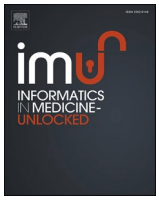
Informatics in Medicine Unlocked 41 (2023) 101317

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning

Rolando Gonzales Martinez [a,*], Daan-Max van Dongen [b]

[a] Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands Interdisciplinary Demographic Institute (NIDI), University of Groningen (RUG), the Netherlands
[b] University College London, United Kingdom

ARTICLE INFO

Keywords:
Deep learning
Machine learning
Breast cancer
Screening
Pre-screening
False detection

ABSTRACT

Deep learning has been widely applied in breast cancer screening to analyze images obtained from X-rays, ultrasound, magnetic resonances, and biopsies. This study suggests that deep learning can also be used to prescreen for cancer by analyzing heterogeneous data obtained from demographic and anthropometric information of patients, biological markers from routine blood samples, and relative risks from meta-analysis and international databases. In this document, feature selection is applied to a database of 64 women diagnosed with breast cancer and a counterfactual group of 52 healthy women, to identify the best predictors of cancer prescreening. The best predictors are used in k-fold Monte Carlo cross-validation experiments that compare deep learning against machine learning. The results indicate that a deep learning architecture that is fine-tuned using feature selection has the lowest false negative rate (i.e., the lowest Type II errors) and can effectively distinguish between patients with and without cancer. Consequently, deep learning—compared to traditional machine learning—promotes a more accurate detection of malignancies, hence reducing the risk of increased tumor size and cancer spreading to nearby or distant lymph nodes, tissues, or organs, due to late detection. Additionally, compared to machine learning, deep learning has the lowest uncertainty in its predictions, as indicated by the lowest standard deviation of its performance metrics. These findings indicate that deep learning algorithms applied to cancer prescreening offer a radiation-free, non-invasive, and an affordable complement to screening methods based on imagery. The implementation of deep learning algorithms in cancer prescreening helps to identify individuals who may require imaging-based screening, can encourage self-examination, and decreases the psychological drawbacks associated with false positives in cancer screening, ultimately leading to earlier detection of malignancy and reducing the healthcare and societal burden associated with cancer treatment.

1. Introduction

Breast cancer is the most commonly diagnosed cancer worldwide. The study by the World Health Organization (WHO) on the current and future burden of breast cancer estimated more than 2.26 million new cases of breast cancer for 2020 [1], with Belgium and the Netherlands having the highest age standardized incidence of breast cancer and developing countries such as Somalia and Syria having the highest mortality from breast cancer. A more recent study [2], indicates that breast cancer is the most common cancer diagnosed in women in 2023, accounting for 31% of female cancers. According to Ref. [2]; female breast cancer incidence rates have increased by approximately 0.5% per year since the mid-2000s, a trend that has been attributed at least in part to increases in excess body weight [3].

Machine learning and deep learning algorithms are frequently applied for breast cancer screening in developed countries [4]. These algorithms are applied to predict the presence of anomalies related to the presence breast cancer in digitalized images [5–7] obtained from magnetic resonance imaging, ultrasounds [8], digital breast tomosynthesis [9], breast density from mammograms [10], tissue images [11], or cell nuclei from fine needle aspirates of breast mass [12,13].

This study proposes that deep learning algorithms can also be applied to the prognosis of breast cancer through the application of deep learning to heterogeneous data obtained from medical records (demographic and anthropometric information), biological markers obtained from routine blood samples, and relative risks obtained from

* Corresponding author.
E-mail address: r.m.gonzales.martinez@rug.nl (R. Gonzales Martinez).

https://doi.org/10.1016/j.imu.2023.101317
Received 3 April 2023; Received in revised form 15 July 2023; Accepted 2 August 2023
Available online 6 August 2023

meta-analysis and publicly available databases. The development and progression of breast cancer has been linked for example to glucose dysmetabolism, insulin resistance and changes in adipokine secretion [14], and hence predictive models of breast cancer based on biological markers (glucose, insulin, HOMA, leptin, adiponectin, resistin and MCP-1) can be applied as prescreening tools for the early detection of breast cancer, particularly when the information of biological markers is combined with the demographic and anthropometric information of patients and their relative risk by age and sex according to body mass index (BMI). Although the carcinogenic mechanisms that related overweight and obesity to breast cancer are still uncertain, the positive association between high BMI and breast cancer risk in postmenopausal women was speculated to be the result of the higher level of estrogen derived from the aromatization of androstenedione within the larger fat reserves of women with high BMI [15]. The tetrad of BMI, leptin, the ratio of leptin/adiponectin and the antigen 15.3 (CA15-3), are considered reliable biomarkers of breast cancer [16], and resistin levels tend to be significantly elevated in postmenopausal breast cancer, after adjusting for demographic, metabolic, and clinicopathological features [17].

This study evaluates the ability of machine learning and deep learning to predict the presence of breast cancer with prescreening information. Machine learning and deep learning algorithms were applied to the information of 116 women, 52 healthy women, and 64 women who were diagnosed with breast cancer at the University Hospital Center of Coimbra (UHCC). The UHCC data set was extended with information on relative risks of breast cancer obtained from the Global Burden of Disease Study 2019 [18] and the relative risks of the meta-analysis of [19]. The novelty of this study is that it shows that the combination of primary and secondary data sets can lead to an improvement in the early detection of breast cancer if feature selection algorithms are combined with deep learning, and if metrics to evaluate the performance of algorithms are used to assess the predictive models not only in a statistical sense, but also in a medical sense, as it is the case for example, of the false negative rate, that signals the rate of cases of cancer that were not detected by the algorithms.

In relation to feature selection, the optimal predictors of breast cancer were chosen with the SULOV-gradient boosting algorithm. The SULOV algorithm reduced the full set of potential explanatory predictors to a subset of features with the highest relevance and the lowest redundancy for predicting breast cancer. The optimal set of predictors was used in a k-fold Monte Carlo cross-validation experiment that compared deep learning against seven traditional machine learning algorithms: support vector machines, neural networks, logistic regression, XGBoost, random forests, naive Bayes, and stochastic gradient algorithms. The results of evaluating the algorithms indicate that—compared to machine learning—deep learning has the highest predictive accuracy, the lowest false negative and false omission rates, and the highest precision in terms of having the lowest standard errors of performance metrics. These findings indicate that a prescreening medical recommendation system based on deep learning algorithms has the potential of being a non-invasive, radiation-free and affordable support for AI-based clinical decision making which can complement and guide cancer screening.

