Physical Education, Health and Social Sciences

https://journal-of-social-education.org

E-ISSN: 2958-5996

P-ISSN: 2958-5988

Comparative study of Deep Learning and BERT-Based Models for Breast Cancer Detection

Hina Javed[1], Ali Saeed[1]

[1,2] Department of Data Science, Faculty of Information Technology, University of Central Punjab, Lahore, Pakistan. Email: L1F24MSDS0019@ucp.edu.pk ali.saeed@ucp.edu.pk

Abstract
Breast cancer is a prominent contributor to female deaths due to its prevalence, nearly twice as high as that of other cancers combined. Early detection of this disease is key to survival. This research examines the performance of the Deep Learning (DL) models such as RNN, LSTM and GRU along with BERT based models such as BERT-Base and BERT-Large in breast cancer detection application on structured medical datasets. The outcomes of the experiments showed that the model with 1-Layer RNN achieved the highest performance in terms of accuracy (97.36%), precision (97.56%), recall (95.24%), F_1 score (96.38%) and AUC (99.21%), while the model with 3-Layer RNN achieved the highest performance in terms of AUC (99.40%). The transformer-based models performed at 92.11% accuracy and 98.88% AUC for BERT-Base (Seed 21) while BERT-Large had inconsistent performance due to its higher complexity. The results show that the light-weight DL models are more reliable and efficient in a structured breast cancer data set.

Keywords: Breast Cancer Detection, Deep Learning, BERT-Base, BERT-Large, RNN, LSTM, GRU, Medical Diagnosis, Artificial Intelligence, Transformer Models.

1. Introduction

Breast cancer is a disease that primarily affects females, accounting for 99% of cases [1] with 2.3 million cases and around 670,000 deaths worldwide in 2022 [2]. Breast cancer is the most common type of cancer in women in Pakistan as reported [3] (2015-2019) showing that breast cancer is the most common cancer among women in Pakistan. Such discrepancies underscore the need to come up with universal risk assessment instruments that would allow early identification and management of problems prior to their escalation.

When breast cells start to grow abnormally, which may result in lumps or tumors, breast cancer begins [4]. These tumors may be benign or malignant and malignant tumors can spread to other body parts [5] [6] [7]. Benign breast changes are the presence of cysts, fibrocystic changes, and tissue lumpiness. A high density in the breast [8] can also pose a problem in detecting cancer because malignant masses are hidden by the dense tissue during mammography.

The objective of the research is to perform a thorough comparative study based on the two breast cancer datasets: WBCD[1] (569 instances, 30 features extracted from FNA images) and BRCA[2] (569 samples, 32 quantitative features of cell nuclei). Consistent data will be preprocessed and model performance will be guaranteed with a strict preprocessing stage. Both DL and transformer-based models will be tested on these data sets to select the most statistically sound, clinically valid and computationally efficient model for automated breast cancer diagnosis.

---

1 https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data
2 https://github.com/RajarshiRay25/Breast-Cancer-Dataset---EDA-Classification-Model-training-with-cross-validation/blob/main/brca.csv

The proposed study aims at building reliable and efficient models for automated breast cancer detection. It demonstrates the performance of the two methods (DL, transformer-based methods) by applying them to a benchmark data set to create the most accurate prediction model, overcoming the limitations of previous studies where the algorithm, data set, and evaluation method were not always the same.

2. Related Work

In 2020[3], there were millions of deaths from cancer all around the globe, making it the second highest cause of non-accidental death on a global scale; it is anticipated to become the leading cause of cancer for women over the next 20 years, with an anticipated increase of 70% [9]. Through the introduction of computer-aided detection (CAD), the utilization of machine learning (ML) algorithms capable of analyzing large volumes of structured medical data allows for the discovery of hidden patterns in the data and for the development of personalized treatment plans [10] Therefore, utilizing these methodologies will continue to elevate the level of service to patients and improve the accuracy of the breast cancer diagnoses being rendered.

