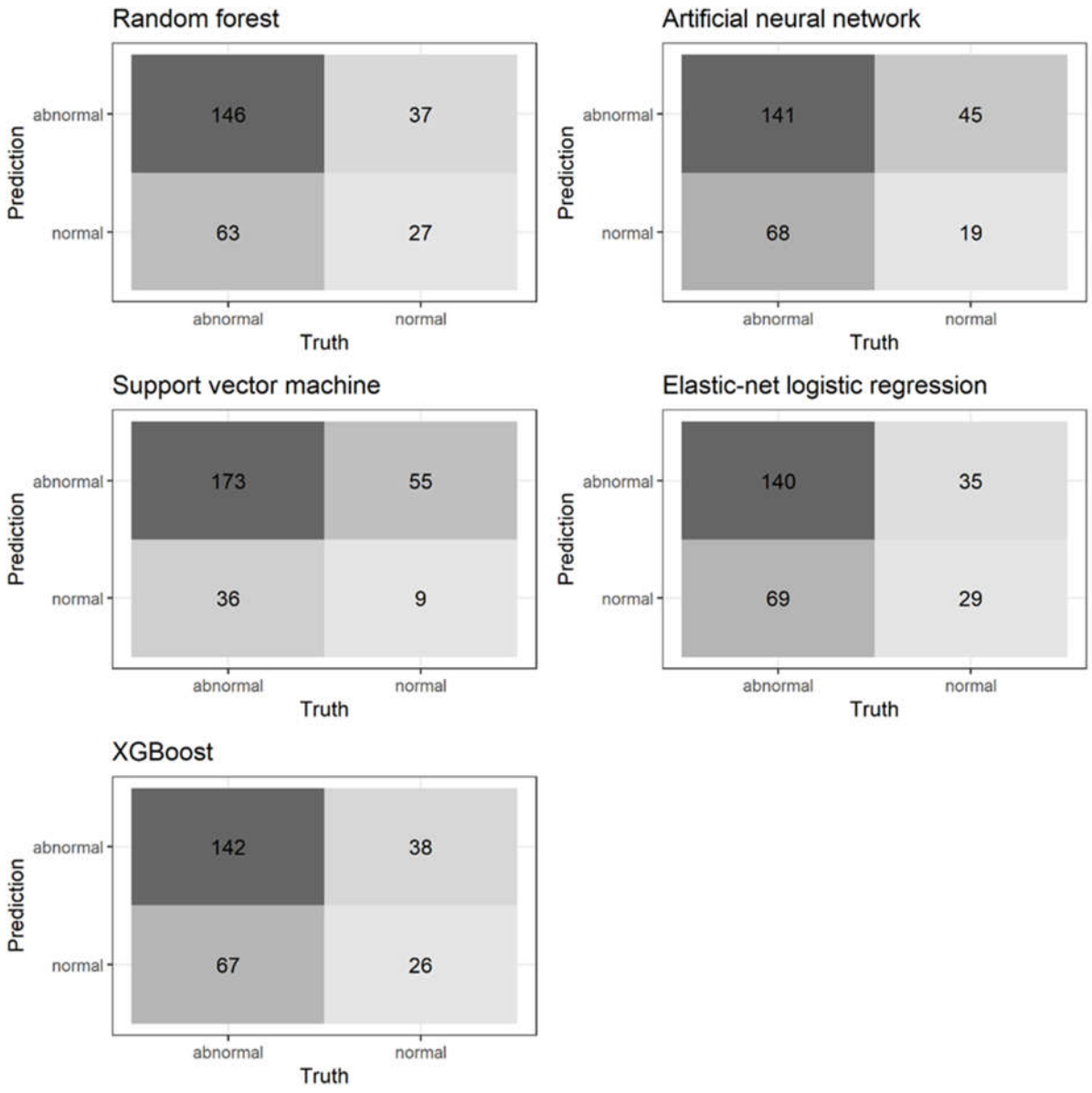

Figure 1. Confusion matrices of the machine learning models on the testing dataset.