2. Materials and methods

2.1. Data

The data of the University Hospital Center of Coimbra (UHCC) contains information of 116 women, 52 healthy women, and 64 women who were diagnosed with breast cancer. The information of UHCC was collected between 2009 and 2013 for a study of biomarkers of breast cancer, based on the results of routine blood analysis. Demographic and anthropometric information from the patients was recorded during the first consultation of the patients with the physician [20]. The demographic information in the dataset is the age of women and the anthropometric information is the BMI of women. The diagnosis of breast cancer was the result of a mammography and was confirmed histologically with samples collected before surgery or treatment. The samples of tumor tissue were obtained by mastectomy or tumourectomy and were evaluated by a pathologist at the Anatomic Pathology Department of UHCC. The counterfactual group of women without breast cancer in the database were healthy volunteers that were enrolled in the study as controls. Both the patients with breast cancer and the control group of healthy women had no prior cancer treatment and were free from infections, acute diseases, or comorbidities at the time of participating in the study.

Blood samples of women with breast cancer and the control group of healthy women were extracted in the Laboratory of Physiology of the Faculty of Medicine of the University of Coimbra from peripheral venous blood vials. Blood samples were collected after an overnight fasting. The fasting blood was centrifuged (2500 g) at 4 °C and stored at −80 °C for biochemical determinations of serum glucose levels (mg/dL), insulin levels (μ U/mL), and serum values of leptin (ng/mL), adiponectin (μ g/mL), resistin (ng/mL), and chemokine monocyte chemoattractant Protein 1 (MCP-1, (pg/dL)). A Homeostasis Model Assessment (HOMA) index that measures insulin resistance was calculated with the information of the fasting insulin level (μ U/mL) and fasting glucose level (mmol/L).

During the data pre-processing stage, the information of the UHCC was extended with data of relative risks of breast cancer obtained from the Global Burden of Disease (GBD) Study 2019 [18] and the relative risks of breast cancer associated to BMI from the dose-response meta--analysis of [19]. The GBD Study 2019 was coordinated by the Institute for Health Metrics and Evaluation, who estimated the burden of diseases, injuries, and risk factors for 204 countries, territories and selected subnational locations worldwide. The GBD database contains information of relative risks of cancer for high BMI (BMI $\geq$ 25), by age and sex. The dose-response meta-analysis of [19] was based on 12 prospective cohort studies comprising 22,728,674 participants. The results of [19] show that every 5 kg/m2 increase in BMI corresponds to a 2% increase in breast cancer risk in women, with differential results for premenopausal women, for which higher BMI could be a protective factor in breast cancer risk (Fig. 1). Table 1 shows the descriptive statistics of the complete data set used in this study; this is, the original UHCC data that was extended with relative risks (RRs) from publicly available databases and studies. The combination of the primary and secondary data sets was performed by using the age of the patients as the key merging variable, since the age of patients is the common information available in the 3 data sets: the UHCC, the GBD data of relative risks, and the relative risks of the dose-response meta-analysis of [19].

2.2. Methods

Experiments using k-fold Monte Carlo cross-validation were conducted to evaluate the ability that machine learning and deep learning classifiers have to differentiate between patients with and without cancer, based on demographic and anthropometric information, as well as biomarkers and relative risks in relation to age and body mass index. Feature selection was applied with the SULOV algorithm to determine the optimal set of predictors for identifying breast cancer.

Deep learning is based on multiple layers of artificial neural networks that are densely-connected. Each layer transforms the input data from the previous layer into a new representation trough non-linear functions connected by synaptic weights [21]. This allows the deep learning model to learn both locally and in the inter-relationships of the whole data, through a hierarchical structure [22]. As deep learning can handle imbalanced, heterogeneous, and high dimensional data, and is robust to changes in the input data due its multiple hidden layers, it is at the center of artificial intelligence and is increasingly being used to mine large and complex relationships hidden in biomedical data [23,24]. Previous applications of deep learning in biomedical science include

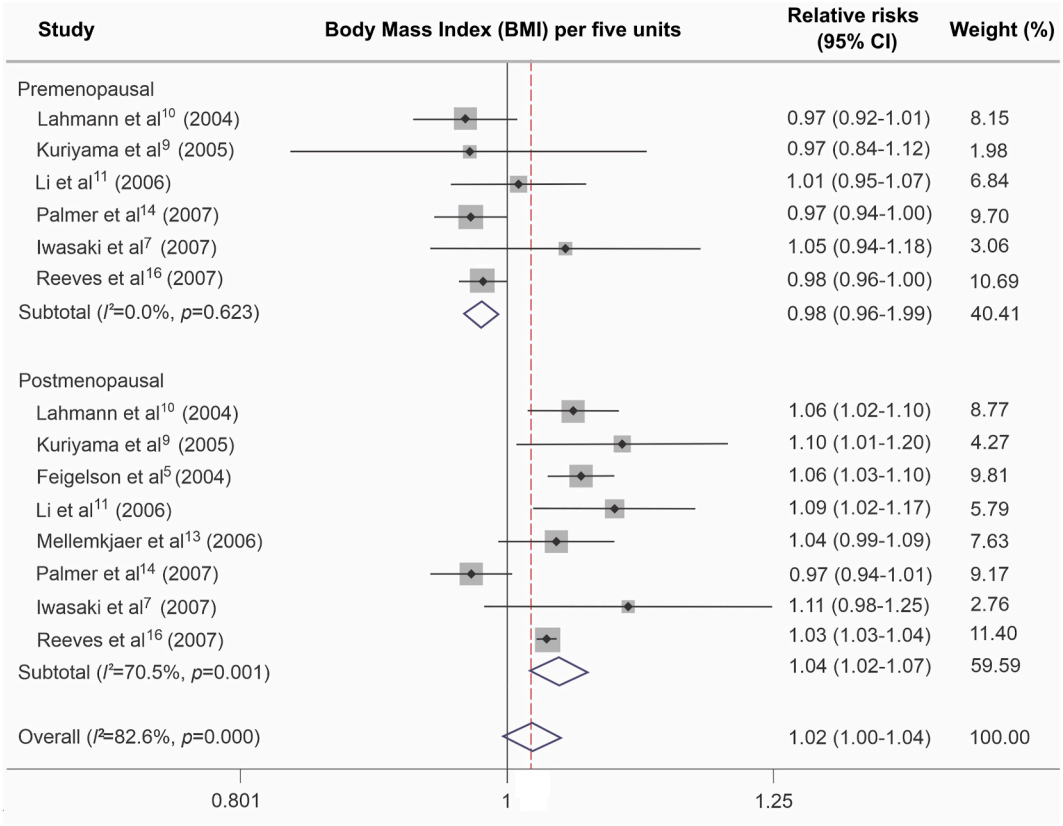


Fig. 1. Subgroup meta-analysis of the association between BMI increment (per five units) and breast cancer risk, by menopausal status. Reproduced from Ref. [19].

