Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 3, December 2023, pp. 1766~1773
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i3.pp1766-1773

Enhancing lung cancer disease diagnosis by employing ensemble deep learning approaches

Manmath Nath Das, Niranjan Panda, Rasmita Rautray
Department of Computer Science and Engineering, Siksha 'O' Anusandhan University, Bhubaneswar, India

Article Info

Article history:

Received Jul 1, 2023
Revised Sep 15, 2023
Accepted Sep 26, 2023

Keywords:

Artificial neural network
Ensemble learning
Lung cancer
Machine learning
Voting techniques

ABSTRACT

Cancer is a disease that results from the unnatural proliferation of aberrant cells that infest the body's healthy cells and spread throughout the body. Lung cancer is characterized by an imbalance in the cells of the affected organs, namely the lungs. The prediction of lung cancer at an early stage is very important, particularly in countries that are densely populated and have lower incomes. Clinically conventional approaches, such as blood tests and other types of treatments, are used by specialists. The age of artificial intelligence (AI) has begun, and today, it is feasible to construct a computer-aided diagnostic mechanism with the assistance of machine learning and deep learning algorithms. In this particular piece of research, one deep learning algorithm, an artificial neural network (ANN), has been investigated to determine whether or not lung cancer could be detected at an earlier stage. In addition to conventional ANN, ensemble ANN with weighted averaging and soft and hard voting ensemble techniques are also considered. In order to achieve this effectiveness, the state-of-the-art parameters for the proposed method using ANN are assessed and evaluated using the lung cancer dataset. The empirical analysis shows that hard voting-enabled ANN shows the highest accuracy at 97.47%.

Corresponding Author:

Manmath Nath Das
Department of Computer Science and Engineering, Siksha 'O' Anusandhan University
Bhubaneswar, Odisha, India
Email: manmathnath.das@gmail.com

1. INTRODUCTION

Cancer research is becoming one of the most competitive medical professions. This disease is distinguished by uncontrollable cell growth that can infiltrate organs and tissues. Cancer can affect DNA. Cancer cells have more hereditary abnormalities than typical cells, even if each person's tumor has distinct genetic variations [1]. According to the NCI and WHO datasheet, 12 million cases were reported in 2012, of which 7 million died due to a lack of diagnosis. Over the next two decades, the number of patients grew by 70%, and one in six died due to a faulty diagnostic model. Estimated 2015 fatalities were 8.8 million, rising to 9.6 million in 2018 [2]. The lack of a trustworthy diagnostic model promotes this death troll. Consequently, cancer categorization is becoming an important medical study area. Due to early cancer classification, detection, and diagnosis, patient survival is favorable. However, accurate identification and therapy can limit cancer mortality. The majority of cancer-related deaths are from lung cancer. Lung abnormal cell growth is the cause. Smokers and individuals with chest sickness are more prone to acquiring lung cancer. Lung cancer risk factors include carcinogenic gases, poverty, and pollution. Early lung cancer detection is vital to averting more devastating outcomes. Artificial intelligence (AI) may boost lung disease survival by diagnosing early. According to computer science, AI research involves "intelligent agents." An intelligent agent is any creature that remembers its surroundings and helps it attain goals [3]–[5].

AI may become more relevant in administrative science and practical research. Intelligence is often characterized as understanding and reasoning about complex topics. In the near future, intelligent computers will imitate human skills in certain fields [6]. Intelligent equipment and software that reason, learn from experience, acquire information, communicate, stretch, and remember are explored and created in AI. Research and development of AI is called "machine learning." This essay discusses how technology, particularly machine learning, is transforming individual conduct and our collective and active organizations and increasing pulmonary disease diagnoses [7]. Machine learning's long-term stability and consistency give it an edge over natural intelligence. It is possible to perform the same amount of work in a fraction of the time most individuals need. Using deep learning algorithms like neural networks [8], accurate predictions require careful data management. It is owing to classifying information, training, validation [9], and evaluation data, which performed effectively in an independent external cohort. After learning about training and testing, data annotations are needed to understand AI, specifically deep learning [10]. Despite the prevalence of conventional approaches, AI systems may accurately diagnose early lung cancer [11].