Figure 2 and Figure 3 present PR-AUC, sensitivity and specificity metrics for each machine learning model across mammographic density on the testing dataset. There was no difference between PR-AUC of dense and non-dense breast women for each machine learning model. In terms of sensitivity and specificity, SVM had a more balanced performance between dense breast women and non-dense breast women compared to the other four machine learning models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

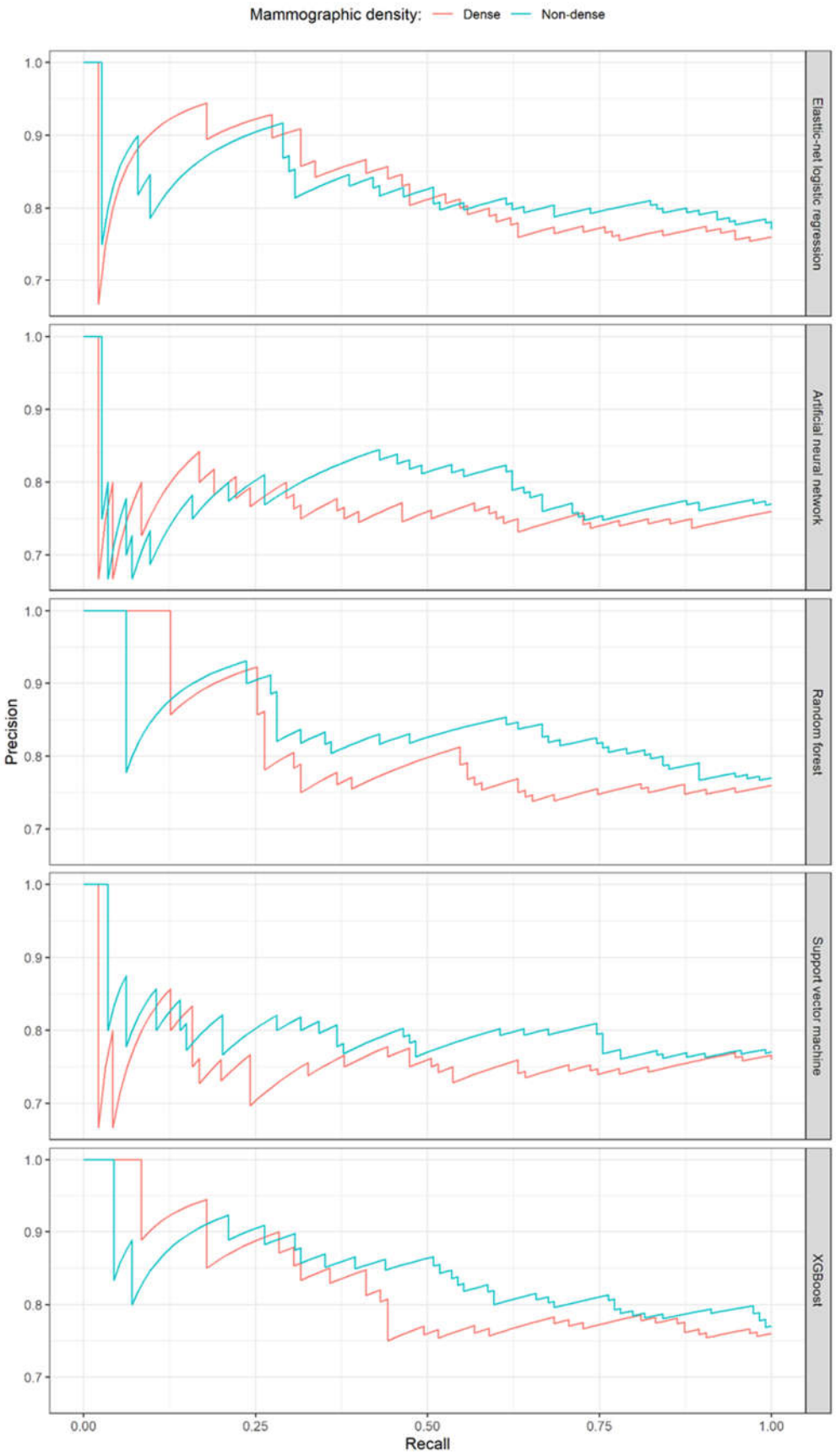

Figure 2. Precision recall – area under the curve for each machine learning model across mammographic density on the testing dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

Figure 3. Specificity and sensitivity for each machine learning model across mammographic density on the testing dataset.