Table 1
Descriptive statistics of the database.

	mean	std.dev.	min	max
Age (years)	57.30	16.11	24	89
BMI	27.58	5.02	18.37	38.58
Glucose (mg/dL)	97.79	22.53	60.00	201.00
Insulin (μ U/mL)	10.01	10.07	2.43	58.46
HOMA	2.69	3.64	0.47	25.05
Leptin (ng/mL)	26.62	19.18	4.31	90.28
Adiponectin (μ g/mL)	10.18	6.84	1.66	38.04
Resistin (ng/mL)	14.73	12.39	3.21	82.10
RRs [19]	1.51	0.48	0.40	2.44
RRs GBD (center)	1.01	0.10	0.89	1.09
RRs GBD (lower)	0.97	0.08	0.87	1.04
RRs GBD (upper)	1.05	0.11	0.91	1.14
High BMI (binary)	0.34	0.47	0	1
Obesity (binary)	0.32	0.47	0	1

image reconstruction and the automated interpretation of downstream images in biomedical optics [25], segmentation of medical images in diseased and healthy regions [22,26], photoacoustic imaging of chromophores [27], gene expression [28], protein structure prediction [29], brain machine interfaces [30], and clinical language processing [31]. The deep learning architectures applied in this study have rectified networks and L1 and L2 regularization in the kernel, bias and activation functions of the dense hidden layers.

A recursive minimum redundancy maximum relevance (MRMR) algorithm of SULOV-gradient boosting was applied for feature selection [32]. Feature selection is the process of reducing the number of input variables when developing a predictive model [33]. Feature selection is a preprocessing step that improves model performance by identifying the variables that are the most relevant to the prediction task, consequentially removing irrelevant and redundant variables from the set of potential predictors. Irrelevant features are those that can be removed without affecting learning performance, while redundant features are correlated features that are individually relevant, but due to their copresence the removal of one of them does not affect learning performance [34]. MRMR finds the minimal-optimal subset of features through the selection of highly predictive but uncorrelated features. The algorithm starts calculating the mutual information score of all the pairs of highly correlated variables and then selects the ones with the highest Information scores and least correlation with each other. This first step is called the SULOV (Searching for Uncorrelated List Of Variables) algorithm. In the second step, XGBoost is used to repeatedly find the best features to predict the target variable in a test sample using the model estimated in the train sample with the variables selected with SULOV in step 1. XGBoost is a regularized gradient boosting algorithm based on boosted tree ensembles [35] that has proven to be highly successful solving a vast array of predictive problems due to its ability to handle the bias-variance trade-o [36]. In biomedical science, the MRMR algorithm has been used before for the feature selection of temporal gene expression data [37,38].

The performance of the machine learning and deep learning classifiers was evaluated using the Area Under the receiver operating characteristic Curve (AUC), besides metrics derived from the confusion matrix. The performance of the deep learning algorithm was compared against the performance of seven machine learning algorithms: XGBoost, stochastic gradient, support vector machines, random forests, neural networks, naive Bayes and logistic regression. AUC measures the ability of the model to distinguish between positive and negative classes; in this case, the presence of breast cancer (positive class) or the absence of breast cancer (negative class). The AUC ranges from 0 to 1, with a value of 1 indicating a perfect classifier and a value of 0.5 indicating a classifier no better than random guessing. AUC is particularly useful when the distribution of classes in the target variable is imbalanced, because in the presence of unbalanced class distribution, a model could achieve high accuracy by simply predicting the majority class most of

the time. AUC is a better metric in these cases because it considers both the true positive rate and the false positive rate of the predictive model.

Besides AUC, the performance of the deep learning algorithms was compared against machine learning models with metrics derived from the confusion matrix. Specifically, the true positive rate (TPR), the true negative rate (TNR), the false negative rate (FNR), the false positive rate (FPR), the false detection rate (FDR), and the false omission rate (FOR). TPR or sensitivity measures the proportion of actual positive cases of breast cancer correctly identified. It is the ratio of true positives to the sum of true positives and false negatives. TPR quantifies the ability of machine learning and deep learning models to correctly identify positive cases of breast cancer. TNR or specificity measures the proportion of actual negative cases correctly identified. It is the ratio of true negatives to the sum of true negatives and false positives. TNR quantifies the ability to correctly identify negative instances, this is, the absence of breast cancer. FNR measures the proportion of actual positive cases of breast cancer that are incorrectly classified as negative cases. It is the ratio of false negatives to the sum of true positives and false negatives. FNR quantifies the rate of missed positives (Type II errors) and it is particularly relevant in the case of breast cancer, since a good model should have the lowest possible FNR for the early detection of breast cancer, since a high missed rate implies late detection of anomalies. The FPR quantifies the proportion of negative cases of breast cancer that are incorrectly classified as positive. It measures the rate at which machine learning and deep learning algorithms produce false positive results. The FPR is calculated as the ratio of false positives to the sum of true negatives and false positives. The FDR measures the proportion of false positive detection of breast cancer. It is the ratio of false positives of breast cancer to the sum of true positives and false positives. FOR measures the proportion of false negative errors or incorrect omissions in a decision-making process. It is calculated as the ratio of false negatives to the sum of true negatives and false negatives. Since FOR captures failures to detect breast cancer, is also relevant in the comparison of machine learning and deep learning algorithms, because missing the detection of a positive condition of breast cancer can have significant health consequences. In summary, TPR and TNR represent the correct identification rates for positive and negative cases of breast cancer, respectively, while FNR and FPR represent the rates of misclassification for negative and positive cases. FDR and FOR specifically focus on the rate of false positives and false negatives, respectively. Having the lowest FNR and the FOR are particularly important in the context of medical decision-making to reduce the implications of a late detection of malignancy.

The performance metrics of the machine learning and deep learning algorithms were evaluated through Monte Carlo experiments based on a k-fold cross-validation that splits the data in k-train and k-test samples. Machine learning and deep learning classifiers are estimated with the data of the train sample, and the predictive ability of the models is tested in the test sample that was not used to estimate the algorithms.

3. Results

3.1. Feature selection with SULOV-gradient boosting

Table 2 shows the results of the selection of the best predictors of breast cancer with the

SULOV algorithm. The recursive MRMR-SULOV algorithm was implemented for values of the correlation threshold of features for values of between = 0:01 and = 0:99, in order to mitigate the impact of different correlation values on the selection of optimal predictors and produce results that are robust to the choice of the correlation threshold between variables. The variables that were selected more frequently by the SULOV algorithm are age (x_1), resistin (x_2), the upper values of the relative risks of breast cancer published by the GBD database (x_3), glucose (x_4), adiponectin (x_5), high BMI (x_6), MCP-1 (x_7), leptin (x_8), the relative risks (x_9) of breast cancer of [19]; obesity (x_{10}) and insulin levels in the blood samples (x_{11}); see Table 2. These are the least correlated variables and at the same time the most relevant features selected by the MRMR SULOV-gradient boosting. A previous independent study based on fuzzy neural networks [39] and stochastic vector machines [20] also found resistin, glucose and age relevant for the prediction of breast cancer. Age is considered a particularly relevant risk factor for breast cancer since the diagnosis of this neoplastic disease is most frequently found in women in menopausal transition and less frequently found in women below 45 years of age [40]. Even for premenopausal women, previous studies [41] have found that each 5-year acceleration in biological age corresponds with a 15% increase in breast cancer risk.

