

A Hybrid Method of 1D-CNN and Machine Learning Algorithms for Breast Cancer Detection

Ahmed Adil Nafea*¹, Manar AL-Mahdawi², Khattab M Ali Alheeti³, Mustafa S. Ibrahim Alsumaidaie³, Mohammed M AL-Ani⁴

¹Department of Artificial Intelligence, College of Computer Science and IT, University of Anbar, Ramadi, Iraq.

²Department of Physics, College of Science, AL-Nahrain University, Baghdad, Iraq.

³Department of Computer Science, University of Anbar Ramadi, Iraq.

⁴Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia.

*Corresponding author.

Received 13/09/2023, Revised 05/12/2023, Accepted 07/12/2023, Published Online First 20/03/2024, Published 01/10/2024



© 2022 The Author(s). Published by College of Science for Women, University of Baghdad.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Breast cancer is a health concern of importance, and it is crucial to detect it early for effective treatment. Recently there has been increasing interest in using artificial intelligence (AI) for breast cancer detection, which has shown results in enhancing accuracy and reducing false positives. However, there are some limitations regarding accuracy in detection. This study introduces an approach that utilizes 1D CNN as feature extraction and employs machine learning (ML) algorithms such as XGBoost, random forests (RF), decision trees (DT) support vector machines (SVM) and k nearest neighbor (KNN) to classify samples as either benign or malignant aiming to enhance accuracy. Our findings reveal that the XGBoost algorithm with feature extraction (1D CNN) achieved an accuracy of 98.24% on the test set. This study highlights the feasibility of employing machine learning algorithms and deep learning (DL). This study uses a dataset of Wisconsin breast cancer (WBC), for detecting breast cancer. The proposed approach has a good detection and improving outcomes via shows accurate and reliable tools for diagnosing breast cancer.

Keywords: Breast cancer diagnosis, Deep learning, Machine learning, Wisconsin, 1D-CNN.

Introduction

Breast cancer is a significant global health issue, with timely identification and diagnosis playing a key role in enhancing patient prediction. In recent developments in technology, ML and DL have shown a good tools in the fight besides breast cancer. These techniques have shown capable results in the prediction of breast cancer, Assisting healthcare

professionals in making well-informed choices regarding patient treatment ¹.

The ML is a subfield of AI and centers on crafting algorithms efficient of developing knowledge from data. This could be applied to statistical models and algorithms to identify complex relationships and models in large datasets. In the

medical domain, ML algorithms have found several employments, covering disease detection and forecasting. In the breast cancer, these algorithms can be learned using large repositories of medical images and patient details, enabling the recognition of breast cancer attributes and potential dangers ².

DL is a type of ML that utilizes artificial neural networks to model complicated relationships between inputs and outputs. Large datasets may be utilized to train DL algorithms, which makes it possible to automatically identify and classify breast cancers this to increases identification accuracy and reduces the need for manual evaluation ³.

The core benefit of utilizing DL and ML for breast cancer detection is the ability to deal with large datasets contained from demographic, clinical, and image data based on a number of risk variables, models that correctly evaluation the possibility of detection breast cancer might be created with this data. These techniques work for the complex and dynamic type of breast cancer as they can also supervise nonlinear connections between variables.

Related work

Arshad. ⁴ this research holds significance in enhancing the precise prediction and assessment of breast cancer, a prevailing form of cancer that ranks among the foremost causes of female mortality across the globe. ML methodologies have demonstrated potential in promptly detecting and foretelling breast cancer. The investigation employs the WBC Diagnostic dataset to assess the efficiency of ensemble classifiers and ML, specifically RF, Logistic Regression (LR), AdaBoost, and Xtreme Gradient Classifier. The primary aim is to ascertain the optimal ensemble and ML classifiers for accurately detecting and diagnosing breast cancer, with a focus on achieving the highest level of Accuracy.

Harika et al. ⁵ the primary emphasis of this investigation revolves around harnessing ML to aid in the diagnosis of cancer, particularly in the anticipation of malignant neoplasms through fine needle aspiration. The study assesses six distinct classification techniques, with an emphasis on precision, objectivity, and reproducibility. These methods encompass Multilayer Perceptron, DT, RF,

ML and DL have shown an effective tool in the detection of breast cancer, so it's important to provide healthcare with valuable information to make informed decisions about patient care. The issue of accuracy stays to be a major challenge in the application of DL and ML in the prediction of breast cancer. However, by combining different ML algorithms, combining previous knowledge and domain-specific information, and evaluating and validating the algorithms in the clinical setting, it is possible to improve the accuracy of these techniques and enhance their reliability and generalizability. Utilizing deep learning and machine learning in breast cancer prediction holds great promise for improving patient outcomes and reducing the burden of breast cancer worldwide. To address the issue of accuracy in breast cancer prediction using deep learning and machine learning, this study proposed. used a combination of different ML algorithms (XGBoost, SVM, DT, RF, and KNN) to achieve better performance and accuracy in this study and proposed a novel 1D CNN with ML algorithms to predict breast cancer.