Figure 4 illustrates the top five influential variables according to the five machine learning models. The most influential features across all the machine learning models were the age at examination, followed by weight and number of children. However, the latter two variables were not included in the top five influential variables of ANN and elastic-net logistic regression.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

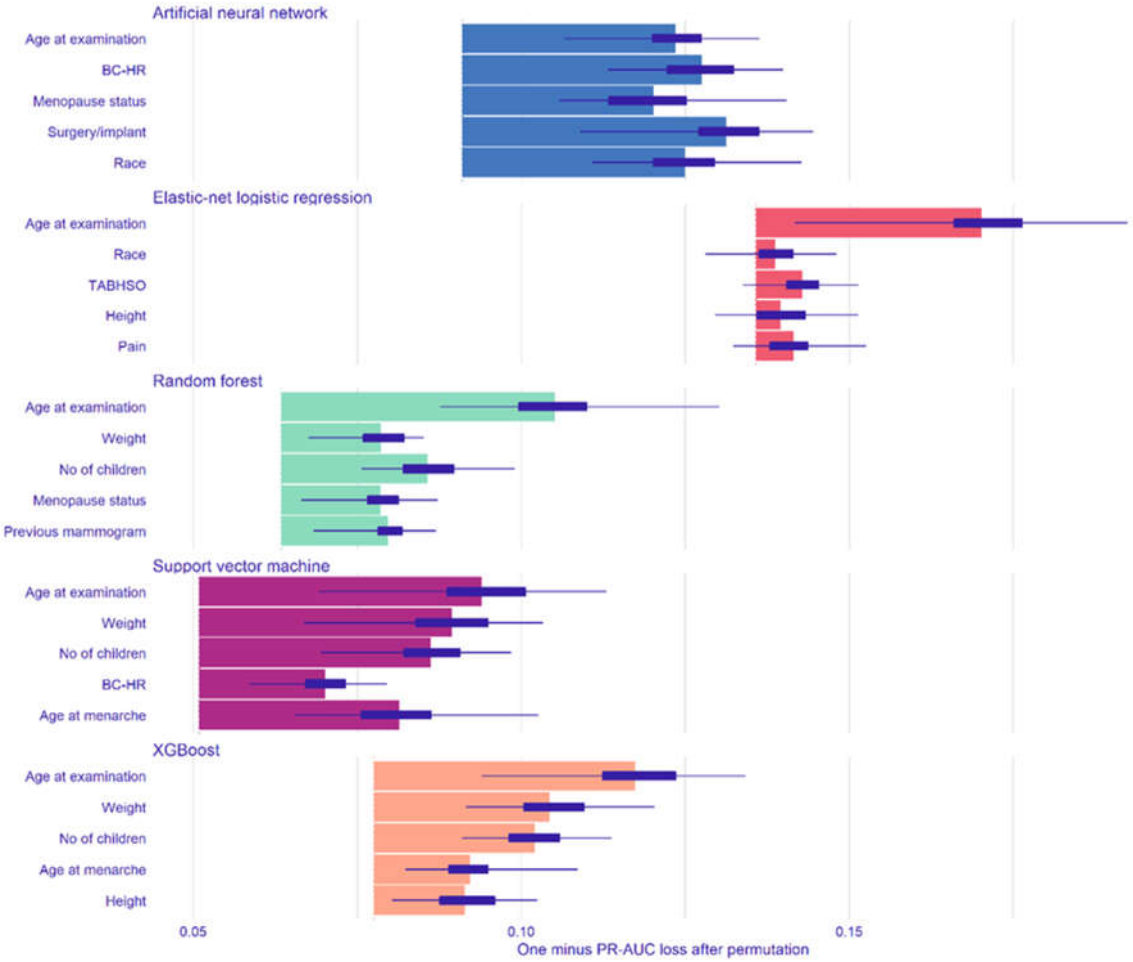

Figure 4. Top five influential variables for each machine learning model. The bar indicates the mean values of one minus PR-AUC and the box plot reflects the distribution of the values of one minus PR-AUC.

