


PAPER • OPEN ACCESS

## Comparative Analysis in Execution of Machine Learning in Breast Cancer Identification: A Review

To cite this article: Ravindran Nadarajan and Noorazliza Sulaiman 2021 *J. Phys.: Conf. Ser.* **1874** 012032


View the [article online](#) for updates and enhancements.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology  
2021 Virtual Education

**Fundamentals of Electrochemistry:**  
Basic Theory and Kinetic Methods  
Instructed by: **Dr. James Noël**  
Sun, Sept 19 & Mon, Sept 20 at 12h–15h ET

Register early and save!



# Comparative Analysis in Execution of Machine Learning in Breast Cancer Identification: A Review

Ravindran Nadarajan, \*Noorazliza Sulaiman

Faculty of Electrical and Electronic Engineering Technology, University Malaysia Pahang, Pekan, Pahang.

ravin.nada@gmail.com, azliza@ump.edu.my(\*corresponding author)

**Abstract.** Carcinoma known as breast cancer is a significant common cancer among women worldwide. In line with the global trends, it accounts for many new cancer cases and cancer-related deaths, giving it a substantial public health issue in today's culture. Early diagnosis is the most effective method to reduce the number of deaths in patients with breast cancer. Effective and early diagnosis of breast cancer ensure like mammography or biopsy to ensure the long-term survival of affected patients. Several conflicts arise in using traditional approaches, such as overdiagnosis or under-diagnosis. Machine learning is used to overcome the issues where it can strengthen the current conventional diagnosing of patients with breast cancer. The application of the classification method for diagnosing breast cancer is reviewed in this paper. Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbour (KNN), Decision Tree, Artificial Neural Network (ANN), and logistic regression are six methods presented in the review. These techniques are integrated with conventional methods, often allow physicians to diagnose breast cancer effectively. In summary, machine learning improvises in diagnosing breast cancer in terms of accuracy, sensitivity, and specificity with excellent performance and quality of patients.

## 1. Introduction

Breast cancer is one of the leading cancers globally, affecting 2.1 million women every year and causes the largest number of cancer-related deaths among women [1]. Table 1 shows that breast cancer is one of the highest cancer-related cases that affect the whole USA populations. In 2017, 250,520 new female breast cancer cases were reported, and 42,000 died of female breast cancer. For every 100,000, 125 new female breast cancer cases have been reported, and 20 died of cancer [2]. In a total of the population in the USA, 15.18% of the population has breast cancer.

**Table 1.** Statistic of cancers at the whole USA in 2017.

Cancer Type	Female Breast	Lung and Bronchus	Colon and Rectum	Corpus and Uterus, NOS	Thyroid	Melanomas of the Skin	Non-Hodgkin Lymphoma	Kidney and Renal Pelvis	Pancreas	Leukemias
Age-adjusted Rate	125.1	49.4	32.3	27.2	19.4	18.1	15.2	11.9	11.3	10.5



Case Count	250,520	107,545	66,869	57,368	33,460	34,983	31,345	24,168	24,551	20,992
Population	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491	165,021,491

Early diagnosis is one of the essential steps to take to reduce the mortality rate. Early diagnosis of carcinoma is essential to treat the disease quickly; therefore, it is necessary to develop techniques that may help physicians generate accurate diagnosis [3]. Detection of non-invasive carcinoma (ductal carcinoma in place, DCIS) shows a relative survival of affected women of > 94%.

## 2. Breast cancer

A breast is composed of three main parts: lobules, ducts, and connective tissue. Most breast cancers begin in the ducts or lobules. Breast cancer can spread outside the breast through blood vessels and lymph vessels. When breast cancer spreads to other parts of the body, it is metastasized [4].

There are two most common types of cancer: in situ (or non-invasive) and invasive. Non-invasive cancers stay in the breast in the milk ducts or lobules. They do not grow or invade normal tissues inside or outside the breast. Non-invasive cancers are sometimes referred to as carcinoma in situ (“in the same place”) or pre-cancers. Invasive cancers are those in which cancer cells migrate beyond the ducts’ basement membrane and lobules to the adjacent normal tissue. Most popular breast cancers are Ductal Carcinoma In Situ (DCIS), Lobular Carcinoma In Situ (LCIS), Invasive Ductal Carcinoma (IDC), and Invasive Lobular Carcinoma (ILC) [5]. IDC accounts for approximately 55 % of the prevalence of breast cancer on diagnosis [6].

