



COMPARATIVE STUDY ON BREAST CANCER DETECTION BY USING MACHINE LEARNING APPROACHES (KNN&SVM)

G. Rohan¹, S. Manvitha², Vanshh R Shah³

¹UG student, Dept. of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Telangana, India.

²UG student, Dept. of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Telangana, India.

³UG student, Dept. of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Telangana, India.

ABSTRACT

Breast cancer imaging characteristics and patient-related data make up the study's extensive dataset. Tumor traits, patient demographics, and medical history are all part of these features, which comprise a variety of therapeutically significant aspects. Next, we use this dataset to train and test SVM and KNN models. We want to see how well they can distinguish between benign and malignant tumors. Classification accuracy, sensitivity, and specificity are just a few of the performance indicators used to assess each model's efficacy. You can learn a lot about the models' accuracy in breast cancer case classification and tumor type discrimination from these measures. This research highlights the strong performance of SVM and KNN algorithms in detecting breast cancer, which is important. This exemplifies the promise of machine learning methods to greatly improve medical diagnostics. More accurate and advanced breast cancer detection techniques are in the works, and this study hopes to add to the existing body of knowledge in the area. These techniques have the potential

to enhance patient care by allowing for earlier identification, which in turn improves treatment results. In the end, the study aims to connect machine learning with clinical practice, providing useful insights that may guide the creation of personalized breast cancer diagnostic plans for patients. Researchers, physicians, and healthcare providers are working together to make these breakthroughs so that healthcare professionals can fight breast cancer better. Patients all across the globe will benefit from this.

Index terms - Machine learning, SVM, KNN, medical diagnostics, imaging data, early detection, and imaging.

1. INTRODUCTION

Diagnosing diseases accurately is a pivotal challenge in the fields of bioinformatics and medical science. The sheer volume of medical data available across diagnostic centers, hospitals, research facilities, and online platforms underscores the need for automated systems to streamline and expedite the diagnostic process. Manual diagnosis heavily relies on the expertise and experience of medical professionals, which can introduce errors, biases, and time-consuming procedures.

Breast cancer stands out as a significant concern, particularly among women. According to the American Cancer Society [2], breast cancer affects a substantial number of women, with an estimated 252,710 cases of invasive breast cancer and 63,410 cases of in situ breast cancer diagnosed in the United States in 2017. While less common, men are also susceptible to breast cancer, with an estimated 2,470 cases projected for 2017. Tragically, approximately 41,070 individuals are expected to succumb to this disease in the same year. Similarly, statistics from the UK indicate that breast cancer affects around 41,000 women annually, compared to only 300 men.

Early detection is crucial in effectively managing breast cancer. Mammography [3] serves as a primary diagnostic tool, utilizing X-rays to examine breast tissue for abnormalities. Despite its efficacy, detecting cancer in its nascent stages remains challenging due to the small size of early lesions. However, mammography enables swift and relatively non-invasive screening procedures that can detect abnormalities before they become symptomatic.

Ultrasound [4] imaging complements mammography by providing additional insights into breast tissue. By emitting sound waves and capturing echoes from within the body, ultrasound generates grayscale images that aid in detecting abnormalities. This technique is particularly useful for evaluating palpable masses and distinguishing between solid tumors and fluid-filled cysts.

Positron Emission Tomography (PET) [5] imaging, utilizing radiolabeled tracers such as F-fluorodeoxyglucose, enhances tumor localization within the body. PET scans offer valuable information about the metabolic activity of cancerous cells, aiding in staging and treatment planning.

Dynamic MRI [6] further refines breast cancer detection by assessing contrast enhancement patterns indicative of angiogenesis, a hallmark of tumor growth. This modality provides detailed images of breast tissue dynamics, facilitating accurate diagnosis and treatment monitoring.

