Breast cancer relapse disease prediction improvements with ensemble learning approaches

Ghanashyam Sahoo[1], Ajit Kumar Nayak[2], Pradyumna Kumar Tripathy[3], Abhilash Pati[1], Amrutanshu Panigrahi[1], Adyasha Rath[4], Bhimasen Moharana[5]

[1]Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be university), Bhubaneswar, India
[2]Department of Computer Science and IT, Siksha 'O' Anusandhan (Deemed to be university), Bhubaneswar, India
[3]Department of Computer Science and Engineering, Silicon Institute of Technology, Bhubaneswar, India
[4]Department of Computer Science and Engineering, C. V. Raman Global University, Bhubaneswar, India
[5]School of Computer Science and Engineering, Lovely Professional University, Phagwara, India

Article Info

Article history:

Received Feb 9, 2024
Revised Feb 27, 2024
Accepted Mar 18, 2024

Keywords:

Breast cancer relapse
Deep learning
Disease prediction
Ensemble learning
Machine learning

ABSTRACT

Diagnosis and prognosis are especially difficult areas of medical research related to cancer due to the high incidence of breast cancer, which has surpassed all other cancers in terms of female mortality. Another factor that has a substantial influence on the quality of life of cancer patients is the fear that they may experience a relapse of their disease. The objective of the study is to give medical practitioners a more effective strategy for using ensemble learning techniques to forecast when breast cancer may recur. This research aimed to investigate the usage of deep neural networks (DNNs) and artificial neural networks (ANNs) in addition to machine learning (ML) based approaches, including bagging, averaging, and voting, to enhance the efficacy of breast cancer relapse diagnosis on two breast cancer relapse datasets. Results from the empirical study demonstrate that the proposed ensemble learning-enabled approach improves accuracies by 96.31% and 95.81%, precisions by 96.70% and 96.15%, sensitivities by 98.88% and 98.68%, specificities by 84.62% in both, F1-scores by 97.78% and 97.40%, and area under the curve (AUCs) of 0.987 and 0.978, with University Medical Centre, Institute of Oncology (UMCIO) and Wisconsin prognostic breast cancer (WPBC) datasets respectively. Consequently, these improved disease outcomes may encourage physicians to use this model to make better treatment choices.

Corresponding Author:

Abhilash Pati
Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be university)
Bhubaneswar, Odisha, India
Email: er.abhilash.pati@gmail.com

1. INTRODUCTION

The second-leading cause of cancer-related deaths worldwide is breast cancer, which primarily affects women and develops when cells multiply uncontrollably, leading to tumors in the breast or other organs [1]. Due to its ability to metastasize, invasive breast cancer is more lethal. This raises the possibility of cancers of the liver and lungs. The breasts are the only sites of noninvasive breast cancer which does not metastasize. This disease is not malignant per se, but it has the potential to metastasize (spread) beyond the breast and become invasive [2], [3]. Recurrence of breast cancer has occurred in certain people even after years of therapy. About 40% of the time, breast cancer comes back. Patients should not assume that recurrence will not occur in the first two or three years after treatment. One of three things can bring breast cancer back. One kind of breast cancer is known as local recurrence (LR) because it happens just where the disease was first found. The axilla, lymph nodes, or collarbone are the sites of the second type of recurrence, known as regional recurrence. Third, when cancer spreads to healthy regions, it's called distant metastasis. Because of their similarities, locoregional recurrence (LLR) is a diagnostic term that may be used to characterize both local and regional recurrence [4]–[6].