Breast cancer detection can be done through ML, as multiple studies have shown. Kochari et al. [11] compared different algorithms and found the Random Forest (RF) algorithm to have the highest classification accuracy at 96.50%. Decision Trees (DT) and Linear Regression (LR) was far less accurate in this study's results. Nashte et al. [12] also studied Support Vector Machine (SVM), RF, Multilayer Perceptron (MLP) and DT models; their results showed an accuracy rate of 96.00% using a Clinical Decision Support System. In [13], Mohammed et al. conducted a study on breast cancer detection using the DT, SVM and Naive Bayes methods with 10-Fold Cross-Validation and data resampling. Their results indicated that accuracy improved significantly with SVM being the most accurate on the Wisconsin dataset. Other studies have also shown positive results when using DL or ML approaches for breast cancer detection and diagnosis.

Osareh and Shadgar showed that feature extraction methods combined with KNN, Probabilistic Neural Networks and SVM yield a high accuracy of diagnosis with SVM reaching 96.33% [14]. Likewise, Al-Hadidi et al. also applied image processing using Logistic Regression and BPNN, with a better accuracy of more than 93% [15], and Maglogiannis et al. also found SVM much better in accuracy and sensitivity than Bayesian and ANN models [16]. Rana et al. [17] made use of SVM, Logistic Regression, KNN, and Naive Bayes to classify tumor and predict recurrence. Prediction was done best by SVM and the best overall performed was done by KNN.

The researcher Liu [18] proved the efficiency of the Logistic Regression with the help of the Sklearn library in the classification of breast cancer and reached 96.50% accuracy, noting that ML helps to minimize the subjectivity of humans. Alanazi et al. [19] suggested an automated classification of aggressive regions of whole slide images of ductal carcinoma using a CNN that had a higher accuracy of 87% than 78% using conventional ML. This technique was tested on a large scale of 275,000 photos and achieved good results. Tyagi et al. [20] used RF, LR, and DT to categorize the stages of breast cancer in order to enhance the level of diagnosis and allow early treatment. Bataineh compared [21] SVM, CART, MLP, KNN, and Naive Bayes to WBCD dataset to enhance classification accuracy, precision and recall to make automated breast cancer detection. The worldwide occurrence of breast cancer, as well as the demand for non-invasive and less expensive detection techniques as opposed to traditional methods such as mammography, has been noted by Salod and Singh [22]

Mehta et al. [23] have compared five classifiers where DT was considered as the most effective with a 95.90% accuracy on the Wisconsin dataset in breast cancer classification. With a maximum accuracy of 96.50% in the categorization of malignancies, Ara [24] found that RF and SVM

[3] https://www.who.int/news-room/fact-sheets/detail/cancer

performed better than other models. Using a variety of advanced techniques (PCA, cross-validation, hyper parameter tuning), S.S. et al. [25] found that SVM was the most accurate of eight classifiers, with 99.10%.

3. Dataset Description

Cancer is diagnosed by examining the breast for abnormalities using imaging tests, like mammography, ultrasound or MRI. The data of patients and imaging characteristics have been identified as being stored in a CSV file for the creation of models to facilitate testing and training of the various methods used for detecting breast cancer.

3.1 Dataset 1

Dataset 1 [26] consists of 569 breast cancer samples and 33 features extracted from cell nuclei images. The classes to be targeted are Malignant (M) and Benign (B). Mean, worst-case and standard error values for important features are included in the dataset: Texture, Smoothness, and Compactness etc. Dataset is frequently used for classification and benchmarking of DL models and BERT-based algorithms because of its dependability and structured style.

3.2 Dataset 2

The second data set, Dataset 2, is another breast cancer diagnostic data set where 569 samples and 32 features are given. As in Dataset 1, it contains numerical data for tumor properties. The target variable is labeled "y," and samples are categorized as either benign or malignant. Features included in the data are Radius Mean, Perimeter Mean, and Area Mean etc. Below is a CSV Model of BRCA [27].