Mannendez *et al.* [12] described a neural network-based breast cancer diagnosis. Before mammography, the model can predict which women would develop a specific tumor. Janghel *et al.* [13] developed an artificial neural network (ANN)-based breast cancer diagnosis, prognosis, and prediction system. This aids in medical diagnosis. Back propagation algorithms, radial basis function (RBF) networks, learning vector quantization, and competitive learning networks are our neural network models. Experiments show learning vector quantization performs best on the test dataset. multi-layer perception (MLP), CL, and radial basis function network (RBFN) follow. High accuracy shows that learning vector quantization (LVQ) can classify breast cancer better than other models. Oustimov and Vu [14] introduction to ANNs in cancer genomics covers genomic data analysis basics. Many academics that adjust and construct new neural network implementations are creative, which boosts their usefulness. Neural networks will be employed for cancer genomics diagnostic, prognostic, and predictive software as genetic data becomes cheaper and more accessible. Using 1 Prasetyo *et al.* [15] constructed an ANN-extreme learning breast cancer detection model. Extreme learning machines (ELM) ANN outperformed BP ANN in generalization classifier modeling. This technology has great potential as an intelligent aspect of healthcare decision support systems. Bertolaccini *et al.* [16] examined lung cancer research data favoring ANN use. ANN was often used incorrectly in medical literature, according to our investigation. Medical professionals and biostatisticians should work together to discover and remedy errors. Nasser and Abu-Naser [17] developed an ANN to detect lung cancer. Lung cancer symptoms include wheezing, coughing, shortness of breath, difficulty swallowing, allergies, anxiety, chronic disease, tiredness, yellow fingers, and chest pain. Our ANN used those and other human data as input variables. Our ANN, "survey lung cancer," was trained and verified using data. A model evaluation showed that the ANN detected lung cancer with 96.67% accuracy.

Muhammad *et al.* [18] trained and evaluated an ANN utilizing 898 pancreatic cancer patients from the pancreatic, lung, colorectal, and ovarian cancer (PLCO) databases. By adding 18 neural network features, pancreatic cancer risk was assessed individually. The ANN model had 87.3% sensitivity, 80.7% specificity, and 80.8% and 80.7% AUC for training and testing cohorts. Our ANN's higher discriminatory capacity in predicting pancreatic cancer risk offers a new way to identify high-risk individuals who may benefit from specialized surveillance and medications. Mao *et al.* [19] examined ANN studies for gastrointestinal and liver cancer detection. They explained ANN development, working philosophy, and characteristics. Even with biased or erroneous data, a good ANN can predict accurately. Finally, training concerns include overfitting, network paralysis, local minimums, and long training timeframes. Alshayeji *et al.* [20], based on the wisconsin breast cancer dataset (WBCD) and wisconsin diagnostic breast cancer (WDBC) dataset, a shallow ANN model with one hidden layer diagnosed and predicted BC. Five-fold cross-validation divides datasets into 20% testing and 80% training. Sensitivity, specificity, precision, accuracy, F1-score, and AUC assess model usefulness and efficiency. The shallow ANN model diagnosed benign and malignant tumors using the WBCD with 99.85% accuracy, 99.72% specificity, 100% sensitivity, 99.69% precision, and 99.84% F1-score. For BC identification, WDBC had 99.47% accuracy, 99.53% specificity, 99.59% sensitivity, 98.71% precision, and 99.13% F1-score. WBCD and WDBC AUCs of 99.86% and 99.56% showed the model's discrimination. Zhao *et al.* [21] developed a comprehensive population EC prediction model. GEO datasets, including GSE (63,678, 106,191, 115,810, and 17,025) and the TCGA EC RNA sequence, were used to generate the training, test, and validation groups. The training group's 96 largest DEGs were assessed for function and pathway enrichment. The random forest found disease-specific genes, and an ANN was diagnosed. Each group's signature was identified through receiver operating characteristic (ROC) curves. Last, immune infiltration was examined for EC time differences. ANN diagnostic accuracy was 0.882, 0.864, and 0.839, with AUCs of 0.928, 0.921, and 0.782. We created a reliable artificial diagnostic model for EC using bioinformatics and uncovered immunological ecology abnormalities that could be diagnostic targets.

In this study, an ANN has been studied to see if it may be used to diagnose lung cancer at an earlier stage than is now possible. In addition to traditional ANN, we also consider ensemble ANN methods such as

weighted averaging and soft and hard voting ensemble techniques. The lung cancer dataset is used to analyze and assess the state-of-the-art parameters for the suggested method utilizing ANN to attain this efficacy.