Almuhaidib *et al.* [7] demonstrated a recurrence prediction accuracy of 76.26% using ensemble learning methods on Wisconsin prognostic breast cancer (WPBC) datasets. Rana *et al.* [8] proposed an machine learning (ML) based model for breast cancer diagnosis model. The model was trained using ML techniques on the wisconsin prognosis and diagnostic breast cancer (WPBC, WDBC) datasets. A breast cancer recurrence prediction algorithm was built using the WPBC dataset by Chakradeo *et al.* [9]. Using ML methods, the system could get a recall of 91%, a precision of 93.36%, and an accuracy of 97.93%. Sakri *et al.* [10] developed a method for predicting breast cancer recurrence using data mining techniques applied to the WPBC datasets. The method achieved an accuracy of 81.3%, sensitivity of 93.4%, and specificity of 63.25. An ML approach for predicting breast cancer recurrence was proposed by Goyal *et al.* [11] on UNC, Institute of Oncology, Ljubljana, Yugoslavia datasets. Their results showed a perfect score of 85.18 percent accuracy, a perfect score of 100% sensitivity, and a perfect score of 100% specificity. Dawangliani *et al.* [12] proposed a prediction system using ML methods applied to breast cancer datasets. These factors influenced the following outcomes: ROC area: 79.6%, accuracy: 81.9%, recall: 82.8%, time to peak: 0.828, and false positive rate: 0.534. Using ensemble learning on the NFSC, Gu *et al.* [13] demonstrated an explainable technique for predicting breast cancer recurrence with an F1-score of 89.39%, an accuracy of 91.62%, and a recall of 90.28%. Rabinovici-Cohen *et al.* [14] achieved 90.0% sensitivity in their predictive model for the women who received Neoadjuvant Chemotherapy. The predictive model employs the ML and DL algorithms using datasets from the Institute Curie. A strategy for predicting the recurrence of breast cancer utilizing ensemble and cost-sensitive learning approaches was developed by Yang *et al* [15]. The results were as follows: with an F-measure of 65.7%, ROC area of 90.7%, specificity of 98.3%, precision of 64.0%, and accuracy of 97.3%. Patients with early-stage non-small cell lung cancer could reach a recurrence prediction accuracy of 76% using ML algorithms on graph datasets, as demonstrated by Janik *et al.* [16]. The breast cancer tumor recurrence date may be predicted with 78.7 percent accuracy by Gupta [17] using ML algorithms on the WPBC and WDBC datasets. By combining both unstructured and structured data from electronic health record (EHRs) (COMB, UNS, and STR databases), González-Castro *et al.* [18] were able to achieve a 90.0% accuracy, 89.7% F1-score, 90.7% recall, and 80.7% AUROC for predicting breast cancer recurrence.

The diversity of breast cancer makes it hard to foretell how the illness will proceed or what the patient's prognosis will be, but new biomarkers and insights are increasing the need for cancer treatments. By sorting through mountains of data, powerful data analysis tools like DL and ML algorithms can aid doctors in developing better patient diagnostic systems. Much remains unknown, even though several studies have investigated ML/ deep learning (DL) algorithms and breast cancer recurrence risk factors. Oncology is shifting its focus from traditional statistical methods to models based on ensemble learning, ML, and DL due to the complexity of cancer data. The fields of oncology and classification both make use of ML to detect and classify tumors. Relapse detection in breast cancer using artificial neural networks (ANNs) and deep neural networks (DNNs) was the focus of this research. Ensemble ANN and DNN approaches are also taken into account, along with classic ANN and DNN methods. These methods encompass ML-based ensemble techniques such as bagging, averaging, and voting. To achieve this success, we use state-of-the-art metrics to evaluate the proposed ensemble ANN and DNN algorithms on two breast cancer relapse datasets. The primary contributions to the reported work can be summarized as follows:

- To assess the relapse classification models for breast cancer using ANN and DNN.
- To develop ANN, DNN-based ensemble models.
- To enhance patient diagnosis outcomes by assisting doctors in making better treatment choices.

2. METHOD

2.1. Dataset employed

This study makes use of two datasets: D1, which is the University Medical Centre, Institute of Oncology (UMCIO) dataset acquired in July 1988 by the Institute of Oncology at the University Medical Centre [19], and D2, which is the WPBC cancer relapse dataset obtained from the UCI-ML repository [20]. The data had duplicate items and missing values. The correct preparation technique handles the parity problem in the dataset. To create a new range, begin by extracting data from an existing one. A nominal dataset is created by normalization. The data is transformed into numbers so it may be processed further. As shown in Table 1, the dataset dimension was pre- and post-processed, with "NR" denoting non-relapse and "R" relapse.

2.2. Methodologies employed

The learning process of ANNs is modeled after that of biological neurons. A DNN is a multi-layered representation of a complex data correlation. By automating the extraction of hierarchical features and complicated patterns from input data, DNNs have radically changed ML [21], [22]. Both ML and DL could benefit from ensemble learning [23]. An ensemble learning (EL) method for classification and regression is the bootstrap aggregator, sometimes called a bagging classifier. The weighted averaging technique averages the initial predictions made by many classifiers to rank the models according to their performance. Voting classifiers take an average of the results. Using them for regression or classification, minority: to create a single prediction, voting combines the results of many refined models trained on the same data. The combined prediction outcome is based on the projected probabilities of the basis learners. Various base learners or classifiers (Bi) have varying beginning prediction probabilities, denoted as λi. As shown in (1) shows that the final ensemble model prediction may be represented by ρ.