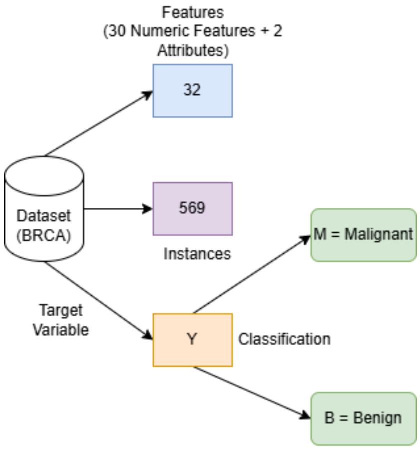

Figure 1: BRCA Dataset Model

4. Model Description

The proposed model uses the DL and BERT models (BERT variants) for classification. The same data splitting scheme, the same preprocessing techniques, and the same evaluation measures have been used for all models, to ensure the models are comparable. The primary goal is to select the most accurate, reliable and efficient model for its use in CAD systems. The diagram of the research method is as follows:

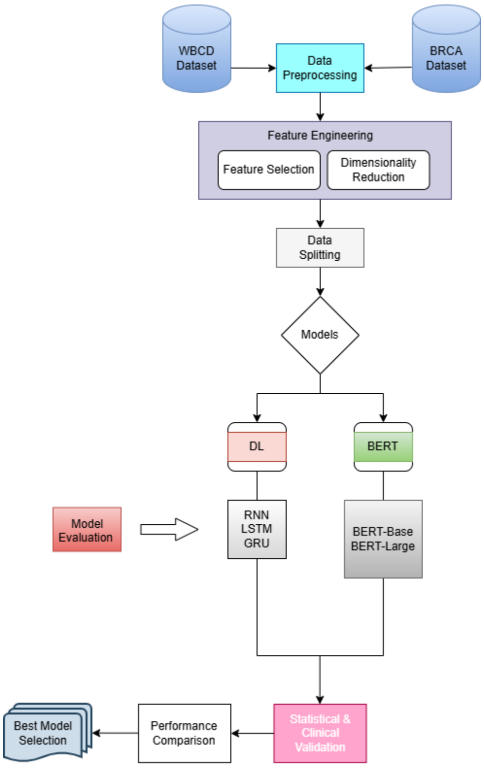

Figure 2: Methodology of Research

5. Evaluation Matrices

DL and BERT based models are assessed under Evaluation Metrics with Accuracy, Precision, Recall, Specificity, AUC, False Negative Rate and F_1-Score. These are used to assess the classification accuracy and model performances. The results are divided into DL and BERT based approaches to see which of these approaches can achieve the best trade-off between precision and recall for breast cancer classification. These performance measures are meant to aid in identifying the most effective and reliable method for detecting breast cancer, as well as helping to evaluate the quality of the models used to create the various performance measures.

6. Results and Analysis

Below results for Dataset 1 are referenced from the published research paper [28] to demonstrate that the DL models, specifically RNN based models, were able to provide accurate and stable results in the detection of breast cancer. DL models showed to be more efficient and reliable in terms of computation in structured datasets, though BERT-based models worked well as well.

DL Algorithms	Accuracy	Precision	Recall	F_1 Score
RNN – 1 Layer	97.37	97.61	95.35	96.47
RNN – 3 Layer	**98.25**	**97.67**	**97.67**	**97.67**
LSTM – 1 Layer	97.37	95.45	97.67	96.55
LSTM – 3 Layer	**98.25**	**97.67**	**97.67**	**97.67**
GRU – 1 Layer	97.37	95.45	97.67	96.55
BERT Models	**Accuracy**	**Precision**	**Recall**	**F_1 Score**
BERT Base	92.98	93.32	92.98	92.86
BERT Large	**92.98**	**93.32**	**92.98**	**92.86**

Table 1: Evaluation of Results for Dataset 1

Experimental results presented for Dataset 2 revealed that the RNN 1-Layer model had the highest performance with an accuracy of 97.36% and AUC score of 99.21%. Simple RNN architectures outperformed DL models on structured breast cancer datasets on average.