4. Discussion

In this study, we evaluated the feasibility of screening models developed from five machine learning methods with the potential use of being the OTC screening model for Asian women in the breast clinic. The model was aimed to predict women with a high probability of developing breast cancer using the information obtained during patients' registration prior to a medical consultation with the physician. Patients with a high probability of getting breast cancer could be prioritised at the screening stage and would be referred to a breast cancer specialist or physician for early consultation and further investigations. Thus, features used for the machine learning models development were limited to the information from the patient's registration form including socio-demographic information and patients' symptoms or complaints. A study conducted to develop machine learning models to predict breast cancer in Chinese women included ten risk factors that achieved the best sensitivity and specificity of 0.66 and 0.69 using extreme gradient boosting [37]. This study achieved the best sensitivity at 0.83 using SVM but with a lower specificity at 0.14. Similarly, XGBoost developed in this study had a higher sensitivity at 0.68 but a lower specificity at 0.41. In our study, the sensitivity metric was more significant as a low sensitivity means a delay in providing medical consultation to patients with a high risk of getting breast cancer. Nonetheless, our study showed that adding patients' symptoms to the features used in the development of the screening model might be beneficial in improving the predictive performance of the screening model. Another study conducted to predict breast cancer using laboratory data showed the best precision

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 June 2022 doi:10.20944/preprints202206.0186.v1

performance at 0.85 using ANN [38] while the best precision performance in our study was at 0.80 using random forest and elastic-net logistic regression. Although the performance of the models in our study was lower, however, getting laboratory data prior to medical consultation was unfeasible and impractical in this study.

Screening models were developed from five machine learning methods in this study. The SVM model showed the best result in terms of sensitivity metrics. Additionally, SVM had a more balanced sensitivity and specificity performance across mammographic density compared to other models (Figure 3). SVM had been shown to work well with imbalanced datasets compared to other machine learning models [39]. Although other machine learning models except for ANN showed a better precision and PR-AUC compared to SVM, we believed the sensitivity metric was more important in the development of the screening model than the precision and PR-AUC. The OTC model aimed to be deployed as an initial screening model prior to the medical consultation. The model would prioritise women with a high probability of breast abnormalities or cancer which in turn accelerates the needed process for those with medical urgency. Previous studies showed that early detection of breast cancer reduces its mortality [40,41]. Additionally, one of the factors of severe breast cancer presentation and poor survival among breast cancer patients was a delay in seeking medical treatment [42–45]. In this study, we developed a screening model with the purpose to minimise the time between a woman first noticing a symptom and getting a medical consultation. At least about 17% of women with breast cancer symptoms in European countries had a delayed medical consultation of at least 3 months or more [46]. In Malaysia, a delay in medical consultation was estimated at 2 months [47]. In general, shortening the delay in getting medical consultations would be beneficial for breast cancer women.

Mammographic density is a known risk factor for developing breast cancer [48]. Asian women had a higher mammographic density than the non-Asian women [49,50], thus, having a higher risk of getting breast cancer. In Malaysia, Chinese women had been shown to have denser breasts than the other races [51,52]. A few studies denoted that the proportion of women who attended mammogram procedures in Malaysia was at least half of them were women with dense breasts [53,54]. A machine learning screening model aimed to be applied to this population should take this information into account. However, it was inappropriate to include the mammographic density as one of the features in the screening model as the density was known at a later process after medical examination. In this study, all machine learning models had a higher sensitivity and lower specificity in a non-dense group compared to a dense group except for SVM for both metrics and XGBoost for specificity (Figure 3). Thus, generally, breast cancer women with non-dense breasts had a better chance to be recognized by the screening model than breast cancer women with dense breasts. In this study, each machine learning model emphasized each feature differently. The most significant feature across machine learning models in this study was the age at examination. Incidence of breast cancer had been shown to increase with age [55]. However, breast cancer presented at a younger age tended to be more aggressive and at a higher stage of cancer [55–58]. Thus, in developing a machine learning screening model, misclassification of abnormal cases as normal cases especially younger women could be a catastrophic error.

This study used secondary data collected from a university- and research-based hospital in Kelantan, Malaysia. The data was further validated by a radiologist and pathologist to ensure the good quality of the data. However, our study still had a few limitations. One of the main limitations of this study was the size of data to develop our screening models. The lack of data was a prevalent issue in the application of machine learning in healthcare [59]. However, this issue was worsened in our study as the dataset had missing values and imbalanced outcome classification. Subsequently, we had used a bagged tree model and ROSE algorithm to overcome these issues, though, undeniably a larger data will further improve our model. Additionally, we had only included one hospital in our study as we primarily used information from patients' registration records which were specific to the BestARi at the time this study was conducted. We believed our

sample was representative of the Malaysian population and surrounding regions, however, future studies should aim to include more hospitals, thus increasing the size of the data. Nonetheless, the challenges and approaches presented in the study reflected a real workflow in the development and application of the machine learning screening model for breast cancer.