Table 1. Number of instances after pre-processing in the dataset

Dataset	Features	Instances					
		Initial dimension			Post pre-processing dimension		
		NR-events	R-events	Overall	NR-events	R-events	Overall
D1 (UMCIO)	10	201	85	286	191	81	272
D2 (WPBC)	35	151	47	198	145	42	187

$$\rho = Max_i \sum_{k=1}^{B} \omega_k \lambda_k \quad (1)$$

The majority voting ensemble considers the forecasts of many models that have been trained using the same dataset. The anticipated class label using the hard voting procedure is denoted as ρ. As shown in (2) may be used to get this expected value. Here, d is the class for the dataset's attributes.

$$\rho = mode\{d(b_1), d(b_2), \dots \dots, d(b_n)\} \quad (2)$$

2.3. Proposed work

The DNNs and ANNs are trained using the breast cancer recurrence dataset in this work. After the training phase, the ensemble ML models are applied to the initial prediction to develop an ensembled predictive model. In order to get the prediction result, the pre-processed dataset is first used using the ANN and DNN algorithms. The bagging classifiers are applied to the obtained outcomes from ANNs and DNNs individually to get the initial predictions in the next. We then apply several EL methods to the pre-processed dataset, including weighted averaging and voting strategies. Algorithm 1 represents the proposed model's pseudocode, and Figure 1 depicts the manuscript's entire process.

Algorithm 1.

```
Input: Raw Breast Cancer Relapse Datasets
Output: Breast Cancer Relapse Prediction
For i 1←k (the number of datasets)
–    Pre-processing the Raw Dataset of Breast Cancer Relapse
1)   Removing rows with the missing values from D, resulting in a clean dataset Dclean with a
     new sample size N′
2)   Apply Standard Scaler to Dclean
3)   For i=1←N′
a.   Mean(μ) = (1/N′) Σ_{i=1}^{N′} f_i
b.   Standard Deviation (σ) = sqrt((1/N′) Σ_{i=1}^{N′} (f_i − μ)^2)
c.   f_i = (f−μ)/σ
d.   Dclean ← f_i
4)   EndFor
–    Split the dataset with a test size of 0.25.
–    Setting ANNF() and DNNF() to the preprocessed dataset with input layer (IL), hidden layer
     (HL)=3, ouput layer (OL)
1)   Set optimizer=ADAM
2)   Initialize the actFun(IL) = "RELU"
3)   Initialize the actFun(HL) = "RELU"
4)   Initializing actFun(OL) = "Sigmoid"
5)   Obtaining Primary_Output (O)
–    Setting bagClassifierFun() to O
1)   Obtaining Initial predictions (IP)
```

```
–    Setting ensApproachFun() to the Initial predictions (IP)
1)   Applying Fun_weightedAveraging()
2)   Applying Fun_minorityVoting()
3)   Applying Fun_majorityVoting()
4)   Obtaining Final predictions (FP)
–    Comparison of outcomes obtained based on ANNF() and DNNF() among Fun_weightedAveraging(),
     Fun_minorityVoting(), and Fun_majorityVoting().
EndFor
```