DL Algorithms	Accuracy	Precision	Recall	F_1 Score	AUC Score
RNN – 1 Layer	**97.36**	**97.56**	**95.24**	**96.38**	**99.21**
RNN – 3 Layer	95.61	93.02	95.24	94.12	99.40
LSTM – 1 Layer	90.35	91.89	80.95	86.08	96.53
LSTM – 3 Layer	87.72	83.33	83.33	83.33	94.51
GRU – 1 Layer	83.33	82.86	69.05	75.32	93.25

Table 2: Evaluation of Deep Learning Results for Dataset 2

The results of BERT-Base showed good and stable performance with the highest result of 98.88% AUC and 92.11% accuracy from Seed 21. In contrast, BERT-Large performed inconsistently and needed much more data and tuning to produce consistent results.

BERT Models	Accuracy	Precision	Recall	Specificity	F_1 Score	AUC Score
BERT Base (Seed 7)	86.84	78.72	88.10	86.11	83.15	89.09
BERT Base (Seed 21)	**92.11**	**90.24**	**88.10**	**94.44**	**89.16**	**98.88**
BERT Base (Seed 42)	91.23	94.44	80.95	97.22	87.18	98.58
BERT Large (Seed 7)	63.15	0.00	0.00	100	0.00	51.09
BERT Large (Seed 21)	89.47	85.71	85.71	91.67	85.71	90.54

Table 3: Evaluation of BERT Results for Dataset 2

Below graph illustrates the accuracy of the BERT and DL models with respect to breast cancer datasets as they each exhibit different levels of performance. This graph shows the strengths of each of these models with respect to their classification algorithm's performance and provides insight into the impact of algorithm performance.

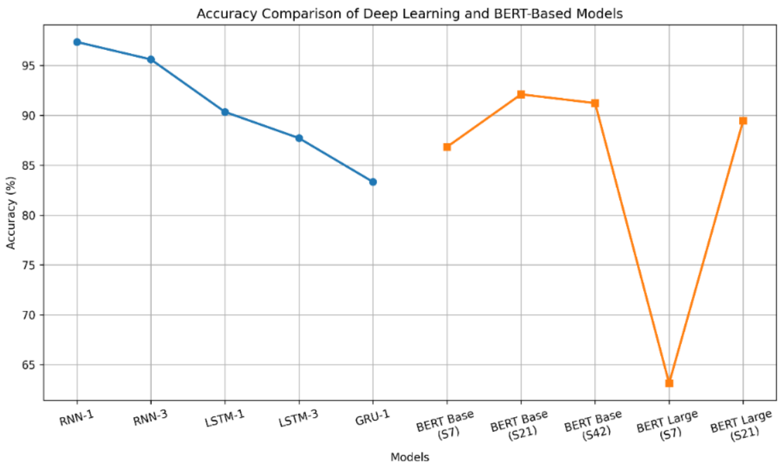

Figure 3: Accuracy Comparison of DL and BERT Models

The outcomes indicate that DL models, particularly RNN 1-Layer model demonstrates best performance for breast cancer classification. Another key finding from the research was that not all models are more complex and better, simpler models performed better on smaller datasets than BERT-Large.

7. Conclusion

In this study we use the breast cancer dataset based on a structured data records to test and evaluate the diagnostic performance of the DL models and the BERT based models on their ability to diagnose breast cancer. The experimental results showed that the RNN 1-Layer model had the best performance with an accuracy of 97.36% and an AUC of 99.21%, showing excellent stability and computation efficiency. BERT-Base was competitive as well and BERT-Large was not as reliable in performance due to the number of parameters and data needed to perform well. Results confirm that the lightweight DL models are indeed more effective and accurate in diagnosing breast cancer using AI.

8. References

[1] W. H. Organisartion, Breast Cancer Fact Sheet, 2024.

[2] B. C. R. Foundation, "Breast Cancer Statistics and Resources," 2025. [Online]. Available: https://www.bcrf.org/breast-cancer-statistics-and-resources/#:~:text=In%202022%2C%20about%202.3%20million,developed%20and%20less%20developed%20world.

[3] N. C. R. o. Pakistan, "First Comprehensive Report of Cancer Statistics (2015-2019)," 2024.

[4] M. Clinic, "Breast Cancer - Symptoms and Causes," 2025. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470.