Table 2
Recursive MRMR feature selection with SULOV-gradient boosting.

	description	frequency	min ρ	mean ρ	std.dev. ρ
x_1	Age (years)	97	0.00	0.49	0.28
x_2	Resistin (ng/mL)	90	0.00	0.51	0.28
x_3	RRs GBD (upper)	77	0.05	0.55	0.27
x_4	Glucose (mg/dL)	76	0.21	0.59	0.22
x_5	Adiponectin (μ g/mL)	75	0.23	0.60	0.22
x_6	High BMI (binary)	65	0.27	0.64	0.20
x_7	MCP-1 (pg/dL)	65	0.26	0.65	0.19
x_8	Leptin (ng/mL)	64	0.30	0.65	0.19
x_9	RRs [19]	59	0.37	0.68	0.17
x_{10}	Obesity (binary)	59	0.11	0.55	0.28
x_{11}	Insulin (μ U/mL)	33	0.05	0.53	0.26
	BMI	30	0.62	0.81	0.11
	HOMA	29	0.69	0.83	0.09
	RRs GBD (center)	8	0.86	0.90	0.04
	RRs GBD (lower)	8	0.21	0.81	0.25

In relation to the biomarkers selected by the SULOV algorithm—resistin, adiponectin, and glucose—, both low serum adiponectin levels and high resistin levels were previously found to be associated with increased breast cancer risk [42]. Resistin has been significantly associated with tumor and inflammatory markers, cancer stage, tumor size, grade and lymph node invasion [17], since resistin facilitates breast cancer progression via TLR4-mediated induction of mesenchymal phenotypes and stemness properties [43]. Furthermore, circulating resistin associated with the presence of breast cancer in a dose-response manner appears to have adiposity-independent roles in breast carcinogenesis [44]. There is also evidence that glucose and other factors related to glucose metabolism, such as insulin and insulin-like growth factors, can contribute to breast cancer development, in pre- and post-menopausal women [45], because malignant cells extensively use glucose for proliferation [46]. In contrast, the homeostasis model assessment index (HOMA) is frequently excluded from the set of features by the SULOV algorithm due to its low relevance and high correlation with other features. This result is expected and suggests that the SULOV gradient boost algorithm correctly excludes redundant features, because HOMA is a measure of insulin resistance measurement that is calculated as a combination of fasting insulin and glucose levels and was also found to be correlated with BMI in previous studies [47].

3.2. Machine learning and deep learning results

The best predictors selected by the SULOV algorithm were used in the deep learning and machine learning algorithms. In the deep learning algorithm, a sigmoid activation function was used for the output layer of the neural networks, due to the binary nature of the target variable (presence or not of breast cancer). The optimization of the deep learning model was performed with adaptive moment estimation [48] during 3×10^2 epochs, with a batch size equal to 1×10, using a binary cross-entropy loss function. Grid search was applied to choose the type of activation function, the number of hidden layers, and the number of nodes in the hidden layers of the neural networks. Fig. 2 shows the architecture of the deep learning algorithm with the highest predictive power that was obtained as a result of the grid search among different

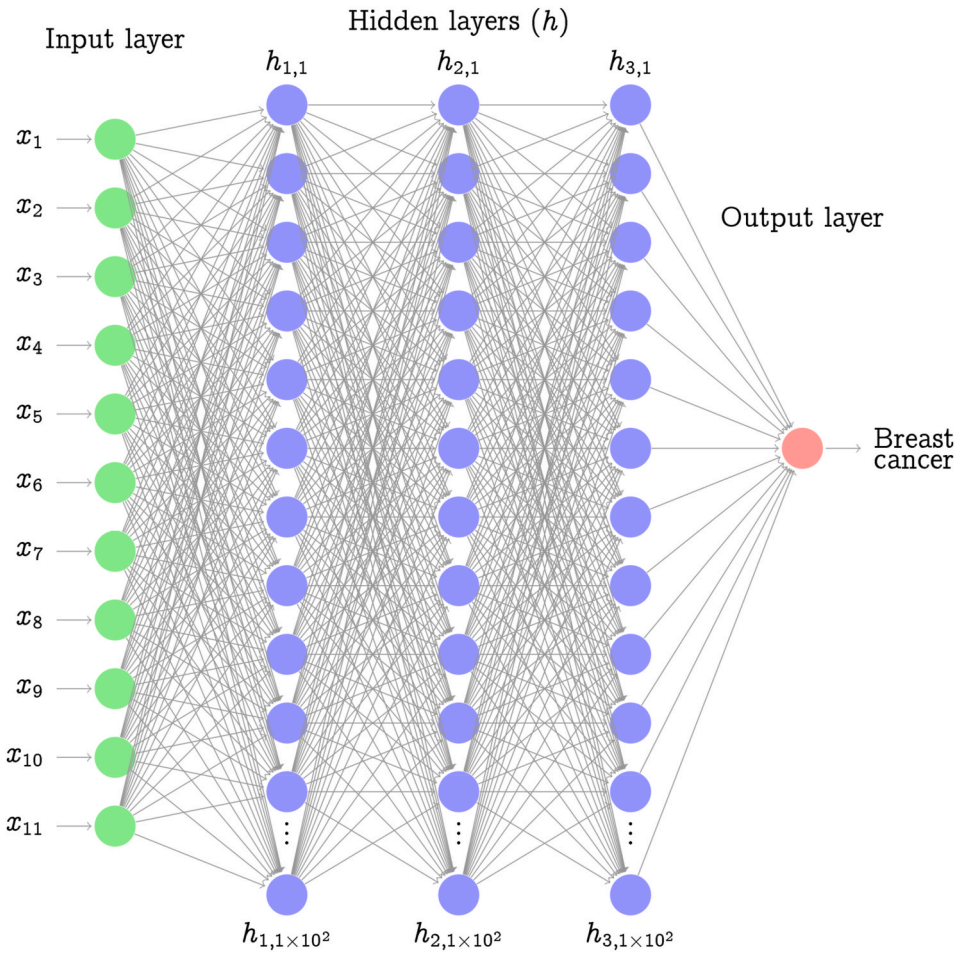


Fig. 2. Illustration of the architecture of the deep learning algorithm. Rectified linear activation function (ReLU) were used in the input layer, with three hidden layers h with 1×10^2 nodes, a L1-L2 regularizer in the kernel function and a L2 regularizer for the bias and the activation functions in the hidden units.

structures, number of hidden layers, nodes in the hidden layers, and activation functions with and without regularization. In the graphical rendition of the best deep learning algorithm illustrated in the architecture of Fig. 2, a rectified linear activation function (ReLU) was used for the eleven variables x_1,x_2, …,x_{11} included in the input layer and for the 1×10^2 nodes with a L1-L2 regularizer in the kernel function, while L2 regularizers for the bias and the activation functions were included in the nodes of three hidden layers. Since deep learning models are data hungry, and the models were trained on a relatively small data set, a limited number of hidden layers were considered in the potential architectures and L1 and L2 regularizers were applied to prevent overfitting. L1 and L2 regularizers were added to the loss function during training to penalize large weights in the network. L1 Regularization encourages sparsity by shrinking some weights towards zero [49]. L2 regularization adds the squared values of the weights to the loss function with the goal of encouraging small weights without enforcing sparsity [50,51].