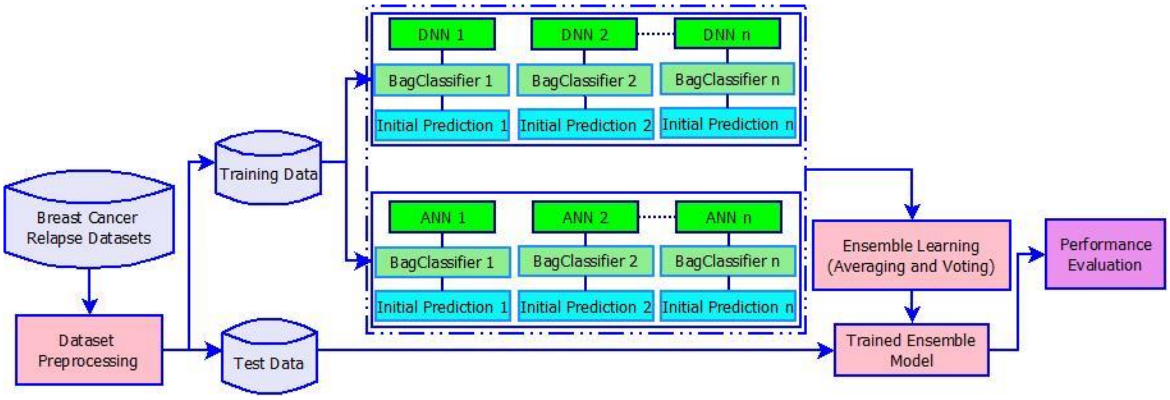

Figure 1. Block diagram of the proposed ensemble-based model

3. RESULTS AND DISCUSSION

A system with Windows 11, an Intel Core i7 CPU with 3.8 GHz clock speed, 16 GB of RAM, and 500 GB SSD is used to evaluate the proposed ensemble models. Following a methodical experimental procedure, these performance measures hope to build a class confusion matrix closer to reality than expectations [24], [25]. Performance evaluations in this research may be done using several performance indicators, such as Accuracy (AC), precision (PR), balanced accuracy (BA), mathew's correlation coefficient (MCC), false positive rate (FPR), false negative rate (FNR), F1-score (FS), specificity (SP), and sensitivity (SN) [26], [27].

In this research, ANN and DNN were first evaluated separately. Three fundamental ensemble methods were used to hone the prediction outcomes, including weighted averaging and minority and majority voting. At first, the experiments are carried out on the D1 dataset, and the results are recorded in Table 2. Figure 2 demonstrates the obtained accuracies in % experimenting on the D1 dataset. It can be observed that, as in Figure 1, DNN, along with bagging classifier and weighted averaging, outperforms all seven other approaches with 96.31% accuracy. Besides, this DNN, along with bagging classifier and weighted averaging, also outperforms all others with 96.70% precision, 98.88% sensitivity (same as with DNN only), 84.62% specificity (same as with DNN with bagging classifier and majority voting), and 97.78% F1-score, as shown in Table 2, which have influenced us to consider this ensemble approach to be the proposed ensemble approach in case of D1 dataset.

Table 2. Obtained outcomes of various recommended approaches considering the D1 dataset

Ensemble models	Outcomes based on performance parameters in (%)									
	AC	MR	PR	SN	FS	SP	FNR	FPR	MCC	BA
ANN	89.86	10.14	90.29	96.93	93.49	68.52	3.07	31.48	71.62	82.73
ANN+BagC+WtAv	91.71	8.29	91.76	98.24	94.89	68.09	1.76	31.91	74.28	83.17
ANN+BagC+MinV	90.78	9.22	91.11	97.62	94.25	67.35	2.38	32.65	72.23	82.49
ANN+BagC+MajV	92.63	7.37	92.70	98.21	95.38	73.47	1.79	26.53	78.06	85.84
DNN	93.55	6.45	93.62	98.88	96.18	69.23	1.12	30.77	76.85	84.06
DNN+BagC+WtAv	96.31	3.69	96.70	98.88	97.78	84.62	1.12	15.38	87.16	91.75
DNN+BagC+MinV	94.47	5.53	95.60	97.75	96.66	79.49	2.25	20.51	80.63	88.62
DNN+BagC+MajV	95.85	4.15	96.69	98.31	97.49	84.62	1.69	15.38	85.60	91.47

Table 3 shows the outcomes of the tests conducted on the D2 dataset. In Figure 3, we can see the results of the accuracy experiments conducted on the D2 dataset. With an accuracy of 95.81%, DNN, bagging classifier, and weighted averaging beat all seven other methods, as shown in Figure 3. In addition, as shown in

Table 3, this DNN, bagging classifier, and weighted averaging outperforms all others. Its precision is 96.15%, sensitivity is 98.68% (the same as with DNN only), specificity is 84.62% (the same as with DNN with bagging classifier and majority voting), and F1-score is 97.40%. Therefore, we recommend this ensemble approach for the D1 dataset.