5. Conclusions

We developed five screening models based on ML methods that would be useful for the OTC screening process. Our screening models could improve the workflow of breast cancer management and reduce patients' delays in getting further investigations and consultations from the breast cancer team. Out of five machine learning models, the SVM model had better performance metrics, especially in terms of sensitivity, thus, best suited for OTC screening purposes in this research.

Author Contributions: Conceptualization, Tengku Muhammad Hanis, Nur Intan Raihana Ruhaiyem, Wan Nor Arifin and Kamarul Imran Musa; Data curation, Tengku Muhammad Hanis and Nur Intan Raihana Ruhaiyem; Formal analysis, Tengku Muhammad Hanis, Wan Nor Arifin and Kamarul Imran Musa; Funding acquisition, Kamarul Imran Musa; Investigation, Juhara Haron and Wan Faiziah Wan Abdul Rahman; Methodology, Tengku Muhammad Hanis, Nur Intan Raihana Ruhaiyem and Wan Nor Arifin; Project administration, Kamarul Imran Musa; Resources, Juhara Haron and Wan Faiziah Wan Abdul Rahman; Supervision, Juhara Haron and Rosni Abdullah; Validation, Juhara Haron, Wan Faiziah Wan Abdul Rahman and Rosni Abdullah; Visualization, Tengku Muhammad Hanis and Kamarul Imran Musa; Writing – original draft, Tengku Muhammad Hanis; Writing – review & editing, Nur Intan Raihana Ruhaiyem, Wan Nor Arifin, Juhara Haron and Kamarul Imran Musa.

Funding: This research was funded by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education, Malaysia (FRGS/1/2019/SKK03/USM/02/1).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the human research ethics committee of Universiti Sains Malaysia (JEPeM) on 19th November 2019 (USM/JEPeM/19090536).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of this study and the use of secondary data.

Data Availability Statement: The data are available upon reasonable request to the corresponding author.

Acknowledgments: We thank all staff and workers in the Department of Radiology, Department of Pathology, and BestARi in HUSM for facilitating the data collection and extraction process.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Toh, T.S.; Dondelinger, F.; Wang, D. Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine* **2019**, *47*, 607–615, doi:10.1016/j.ebiom.2019.08.027.
2. Blasiak, A.; Khong, J.; Kee, T. CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence. *SLAS Technol.* **2020**, *25*, 95–105, doi:10.1177/2472630319890316.
3. Poon, A.I.F.; Sung, J.J.Y. Opening the black box of AI-Medicine. *J. Gastroenterol. Hepatol.* **2021**, *36*, 581–584, doi:10.1111/jgh.15384.
4. Johnson, K.B.; Wei, W.Q.; Weeraratne, D.; Frisse, M.E.; Misulis, K.; Rhee, K.; Zhao, J.; Snowdon, J.L. Precision Medicine, AI, and the Future of Personalized Health Care. *Clin. Transl. Sci.* **2021**, *14*, 86–93, doi:10.1111/cts.12884.
5. DeSantis, C.E.; Bray, F.; Ferlay, J.; Lortet-Tieulent, J.; Anderson, B.O.; Jemal, A. International variation in female breast cancer incidence and mortality rates. *Cancer Epidemiol. Biomarkers Prev.* **2015**, *24*, 1495–1506, doi:10.1158/1055-9965.EPI-15-0535.
6. WHO Breast cancer Available online: https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed on May 24, 2022).
7. Unger-Saldaña, K. Challenges to the early diagnosis and treatment of breast cancer in developing countries. *World J. Clin. Oncol.* **2014**, *5*, 465–477, doi:10.5306/wjco.v5.i3.465.
8. Andersen, B.L.; Cacioppo, J.T.; Roberts, D.C. Delay in seeking a cancer diagnosis: Delay stages and psychophysiological comparison processes. *Br. J. Soc. Psychol.* **1995**, *34*, 33–52, doi:10.1111/j.2044-8309.1995.tb01047.x.