SVM, and Deep Neural Network (DNN). To conduct this evaluation, the research leverages the University of Wisconsin Hospital database, a repository containing thirty attributes that intricately delineate the nucleus properties of breast masses.

Elsadig et al. ⁶ this research delves into the contribution of AI in enhancing the prompt identification of breast cancer. The investigation examines a range of eight classification models, comprising both individual and ensemble classifiers, while also employing five distinct techniques for feature selection. This process culminates in the creation of a reliable dataset containing a mere 17 features. The experimental findings reveal that among the classifiers assessed, namely the multi-layer perceptron, SVM, and stack models, three exhibit superior classification accuracy in comparison to their counterparts.

Chen et al. ⁷ this study aims to establish various ML models, including XGBoost, RF, LR, and the K-NN, to classify and predict breast cancer for early diagnosis. The evaluation index is the recall,

with precision, accuracy, and the F1 score also considered. The dataset was standardized, and 15 features were selected using the Pearson correlation test. The K-NN model used cross-validation to select the optimal k value. The hierarchical sampling method was used to address the problem of positive and negative sample imbalance. The division of different data sets affected the prediction performance of the same model.

Sakib et al.⁸ This study discusses the importance of treatment and early detection of breast cancer and how ML and DL techniques can help medical professionals in this process. The study compares five popular supervised ML techniques (SVM, DT, LR, RF, KNN) and a DL technique for breast cancer detection using the WBC dataset. The assessment criteria employed to gauge the efficacy and efficiency of the models encompass a range of evaluation metrics, encompassing accuracy, recall, precision, F1 score.

Abiodun et al.⁹ this research underscores the significance of early detection and preventive

Research Methodology

This research paper explores the efficacy of employing 1D-CNN architectures as feature extractors in tandem with diverse ML classifiers. The aim is to assess their collective potential for classifying and predicting breast cancer within a given dataset. This study starts with data acquisition, followed by a preprocessing stage that encompasses three sequential steps: data cleansing, attribute selection, and target role assignment. Subsequently, the focus shifts to feature extraction utilizing the 1D-CNN technique. The extracted features are then harnessed to create machine learning algorithms qualified of predicting breast cancer based on new measurements. For the purpose of evaluating

measures for breast cancer through the utilization of data mining methodologies. The study conducts a comparative analysis of classification accuracy among four distinct ML algorithms: KNN, DT, Naive Bayes (NB), and SVM. The primary aim is to identify the most precise supervised ML algorithm for diagnosing breast cancer. The findings demonstrate that, within the given dataset, NB exhibits the highest accuracy, surpassing KNN, SVM, and DT. In light of these outcomes, the research proposes that the integration of data mining and ML techniques can empower practitioners in formulating tools for early breast cancer detection.

The key limitation behind their studies lies in the dependency on using machine learning or deep learning focused on accuracy performance detection, and it needs to improve accuracy. This study proposed 1D-CNN as a feature extraction with machine learning algorithms such as SVM, KNN, DT, RF, and XGBoost. This finding implies the effectiveness of proposing 1D-CNN with a ML algorithm, for the detection of breast cancer.

algorithm performance, the model is subjected to new data with connected labels. This evaluation typically involves partitioning the labeled dataset into two segments using the Train_test_split method. The data is employed to create the machine learning model, establishing the training set 70%, while the staying 30% is reserved for evaluating model efficiency, forming the test set. Upon accurate testing of the models, the results are compared to discern the algorithm that yields the highest accuracy, thereby identifying the most predictive approach for breast cancer detection. The proposed method's workflow is shown in Fig. 1.