The next step is to calculate the area under the receiver operating characteristic (ROC) area under the curve (AUC), which is dependent on the TPR and FPR. Figures 4 and 5 display the AUC and ROC of the suggested ensemble model on the two uncommon cancer relapse datasets, D1 and D2, respectively. With an AUC of 0.987 and 0.978, respectively, when applied to the D1 and D2 breast cancer relapse datasets, this suggested ensemble method was extremely relevant. Table 4 shows a critical analysis of the proposed ensemble model with some existing literature to back up the novelty and efficacy of the proposed model. Compared to existing research, the proposed approach clearly outperforms it in all maximum case scenarios across all evaluative factors.

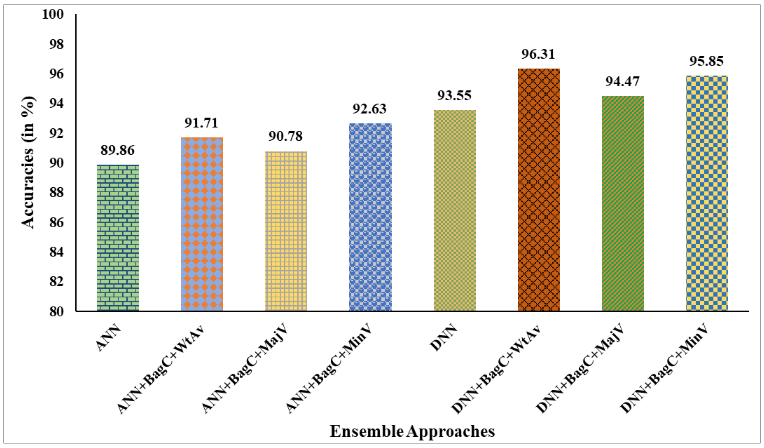

Figure 2. Obtained accuracy in percentage considering the D1 dataset

Table 3. Obtained outcomes of various recommended approaches considering the D2 dataset

Ensemble models	Outcomes based on performance parameters in (%)									
	AC	MR	PR	SN	FS	SP	FNR	FPR	MCC	BA
ANN	88.48	11.52	88.59	96.35	92.31	68.52	3.65	31.48	70.53	82.44
ANN+BagC+WtAv	90.58	9.42	90.38	97.92	94	68.09	2.08	31.91	73.48	83.01
ANN+BagC+MinV	89.53	10.47	89.61	97.18	93.24	67.35	2.82	32.65	71.31	82.27
ANN+BagC+MajV	91.62	8.38	91.45	97.89	94.56	73.47	2.11	26.53	77.31	85.68
DNN	92.67	7.33	92.59	98.68	95.54	69.23	1.32	30.77	76.29	83.96
DNN+BagC+WtAv	95.81	4.19	96.15	98.68	97.4	84.62	1.32	15.38	86.8	91.65
DNN+BagC+MinV	93.72	6.28	94.87	97.37	96.1	79.49	2.63	20.51	80.08	88.43
DNN+BagC+MajV	95.29	4.71	96.13	98.03	97.07	84.62	1.97	15.38	85.18	91.33