The better performance of ReLU activation functions compared to tangenthyperbolic or (S)eLU activations functions was expected [52,53], since the use of piece-wise linear hidden units based on the ReLU activation functions is considered a major algorithmic change that improves the performance of feed-forward networks [54] and reduces the computational burden of calculating the exponential function in activation functions [55].

Table 3 and Fig. 3 show the results of predicting the presence of breast cancer with machine learning and deep learning algorithms. On average, the highest predictive ability to differentiate between women with and without breast cancer is obtained with deep learning.

Table 3
AUC of the machine learning and deep learning algorithms.

Model	mean	std.dev.	p2.5	p97.5
Deep learning	0.8699	0.0345	0.8190	0.9138
Support vector machines	0.8344	0.0711	0.6768	0.9527
Neural network	0.8232	0.0720	0.6667	0.9476
Logistic regression	0.8078	0.0741	0.6569	0.9334
XGBoost	0.7834	0.0755	0.6263	0.9191
Random forest	0.7763	0.0804	0.6035	0.9192
Naive bayes	0.7504	0.0869	0.5686	0.9001
Stochastic gradient	0.6861	0.0860	0.5120	0.8553

The deep learning algorithm has on average an AUC equal to 87%, with a 95% confidence interval between 82% and 91% (Table 3). Support vector machines, in turn, have a lower average AUC (83%) with a wider confidence interval (95% CI = [68%, 95%]), and neural networks with a simple architecture have an average AUC of 82%, also with a wider confidence interval (95% CI = [67%, 95%]). Compared to the traditional machine learning algorithms, the deep learning algorithms also have the lowest predictive uncertainty, measured by the standard deviation and the percentiles of the distribution of the AUC obtained with the k-fold Monte Carlo experiments (Table 3). The lowest dispersion of the AUC (Fig. 3) is obtained with the deep learning algorithm (AUC standard deviation = 0.0345), followed by support vector machines (AUC standard deviation = 0.0711) and neural networks (AUC standard deviation = 0.0720). This last result shows the high precision of predictions obtained with deep learning compared to machine

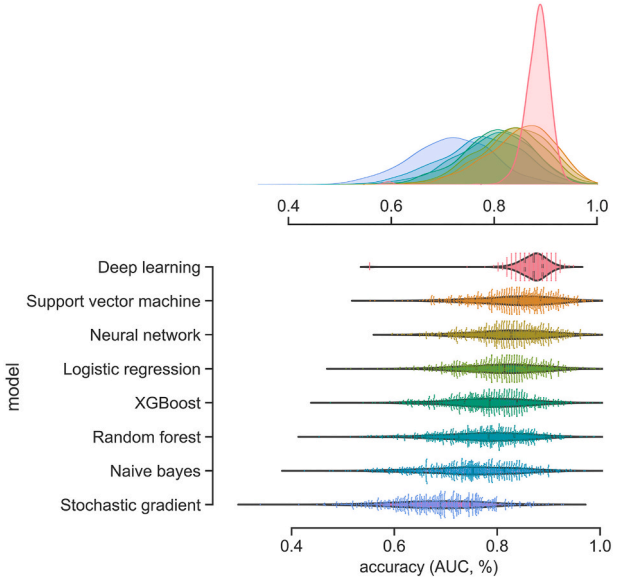


Fig. 3. AUC of machine learning and deep learning classifiers in the Monte Carlo cross-validation.

learning.

Table 4 shows additional evaluation metrics based on the confusion matrix. Compared to machine learning models, the deep learning algorithm has the highest true positive detection rate (TPR), equal to TPR = 0.922, above the best TPR of machine learning models (0.7930, obtained with support vector machines). Moreover, the TPR of the deep learning algorithm has the lowest standard error of TPR (equal to 0.0597). The true negative rate (TNR) is also one the highest for the deep learning algorithm (TNR = 0.8021, standard error = 0.1118), compared to the TNR obtained with a machine learning based on a Naive Bayes algorithm (TNR = 0.8023, standard error = 0.1131), a model that however has a low TPR (equal to 0.5313). These results indicate that—compared to traditional machine learning models—deep learning algorithms have the highest predictive ability and the highest precision to identify both the presence and the absence of breast cancer.

In terms of false positives of breast cancer, the lowest false positive rate (FPR) Is obtained for the deep learning algorithm (FPR = 0.1979) and for a machine learning model based on a Naive Bayes algorithm (FPR = 0.1977). In the case of the false detection rate (FDR), the lowest FDR is obtained only with a deep learning algorithm (FDR = 0.1425, standard error = 0.0522). This finding indicates that deep learning—compared to several machine learning algorithms—has the lowest chance of suggesting the presence of breast cancer when in fact there is not breast cancer malignancy in the patients.

Finally, in the case of cancer prescreening it is particularly important that medical recommendation systems for the early detection of breast cancer and other types of cancer have the lowest rates of false negative detection. In terms of false negatives, the deep learning algorithm has the lowest false negative rate (FNR) and the lowest false omission rate (FOR), compared to the FNRs and the FORs of any of the machine learning models. The FNR of deep learning algorithms is equal to 0.078 (with a standard error of 0.0597), and the FOR of deep learning algorithms is equal to 0.0983 (with a standard error of 0.0655). In comparison, the lowest FNR and FOR of the machine learning models are obtained with support vector machines, that have a FNR equal to 0.2061 (with a FNR standard error of 0.1183), and a FOR of 0.2672 (with a FOR standard error of 0.1399). The lowest false negative rates and the smallest standard errors associated to the FNR and the FOR obtained for the deep learning models with k-fold cross validation indicate that—compared to machine learning—deep learning has the highest precision and the lowest chance of not identifying the presence of breast cancer when in fact there is the presence of cancer in women.