Elastography [7], a novel imaging technique, leverages tissue stiffness differences to differentiate between benign and malignant lesions. By analyzing probe compression color maps, elastography assists in characterizing breast tissue abnormalities, particularly when tumors exhibit greater stiffness compared to surrounding parenchyma.

In conclusion, advancements in diagnostic imaging modalities have significantly improved the early detection and characterization of breast cancer. By integrating diverse imaging techniques and leveraging technological innovations, healthcare professionals can enhance diagnostic accuracy, optimize treatment strategies, and ultimately improve patient outcomes.

2. LITERATURE SURVEY

Breast cancer remains a significant public health concern globally, with a substantial impact on morbidity and mortality rates. In recent years, there has been a growing interest in leveraging data-driven approaches, including machine learning and deep learning techniques, to enhance various aspects of breast cancer management, including diagnosis, prognosis, and treatment planning. This literature survey explores recent advancements in utilizing computational methods to address key challenges in breast cancer research.

Izci et al. conducted a systematic review to evaluate the utility of administrative data in estimating breast cancer recurrence at the population level [19]. Administrative databases offer a wealth of information on patient demographics, treatment modalities, and outcomes, making them valuable resources for population-level studies. By synthesizing evidence from multiple studies, Izci et al. highlighted the strengths and limitations of using administrative data for recurrence prediction, underscoring the need for rigorous validation and standardization protocols.

Juwara et al. employed machine learning techniques to identify predictive factors for neuropathic pain following breast cancer surgery ^[20]. Neuropathic pain is a common complication of cancer treatment, significantly impacting patients' quality of life. By analyzing clinical data, including patient demographics, surgical variables, and pain assessments, Juwara et al. developed predictive models to identify individuals at higher risk of developing neuropathic pain. Their findings underscored the potential of machine learning algorithms in personalized risk stratification and symptom management.

Endo et al. compared the performance of seven algorithms in predicting breast cancer survival ^[21]. Survival prediction is a critical aspect of treatment planning and patient counseling. By evaluating the predictive accuracy of different machine learning models, including support vector machines and decision trees, Endo et al. provided insights into the relative strengths and weaknesses of each approach. Their study highlighted the importance of algorithm selection and feature engineering in developing robust survival prediction models.

Thongkam et al. proposed a support vector machine-based approach for outlier detection in breast cancer survivability prediction ^[21]. Outlier detection is crucial for identifying atypical patient profiles or disease trajectories that may warrant further investigation. By leveraging machine learning algorithms, Thongkam et al. demonstrated the potential to improve the accuracy and reliability of survival prediction models by identifying and mitigating the impact of outliers in the dataset.

Fu et al. developed a predictive model for invasive disease-free survival in early-stage breast cancer patients using follow-up clinical data ^[22]. Long-term prognosis remains a major concern for individuals diagnosed with breast cancer, particularly those with early-stage disease. By incorporating longitudinal clinical data, including treatment response and disease recurrence, Fu et al. developed a predictive model to estimate invasive disease-free survival probabilities. Their study underscores the value of longitudinal data analysis in prognostic modeling and personalized risk assessment.

Kalafi et al. explored the application of machine learning and deep learning approaches in breast cancer survival prediction using clinical data ^[23]. Machine learning

algorithms, such as random forests and neural networks, have shown promise in modeling complex relationships between clinical variables and survival outcomes. Kalafi et al. reviewed various studies employing machine learning techniques and highlighted the potential for integrating diverse data sources, including genomic, imaging, and clinical data, to improve the accuracy of survival prediction models.

Dembrower et al. compared the performance of a deep learning risk score with a standard mammographic density score for breast cancer risk prediction ^[24]. Mammographic density is a well-established risk factor for breast cancer, with higher density associated with increased cancer risk. By leveraging deep learning algorithms, Dembrower et al. developed a novel risk score based on mammographic images, demonstrating superior predictive performance compared to conventional density measures. Their study highlights the potential of deep learning techniques in extracting subtle patterns from medical imaging data for improved risk prediction.