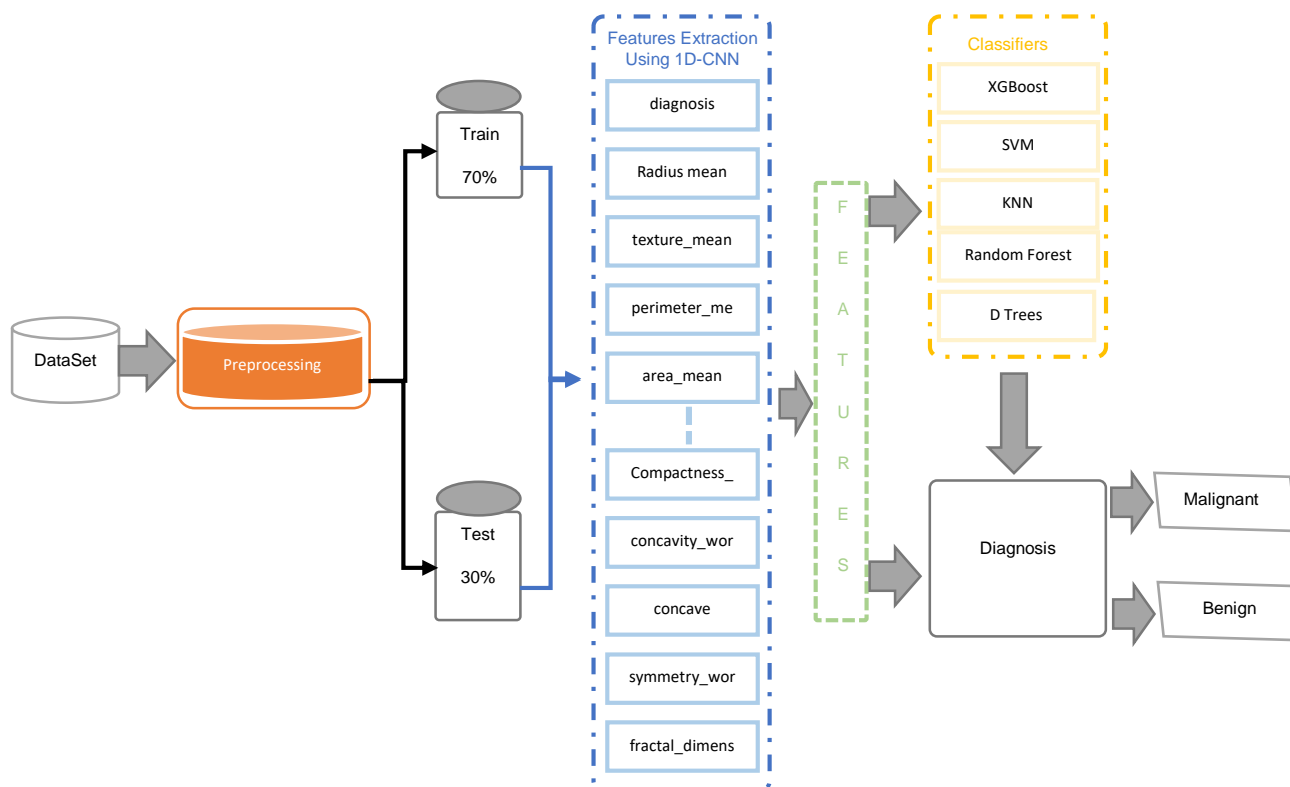


Figure 1. Research Methodology

Dataset

Within this study, utilized the WBC Diagnostic dataset sourced from the University of Wisconsin Hospitals Madison Breast Cancer Database¹⁰. The dataset's attributes are derived from digitized images of breast cancer samples obtained through fine-needle aspiration (FNA). These attributes encapsulate the characteristics of cell nuclei evident in the images. The WBC Diagnostic dataset comprises 569 instances, with 357 instances

categorized as benign and 212 instances as malignant. This equates to a distribution of 62.74% for the benign class and 37.26% for the malignant class. The dataset encompasses two distinct classes and is composed of 11 integer-valued attributes, including parameters like -Id, -Diagnosis, -Radius, -Texture, -Area, -Perimeter, -Smoothness, -Compactness, -Concavity, -Concave points, -Symmetry, and -Fractal dimension. As shown in Table 1 description of features on the dataset.

Table 1. Describe features of the dataset

Feature Name	Description	Data Type
id	Unique identification number for each patient	Numeric
diagnosis	Diagnosis (M = Malignant, B = Benign)	Categorical
radius_mean	Mean radius of the tumor	Numeric
texture_mean	Mean texture of the tumor	Numeric
perimeter_mean	Mean perimeter of the tumor	Numeric
area_mean	Mean area of the tumor	Numeric
smoothness_mean	Mean smoothness of the tumor	Numeric
compactness_mean	Mean compactness of the tumor	Numeric
concavity_mean	Mean concavity of the tumor	Numeric
... (and so on)	... (other descriptions of feature names)	...

Pre-processing

Before training machine learning models, this proposed applied min-max scaling to input features to ensure that they were on a similar scale. Min-Max scaling is a common pre-processing technique that scales input features to a specified range (usually [0, 1])¹¹. This can be important in machine learning, as features that are on vastly different scales can cause problems during training and may result in some features having a disproportionately large influence on the model's predictions. Through the process of scaling proposed input features to a uniform range, this study achieved the equitable consideration of each feature during training. This approach ensured that the proposed models could effectively discern significant patterns within the data without undue bias stemming from any single feature.

Splitting dataset

In this study, the dataset into 70% training and 30% testing. This proposed utilized stratified sampling to ensure the training and testing subsets were representative of the entire dataset. This technique is utilized in ML to keep the target variable distribution in both datasets, so enhancing facility and position in the data analysis process.

Features extraction

In this proposed utilizing 1D-CNN for features extraction from the input data rather than utilizing for classification or prediction. These features extraction can later be fed into another ML model, providing valuable input for its performance¹². The 1D-CNN is trained on a dataset utilizing a supervised learning approach to learn relevant relationships in the data during training, the network automatically learns a

set of filters that capture local patterns which identify important features in the data¹³.