[5] C. Clinic, "Tumor: Overview, Types, and Treatment," 2025. [Online]. Available: https://my.clevelandclinic.org/health/diseases/21881-tumor.

[6] M. Clinic, "Breast cancer: Symptoms and causes," 2024.

[7] S. B. Medicine, "Breast Lumps: What You Need to Know," 2025. [Online]. Available: https://cancer.stonybrookmedicine.edu/breast-cancer-team/patients/bse/breastlumps.

[8] M. A. A. A. A. B. A. F. A. H. C. G. Khalid A, " Breast Cancer Detection and Prevention Using Machine Learning," Diagnostics (Basel), vol. 13, no. 19, p. 3113, 2023.

[9] J. F. I. S. R. L. S. L. A. T. a. A. J. F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209-249, 2021.

[10] S. O. E. M. E. H. H. A. Nisreen I.R. Yassin, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260717306405.

[11] V. Kochari, "Detection of Breast Cancer Using Machine Learning Algorithms," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, 2021.

[12] J. N. A. D. S. S. a. M. B. C. Nashte, "Differentiation of Malignant and Benign Breast Lesions Using Machine Learning Algorithms," Differentiation, vol. 4, no. 5, 2017.

[13] S. D. S. A. N. a. G. S. S. A. Mohammed, "Analysis of breast cancer detection using different machine learning techniques," in Data Mining and Big Data: 5th International Conference, DMBD 2020, Belgrade, Serbia, July 14–20, 2020, Proceedings 5, Springer Singapore, 2020, pp. 108-117.

[14] A. O. a. B. Shadgar, "Machine learning techniques to diagnose breast cancer," in 2010 5th International Symposium on Health Informatics and Bioinformatics, 2010.

[15] A. A. a. M. Alhanahnah, "Breast cancer detection using k-nearest neighbor machine learning algorithm," in 2016 9th International Conference on Developments in eSystems Engineering (DeSE), 2016.

[16] E. Z. a. I. A. I. Maglogiannis, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM-based classifiers," Applied Intelligence, vol. 30, pp. 24-36, 2009.

[17] P. C. A. D. a. N. K. M. Rana, "Breast cancer diagnosis and recurrence prediction using machine learning techniques," International Journal of Research in Engineering and Technology, vol. 4, no. 4, pp. 372-376, 2015.

[18] L. Liu, "Research on logistic regression algorithm of breast cancer diagnose data by machine learning," in 2018 International Conference on Robots & Intelligent System (ICRIS), 2018.

[19] M. M. K. M. N. I. S. M. A. Y. A. N. A. a. M. H. S. S. A. Alanazi, "Boosting breast cancer detection using convolutional neural network," Journal of Healthcare Engineering, vol. 2021, no. 1, p. Article ID 5528622, 2021.

[20] T. M. P. S. a. R. K. V. Tyagi, "Using Machine Learning algorithms for breast cancer classification and risk prediction," 2020.

[21] A. A. Bataineh, "A comparative analysis of nonlinear machine learning algorithms for breast cancer detection," International Journal of Machine Learning and Computing, vol. 9, no. 3, pp. 248-254, 2019.

[22] Z. S. a. Y. Singh, "Comparison of the performance of machine learning algorithms in breast cancer screening and detection: A protocol," Journal of Public Health Research, vol. 8, no. 3, jphr-2019.

[23] A. A. a. T. C. S. Sharma, "Breast cancer detection using machine learning algorithms," in 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2018.

[24] S. a. D. A. a. D. A. Ara, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms," in 2021 International Conference on Artificial Intelligence (ICAI), 2021.

[25] C. S. M. S. a. S. V. S. Suthagar, "Analysis of breast cancer classification using various algorithms," in 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 2022 .

[26] "Kaggle Dataset 1".

[27] "Kaggle Dataset 2".

[28] H. Javed, A. Saeed, H. M. ul Haque, Z. Fatima, J. L. O. Rodríguez and C. G. Sánchez-Mejorada, "Automated Detection of Breast Cancer using Machine Learning and Deep Learning Models from Tabular Data," Letters in Drug Design & Discovery, vol. 22, no. 12, p. 100295, 2026.