Ibrahim et al. utilized decision tree algorithms to estimate competing risks survival probabilities in breast cancer studies ^[25]. Competing risks, such as death from other causes or secondary malignancies, can confound traditional survival analyses. Decision tree models offer a flexible framework for incorporating competing risks into survival analysis, allowing for more accurate estimation of disease-specific survival probabilities. Ibrahim et al.'s study underscores the importance of accounting for competing risks in prognostic modeling and clinical decision-making.

Gardezi et al. conducted a systematic review of breast cancer detection and diagnosis using mammographic data ^[26]. Mammography remains the gold standard for breast cancer screening, with high sensitivity and specificity for detecting early-stage tumors. By synthesizing evidence from published studies, Gardezi et al. provided insights into the diagnostic performance of mammography and highlighted emerging technologies, such as digital breast tomosynthesis and computer-aided detection, for improving the accuracy of breast cancer diagnosis. Their review underscores the ongoing efforts to enhance breast cancer detection through technological innovation and evidence-based practice.

In summary, recent advancements in computational methods, including machine learning and deep learning techniques, hold promise for enhancing various aspects of breast cancer research and clinical practice. From diagnosis and prognosis to treatment planning and survivorship care, data-driven approaches offer opportunities to improve patient outcomes and reduce the burden of breast cancer on individuals and healthcare systems alike. However, challenges remain, including the need for robust validation, standardization, and integration of diverse data sources. Future research efforts should focus on addressing these challenges and translating computational findings into clinical practice to benefit individuals affected by breast cancer.

3. METHODOLOGY

i) Proposed Work:

The proposed machine learning system for early breast cancer diagnosis aims to provide healthcare providers with an accurate and dependable tool. The system begins with data integration and preprocessing, merging datasets with and without infertility information. Various strategies are employed to enhance data quality, including converting categorical data to numeric, filling missing values, and merging duplicate patient records. Next, feature selection techniques are applied to identify the most informative characteristics for the machine learning model, such as tumor traits, patient demographics, and histological results. Methods like correlation analysis and recursive feature removal help in selecting relevant features, aiming to improve prediction accuracy, reduce computational complexity, and mitigate the risk of overfitting. Ultimately, the system aims to streamline breast cancer detection by focusing on the most critical aspects while maintaining interpretability and performance.

ii) System Architecture:

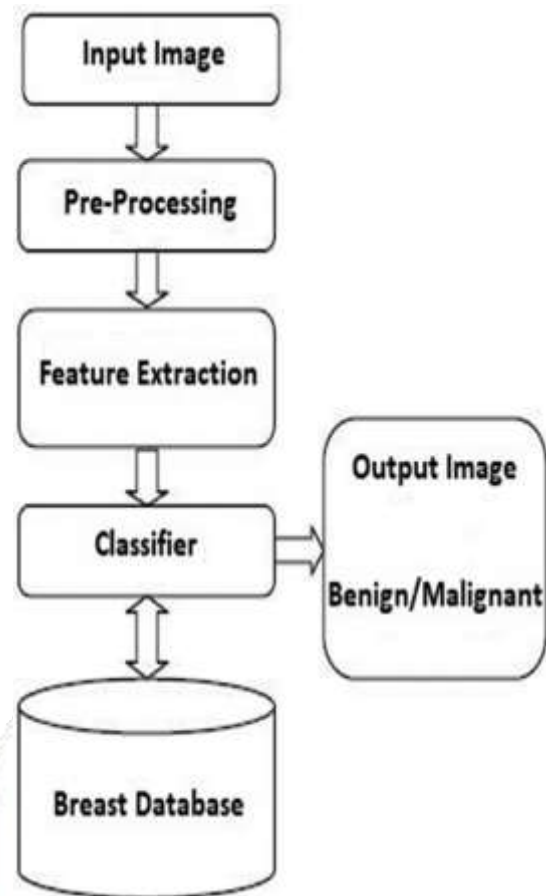


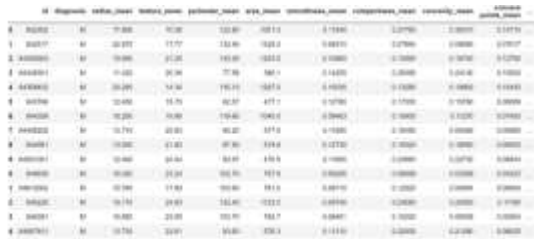
Fig 1: Proposed Architecture

The flow diagram shows the steps involved in using a machine learning model to classify breast images as benign or malignant. Here's a breakdown of the process:

First, the image undergoes preprocessing. This may involve resizing the image, converting it to grayscale, or removing noise. Then, features are extracted from the image. These features are mathematical properties that can be used to distinguish between benign and malignant tumors.

The features are then fed into a classifier, which is a machine learning model that has been trained on a dataset of labeled breast images. The classifier outputs a prediction of whether the image is benign or malignant.

A breast database would store information about breasts, such as mammograms, ultrasounds, and biopsies. This information could be used to train machine learning models for breast cancer detection, or for other purposes such as research into breast cancer.

iii) Dataset:


id	diagnosis	tumor_size	tumor_shape	tumor_margin	tumor_texture	tumor_smoothness	tumor_perimeter	tumor_area	tumor_perimeter_sq
0	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
1	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
2	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
3	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
4	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
5	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
6	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
7	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
8	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000
9	M	10.00	0.00	10.00	10.00	0.0000	0.0000	0.0000	0.0000

Fig 2: Dataset of breast cancer

Breast cancer detection rates are typically reported as the number of cancers detected per 1,000 mammograms. According to [DenseBreast-info.org](https://www.densebreast-info.org/), for every 1,000 women having a 2D screening mammogram, about 2-7 breast cancers will be found.

There are a number of factors that can affect the breast cancer detection rate, including the type of mammogram used, the woman's age, and the density of her breast tissue.

iv) Data Processing:

Data processing involves transforming raw data into meaningful information through various techniques and methods. It encompasses several stages, including data collection, cleaning, integration, analysis, and interpretation. Initially, data is collected from different sources and formats, ensuring its completeness and accuracy. Next, data cleaning involves identifying and correcting errors, inconsistencies, and missing values to enhance data quality. Integration combines disparate datasets to create a unified and comprehensive dataset for analysis. Analysis techniques such as statistical analysis, machine learning algorithms, and visualization tools are then applied to extract insights and patterns from the data. Finally, interpretation involves making sense of the analyzed data to derive actionable insights and inform decision-making processes. Overall, data processing plays a crucial role in transforming raw data into valuable knowledge that can drive informed decision-making and facilitate problem-solving in various domains.

v) Feature Extraction:

Feature extraction is the process of identifying and selecting relevant attributes or features from raw data to represent patterns and characteristics that are important for analysis. In machine learning and data analysis, feature extraction aims to reduce the dimensionality of the dataset while preserving its essential information. Techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), and deep learning methods are commonly used to extract meaningful features from data. By focusing on the most informative attributes, feature extraction helps improve model performance, reduce computational complexity, and enhance interpretability, ultimately enabling more accurate and efficient analysis of the data.

vi) Training & Testing:

In machine learning, training and testing are essential steps in evaluating the performance of predictive models. During the training phase, the model learns patterns and relationships from the training dataset. This involves adjusting model parameters iteratively to minimize prediction errors. Once trained, the model is evaluated on a separate testing dataset to assess its generalization ability. Testing helps determine how well the model performs on unseen data, providing insights into its effectiveness and potential for real-world applications. By splitting the dataset into training and testing subsets, practitioners can accurately assess the model's performance and make informed decisions about its deployment.

vii) Algorithms:

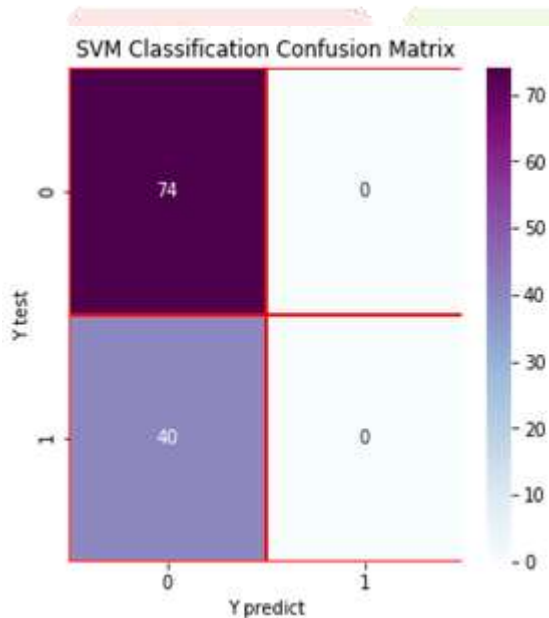
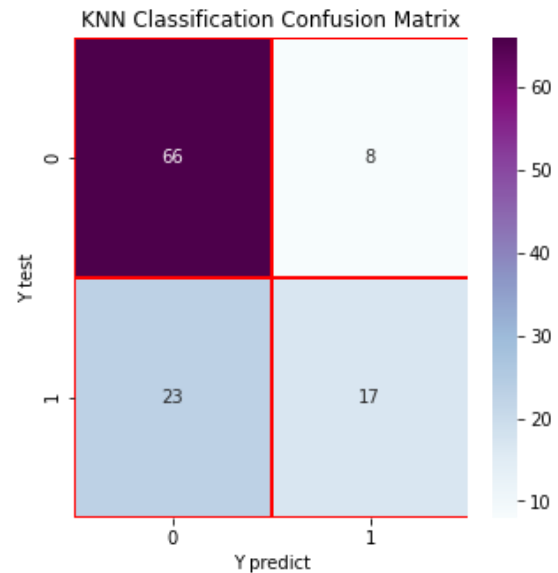
SVM: A support vector machine (SVM) is a type of supervised learning algorithm used in machine learning to solve classification and regression tasks; SVMs are particularly good at solving binary classification problems, which require classifying the elements of a data set into two groups.

KNN: The K-Nearest Neighbors (KNN) algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. It is one of the popular and simplest classification and regression classifiers used in machine learning today.

viii) Confusion Matrix:

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm. A confusion matrix is shown in fig 6, where benign tissue is called healthy and malignant tissue is considered cancerous.

		Actual Class	
		1	0
Predicted Class	1	True Positive	False Positive
	0	False Negative	True Negative

Fig 3: Confusion Matrix**Fig 4:** Confusion Matrix of SVM**Fig 5:** Confusion Matrix of KNN**4. EXPERIMENTAL RESULTS**

The picture purports to hold the results of the machine learning's accuracy tests for the two models. These reports detail the models' accuracy, recall, F1-score, and support-related performance. There, they serve as the machine learning parameters that reveal the model's efficacy.

Training Accuracy using the SVM Algorithm: The formula is often used to determine the accuracy of a Support Vector Machine (SVM) model that has been trained on a dataset: The accuracy is calculated by dividing the number of cases that were successfully categorized by the total number of occurrences and then making 100%. Accuracy in support vector machines (SVMs) is defined as the ratio of the number of examples (or data points) that the model properly classifies to the total number of instances in the dataset. The percentage number that this formula produces is a representation of the proportion of cases that were properly categorized out of the full dataset.