Once the 1D-CNN is trained, this study can utilize it to extract features from new data that wanted to classify or predict. After that, the new data into the 1D-CNN and the output of one or more of the intermediate layers, which represent the features learned by the network. These features can then be used as input to another ML model, such as an SVM, XG-Boost, KNN, DT and RF as a classifier, to perform the final classification or prediction. Finally, by using a pre-trained 1D-CNN for feature extraction, this can benefit from the knowledge learned by the network on similar tasks, which can improve the performance of the proposed ML model based on the new features, ML algorithms classify breast cancer from the diagnosis dataset into Malignant or Benign.

As shown in Fig. 2 feature extraction via 1D CNN model. This study utilizes TensorFlow and sets random seeds for reproducibility. Define the input layer based on the shape of the training data. Build a 1D Convolutional Neural Network (CNN) model with the following layers:

- Conv1D layer with 64 filters and a kernel size of 3, using ReLU activation.
- MaxPooling1D layer with a pool size of 2.
- Flatten the layer to transform the 1D feature map into a 1D vector.
- Dense layer with 64 units and ReLU activation.

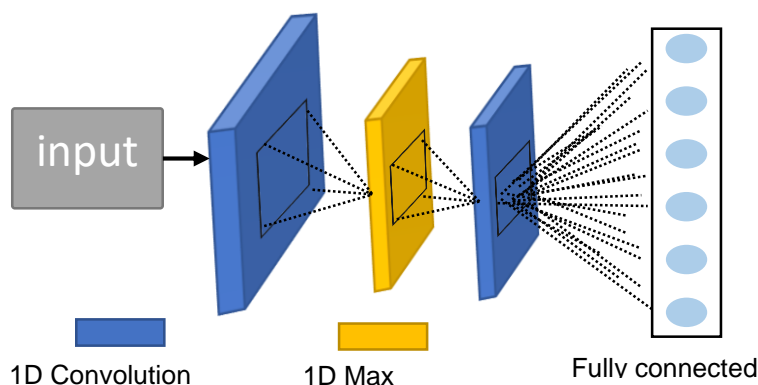


Figure 2. Feature extraction using 1D CNN

The input data consists of 30 columns of numeric features. Each column represents a specific measurement or characteristic related to the cell nuclei samples in our dataset. These features are used as input to a 1D Convolutional Neural Network (1D-CNN) for the purpose of feature extraction.

Machine learning classifiers

This section describes the ML classifiers proposed to classify breast cancer. This study used a combination of different ML algorithms (XGBoost, SVM, DT, RF, and KNN) algorithms using `random_state=42` this used to set the random seed for reproducibility. It ensures that the same results can be obtained when the code is run multiple times with the same dataset. This study uses the default settings for the Decision Tree Classifier and XG-Boost and SVM, which means you are not specifying any hyperparameters explicitly. While `n_neighbors` is set to 3, indicating that the classifier considers the labels of the three nearest neighbors to make predictions. No additional hyperparameters or settings, so the default settings for KNN, This study uses several hyperparameters that control the behavior of the Random Forest classifier. `n_estimators` is set to 10, meaning that your Random Forest consists of 10 decision trees and `max_depth` is set to 15, which means that each tree can grow to a maximum depth of 15 nodes.

Decision Tree

DT constitutes a graphical representation employing branching techniques to portray potential courses of action and their respective outcomes. This technique accommodates both categorical and numerical variables, eliminating the necessity for presumptions about data distribution or classifier configuration. DT excels in furnishing precise and streamlined classifications, even when handling extensive datasets^{14,15}.

XG-Boost

XGBoost stands as an ensemble technique that amalgamates numerous decision trees for prediction purposes. Its operational principle involves a stepwise inclusion of decision trees into a model, aiming to rectify errors introduced by prior trees. This sequential progression persists until the targeted level of accuracy is attained. Notably, XGBoost distinguishes itself by its capacity to enhance the performance of each decision tree through the application of gradient boosting techniques¹⁶.

Support Vector Machines

SVM is a widely adopted machine learning method employed primarily for binary classification endeavors. SVM strives to determine the hyperplane that optimizes the separation between two classes, with the margin denoting the distance between this hyperplane and the nearest data points from each class. SVM has demonstrated efficacy in breast cancer binary classification undertakings. In the context of breast cancer, SVM aims to precisely forecast whether a tumor possesses malignancy or benign characteristics, leveraging diverse tumor attributes^{17,18}.

K-Nearest Neighbor Classifiers

KNN algorithm categorizes unlabeled data by associating it with the closest labeled data of similar characteristics. Renowned for its straightforwardness and effectiveness, KNN is extensively utilized for supervised classification in scenarios involving multiple variables. The KNN classifier is solely influenced by one parameter: the choice of the number of nearest neighbors to be taken into account, denoted as *K* aimed at mitigating challenges like overfitting and underfitting¹⁹.

Random Forest Classifier

RF is a common ensemble ML technique suitable for both classification and regression assignments operating on the foundation of DT principles. It combines many decision trees, each enhanced on a separate subset of training data, to prepare predictions. It's can helping high-dimensional datasets full with various features, rendering it notably accurate in contrast to conventional classification methods²⁰.