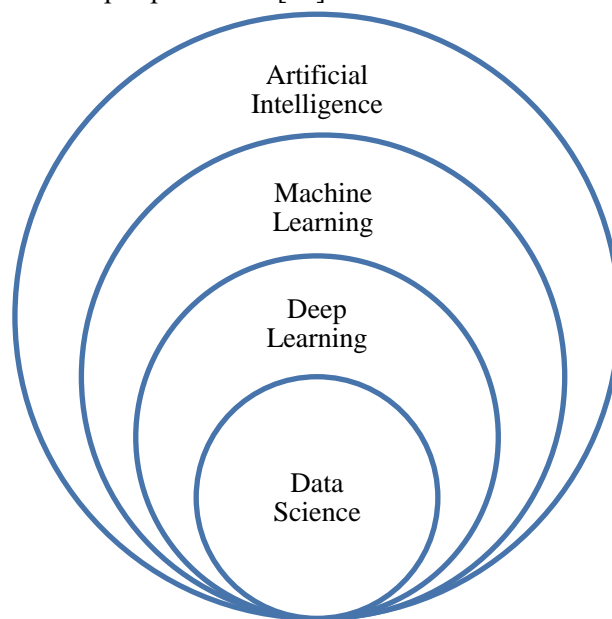
This one of the reasons why early diagnosis is essential. Early diagnose is to ensure all women to detect any anomalies in their breast to prevent severe cases. One of the most common methods used to diagnose cancer is mammography. Mammography is the best screening method that uses low-dose x-rays to detect tumours in the breast picture [7]. There are some setbacks in this diagnostic type, one of which is overdiagnosis[8]. Mammograms are the image that mammography creates. The mammograms are among the most complex diagnostic objects to understand owing to their low contrast and discrepancies between tissue forms. Significant visual clues to breast cancer include tentative signs of mass and calcification clusters [9].

Besides that, breast cancer Dynamic Magnetic Resonance Imaging (MRI) has appeared as a useful diagnostic method for detecting breast cancer due to its high sensitivity. It has shown a position where the results of traditional mammography techniques are inaccurate. MRI currently has a high detection sensitivity of breast cancer estimated as high as 94-100 %. However, a lower specificity is estimated at 37-97 %, best performed by mammography in specificity [10].

A biopsy is one of the conventional methods that are used to diagnosing breast cancer. Cell or tissue biopsy is required to provide a definitive diagnosis of potentially malignant breast lesions in the vast majority of patients. Core-needle biopsy (CNB) is currently known as the most widely used diagnostic procedure in the world. In contrast, fine-needle aspiration biopsy (FNAB) and surgical biopsy are less commonly used [11]. The purpose of the biopsy is to obtain diagnostic tissue while minimizing morbidity, restricting the possible spread of the tumour, and preventing interference with future treatments. Techniques that have developed to achieve these goals include open surgical biopsy, core biopsy, and fine-needle aspiration (FNA). Open (incisional) biopsy has long been the gold standard for soft tissue mass diagnosis, with a diagnostic accuracy of 94% to 99%.[12]. There is a drawback in using the biopsy method. A breast biopsy can cause architectural changes in the breast, such as scarring and tissue distortion [13].

The other conventional method used to diagnose breast cancer is using ultrasound. Ultrasound imaging is also an effective method for clinical diagnosis. The patient's health risk is minimal, and procurement costs are comparatively minimal. Speckle is a common phenomenon in ultrasonic pulse-echo measurements. It occurs when a coherent source and non-coherent detector are used to interrogate a rough medium on the wavelength's scale. Other ultrasound data properties that make visualization and registration difficult are noise inside organ, indistinct or blurred boundary, and signal drop-out [14][15]. By using this method may cause misdiagnosis or deficiency.

These drawbacks can be overcome by integrating with artificial intelligence or AI to reduce the risk of false alarm and overdiagnosis. AI refers to a field of computing dedicated to the creation of systems performing tasks that typically require human intelligence, branching off into different techniques [16]. Several variables are used to calculate the efficiency of the techniques/algorithm. Sensitivity and specificity are statistical measures of the performance of a binary classification test. In mammography, diagnosis sensitivity tests the proportion of real positive results that are correctly detected when it has cancer tissues. Specificity tests the proportion of negatives correctly showed when cancer is not present in the mammogram[17]. To achieve the required classification accuracy, which is expressed as the percentage of patients in the test set that were correctly identified. Besides that cut-off prediction between 0 and 1 had to be selected before any network output (ranging from 0 to 1) could be interpreted as a breast cancer relapse prediction [18].