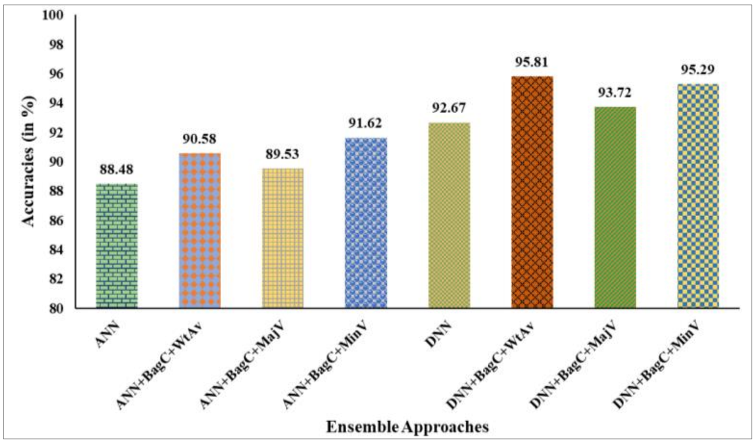

Figure 3. Obtained accuracy in percentage considering the D2 dataset

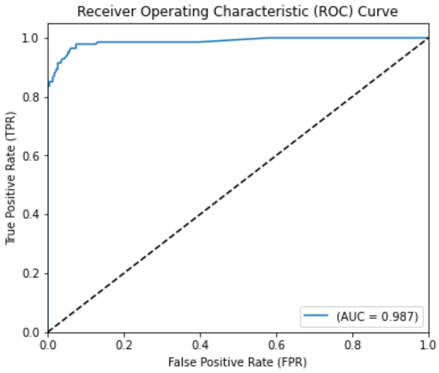

Figure 4. AUC obtained from ROC plotted for recommended ensemble models

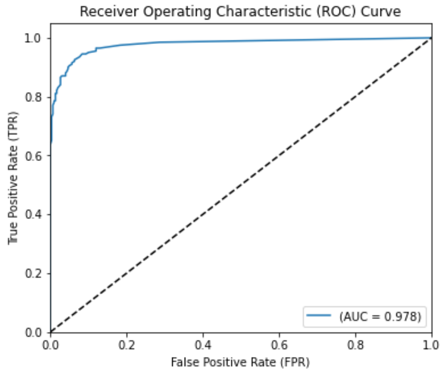

Figure 5. AUC obtained from ROC plotted for recommended ensemble models

Table 4. Obtained outcomes of various recommended approaches

Ref	Parameters based comparison					
	AC in (%)	PR in (%)	SN in (%)	SP in (%)	FS in (%)	AUC
[7]	76.26	-	-	-	-	
[8]	95.68, and 72	-	-	-	-	
[9]	97.93	93.36	91.00	-	-	
[10]	81.3	-	93.4	63.25	-	
[11]	85.18	100.0	100.0	100.0		
[12]	82.80	81.9	82.8	-	82.3	0.796
[13]	91.62	-	90.28	-	89.39	
[14]	-	-	90.0	57.0	-	0.75
[15]	97.3	64.0	97.7	98.3	65.7	0.907
[16]	76	-	-	-	-	
[17]	78.7	-	-	-	-	
[18]	-	90.0	90.7	-	89.7	0.807
Proposed approach [D1]	96.31	96.70	98.88	84.62	97.78	0.987
[D2]	95.81	96.15	98.68	84.62	97.40	0.978

4. CONCLUSION

The recommended ensemble approach is evaluated on the two relapse datasets of breast cancer. Initially, the datasets undergo a data preprocessing phase to deal with data imbalance. Then, ANN and DNN are implemented to get the initial prediction. Various ML-based ensembled methods, including bagging, averaging, and voting (majority and minority), are applied. The empirical analysis shows that the DNN, along with the bagging classifier and weighted averaging, improves accuracies by 96.31% and 95.81%, precisions by 96.70% and 96.15%, sensitivities by 98.88% and 98.68%, specificities by 84.62% in both, F1-scores by 97.78% and 97.40%, and AUCs of 0.987 and 0.978, with UMCIO and WPBC datasets respectively. The analysis clearly shows that this proposed ensemble approach outperforms other proposed models. Additionally, other recurrent datasets may be employed with different ML and DL algorithms, expanding the breadth of this study. A breast cancer relapse imaging dataset is also in the works for the future.

REFERENCES

[1] A. Panigrahi *et al.*, "En-MinWhale: an ensemble approach based on MRMR and whale optimization for cancer diagnosis," *IEEE Access*, vol. 11, pp. 113526–113542, 2023, doi: 10.1109/ACCESS.2023.3318261.

[2] G. Sahoo, A. K. Nayak, P. K. Tripathy, and J. Tripathy, "A novel machine learning based hybrid approach for breast cancer relapse prediction," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 3, pp. 1655–1663, Dec. 2023, doi: 10.11591/ijeecs.v32.i3.pp1655-1663.

[3] A. Pati *et al.*, "Breast cancer diagnosis based on IoT and deep transfer learning enabled by fog computing," *Diagnostics*, vol. 13, no. 13, Jun. 2023, doi: 10.3390/diagnostics13132191.

[4] A. Hamza, "An enhanced breast cancer diagnosis scheme based on two-Step-SVM technique," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, 2017, doi: 10.14569/IJACSA.2017.080423.

[5] A. Pati *et al.*, "FOHC: Firefly optimizer enabled hybrid approach for cancer classification," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 7s, pp. 118–125, Jul. 2023, doi: 10.17762/ijritcc.v11i7s.6983.