The process of RF algorithm contains the deliberate selection of subsets from the training data and features in a random mode to build multiple decision trees. These trees are combined to get a final prediction through a consensus of their individual forecasts. This strategy effectively mitigates overfitting concerns and enhances the overall adaptability of the model for more generalized outcomes.

Evaluation

Accuracy, F-score, precision, and Recall are usually used for metrics evaluating the efficiency of machine learning models. These metrics provide valuable insights into different aspects of model performance.

Precision assesses the model's accuracy in finding positive instances between the whole predicted true positives(TP), showing a short false positives(FP) rate ²¹.

$$\text{Precision} = \frac{TP}{TP + FP} \quad 1$$

Recall estimates the model's capability to detect all positive instances out of the total TP, showing a low false negatives(FN) rate ²².

$$\text{Recall} = \frac{TP}{TP + FN} \quad 2$$

Accuracy, measures the whole correctness of the model via calculating the ratio of correct predictions

as a true negative (TN) and TP to the total number of predictions made ²³.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad 3$$

Lastly, **F-measure** combines precision and recall to provide a single metric that balances together measures ²⁴.

$$F - \text{measure} = \frac{2 \times (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad 4$$

In the presented equations, the different metrics are expressed via individual calculations that rely on the values of FN, TN, FP, and TP. This values are found from the models the actual truth and predictions ²⁵.

Results and Discussion

Breast cancer was classified into benign and malignant tumours using five classification methods. XgBoost and RF, SVM, DT, and KNN, and found that XGBoost was the best in diagnosing breast cancer, with XGBoost being the dominant classifier. XGBoost achieved the highest classification accuracy. XGBoost outperformed all other classifiers in all testing scenarios. Table 2 presents the results obtained of sensitivity. It can be seen that the accuracy of the XGBoost model had the highest value of 98.24%.

Breast cancer was classified into benign and malignant tumors by using 1D-CNN as feature extraction with multiple classification methods like XGBoost, SVM, DT, KNN, and RF which are used to evaluate the performance accuracy. The five classifiers are trained on the extracted patterns produced by the proposed 1D-CNN. As shown in Table 8, the accuracy results of all classifiers using the proposed 1D-CNN with XGBoost, SVM, DT, KNN and RF. this proposal found that XGBoost was the best in diagnosing breast cancer, with XGBoost being the dominant classifier. XGBoost achieved the highest classification accuracy. XGBoost outperformed all other classifiers in all testing scenarios. It can be seen that the accuracy of the XGBoost model had the highest value of 98.24% while the performance of SVM based on accuracy

has achieved 95.61%, DT achieved 92.98, KNN achieved 95.61, and RF achieved 94.70.

Table 2. Comparison algorithms

	Accuracy	Precision	Recall	F1-score
KNN	95.61	97	90	94
DT	92.98	93	88	90
SVM	95.61	1.00	88	94
RF	94.7	97	88	93
XGBoost	98.24	98	98	98

The outcomes of the accuracy assessment can be juxtaposed in Fig. 3, encompassing all employed models. Notably, a considerable uptick in accuracy was observed across nearly all models, underscoring their efficacy. These findings highlight that the XGBoost model achieved the pinnacle of accuracy at 98.24%, while the Decision Tree (DT) model exhibited a comparatively lower precision, standing at 92.98%.