**Figure 1.** Illustration of AI and its subsets.

### 3. Artificial intelligence and machine learning

Artificial Intelligence (AI) may be a general term that uses a computer to model intelligent behaviour with minimal human intervention [19]. Artificial intelligence is one of the innovations that helps many applications in the medical field. One of them is medical imaging, interpretation, and processing. Besides that, AI also contributes a lot to aided reporting, follow-up planning, data storage, data mining, and many others. Artificial intelligence also eases the health workers to diagnose the patients faster and exact results. Machine learning is an AI that allows computers to learn from the data without being explicitly programmed and has been extensively applied to medical imaging [16][20][21]. Figure 1 above illustrates the relationship between each category of subsets with artificial intelligence.

Numerous classification models in data mining domains are adapted to breast cancer diagnosis based on patients' historical medical records [22]. Support Vector Machine or SVM is one of the classification models commonly used in a wide variety of applications. The usages are packed in SVM

and can-do face and speech recognition, face detection, and image recognition. SVM is one of the most useful algorithm dates to the 1990s. It is a mathematical, technological, and orthogonal transformation used to translate a set of observations of correlated variables into a collection of values of linearly uncorrelated variables [23][24].

The other classification algorithm that is universally used is Naïve Bayes. Naïve Bayes is a subset of Bayesian decision theory. It is called naive since the formulation gives some naïve suppositions. The text-processing capabilities of Python, which divide a document into a vector, are used. It can be used to find the text. Classifies can be placed in a human-readable form. It is a standard method of classification and conditional independence, over-fitting, and Bayesian methods [25]. However, few drawbacks exist in Naïve Bayes, which is an independent assumption among the features that are not practical in real datasets[26].

One of the simplest algorithms is the K-Nearest Neighbours Algorithm (KNN); it is commonly used in the predictive analysis [27]. When classifying, the basic principle of KNN is to make the closest neighbour instances in the form of a predefined vote on space. Then the new instance class is defined by the most common class of the nearest k neighbours. It is vital to choose the value of Ka Priori; various techniques have been proposed for selecting it, such as cross-validation and heuristics. To prevent tie votes, this value should not be a multiple of the number of groups [21].

Decision tree techniques have been widely used to build classification models. Such models closely resemble human reasoning and are easy to understand [28]. Decision trees, either classification or regression trees, are extremely appealing models for three reasons: they have an intuitive representation. The resulting model is easily interpreted and assimilated by humans. Second, decision trees are non-parametric models; no input is needed by the consumer and are therefore exceptionally good for exploratory information exploration. Finally, efficiency is degraded gracefully by increasing the size of training results [29].

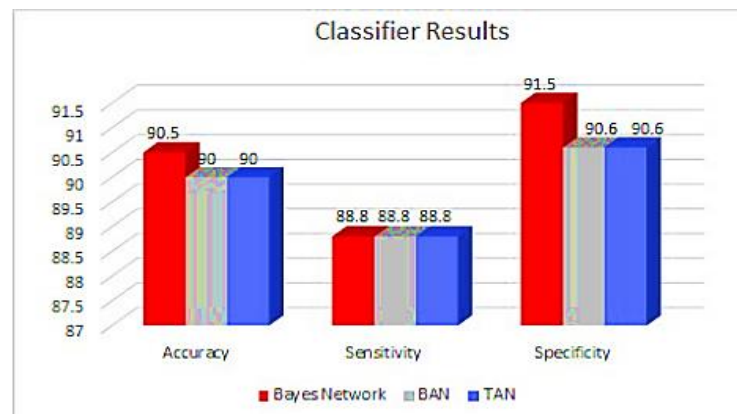
One of the most universally used classification algorithms is an artificial neural network. Artificial Neural Network (ANN) has been widely applied in breast cancer diagnosis using distinctive features. The network can separate non-determined data into the different classes and calculate the probability of belonging to each class [10]. The architecture of the neural network is organized into layers composed of interconnected nodes. Each network node performs a weighted sum of the input data that is then transferred to an activation function. During the training point, weights are dynamically optimized. To improve the performance of traditional ANN when using with the deep architectures, the deep learning (DL) strategy was developed [16].