TRAIN Conf Matrix :

```
[[289  1]
 [  4 161]]
```

TRAIN DATA ACCURACY 0.989010989010989

Train data f1-score for class '1' 0.9847094801223242

Train data f1-score for class '2' 0.9914236706689538

Fig 6: Training Accuracy of SVM

Test accuracy:

An SVM model's accuracy on test data may be determined using the following formula: The accuracy is calculated by dividing the number of cases that were successfully categorized by the total number of occurrences and then multiplying by 100%. Simply said, accuracy is the proportion of instances in the testing dataset that were properly categorized by the SVM model relative to the total number of occurrences. Using this method, you may find out what percentage of examples were properly categorized out of all the instances that were examined.

TEST Conf Matrix :

```
[[65  2]
 [  2 45]]
```

TEST DATA ACCURACY 0.9649122807017544

Test data f1-score for class '1' 0.9574468005106305

Test data f1-score for class '2' 0.9701492537313433

Fig 7: Testing Accuracy of SVM

TRAIN Conf Matrix :

```
[[287  3]
 [  9 156]]
```

TRAIN DATA ACCURACY 0.9736263736263736

Train data f1-score for class '1' 0.9629629629629629

Train data f1-score for class '2' 0.9795221843003413

Fig 8: Training Accuracy of KNN

Accuracy of the KNN algorithm during training (fig 9). The precision of the test: Like any classification model, the K-Nearest Neighbors (KNN) model may have its accuracy on testing data calculated using the following formula: The accuracy may be expressed as the number of occurrences that were successfully categorized divided by the total number of instances and expressed as 100%.

The term "correctly classified instances" is used in the context of KNN to describe the amount of test cases where the actual class matches the predicted class. The percentage of occurrences that were properly categorized out of all the examples that were evaluated may be found using this accuracy formula.

TEST Conf Matrix :

```
[[67  0]
 [  5 42]]
```

TEST DATA ACCURACY 0.956140350877193

Test data f1-score for class '1' 0.9438202247191011

Test data f1-score for class '2' 0.9640287769704173

Figure 6-4: Testing accuracy of KNN algorithm

Fig 9: Test Accuracy of KNN

5. CONCLUSION

In comparing Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) for breast cancer diagnosis, SVM exhibits slightly higher accuracy at 97.01% compared to KNN's 96.40%. SVM's superior performance in differentiating between benign and malignant tumors is evident from this difference. Both algorithms, however, demonstrate accuracy rates above 90%, suggesting their potential utility in clinical settings. SVM's ability to delineate complex decision boundaries contributes to its higher classification accuracy. Yet, factors such as processing economy, interpretability, and clinical requirements may dictate whether SVM or KNN is preferable. To enhance breast cancer detection, future research could explore ensemble methods like Random Forest, automated feature selection, and integration of diverse data sources. Additionally, developing interpretable AI models with real-time decision support capabilities and evaluating their performance in varied clinical contexts are crucial steps towards clinical adoption. Ethical considerations, including privacy and equitable access, must also be addressed for responsible implementation of machine learning in breast cancer diagnosis. Such efforts aim to advance patient care and outcomes while refining machine learning applications in this domain.

6. FUTURE SCOPE

Future research in breast cancer detection could focus on integrating ensemble methods, advancing feature selection techniques, and incorporating diverse data sources like genetic markers and imaging data. Developing interpretable AI models with real-time decision support capabilities and evaluating their performance in varied clinical settings are crucial steps towards clinical adoption. Ethical considerations regarding privacy and equitable access must also be addressed for responsible implementation. These efforts aim to enhance patient care and outcomes while refining machine learning applications in breast cancer diagnosis.