4. Conclusion

Breast cancer is a prevalent form of cancer among women, and early detection is crucial to reduce mortality. The screening of breast cancer is typically performed through methods such as mammograms, ultrasound, magnetic resonance imaging (MRI), self-examination, or examination by a clinician. Advances in deep learning have led to the development of algorithms that can detect abnormalities related to breast cancer in images obtained from patients, particularly for the segmentation and classification of normal and abnormal breast tissue from thermograms [4,56].

This study explored the application of deep learning for the prescreening of breast cancer. Machine learning and deep learning were applied to heterogeneous data that contain demographic and anthropometric information of cancer patients and a control group of healthy women, in addition to biological markers obtained from routine blood samples and relative risks of cancer obtained from publicly available databases published by international studies. The findings indicate that deep learning algorithms have the highest predictive ability to distinguish between women with and without breast cancer. Deep learning also has the lowest uncertainty—the highest precision—and the lowest rate of false negatives in its predictions, compared to machine learning. Based on these results, it can be concluded that prescreening with deep learning algorithms offers a non-invasive, radiation-free, and affordable alternative for the early detection of breast cancer. A deep learning-based medical prescreening recommendation system can help AI-based clinical decision-making that supports the early detection of mass anomalies that may not yet be detectable through self-examination or have not yet caused symptoms but are at a stage where they are easier to treat. This is particularly relevant for breast cancer, as previous studies have shown that women tend to have an already developed metastatic breast cancer when they are first diagnosed, and they have a

Table 4
Performance metrics based on the confusion matrix.

Model	Derivations from confusion matrix*					
	TPR	TNR	FNR	FPR	FDR	FOR
Deep learning	0.9220 (0.0597)	0.8021 (0.1118)	0.0780 (0.0597)	0.1979 (0.1118)	0.1425 (0.0522)	0.0983 (0.0655)
SVM	0.7939 (0.1183)	0.6747 (0.1381)	0.2061 (0.1183)	0.3253 (0.1381)	0.2501 (0.1065)	0.2672 (0.1399)
Neural network	0.7638 (0.1148)	0.7135 (0.1324)	0.2362 (0.1148)	0.2865 (0.1324)	0.2331 (0.1051)	0.2851 (0.1284)
Logistic regression	0.7329 (0.1243)	0.7311 (0.1292)	0.2671 (0.1243)	0.2689 (0.1292)	0.2291 (0.108)	0.3055 (0.1319)
XGBoost	0.7537 (0.1158)	0.6723 (0.1342)	0.2463 (0.1158)	0.3277 (0.1342)	0.2604 (0.1029)	0.3061 (0.1327)
Random forest	0.6892 (0.1267)	0.7038 (0.1365)	0.3108 (0.1267)	0.2962 (0.1365)	0.2575 (0.1144)	0.3477 (0.1263)
Naive bayes	0.5313 (0.1296)	0.8023 (0.1131)	0.4687 (0.1296)	0.1977 (0.1131)	0.2303 (0.126)	0.4147 (0.1083)
Stochastic gradient	0.7126 (0.1394)	0.6518 (0.1587)	0.2874 (0.1394)	0.3482 (0.1587)	0.2792 (0.1106)	0.3412 (0.1409)

(*) TPR: True positive rate, TNR: true negative rate, FNR: false negative rate, FPR: false positive rate, FDR: false detection rate, FOR: false omission rate. Rates are averages of the 1×10^3 k-fold cross validations. Standard deviations in brackets below each average rate. SVM: Support Vector Machines.

poor prognosis regardless of their menopausal status [42]. The integration of deep learning algorithms for cancer prescreening in medical recommendation systems can guide the frequency and follow-up of radiographic imaging and tissue testing, particularly in developing countries, where access to traditional cancer screening and medical personnel is limited for low-income populations in rural areas, making breast cancer a leading cause of death due to late detection [57]. Furthermore, prescreening for breast cancer could reduce the risks associated with traditional screening methods, such as false positive test results and unnecessary tests that are expensive, invasive, time-consuming, and can cause anxiety in patients [58,59].

In the case of cancer prescreening, it is particularly important that any medical recommendation system, based on machine learning or deep learning, has a high detection rate for positive cases while maintaining the lowest possible false negative rate. Rejecting the presence of breast cancer when it actually does exist will hinder early detection of malignant abnormalities, thus amplifying the likelihood of metastasis, increased tumor size, and spread to nearby lymph nodes. False positives and false detection rates can have negative psychological consequences in cancer screening, such as changes in existential values and inner calmness [60]. However, in the case of cancer prescreening, if patients are properly informed that the recommendation system is only designed to suggest the need for a further analysis to con rm or rule out the presence of malignancy, the psychological effects will be less pronounced. Moreover, previous evidence indicated that false-positives encourages women to engage more frequently in future screening [61].

The strength of this study is that it showed that the combination of data from multiple sources can be used for the early detection of breast cancer if feature selection algorithms are combined with deep learning and if predictive models are evaluated with metrics that are statistically relevant and at the same time relevant for medical recommendation systems, such as the false omission rate. This study is limited, however, because the algorithms only detect the potential existence of malignancy but cannot differentiate between types of malignancy. Future studies should investigate the potential of deep learning algorithms to detect the type of malignancy by incorporating additional cancer risk factors, such as those associated with smoking and alcohol consumption, since a recent analysis of the impact of 34 risk factors for 23 cancer types suggests that smoking, alcohol use, and high BMI are the leading contributors to 4.45 million cancer deaths (44.4% of all cancer deaths) globally in 2019 [62].

Declaration of competing interest

We have no conflicts of interest to disclose.

Acknowledgement

The authors gratefully acknowledge the financial support provided by the Bayesian Institute for Research & Development. We also thank the comments, feedback and suggestions provided by the reviewers, as well as the comments received during the presentation of an earlier version of the study during a struggle seminar at the University of Groningen.