[6] Y. Lim *et al.*, "Background parenchymal enhancement on breast MRI: association with recurrence-free survival in patients with newly diagnosed invasive breast cancer," *Breast Cancer Research and Treatment*, vol. 163, no. 3, pp. 573–586, Jun. 2017, doi: 10.1007/s10549-017-4217-5.

[7] D. A. Almuhaidib *et al.*, "Ensemble learning method for the prediction of breast cancer recurrence," in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, Apr. 2018, pp. 1–6. doi: 10.1109/CAIS.2018.8442017.

[8] M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, "Breast cancer diagnosis and recurrence prediction using machine learning techniques," *International Journal of Research in Engineering and Technology*, vol. 04, no. 04, pp. 372–376, Apr. 2015, doi: 10.15623/ijret.2015.0404066.
[9] K. Chakradeo, S. Vyawahare, and P. Pawar, "Breast cancer recurrence prediction using machine learning," in *2019 IEEE Conference on Information and Communication Technology*, Dec. 2019, pp. 1–7. doi: 10.1109/CICT48419.2019.9066248.
[10] S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction," *IEEE Access*, vol. 6, pp. 29637–29647, 2018, doi: 10.1109/ACCESS.2018.2843443.
[11] K. Goyal, P. Aggarwal, and M. Kumar, "Prediction of breast cancer recurrence: A machine learning approach," in *Computational Intelligence in Data Mining*, 2020, pp. 101–113. doi: 10.1007/978-981-13-8676-3_10.
[12] M. S. Dawngliani, N. Chandrasekaran, S. Lalmuanawma, and H. Thangkhanhau, "Prediction of breast cancer recurrence using ensemble machine learning classifiers," 2020, pp. 232–244. doi: 10.1007/978-3-030-46828-6_20.
[13] D. Gu, K. Su, and H. Zhao, "A case-based ensemble learning system for explainable breast cancer recurrence prediction," *Artificial Intelligence in Medicine*, vol. 107, Jul. 2020, doi: 10.1016/j.artmed.2020.101858.
[14] S. Rabinovici-Cohen *et al.*, "Multimodal prediction of five-year breast cancer recurrence in women who receive neoadjuvant chemotherapy," *Cancers*, vol. 14, no. 16, Aug. 2022, doi: 10.3390/cancers14163848.
[15] P.-T. Yang, W.-S. Wu, C.-C. Wu, Y.-N. Shih, C.-H. Hsieh, and J.-L. Hsu, "Breast cancer recurrence prediction with ensemble methods and cost-sensitive learning," *Open Medicine*, vol. 16, no. 1, pp. 754–768, May 2021, doi: 10.1515/med-2021-0282.
[16] A. Janik *et al.*, "Machine learning-assisted recurrence prediction for early-stage non-small-cell lung cancer patients," *Prepr. arXiv.2211.09856*, Nov. 2022.
[17] S. R. Gupta, "Prediction time of breast cancer tumor recurrence using machine learning," *Cancer Treatment and Research Communications*, vol. 32, 2022, doi: 10.1016/j.ctarc.2022.100602.
[18] L. González-Castro *et al.*, "Machine learning algorithms to predict breast cancer recurrence using structured and unstructured sources from electronic health records," *Cancers*, vol. 15, no. 10, p. 741, May 2023, doi: 10.3390/cancers15102741.
[19] M. Zwitter and M. Soklic, "Breast cancer data. Institute of oncology," University Medical Centre Ljubljana, Yugoslavia, 1988.
[20] "UCI machine learning repository," *archive.ics.uci.edu*. https://archive.ics.uci.edu/dataset/14/breast+cancer (accessed Mar. 22, 2023).
[21] A. A. A. Al-Ameer, G. A. Hussien, and H. A. Al Ameri, "Lung cancer detection using image processing and deep learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 2, pp. 987–993, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp987-993.
[22] R. Pandian, D. N. S. R. Kumar, and R. R. Kumar, "Development of algorithm for identification of maligant growth in cancer using artificial neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 5709–5713, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5709-5713.
[23] S. S. Reddy, N. Pilli, P. Voosala, and S. R. Chigurupati, "A comparative study to predict breast cancer using machine learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 27, no. 1, pp. 171–180, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp171-180.
[24] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4217–4225, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4217-4225.
[25] A. Pati, A. Panigrahi, D. S. K. Nayak, G. Sahoo, and D. Singh, "Predicting pediatric appendicitis using ensemble learning techniques," *Procedia Computer Science*, vol. 218, pp. 1166–1175, 2023, doi: 10.1016/j.procs.2023.01.095.
[26] S. K. Rout, B. Sahu, A. Panigrahi, B. Nayak, and A. Pati, "Early detection of sepsis using LSTM neural network with electronic health record," in *Ambient Intelligence in Health Care*, 2023, pp. 201–207. doi: 10.1007/978-981-19-6068-0_19.
[27] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 815–821, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14785.