The following is a summary of the main contributions of the recent research:

- To evaluate the ANN model for lung cancer classification.
- To propose an ensemble ANN based on weighted averaging and soft and hard voting ensemble techniques.

The remaining portions of the paper are organized as follows: The information and techniques employed in the present research endeavors, along with the suggested work, are represented in section 2. In section 3 summarizes the analysis of the results obtained. Finally, section 4 provides this research's ultimate conclusion with possible future directions.

2. METHOD

The dataset descriptions utilized for the empirical study of the suggested model are presented in this section. The lung cancer microarray dataset has been considered for the empirical study. Initially, the dataset is not normalized. Therefore, before considering this dataset, the dataset is normalized. Then, the ANN is applied to the normalized dataset to make the initial predictions. Ensemble approaches such as Weighted averaging and Voting are applied to the initial prediction. The two different approaches, hard and soft voting, are used in this research work. The proposed model is discussed with a suitable Algorithm, and the block diagram is presented in the last sub-section.

2.1. Dataset description

A lung cancer microarray data set was considered from the UCI ML data repository for this work [22]. Table 1 briefly summarizes the dataset used and its 72 sample details and 7,129 attributes. Acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML) are the classes of interest in the dataset under consideration.

Table 1. Dataset description

Attribute description	Number
Dimension of dataset	72×7,129
Samples	72
Genes	7,129

2.2. Artificial neural network (ANN)

ANNs use human brain-inspired approaches to comprehend nonlinear data. ANNs are made up of neurons stacked in layers between data input and output. Similar to biological neurons, the connections and patterns between these units form the network's behavior, which may be learned by backpropagation. Backpropagation feeds the network with known-good data. After comparing the ANN's output with the known output, component patterns are changed to lower the local motion similarity degree (LMSD). After several iterations, the network gets more accurate and can execute complex tasks with less computing power. ANNs are a crucial computing resource for biological research because nonlinear systems must increasingly mimic biomedical systems. ANNs have long been employed in cancer treatment. Recent research has improved our grasp of cancer's molecular features. As a result, computational techniques are more efficient [23], [24].

2.3. Ensemble approaches employed

Machine learning outcomes may be improved via ensemble learning by merging multiple different models. Compared to using only one model, this strategy enables the generation of far more accurate predicted performance. The fundamental concept is to acquire a group of classifiers and then lend those classifiers to vote. For the current work, four different kinds of ensemble approaches have been considered [25], [26]. An ensemble method for machine learning known as "weighted average" or "weighted sum ensemble" integrates the predictions from several models in such a way that the contribution made by each model is given a weighting that is proportionate to its level of capacity.

To forecast the output class, it simply takes the class with the largest margin of victory. It averages the results from all the classifiers that were supplied into the voting classifier. Instead of building several separate specialized models and comparing their performance, we may instead train a single model using the data from the other models and have it made output predictions based on the models' combined majority vote. Soft voting and hard voting are two separate voting procedures. Soft voting may be used for both classification and regression tasks, and it averages the predictions generated by several different finely adjusted models

trained on the same data. The predicted probabilities of the base learners are combined to get the final prediction result. Let λi be the initial prediction probability of different base learners or classifiers (Bi). The final prediction of the ensemble model can be represented by, ρ as shown in (1).

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

The hard voting or majority voting principle is used in the hard voting ensemble, which is used for classification tasks. This ensemble integrates the predictions from different trained models that are trained on the same dataset. Let ρ be the predicted class label through the hard voting technique. This predicted value can be calculated by (2):

$$\rho = mode\{c(a_1), c(a_2), \ldots\ldots, c(a_n)\} \quad (2)$$

where c is the class for the attribute a_1 of the dataset D.

2.4. Proposed work

The current work employs the ANN model with the lung cancer dataset. The ANN model is initially applied to the normalized dataset to obtain the prediction result. Then, different ensemble approaches, such as weighted averaging and voting techniques, are applied to the normalized dataset. The pseudocode for the proposed model is represented in Algorithm 1, and the detailed workflow of the manuscript is depicted in Figure 1.