Besides that, logistic regression is one of the commonly used classification system types. The logit-the normal logarithm of an odds ratio is the fundamental mathematical principle that underlies logistic regression. The most straightforward instance of a logit is extracted from a contingency table of 2 by 2 [30]. Logistic regression works like linear regression, but with a binational response variable. It is easier to handle more than two explanatory variables simultaneously [31].

#### **4. Application of machine learning algorithm in breast cancer diagnosis**

There are several machine learning models used to diagnose and quantify breast cancer efficiency. In a study done by Maglogiannis *et al.* in 2009, they used SVM to diagnose breast cancer using datasets from fine-needle aspiration. The datasets are from Wisconsin Diagnostic Breast Cancer and as well from Wisconsin Prognostic Breast Cancer. The role of the diagnosis is to find the malignant and benign breast masses. The performance results show that, the optimized SVM has accuracy of 96.91%, specificity (up 97,67%) and sensitivity (up to 97,84%) [32]. Wang *et al.* did another similar study. In 2018 using a support vector machine-based ensemble algorithm for breast cancer diagnosis. The study implements the Weighted Area's hybridization under the Receiver Operating Characteristic Curve Ensemble and with SVM. The model reduces the variance by 97.89% and increases accuracy by 33.34%, compared to the best single SVM model on The Surveillance, Epidemiology, and End Results (SEER) dataset [22].

In the early year 2014, a simple naïve Bayes used for diagnosing breast cancer by Kharya *et al.* shows 93% of accuracy [33]. In 2015 a study made by Karabatak used an improvised naïve Bayes to diagnose breast cancer. They used a weighted naïve Bayes classifier to overcome the drawbacks of original naïve Bayes, crisp classes assigned to the training data. The weighted naïve Bayes classifier shows obtain 99.11% sensitivity, 98.25% specificity, and 98.54% the accuracy values, respectively [34]. A performance comparison was made with Tree Augmented Naive Bayes (TAN), Boosted Augmented Naive Bayes (BAN), and Bayes Belief Network (BBN) by Bazila Banu & Thirumalaikolundusubramanian in 2018. The comparison study's findings using Bayes classifiers such as Bayes network, BAN, and TAN are based on two parameters: patients with benign and malignant cancers [35]. Figure 2 below shows the comparison of classifier result of Bayes network, TAN and BAN.



**Figure 2.** Comparative Results of Classifiers by Accuracy, Specificity and Sensitivity[35]

In the year 2000, a research study by Sarkar & Leong shows an implementation of the K nearest neighbour algorithm to diagnose breast cancer. It produces the overall classification result 1.17% better than the best result known for this problem. These studies also say the limitation in using KNN, which is the requirement of enormous storage for storing a more extensive set of data [36]. A similar study was made by El-Baz in 2014 using a hybrid intelligent system-based rough set and ensemble classifier. The combined classifier is based on the KNN classifier [37]. A comparative analysis between K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm was made by Okfalisa *et al.* in 2017. The analysis concluded that the accuracy ratio was evaluated to find that the maximum accuracy of KNN was 94.95 %. The average accuracy during the test is 93.94 %, while the highest accuracy of MKNN being 99.51 %. The average accuracy during the test was 99.20 % [27].

A study in 2003 by Jerez-Aragonés *et al* uses decision trees for diagnosing breast cancer. In the paper put into effect a joint neural network and decision trees model for prognosis of breast cancer relapse [18]. Lavanya, in 2012 did a study about an ensemble decision tree classifier for breast cancer data. In the study, decision trees variants used was Classification and regression trees or CART. The classifier is further in comparison to the CART with Feature Selection Method and hybrid approach. It concludes that the hybrid approach produces the highest accuracy of around 95.96%, and the CART with Feature Selection Method has 94.72% accuracy; however, the CART itself has 92.97 accuracies. Those comparisons are made using the datasets of Breast Cancer Wisconsin (Diagnostic)[38].