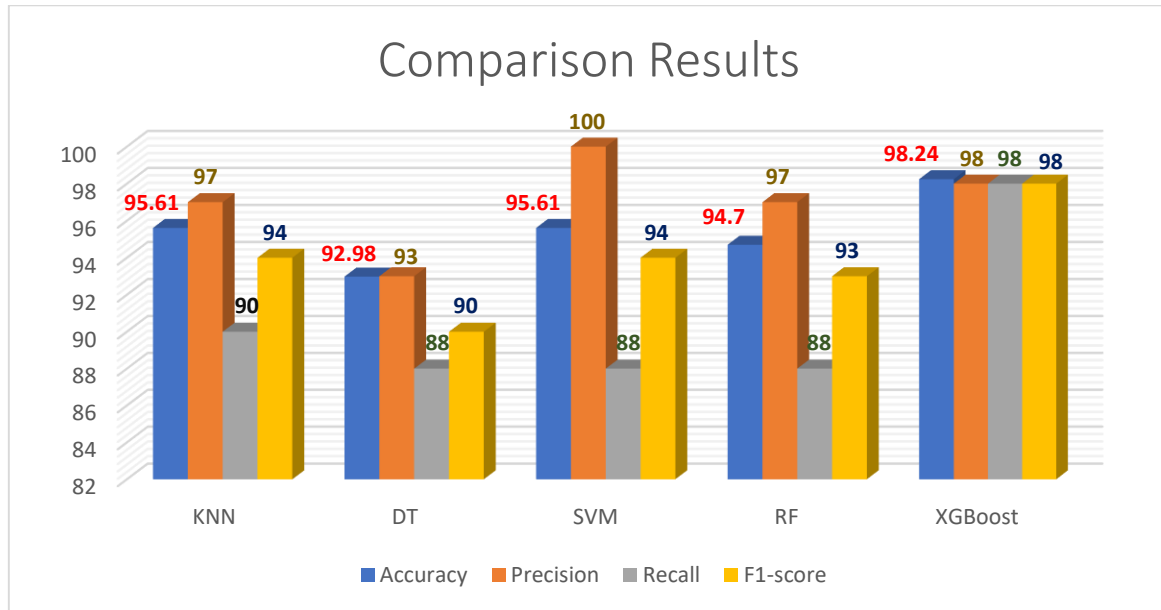


Figure 3. A Comparison Results

It is necessary to compare the proposed method against state-of-the-art methods as shown in Table 3 and Fig. 4.

Table 3. A comparison related work

Author's	Accuracy
Fabiano Teixeira (2019) ⁵	92
Shadman Sakib1 (2022) ⁸	96.66
Hua Chen (2023) ⁷	97.4
Muawia A. Elsadig1 (2023) ⁶	97.7
Waqas (2023) ⁴	98.1
Proposed Model	98.24

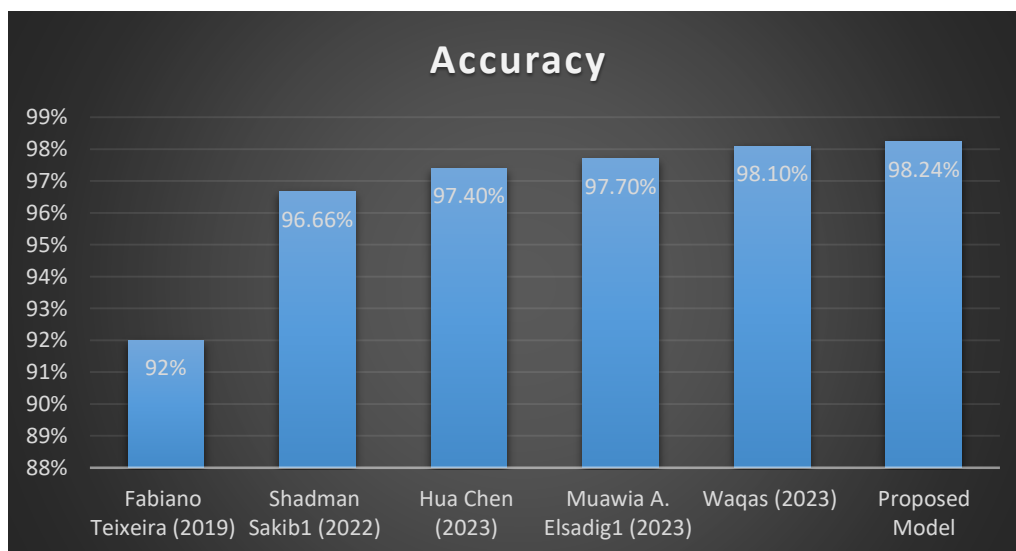


Figure. 4 A comparison related works

Arshad⁴ used an ensemble classifier and machine learning to detect breast cancer, which was proposed to have achieved an accuracy of 98.1%. Fabiano Teixeira et al. ⁵ used MLP, SVM, RF, DT and DNN to extract Breast cancer and acquired an accuracy of

0.92%. Elsadig1 et al. ⁶ have used an MLP, SM, and stack to extract Breast cancer and acquired an inaccuracy of 97.7%. Chen et al. ⁷ which utilized various machine learning models has obtained an accuracy of 97.4%. Sakib1 et al. ⁸, this study

proposed ML and DL techniques to detect breast cancer and acquired an accuracy of 96.66%. The previous study used the same dataset. This study proposed a hybrid method as an ID-CNN feature extraction and machine learning (ML) algorithms such as KNN, DT, NB, and SVM classifier, to detect breast cancer and acquired an accuracy of 98.24%.

The accuracy of breast cancer detection utilizing ML models is affected via factors such as model selection, dataset quality, size, and preprocessing

Conclusion

The early detection of breast cancer stays an important and constant focus in the field of scientific research. This study applied an assessment of classification accuracy through the evaluation of five ML algorithms like KNN, DT, RF, XGBoost, and SVM. The principal aim was to enhance the precision and effectiveness of classification algorithms. The findings underscore the pivotal role of ML in augmenting the prediction and diagnosis of breast cancer, a paramount outcome of this investigation. Notably, the research reveals that the

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for

Authors' Contribution Statement

A.A.N Carried out the investigation into techniques of breast cancer detection and he implemented the proposed work and performed the experiments.

M.A Advised the investigation and writing of this manuscript, in which all authors had approved the final version.