9. Taib, N.A.; Yip, C.H.; Low, W.Y. A grounded explanation of why women present with advanced breast cancer. *World J. Surg.* **2014**, *38*, 1676–1684, doi:10.1007/s00268-013-2339-4.
10. McKenzie, F.; Zietsman, A.; Galukande, M.; Anele, A.; Adisa, C.; Parham, G.; Pinder, L.; Cubasch, H.; Joffe, M.; Kidaaga, F.; et al. Drivers of advanced stage at breast cancer diagnosis in the multicountry African breast cancer – disparities in outcomes (ABC-DO) study. *Int. J. Cancer* **2018**, *142*, 1568–1579, doi:10.1002/ijc.31187.
11. Ren, S.; Zhang, Y.; Qin, P.; Wang, J. Factors Influencing Total Delay of Breast Cancer in Northeast of China. *Front. Oncol.* **2022**, *12*, 10–15, doi:10.3389/fonc.2022.841438.
12. Norsa'adah, B.; Rampal, K.G.; Amin, R.M. Time Taken For Symptom Recognition, First Consultation, Diagnosis And First Definitive Treatment And Its Associated Factors Among Women With Breast Cancer. *Asian Pacific J. Cancer Prev.* **2021**, *22*, 3623–3631, doi:10.31557/APJCP.2021.22.11.3623.
13. Warner, E.T.; Rice, M.S.; Zeleznik, O.A.; Fowler, E.E.; Murthy, D.; Vachon, C.M.; Bertrand, K.A.; Rosner, B.A.; Heine, J.; Tamimi, R.M. Automated percent mammographic density, mammographic texture variation, and risk of breast cancer: a nested case-control study. *npj Breast Cancer* **2021**, *7*, doi:10.1038/s41523-021-00272-2.
14. Burton, A.; Maskarinec, G.; Perez-Gomez, B.; Vachon, C.; Miao, H.; Lajous, M.; López-Ridaura, R.; Rice, M.; Pereira, A.; Garmendia, M.L.; et al. Mammographic density and ageing: A collaborative pooled analysis of cross-sectional data from 22 countries worldwide. *PLoS Med.* **2017**, *14*, 1–20, doi:10.1371/journal.pmed.1002335.
15. Sherratt, M.J.; McConnell, J.C.; Streuli, C.H. Raised mammographic density: Causative mechanisms and biological consequences. *Breast Cancer Res.* **2016**, *18*, 1–9, doi:10.1186/S13058-016-0701-9.
16. Nazari, S.S.; Mukherjee, P. An overview of mammographic density and its association with breast cancer. *Breast Cancer* **2018**, *25*, 259–267, doi:10.1007/s12282-018-0857-5.
17. Anisha, P.R.; Kishor Kumar Reddy, C.; Apoorva, K.; Meghana Mangipudi, C. Early Diagnosis of Breast Cancer Prediction using Random Forest Classifier. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1116*, 012187, doi:10.1088/1757-899X/1116/1/012187.
18. Yue, W.; Wang, Z.; Chen, H.; Payne, A.; Liu, X. Machine learning with applications in breast cancer diagnosis and prognosis. *Designs* **2018**, *2*, 1–17, doi:10.3390/designs2020013.
19. Kabiraj, S.; Raihan, M.; Alvi, N.; Afrin, M.; Akter, L.; Sohagi, S.A.; Podder, E. Breast cancer risk prediction using XGBoost and random forest algorithm. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT); IEEE, 2020; pp. 1–4.