Besides, the neural network has improved the breast cancer detection process to achieve reliable and efficient outcomes. Dheeba *et al.* in 2004 investigate a new classification method for finding breast deformities in digital mammograms using the PSOWNN (Particle Swarm Optimized Wavelet Neural Network). The specificity of the PSOWNN is 94.167%, sensitivity is 92.15%, and accuracy of 93.671% [39]. In 2015, a paper published by Bhardwaj & Tiwari used the Genetically Optimized Neural Network (GONN) algorithm for solving classification problems in diagnosing breast cancer.

Their proposed approach has a classification accuracy of 98.24 %, 99.63 % and 100 % for 50-50, 60-40, 70-30 partition tests and 100 % for 10 cross-validation [40].

Logistic regression also enhances or strengthens the performance of breast cancer diagnosis. In 2015, a paper published by Seddik & Shawky used the logistic regression approach to diagnose breast cancer. The built-in model has an average classification accuracy of 98.9 % of the data set used, with a sensitivity and specificity of 98.5 % and 99.1 %, respectively [41]. A paper is done in 2009 by Chhatwal et al. used logistic regression model based on the national mammography database format to aid breast cancer diagnosis. The result shows that 90% specificity the sensitivity of the model is 90% [42].

**Table 2.** Comparisons of the classification method.

Classification Method	SVM	Naïve Bayes	Naïve Bayes	KNN	KNN	Decision Tree	Decision Tree	ANN	ANN	Logistic Regression
Hybridization/Optimized	Optimized	None	Weighted	None	Modified	CART with feature selection	CART	PSOWNN	GONN	None
Specificity (%)	97.67	Null	98.25	Null	Null	Null	Null	94.167	Null	99.1
Sensitivity (%)	97.84	Null	99.11	Null	Null	Null	Null	92.15	Null	98.5
Accuracy (%)	96.91	93	98.54	94.95	99.51	94.72	95.96	93.671	98.24	98.9

Table 2 shows a comparison of the classification method and its specificity, sensitivity, and accuracy. Based on the comparison, we can assume that logistic regression does give the highest accuracy among all the classification methods.

## 5. Conclusion

In conclusion, many machine learning approaches can be used to diagnose breast cancer. Machine learning allows greater precision, sensitivity, and specificity to diagnose breast cancer, showing reliable performance through general. Different approaches may lead to enhance the diagnosis of breast cancer. One way is by hybridizing machine learning with other types of algorithms. This approach can lead to better and reliable outcomes. In future by hybridizing algorithm also can be implemented or integrated to unlock several algorithms that can be experimented, which allow doctors to obtain a complete result and solve the widespread problem faced during a breast cancer diagnosis.

## Acknowledgement

The authors acknowledge MOHE FRGS Grant No: (RDU190180) and Universiti Malaysia Pahang for financial and facility support.

## References

- [1] W. H. Organization, "WHO | Breast cancer," *WHO*, 2018. <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> (accessed Oct. 07, 2019).
- [2] CDC, "USCS Data Visualizations - CDC," 2016.