References

1. Waks AG, Winer EP. Breast cancer treatment: a review. *Jama*. 2019; 321(3): 288–300. <https://doi.org/10.1001/jama.2018.19323>
2. Averbuch T, Sullivan K, Sauer A, Mamas MA, Voors AA, Gale CP, et al. Applications of artificial intelligence and machine learning in heart failure. *Eur heart J health*. 2022; 3(2): 311–22.

techniques so the differences in reported accuracies can be recognized to these variations. ML advancements and larger datasets may contribute to improved accuracy rates.

ML and DL techniques are valuable tools for breast cancer detection, and the proposed method using 1D-CNN stays competitive in the field, depending on variations in reported accuracies among studies.

XGBoost algorithm surpasses 1D-CNN in terms of feature selection accuracy, when compared to the other algorithms utilized. Furthermore, the proposed approach demonstrates its efficacy in both the identification and prognosis of breast cancer, achieving the highest accuracy of 98.24% through the utilization of XGBoost. In the Future work uses Ensemble Models to Investigate the potential of ensemble models that combine the strengths of multiple machine learning algorithms to achieve even higher accuracy in breast cancer detection.

re-publication, which is attached to the manuscript.

- Ethical Clearance: The project was approved by the local ethical committee at University of Anbar.

k.M.A Advised the investigation and writing of this manuscript.

M.S.A Assisted in the realization of the ideas and writing of this manuscript.

M.M.A Assisted in the realization of the ideas and writing of this manuscript.

3. Chen K, Zhai X, Wang S, Li X, Lu Z, Xia D, et al. Emerging trends and research foci of deep learning in spine: bibliometric and visualization study. *Neurosurg Rev*. 2023; 46(1): 81. <https://doi.org/10.1007/s10143-023-01987-5>
4. Arshad MW. Prediction and diagnosis of breast cancer

- using machine learning and ensemble classifiers. *Cent Asian J Math Theory Comput Sci.* 2023; 4(1): 49–56. <https://doi.org/10.17605/OSF.IO/9CFN6>
5. Teixeira F, Montenegro JLZ, da Costa CA, da Rosa Righi R. An analysis of machine learning classifiers in breast cancer diagnosis. In: 2019 XLV Latin American computing conference (CLEI). IEEE; 2019: 1–10. <https://doi.org/10.1109/CLEI47609.2019.235094>
 6. Elsadig MA, Altigani A, Elshoush HT. Breast cancer detection using machine learning approaches: a comparative study. *Int J Electr Comput Eng.* 2023; 13(1): 736-45. <https://doi.org/10.11591/ijece.v13i1.pp736-745>
 7. Chen H, Wang N, Du X, Mei K, Zhou Y, Cai G. Classification prediction of breast cancer based on machine learning. *Comput Intell Neurosci.* 2023; 2023. <https://doi.org/10.1155/2023/6530719>
 8. Sakib S, Yasmin N, Tanzeem AK, Shorna F, Hasib K MD, Alam SB. Breast cancer detection and classification: A comparative analysis using machine learning algorithms. In: *Proceedings of Third International Conference on Communication, Computing and Electronics Systems: ICCCES 2021.* Springer; 2022: 703–17. https://doi.org/10.1007/978-981-16-8862-1_46
 9. Abiodun MK, Misra S, Awotunde JB, Adewole S, Joshua A, Oluranti J. Comparing the performance of various supervised machine learning techniques for early detection of breast cancer. In: *International Conference on Hybrid Intelligent Systems.* Springer; 2021: 473–82. https://doi.org/10.1007/978-3-030-96305-7_44
 10. Sridevi T, Murugan A. An intelligent classifier for breast cancer diagnosis based on K-Means clustering and rough set. *Int J Comput Appl.* 2014; 85(11). <https://doi.org/10.5120/14889-3336>
 11. Henderi H, Wahyuningsih T, Rahwanto E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *Int J Informatics Inf Syst.* 2021; 4(1): 13–20. <https://doi.org/10.47738/ijis.v4i1.73>
 12. Kareem AK, Al-ani MM, Nafea AA. Detection of Autism Spectrum Disorder Using A 1-Dimensional Convolutional Neural Network. *Baghdad Sci J.* 2023; 20: 1182–93. <https://doi.org/10.21123/bsj.2023.8564>
 13. Alajanbi M, Malerba D, Liu H. Distributed reduced convolution neural networks. *Mesopotamian J Big Data.* 2021; 2021: 26–9. <https://doi.org/10.58496/MJBD/2021/005>
 14. Dabiri H, Farhangi V, Moradi MJ, Zadehmohamad M, Karakouzian M. Applications of Decision Tree and Random Forest as Tree-Based Machine Learning Techniques for Analyzing the Ultimate Strain of Spliced and Non-Spliced Reinforcement Bars. *Appl Sci.* 2022; 12(10): 4851. <https://doi.org/10.3390/app12104851>
 15. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. *J Appl Sci Technol Trends.* 2021; 2(01): 20–8. <https://doi.org/10.38094/jastt20165>
 16. Chandrhas NS, Choudhary BS, Teja MV, Venkataramayya MS, Prasad NSRK. XG boost algorithm to simultaneous prediction of rock fragmentation and induced ground vibration using unique blast data. *Appl Sci.* 2022; 12(10): 5269. <https://doi.org/10.3390/app12105269>
 17. Roy A, Chakraborty S. Support vector machine in structural reliability analysis: A review. *Reliab Eng Syst Saf.* 2023; 233:109126. <https://doi.org/10.1016/j.res.2023.109126>
 18. Kurani A, Doshi P, Vakharia A, Shah M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Ann Data Sci.* 2023;10(1):183–208. <https://doi.org/10.1007/s40745-021-00344-x>
 19. Lubis AR, Lubis M. Optimization of distance formula in K-Nearest Neighbor method. *Bull Electr Eng Informatics.* 2020; 9(1) :326–38. <https://doi.org/10.11591/eei.v9i1.1464>
 20. ElSahly O, Abdelfatah A. An incident detection model using random forest classifier. *Smart Cities.*2023;6(4): 1786–813. <https://doi.org/10.3390/smartcities6040083>
 21. Mukhlif AA, Al-Khateeb B, Mohammed M. Classification of breast cancer images using new transfer learning techniques. *Iraqi J Comput Sci Math.* 2023; 4(1): 167–80. <https://doi.org/10.52866/ijcsm.2023.01.01.0014>
 22. Nafea AA, Omar N, Al-qfail ZM. Artificial Neural Network and Latent Semantic Analysis for Adverse Drug Reaction Detection. *Baghdad Sci J.* 2024; 21(1), pp.0226-0233. <https://doi.org/10.21123/bsj.2023.7988>
 23. Li JP, Haq AU, Din SU, Khan J, Khan A, Saboor A. Heart disease identification method using machine learning classification in e-healthcare. *IEEE access.* 2020; 8 :107562–82. <https://doi.org/10.1109/ACCESS.2020.3001149>
 24. Trivedi NK, Gautam V, Anand A, Aljahdali HM, Villar SG, Anand D, et al. Early detection and classification of tomato leaf disease using high-performance deep neural network. *Sensors.* 2021;21(23):7987. <https://doi.org/10.3390/s21237987>
 25. Nafea AA, Mishlish M, Muwafaq A, Shaban S, Al-ani MM, Alheeti KMA, et al. Enhancing Student 's Performance Classification Using Ensemble Modeling. *Iraqi J Comput Sci Math.* 2023; 4(4): 204–14. <https://doi.org/10.52866/%20ijcsm.2023.04.04.016>