20. Nindrea, R.D.; Aryandono, T.; Lazuardi, L.; Dwiprahasto, I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: A meta-analysis. *Asian Pacific J. Cancer Prev.* **2018**, *19*, 1747–1752, doi:10.22034/APJCP.2018.19.7.1747.
21. Akrami, M.; Arasteh, P.; Eghbali, T.; Shahraki, H.R.; Tahmasebi, S.; Zangouri, V.; Rezaianzadeh, A.; Talei, A. Introducing novel and comprehensive models for predicting recurrence in breast cancer using the group LASSO approach: Are estimates of early and late recurrence different? *World J. Surg. Oncol.* **2018**, *16*, doi:10.1186/s12957-018-1489-0.
22. Kuhn, M.; Vaughan, D. parsnip: A Common API to Modeling and Analysis Functions 2022.
23. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. kernlab - An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, doi:10.18637/jss.v011.i09.
24. Wright, M.N.; Ziegler, A. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, doi:10.18637/jss.v077.i01.
25. Ripley, B.; Venables, W. nnet: Feed-forward neural networks and multinomial log-linear models 2022.
26. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, doi:10.18637/jss.v033.i01.
27. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. xgboost: Extreme Gradient Boosting 2022.
28. R Core Team R: A Language and Environment for Statistical Computing 2022.
29. Lunardon, N.; Menardi, G.; Torelli, N. ROSE: A package for binary imbalanced learning. *R J.* **2014**, *6*, 79–89, doi:10.32614/rj-2014-008.
30. Yeo, I.-K.; Johnson, R.A. A new family of power transformations to improve normality or symmetry. *Biometrika* **2000**, *87*, 954–959, doi:10.1093/biomet/87.4.954.
31. Kuhn, M.; Johnson, K. *Feature engineering and selection: a practical approach for predictive models*; 1st ed.; CRC Press, 2020; ISBN 9781315108230.
32. Hvitfeldt, E. themis: Extra Recipes Steps for Dealing with Unbalanced Data 2022.
33. Kuhn, M.; Wickham, H. recipes: preprocessing and feature engineering steps for modeling 2022.
34. Kuhn, M. finetune: additional functions for model tuning 2022.
35. Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *J. Mach. Learn. Res.* **2018**, *19*, 1–5.
36. Maksymiuk, S.; Gosiewska, A.; Biecek, P. Landscape of R packages for eXplainable Artificial Intelligence. *arXiv Prepr. arXiv2009.13248* **2020**, doi:10.48550/arXiv.2009.13248.
37. Hou, C.; Zhong, X.; He, P.; Xu, B.; Diao, S.; Yi, F.; Zheng, H.; Li, J. Predicting breast cancer in chinese women using machine learning techniques: algorithm development. *JMIR Med. Informatics* **2020**, *8*, doi:10.2196/17364.
38. Khatun, T.; Utsho, M.M.R.; Islam, M.A.; Zohura, M.F.; Hossen, M.S.; Rimi, R.A.; Anni, S.J. Performance Analysis of Breast Cancer: A Machine Learning Approach. *Proc. 3rd Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2021* **2021**, 1426–1434, doi:10.1109/ICIRCA51532.2021.9544879.