References

[1] Wilkinson Louise, Gathani Toral. Understanding breast cancer as a global health concern. Br J Radiol 2022;95(1130):20211033.
[2] Siegel Rebecca L, Miller Kimberly D. Nikita sandeep wagle, and ahmedin jemal. Cancer statistics, 2023. CA: A Cancer Journal for Clini- cians 2023;73(1):17–48. https://doi.org/10.3322/caac.21763. https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21763. URL.
[3] Pfei er Ruth M, Webb-Vargas Yenny, Wheeler William, Mitchell H Gail. Proportion of us trends in breast cancer incidence attributable to long-term changes in risk factor distributions. Cancer Epidemiol Biomarkers Prev 2018;27(10):1214–22.
[4] Torres-Galvan Juan Carlos, Guevara Edgar, Javier Gonzalez Francisco. Comparison of deep learning architectures for pre-screening of breast cancer thermo- grams. In: 2019 photonics north (PN). IEEE; 2019. pages 1{2.
[5] Debelee Taye Girma, Schwenker Friedhelm, Ibenthal Achim, Yohannes Dereje. Survey of deep learning in breast cancer image analysis. Evolving Systems 2020;11(1):143–63.
[6] Yu Keping, Tan Liang, Lin Long, Cheng Xiaofan, Zhang Yi, Sato Takuro. Deeplearning-empowered breast cancer auxiliary diagnosis for 5gb remote e-health. IEEE Wireless Commun 2021;28(3):54–61.
[7] Zhou Li-Qiang, Wu Xing-Long, Huang Shu-Yan, Wu Ge-Ge, Ye Hua-Rong, Qi Wei, Bao Ling-Yun, Deng You-Bin, Li Xing-Rui, Cui Xin-Wu, et al. Lymph node metastasis prediction from primary breast cancer us images using deep learning. Radiology 2020;294(1):19–28.
[8] Zheng Jing, Lin Denan, Gao Zhongjun, Wang Shuang, He Mingjie, Fan Jipeng. Deep learning assisted e cient adaboost algorithm for breast cancer detection and early diagnosis. IEEE Access 2020;8:96946–54.
[9] Bai Jun, Posner Russell, Wang Tianyu, Yang Cli ord, Nabavi Sheida. Ap plying deep learning in digital breast tomosynthesis for automatic breast cancer detection: a review. Med Image Anal 2021;71:102049.
[10] Yala Adam, Lehman Constance, Schuster Tal, Portnoi Tally, Barzilay Regina. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019;292(1):60–6.
[11] Naik Nikhil, Ali Madani, Esteva Andre, Keskar Nitish Shirish, Press Michael F, Ruderman Daniel, Agus David B, Socher Richard. Deep learning-enabled breast cancer hormonal receptor status determination from base-level h&e stains. Nat Commun 2020;11(1):1–8.
[12] Khuriwal Naresh, Mishra Nidhi. Breast cancer diagnosis using deep learning algorithm. In: 2018 international conference on advances in computing, communication control and networking (ICACCCN). IEEE; 2018. p. 98–103.
[13] Street W Nick, Wolberg William H, Mangasarian Olvi L. Nuclear feature extraction for breast tumor diagnosis. In: Biomedical image processing and biomedical visualization. vol. 1905. SPIE; 1993. p. 861–70.
[14] Crisostomo Joana, Matafome Paulo, Santos-Silva Daniela, Gomes Ana L, Gomes Manuel, cio Miguel Patr, Letra Liliana, Sarmento-Ribeiro Ana B, Santos Lelita, Seica Raquel. Hyperresistinemia and metabolic dysregulation: a risky crosstalk in obese breast cancer. Endocrine 2016;53(2):433–42.
[15] Lake JK, Power C, Cole TJ. Women's reproductive health: the role of body mass index in early and adult life. Int J Obes 1997;21(6):432–8.
[16] Santillan-Ben tez Jonnathan G, Mendieta-Zeron Hugo, Gomez-Olivan Leobardo M, Torres-Juarez Juan J, Gonzalez-Bañales Juan M, Hernandez-Peña Lorena V, Ordoñez-Quiroz Angel. The tetrad bmi, leptin, leptin/adiponectin (l/a) ratio and ca 15-3 are reliable biomarkers of breast cancer. J Clin Lab Anal 2013;27(1):12–20.
[17] Dalamaga Maria, George Sotiropoulos, Karmaniolas Konstantinos, Pelekanos Nicolaos, Papadavid Evangelia, Lekka Antigoni. Serum resistin: a biomarker of breast cancer in postmenopausal women? association with clinicopathological characteristics, tumor markers, in ammatory and metabolic parameters. Clin Biochem 2013;46(7–8):584–90.
[18] Vos Theo, Lim Stephen S, Abbafati Cristiana, Abbas Kaja M, Abbasi Mohammad, Mitra Abbasifard, Abbasi-Kangevari Mohsen, Abbastabar Hedayat, AbdAllah Foad, Ahmed Abdelalim, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990{2019: a systematic analysis for the global burden of disease study 2019. Lancet 2020;396(10258):1204–22.
[19] Liu Kang, Zhang Weining, Dai Zhiming, Wang Meng, Tian Tian, Liu Xinghan, Kang Huafeng, Guan Haitao, Zhang Shuqun, Dai Zhijun. Association between body mass index and breast cancer risk: evidence based on a dose{response meta-analysis. Cancer Manag Res 2018;10:143.
[20] Patrcio Miguel, Pereira Jose, Crisostomo Joana, Matafome Paulo, Gomes Manuel, Seica Raquel, Caramelo Francisco. Using resistin, glucose, age and bmi to predict the presence of breast cancer. BMC Cancer 2018;18(1):1–8.
[21] Shen Dinggang, Wu Guorong, Suk Heung-Il. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221.
[22] Haque Intisar Rizwan I, Neubert Jeremiah. Deep learning approaches to biomedical image segmentation. Inform Med Unlocked 2020;18:100297.
[23] Dash Sujata, Subhendu Kumar Pani, Joel Jpc Rodrigues, Majhi Babita. Deep learning, machine learning and IoT in biomedical and health informatics: techniques and applications. CRC Press; 2022.
[24] Baldi Pierre. Deep learning in biomedical data science. Annual review of biomedical data science 2018;1:181–205.
[25] Tian Lei, Hunt Brady, A Lediju Bell Muyinatu, Yi Ji, Smith Jason T, Ochoa Marien, Intes Xavier, J Durr Nicholas. Deep learning in biomedical optics. Laser Surg Med 2021;53(6):748–75.
[26] Isensee Fabian, Jaeger Paul F, Kohl Simon AA, Petersen Jens, MaierHein Klaus H. nnu-net: a self-con guring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203–11.
[27] Grohl Janek, Schellenberg Melanie, Dreher Kris, Maier-Hein Lena. Deep learning for biomedical photoacoustic imaging: a review. Photoacoustics 2021;22:100241.
[28] He Bryan, Bergenstrahle Ludvig, Stenbeck Linnea, Abid Abubakar, Andersson Alma, Borg Ake, Jonas Maaskola, Lundeberg Joakim, Zou James. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng 2020;4(8):827–34.
[29] Torrisi Mirko, Pollastri Gianluca, Quan Le. Deep learning methods in protein structure prediction. Comput Struct Biotechnol J 2020;18:1301–10.
[30] Bozhkov Lachezar, Georgieva Petia. Deep learning models for brain machine interfaces. Annals of Mathematics and Arti cial Intelligence 2020;88(11):1175–90.
[31] Ionescu Daniela. Deep learning algorithms and big health care data in clinical natural language processing. Ling Phil Invest 2020;19:86–92.