- <https://gis.cdc.gov/Cancer/USCS/DataViz.html> (accessed Aug. 25, 2020).
- [3] B. M. Abed *et al.*, “A hybrid classification algorithm approach for breast cancer diagnosis,” *IEACon 2016 - 2016 IEEE Ind. Electron. Appl. Conf.*, pp. 269–274, Oct. 2017, doi: 10.1109/IEACON.2016.8067390.
- [4] “What is breast cancer?,” *J. Okla. State Med. Assoc.*, vol. 105, no. 10, pp. 409–10, Oct. 2012, Accessed: Sep. 22, 2020. [Online]. Available: [https://www.cdc.gov/cancer/breast/basic\\_info/what-is-breast-cancer.htm](https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm).
- [5] G. Sree, V. K. Velvizhi, and R. Sundararajan, “Electric field distribution of malignant breast tissue under needle electrode configuration,” *Annu. Rep. - Conf. Electr. Insul. Dielectr. Phenomena, CEIDP*, no. 11c, pp. 267–270, 2012, doi: 10.1109/CEIDP.2012.6378772.
- [6] F. Siddiqui, S. Gupta, S. Dubey, S. Murtuza, and A. Jain, “Classification and diagnosis of invasive ductal carcinoma using deep learning,” *Proc. Conflu. 2020 - 10th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 242–247, 2020, doi: 10.1109/Confluence47617.2020.9058077.
- [7] F. Shirazi and E. Rashedi, “Feature weighting for cancer tumor detection in mammography images using gravitational search algorithm,” in *2016 6th International Conference on Computer and Knowledge Engineering, ICCKE 2016*, Dec. 2016, pp. 310–313, doi: 10.1109/ICCKE.2016.7802158.
- [8] S. H. Heywang-Köbrunner, A. Hacker, and S. Sedlacek, “Advantages and disadvantages of mammography screening,” *Breast Care*, vol. 6, no. 3. Karger Publishers, pp. 199–207, 2011, doi: 10.1159/000329005.
- [9] A. T. Azar and S. A. El-Said, “Performance analysis of support vector machines classifiers in breast cancer mammography recognition,” *Neural Comput. Appl.*, vol. 24, no. 5, pp. 1163–1177, Apr. 2014, doi: 10.1007/s00521-012-1324-4.
- [10] F. Keivanfard, M. Teshnehlab, M. A. Shoorehdeli, K. Nie, and M.-Y. Su, “Feature selection and classification of breast cancer on dynamic Magnetic Resonance Imaging by using artificial neural networks,” in *2010 17th Iranian Conference of Biomedical Engineering (ICBME)*, Nov. 2010, pp. 1–4, doi: 10.1109/ICBME.2010.5704942.
- [11] R. B. Sennerstam, B. S. H. Franzén, H. O. T. Wiksell, and G. U. Auer, “Core-needle biopsy of breast cancer is associated with a higher rate of distant metastases 5 to 15 years after diagnosis than FNA biopsy,” *Cancer Cytopathol.*, vol. 125, no. 10, pp. 748–756, Oct. 2017, doi: 10.1002/cncy.21909.
- [12] S. Kasraeian, D. C. Allison, E. R. Ahlmann, A. N. Fedenko, and L. R. Menendez, “A comparison of fine-needle aspiration, core biopsy, and surgical biopsy in the diagnosis of extremity soft tissue masses,” in *Clinical Orthopaedics and Related Research*, 2010, vol. 468, no. 11, pp. 2992–3002, doi: 10.1007/s11999-010-1401-x.
- [13] S. H. Taplin *et al.*, “Effect of previous benign breast biopsy on the interpretive performance of subsequent screening mammography,” *J. Natl. Cancer Inst.*, vol. 102, no. 14, pp. 1040–1051, 2010, doi: 10.1093/jnci/djq233.
- [14] C.-J. Chen, R.-F. Chang, and D.-R. Chen, “Automatic Tumor Diagnosis for Breast Ultrasound Using 3D Sub-volume Registration,” in *2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering*, Jun. 2009, pp. 61–68, doi: 10.1109/BIBE.2009.17.
- [15] L. R. Sultan, S. M. Schultz, T. W. Cary, and C. M. Sehgal, “Machine Learning to Improve Breast Cancer Diagnosis by Multimodal Ultrasound,” in *2018 IEEE International Ultrasonics Symposium (IUS)*, 2018, pp. 1–4, doi: 10.1109/ULTSYM.2018.8579953.
- [16] F. Pesapane, M. Codari, and F. Sardanelli, “Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine,” *European Radiology Experimental*, vol. 2, no. 1. Springer, Dec. 01, 2018, doi: 10.1186/s41747-018-0061-6.
- [17] X.-H. Zhou, D. K. McClish, and N. A. Obuchowski, *Statistical methods in diagnostic medicine*, vol. 569. John Wiley & Sons, 2009.
- [18] J. M. Jerez-Aragónés, J. A. Gómez-Ruiz, G. Ramos-Jiménez, J. Muñoz-Pérez, and E. Alba-