39. Sun, Y.; Wong, A.K.C.; Kamel, M.S. Classification of imbalanced data: a review. *Int. J. Pattern Recognit. Artif. Intell.* **2009**, *23*, 687–719, doi:10.1142/S0218001409007326.
40. Malvezzi, M.; Carioli, G.; Bertuccio, P.; Boffetta, P.; Levi, F.; La Vecchia, C.; Negri, E. European cancer mortality predictions for the year 2019 with focus on breast cancer. *Ann. Oncol.* **2019**, *30*, 781–787, doi:10.1093/annonc/mdz051.
41. Tahmooresi, M.; Afshar, A.; Bashari Rad, B.; Nowshath, K.B.; Bamiah, M.A. Early detection of breast cancer using machine learning techniques. *J. Telecommun. Electron. Comput. Eng.* **2018**, *10*, 21–27.
42. Khan, T.M.; Leong, J.P.Y.; Ming, L.C.; Khan, A.H. Association of knowledge and cultural perceptions of Malaysian women with delay in diagnosis and treatment of breast cancer: A systematic review. *Asian Pacific J. Cancer Prev.* **2015**, *16*, 5349–5357, doi:10.7314/APJCP.2015.16.13.5349.
43. Mujar, N.M.M.; Dahlui, M.; Emran, N.A.; Hadi, I.A.; Wai, Y.Y.; Arulanantham, S.; Hooi, C.C.; Taib, N.A.M. Complementary and alternative medicine (CAM) use and delays in presentation and diagnosis of breast cancer patients in public hospitals in Malaysia. *PLoS One* **2017**, *12*, doi:10.1371/journal.pone.0176394.
44. Caplan, L. Delay in breast cancer: Implications for stage at diagnosis and survival. *Front. Public Heal.* **2014**, *2*, 1–5, doi:10.3389/fpubh.2014.00087.
45. Freitas, A.G.Q.; Weller, M. Patient delays and system delays in breast cancer treatment in developed and developing countries. *Cien. Saude Colet.* **2015**, *20*, 3177–3189, doi:10.1590/1413-812320152010.19692014.
46. Innos, K.; Padrik, P.; Valvere, V.; Eelma, E.; Kütner, R.; Lehtsaar, J.; Tekkel, M. Identifying women at risk for delayed presentation of breast cancer: A cross-sectional study in Estonia. *BMC Public Health* **2013**, *13*, doi:10.1186/1471-2458-13-947.
47. Norsa'adah, B.; Rampal, K.G.; Rahmah, M.A.; Naing, N.N.; Biswal, B.M. Diagnosis delay of breast cancer and its associated factors in Malaysian women. *BMC Cancer* **2011**, *11*, doi:10.1186/1471-2407-11-141.
48. Johnson, R.H.; Anders, C.K.; Litton, J.K.; Ruddy, K.J.; Bleyer, A. Breast cancer in adolescents and young adults. *Pediatr. Blood Cancer* **2018**, *65*, 1–9, doi:10.1002/pbc.27397.
49. Rajaram, N.; Mariapun, S.; Eriksson, M.; Tapia, J.; Kwan, P.Y.; Ho, W.K.; Harun, F.; Rahmat, K.; Czene, K.; Taib, N.A.M.; et al. Differences in mammographic density between Asian and Caucasian populations: a comparative analysis. *Breast Cancer Res. Treat.* **2017**, *161*, 353–362, doi:10.1007/s10549-016-4054-y.
50. Yap, Y.S.; Lu, Y.S.; Tamura, K.; Lee, J.E.; Ko, E.Y.; Park, Y.H.; Cao, A.Y.; Lin, C.H.; Toi, M.; Wu, J.; et al. Insights into Breast Cancer in the East vs the West: A Review. *JAMA Oncol.* **2019**, *5*, 1489–1496, doi:10.1001/jamaoncol.2019.0620.
51. Kumari Chelliah, K.; Shatirah Mohd Fandi Voon, N.; Ahamad, H. Breast Density: Does It Vary among the Main Ethnic Groups in Malaysia? *Open J. Med. Imaging* **2013**, *03*, 105–109, doi:10.4236/ojmi.2013.34017.
52. Mariapun, S.; Li, J.; Yip, C.H.; Taib, N.A.M.; Teo, S.H. Ethnic differences in mammographic densities: An Asian cross-sectional study. *PLoS One* **2015**, *10*, 1–13, doi:10.1371/journal.pone.0117568.
53. Hanis, T.M.; Arifin, W.N.; Haron, J.; Wan Abdul Rahman, W.F.; Ruhaiyem, N.I.R.; Abdullah, R.; Musa, K.I. Factors Influencing Mammographic Density in Asian Women: A Retrospective Cohort Study in the Northeast Region of Peninsular Malaysia. *Diagnostics* **2022**, *12*, 860, doi:10.3390/diagnostics12040860.
54. Rahayu, A.; Zaharuddin, B.; Le, T.Q.; Rifhana, I.; Muhamad, B.; Mahmud, R.; Hamid, S.A. Relation of Breast Density with Age and Ethnicity in Malaysia. *Front. Heal. Informatics* **2013**, *2*, 1–4.
55. McGuire, A.; Brown, J.A.L.; Malone, C.; McLaughlin, R.; Kerin, M.J. Effects of age on the detection and management of breast cancer. *Cancers (Basel).* **2015**, *7*, 908–929, doi:10.3390/cancers7020815.
56. Murphy, B.L.; Day, C.N.; Hoskin, T.L.; Habermann, E.B.; Boughey, J.C. Adolescents and Young Adults with Breast Cancer have More Aggressive Disease and Treatment Than Patients in Their Forties. *Ann. Surg. Oncol.* **2019**, *26*, 3920–3930, doi:10.1245/s10434-019-07653-9.
57. Momenimovahed, Z.; Salehiniya, H. Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer Targets Ther.* **2019**, *11*, 151–164, doi:10.2147/BCTT.S176070.
58. Tao, Z.Q.; Shi, A.; Lu, C.; Song, T.; Zhang, Z.; Zhao, J. Breast Cancer: Epidemiology and Etiology. *Cell Biochem. Biophys.* **2015**, *72*, 333–338, doi:10.1007/s12013-014-0459-6.
59. Ahmed, Z.; Mohamed, K.; Zeeshan, S.; Dong, X.Q. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database* **2020**, *2020*, 1–35, doi:10.1093/database/baaa010.