[32] Ram rez-Gallego Sergio, Lastra Iago, Mart nez-Rego David, BolonCanedo Veronica, Ben tez Jose Manuel, Herrera Francisco, Alonso-Betanzos Amparo. Fast-mrmr: fast minimum redundancy maximum relevance algorithm for highdimensional big data. Int J Intell Syst 2017;32(2):134–52.
[33] Brownlee Jason. Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python. Machine Learning Mastery; 2020.
[34] Liu Huan, Motoda Hiroshi. Computational methods of feature selection. CRC Press; 2007.
[35] Chen Tianqi, Guestrin Carlos. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785–94.
[36] Nielsen Didrik. Tree boosting with xgboost-why does xgboost win \every" machine learning competition? Master's thesis. NTNU; 2016.
[37] Ding Chris, Peng Hanchuan. Minimum redundancy feature selection from microarray gene expression data. J Bioinf Comput Biol 2005;3(2):185–205.
[38] Radovic Milos, Ghalwash Mohamed, Filipovic Nenad, Obradovic Zoran. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinf 2017;18(1):1–14.
[39] Silva Araujo Vin cius Jonathan, Guimar~aes Augusto Junio, Vitor de Campos Souza Paulo, Silva Rezende Thiago, Souza Araujo Vanessa. Using resistin, glucose, age and bmi and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer. Machine Learning and Knowledge Extraction 2019;1(1):466–82.
[40] Kaminska Marzena, Ciszewski Tomasz, opacka-Szatan Karolina L, Pawe l Miot la, lawska Elz_bieta Staros. Breast cancer risk factors. Menopausal Rev 2015;14(3): 196–202.
[41] Kresovich Jacob K, Xu Zongli, O'Brien Katie M, Weinberg Clarice R, Sandler Dale P, Taylor Jack A. Methylation-based biological age and breast cancer risk. JNCI: J Natl Cancer Inst 2019;111(10):1051–8.
[42] Kang Jee-Hyun, Yu Byung-Yeon, Youn Dae-Sung. Relationship of serum adiponectin and resistin levels with breast cancer risk. J Kor Med Sci 2007;22(1): 117–21.
[43] Wang CH, Wang PJ, Hsieh YC, Lo S, Lee YC, Chen YC, Tsai CH, Chiu WC, Hu S Chu-Sung, Lu CW, et al. Resistin facilitates breast cancer progression via tlr4-mediated induction of mesenchymal phenotypes and stemness properties. Oncogene 2018;37 (5):589–600.
[44] Sun Chien-An, Wu Mei-Hsuan, Chu Chi-Hong, Chou Yu-Ching, Hsu Giu-Cheng, Yang Tsan, Chou Wan-Yun, Yu Cheng-Ping, Yu Jyh-Cherng. Adipocytokine resistin and breast cancer risk. Breast Cancer Res Treat 2010;123(3):869–76.
[45] Sieri Sabina, Muti Paola, Claudia Agnoli, Franco Berrino, Pala Valeria, Grioni Sara, Abagnato Carlo Alberto, Blandino Giovanni, Contiero Paolo, Schunemann Holger J, et al. Prospective study on the role of glucose metabolism in breast cancer occurrence. Int J Cancer 2012;130(4):921–9.
[46] Muti Paola, Quattrin Teresa, Grant Brydon JB, Krogh Vittorio, Micheli Andrea, Schunemann Holger J, Ram Malathi, Freudenheim Jo L, Sieri Sabina, Trevisan Maurizio, et al. Fasting glucose is a risk factor for breast cancer: a prospective study. Cancer Epidemiol Biomark Prev 2002;11(11):1361–8.
[47] Timoteo Ana Teresa, Miranda Fernando, Mota Carmo Miguel, Ferreira Rui Cruz. Optimal cut-o value for homeostasis model assessment (homa) index of insulin-resistance in a population of patients admitted electively in a Portuguese cardiology ward. Acta Med Port 2014;27(4):473–9.
[48] Kingma Diederik P, Jimmy Ba. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
[49] Tibshirani Robert. Regression shrinkage and selection via the lasso. J Roy Stat Soc B 1996;58(1):267–88.
[50] Hoerl Arthur E, Kennard Robert W. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 1970;12(1):55–67.
[51] Hinton Geo rey, Vinyals Oriol, Dean Je. Distilling the knowledge in a neural network. 2015. arXiv preprint arXiv:1503.02531.
[52] Clevert Djork-Arne, Unterthiner Thomas, Hochreiter Sepp. Fast and accurate deep network learning by exponential linear units (elus). 2015. arXiv preprint arXiv: 1511.07289.
[53] Klambauer Gunter, Unterthiner Thomas. Andreas mayr, and sepp hochreiter. Selfnormalizing neural networks. Adv Neural Inf Process Syst 2017;30.
[54] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning (adaptive computation and machine learning series). 2017. Cambridge Massachusetts.
[55] Glorot Xavier, Antoine Bordes, Bengio Yoshua. Deep sparse recti er neural networks. In: Proceedings of the fourteenth international conference on arti cial intelligence and statistics. JMLR Workshop and Conference Proceedings; 2011. p. 315–23.
[56] Mohamed Esraa A, Rashed Essam A, Gaber Tarek, Karam Omar. Deep learning model for fully automated breast cancer detection system from thermograms. PLoS One 2022;17(1):e0262349.
[57] Kakileti Siva Teja, Madhu Himanshu J, Manjunath Geetha, Leonard Wee, Dekker Andre, Sampangi Sudhakar. Personalized risk prediction for breast cancer pre-screening using arti cial intelligence and thermal radiomics. Arti cial Intelligence in Medicine 2020;105:101854.
[58] Lerman Caryn, Bruce Trock, Rimer Barbara K, Jepson Christopher, Brody David, Boyce Alice. Psychological side e ects of breast cancer screening. Health Psychol 1991;10(4):259.
[59] Mathioudakis Alexander G, Salakari Minna, Pylkkanen Liisa, SazParkinson Zuleika, Bramesfeld Anke, Deandrea Silvia, Lerda Donata, Neamtiu Luciana, Pardo-Hernandez Hector, Sola Ivan, et al. Systematic review on women's values and preferences concerning breast cancer screening and diagnostic services. Psycho Oncol 2019;28(5):939–47.
[60] Brodersen John, Siersma Volkert Dirk. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med 2013;11(2):106–15.
[61] Taksler Glen B, Keating Nancy L, Rothberg Michael B. Implications of falsepositive results for future cancer screenings. Cancer 2018;124(11):2390–8.
[62] Tran Khanh Bao, Lang Justin J, Compton Kelly, Xu Rixing, Acheson Alistair R, Henrikson Hannah Jacqueline, Kocarnik Jonathan M, Penberthy Louise, Aali Amirali, Abbas Qamar, et al. The global burden of cancer attributable to risk factors, 2010{19: a systematic analysis for the global burden of disease study 2019. Lancet 2022;400(10352):563–91.