- Conejo, "A combined neural network and decision trees model for prognosis of breast cancer relapse," *Artif. Intell. Med.*, vol. 27, no. 1, pp. 45–63, Jan. 2003, doi: 10.1016/S0933-3657(02)00086-6.
- [19] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism.*, vol. 69, pp. S36–S40, Apr. 2017, doi: 10.1016/j.metabol.2017.01.011.
- [20] S. Kharya, D. Dubey, and S. Soni, "Predictive Machine Learning Techniques for Breast Cancer Detection," *Int. J. Comput. Sci. Inf. Technol.*, 2013.
- [21] W. Cherif, "Optimization of K-NN algorithm by clustering and reliability coefficients: Application to breast-cancer diagnosis," *Procedia Comput. Sci.*, vol. 127, pp. 293–299, 2018, doi: 10.1016/j.procs.2018.01.125.
- [22] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, Jun. 2018, doi: 10.1016/j.ejor.2017.12.001.
- [23] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan, and M. J. Rajabi, "Advantage and drawback of support vector machine functionality," in *I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology, Proceedings*, 2014, no. I4ct, pp. 63–65, doi: 10.1109/I4CT.2014.6914146.
- [24] W. K. Estes, "Learning Theory," *Annu. Rev. Psychol.*, vol. 13, no. 1, pp. 107–144, Jan. 1962, doi: 10.1146/annurev.ps.13.020162.000543.
- [25] D. Parmar, "International Journal of Advance Engineering and Research SURVEY ON WEB USAGE MINING AND PRE-FETCHING," pp. 222–224, 2015.
- [26] S. L. Ang, H. C. Ong, and H. C. Low, "Classification using the general bayesian network," *Pertanika J. Sci. Technol.*, vol. 24, no. 1, pp. 205–211, 2016.
- [27] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017*, vol. 2018-Janua, pp. 294–298, 2018, doi: 10.1109/ICITISEE.2017.8285514.
- [28] S. B. Kotsiantis, "Decision trees: A recent overview," *Artificial Intelligence Review*. 2013, doi: 10.1007/s10462-011-9272-4.
- [29] C. Bulac and A. Bulac, "Decision Trees," in *Advanced Solutions in Power Systems: HVDC, FACTS, and AI Techniques*, 2016.
- [30] C. Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002, doi: 10.1080/00220670209598786.
- [31] S. Sperandei, "Understanding logistic regression analysis," *Biochem. Medica*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.
- [32] I. Maglogiannis, E. Zafiroopoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," *Appl. Intell.*, vol. 30, no. 1, pp. 24–36, Feb. 2009, doi: 10.1007/s10489-007-0073-z.
- [33] S. Kharya, S. Agrawal, and S. Soni, "Using Bayesian Belief Networks for Prognosis & Diagnosis of Breast Cancer," ... *Adv. Res. Comput. ...*, 2014.
- [34] M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian," *Meas. J. Int. Meas. Confed.*, vol. 72, pp. 32–36, 2015, doi: 10.1016/j.measurement.2015.04.028.
- [35] A. Bazila Banu and P. Thirumalaikolundusubramanian, "Comparison of bayes classifiers for breast cancer classification," *Asian Pacific J. Cancer Prev.*, vol. 19, no. 10, pp. 2917–2920, Oct. 2018, doi: 10.22034/APJCP.2018.19.10.2917.
- [36] M. Sarkar and T. Y. Leong, "Application of K-nearest neighbors algorithm on breast cancer diagnosis problem.," *Proc. AMIA Symp.*, pp. 759–763, 2000.
- [37] A. H. El-Baz, "Hybrid intelligent system-based rough set and ensemble classifier for breast cancer diagnosis," *Neural Comput. Appl.*, vol. 26, no. 2, pp. 437–446, 2014, doi: 10.1007/s00521-014-1731-9.

- [38] D. Lavanya, "Ensemble Decision Tree Classifier For Breast Cancer Data," *Int. J. Inf. Technol. Converg. Serv.*, vol. 2, no. 1, pp. 17–24, Feb. 2012, doi: 10.5121/ijitcs.2012.2103.
- [39] J. Dheeba, N. Albert Singh, and S. Tamil Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *J. Biomed. Inform.*, vol. 49, pp. 45–52, 2014, doi: 10.1016/j.jbi.2014.01.010.
- [40] A. Bhardwaj and A. Tiwari, "Breast cancer diagnosis using Genetically Optimized Neural Network model," *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4611–4620, Jun. 2015, doi: 10.1016/j.eswa.2015.01.065.
- [41] A. F. Seddik and D. M. Shawky, "Logistic regression model for breast cancer automatic diagnosis," *IntelliSys 2015 - Proc. 2015 SAI Intell. Syst. Conf.*, pp. 150–154, 2015, doi: 10.1109/IntelliSys.2015.7361138.
- [42] S. Murugan, B. Muthu Kumar, and S. Amudha, "Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest," *Int. Conf. Curr. Trends Comput. Electr. Electron. Commun. CTCEEC 2017*, pp. 763–766, 2018, doi: 10.1109/CTCEEC.2017.8455058.