BIOGRAPHIES OF AUTHORS

Mr. Ghanashyam Sahoo is currently a Research Scholar in the Department of Computer Science and Engineering, ITER, Siksha 'O' Anusandhan (Deemed to be University), Odisha, India. He completed his M. Tech. in CSE from KIIT University Odisha, India, in 2008. His research interests include Machine Learning, statistical computing, Deep Learning, Wireless sensor networks, and the Internet of Things. He has authored a book, "Data Structure Using C," and has three publications. He can be contacted at email: ghanarvind@gmail.com.

Dr. Ajit Kumar Nayak is the professor and HoD of the Department of Computer Science and Information Technology, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha. He graduated in Electrical Engineering from the Institution of Engineers, India, in 1994, and he earned his M. Tech and Ph.D. in Computer Science from Utkal University in 2001 and 2010, respectively. His research interests include computer networking, ad hoc and sensor networks, ML, natural language computing, speech and image processing. He has published about 70 research papers in various journals and conferences. He can be contacted at email: ajitnayak@soa.ac.in.

Dr. Pradyumna Kumar Tripathy completed his M.Tech. and Ph.D. in Computer Science from Utkal University, India, in 2007 and 2015 respectively. He is an Associate Professor in the Computer Science and Engineering Department at Silicon Institute of Technology, Bhubaneswar, India. His research interests include reliability analysis of interconnection networks, parallel distributed systems, topological optimization of interconnection networks, internet of things (IoT), and ML. He can be contacted at email: pradyumnatripathy@gmail.com.

Dr. Abhilash Pati is currently working as an Assistant Professor in the Department of CSE, FET-ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India. He completed his B.Tech. and M.Tech. in CSE with the BPUT, Odisha, India, in 2009 and 2012, respectively, and his Ph.D. in CSE with Siksha 'O' Anusandhan University, Bhubaneswar, India in 2023. His research interests include the IoT, Fog Computing, ML, and DL, and he has more than 35 publications in SCIE, ESCI, Scopus indexed journals and/or conferences to his credit. He can be contacted at email: er.abhilash.pati@gmail.com.

Dr. Amrutanshu Panigrahi is currently working as an Assistant Professor at the Dept. of CSE, Siksha O Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. He obtained M.Tech in Information Technology from the College of Engineering and Technology, Govt. of Odisha, and a B.Tech from BPUT Odisha in 2014 and 2011, respectively. He completed his Ph.D. in CSE with Siksha 'O' Anusandhan University in 2023. His research interests include blockchain, ML, and DL. He has over 40 publications in SCIE, Scopus journals, and conferences. He can be contacted at email: amrutansup89@gmail.com.

Dr. Adyasha Rath is currently working as an Assistant Professor in the Department of CSE, C. V. Raman Global University, Odisha, India. She was awarded her Ph.D. degree in CSE from Siksha O Anusandhan (deemed to be) University, Bhubaneswar, Odisha, in April 2023. She has several publications in various Journals and International Conferences. Her research interests are in the areas of data analytics, soft computing, evolutionary computing, ML and DL techniques, and their applications to healthcare, finance, and network security. She can be contacted at email: adyasha.rath@cgu-odisha.ac.in.

Dr. Bhimasen Moharana is currently working as an Assistant Professor at the School of CSE, Lovely Professional University, Phagwara, Punjab, India. He has completed his B.Tech. M.Tech. in CSE from BPUT, Odisha, and NITTTR Chandigarh, in 2004 and 2010, respectively. He was awarded a Ph.D. in CSE from SSSUTMS, Bhopal, in 2022. His research interests include deep learning, IoT, and machine learning. He can be contacted at email: bhimasen.moharana@gmail.com.