طريقة هجينة لخوارزمية الـ 1D-CNN والتعلم الآلي للكشف عن سرطان الثدي

احمد عادل نافع¹، منار المهداوي²، خطاب معجل علي الهيتي³، مصطفى صفوك ابراهيم³، محمد ماهر العاني⁴

¹قسم الذكاء الاصطناعي، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة الأنبار، الرمادي، العراق.

²كلية العلوم، جامعة النهدين، بغداد، العراق.

³قسم علوم الحاسوب، جامعة الأنبار، الرمادي، العراق.

⁴مركز تكنولوجيا الذكاء الاصطناعي، كلية علوم وتكنولوجيا المعلومات، جامعة كيبانجسان الماليزية، بانجي، سيلانجور، ماليزيا.

الخلاصة

يعد سرطان الثدي من المخاوف الصحية ذات الأهمية، ومن الضروري اكتشافه مبكراً للحصول على علاج فعال. في الأونة الأخيرة، كان هناك اهتمام متزايد باستخدام الذكاء الاصطناعي (AI) للكشف عن سرطان الثدي، مما أظهر نتائج في تعزيز الدقة وتقليل النتائج الإيجابية الخاطئة. ومع ذلك، هناك بعض القيود فيما يتعلق بدقة الكشف. تقدم هذه الدراسة منهجاً هجيناً يستخدم 1D CNN لاستخراج الميزات ويستخدم خوارزميات التعلم الآلي مثل XGBoost والغابات العشوائية (RF) وأشجار القرار (DT) وآلات ناقلات الدعم (SVM) و أقرب جار (KNN) لتصنيف العينات إما حميدة أو خبيثة تهدف إلى تعزيز الدقة في الكشف. تكشف النتائج التي توصلنا إليها أن خوارزمية XGBoost مع استخراج الميزات (1D CNN) حققت دقة قدرها 98.24% في مجموعة الاختبار. تسلط هذه الدراسة الضوء على جدوى استخدام خوارزميات التعلم الآلي والتعلم العميق في هذه الدراسة تم استخدام مجموعة بيانات سرطان الثدي في ولاية ويسكونسن (WBC)، للكشف عن سرطان الثدي. يبشر نهجنا بالوعد في تسهيل الكشف وتحسين النتائج من خلال توفير أدوات دقيقة وموثوقة لتشخيص سرطان الثدي.

الكلمات المفتاحية: تشخيص سرطان الثدي، التعلم العميق، التعلم الآلي، ويسكونسن، 1D-CNN.