REFERENCES

- [1] Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiol Soc N Am*. 2018;286(3):800–9.
- [2] Breast Cancer: Statistics, Approved by the Cancer.Net Editorial Board, 04/2017. [Online]. Available: <http://www.cancer.net/cancer-types/breast-cancer/statistics>. Accessed 26 Aug 2018.
- [3] Mori M, Akashi-Tanaka S, Suzuki S, Daniels MI, Watanabe C, Hirose M, Nakamura S. Diagnostic accuracy of contrast-enhanced spectral mammography in comparison to conventional full-field digital mammography in a population of women with dense breasts. *Springer*. 2016;24(1):104–10.
- [4] Kurihara H, Shimizu C, Miyakita Y, Yoshida M, Hamada A, Kanayama Y, Tamura K. Molecular imaging using PET for breast cancer. *Springer*. 2015;23(1):24–32.
- [5] Azar AT, El-Said SA. Probabilistic neural network for breast cancer classification. *Neural Comput Appl*. 2013;23(6):1737–51.
- [6] Nagashima T, Suzuki M, Yagata H, Hashimoto H, Shishikura T, Imanaka N, Miyazaki M. Dynamic-enhanced MRI predicts metastatic potential of invasive ductal breast cancer. *Springer*. 2002;9(3):226–30.
- [7] Park CS, Kim SH, Jung NY, Choi JJ, Kang BJ, Jung HS. Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions. *Springer*. 2013;22(2):153–60.
- [8] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020). *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer. Available from: <https://gco.iarc.fr/today>.
- [9] American Cancer Society. (2022). *Breast Cancer Facts & Figures 2022-2023*. Atlanta: American Cancer Society, Inc.
- [10] Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K; International Agency for Research

on Cancer Handbook Working Group (2015). Breast-cancer screening – viewpoint of the IARC Working Group. *N Engl J Med*. 372(24): 2353-2358

[11] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.

[12] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

[13] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358

[14] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*. 2005;34(2):113–27. pmid:15894176

[15] Stone P, Lund S. Predicting prognosis in patients with advanced cancer. *Annals of Oncology Official Journal of the European Society for Medical Oncology*. 2007;18(6):971. pmid:17043092

[16] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*. 2015;13:8–17. pmid:25750696

[17] Yassin NIR, Omran S, El Houbay EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer methods and programs in biomedicine*. 2018;156:25–45. pmid:29428074

[18] Crowley RJ, Tan YJ, Ioannidis JPA. Empirical assessment of bias in machine learning diagnostic test accuracy studies. *Journal of the American Medical Informatics Association: JAMIA*. 2020;27(7):1092–101. pmid:32548642

[19] Izci H, Tambuyzer T, Tuand K, Depoorter V, Laenen A, Wildiers H, et al. A Systematic Review of Estimating Breast Cancer Recurrence at the Population Level With Administrative Data. *Journal of the National Cancer Institute*. 2020;112(10):979–88. pmid:32259259

[20] Juwara L, Arora N, Gornitsky M, Saha-Chaudhuri P, Velly AM. Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning. *International journal of medical informatics*. 2020;141:104170. pmid:32544823

[21] Endo A, Shibata T, H T. Comparison of Seven Algorithms to Predict Breast Cancer Survival. *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association*. 2008;13(2):11–6.

[21] Thongkam J, Xu G, Zhang Y, Huang F, editors. Support Vector Machine for Outlier Detection in Breast Cancer Survivability Prediction. *Advanced Web & Network technologies, & Applications*; 2008.

[22] Fu B, Liu P, Lin J, Deng L, Hu K, Zheng H. Predicting Invasive Disease-Free Survival for Early-stage Breast Cancer Patients Using Follow-up Clinical Data. *IEEE Trans Biomed Eng*. 2018. pmid:30475709

[23] Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. *Folia biologica*. 2019;65(5–6):212–20. pmid:32362304

[24] Dembrower K, Liu Y, Azizpour H, Eklund M, Strand F. Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction. *Radiology*. 2019;294(2):190872. pmid:31845842

[25] Ibrahim N, Kudus A, Daud I, Abu Bakar M. Decision Tree for Competing Risks Survival Probability in Breast Cancer Study. *Proc Wrlld Acad Sci Eng Tech*. 2008.

[26] Gardezi SJS, Elazab A, Lei B, Wang T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. *Journal of medical Internet research*. 2019;21(7):e14464. pmid:31350843