Algorithm 1. The pseudocode models

- Raw Dataset input
- Dataset Normalization
- Perform Dataset Splitting operation with distribution ratio (d) as 0.25
- Initialize ANN () to the normalized dataset
    - Set the number of input layer (I)
    - Set the number of hidden layer (H)=5
    - Set optimizer = 'adam'
    - Set activation function I='relu'
    - Set activation function H= 'relu'
    - Set activation function for output layer= 'sigmoid
    - Obtain output (O)
- Initialize Ensemble ()
    - Set the number of epoch (E)
    - Obtain Initial prediction
        - Invoke Weighted Avg ()
        - Invoke Hard Voting ()
        - Invoke Soft Voting ()
- Result comparison among weighted Avg (), Hard Voting (), Soft Voting()

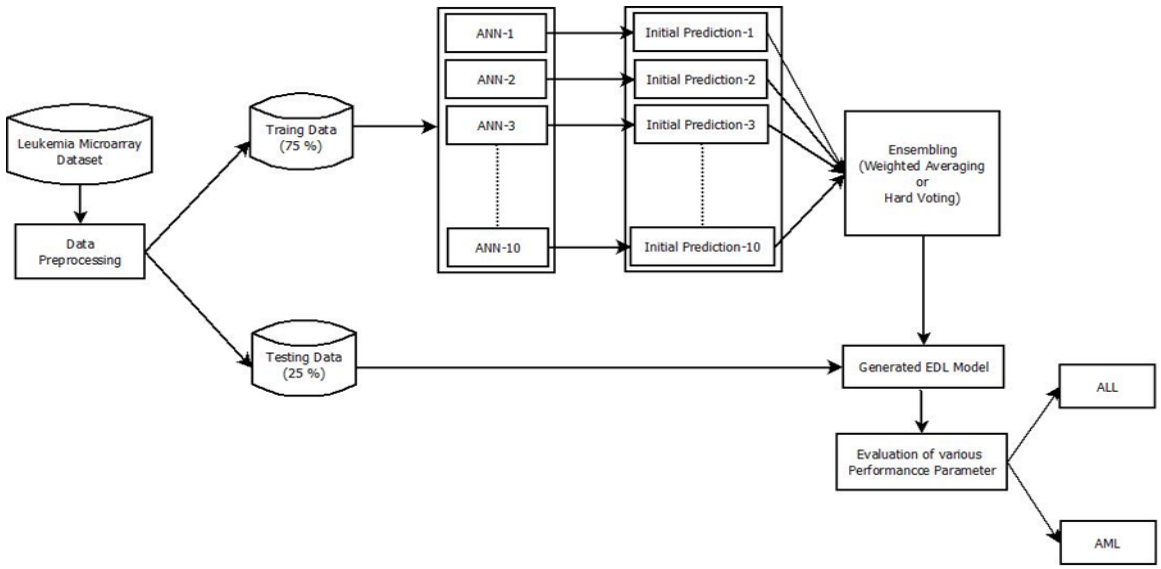

Figure 1. Workflow of the proposed work

3. RESULTS AND DISCUSSION

There are various presumptions involved in the assessment of this suggested model. To begin, some experiments focus just on ANN. The prediction findings were then improved using the three fundamental ensemble approaches of weighted averaging, soft voting, and hard voting. Additionally, a workstation equipped with 8 GB of RAM, a 500 GB SSD, 1 TB HDD, an Intel Core i5 processor with 3.6 GHz clock speed, and Ubuntu 20.04 has been used to test the proposed system. The proposed work is evaluated over several parameters, including accuracy (A_{CC}), misclassification rate (M_{CR}), precision (P_{RE}), recall (R_{EC}) or sensitivity (S_{EN}) or true positive rate (T_{PR}), F1-score (F-1S), specificity (S_{PE}) or true negative rate (T_{NR}), false-positive rate (F_{PR}), false-negative rate (F_{NR}), and Mathew's correlation coefficient (M_{CC}) as shown in (3)-(11). Here, T_A and F_A stand for true and false positives, respectively. T_B and F_B stand for true and false negatives, respectively.

$$A_{CC} = \frac{T_A + T_B}{T_A + T_B + F_A + F_B} \tag{3}$$

$$M_{CR} = \frac{F_A + F_B}{T_A + T_B + F_A + F_B} \tag{4}$$

$$P_{RE} = \frac{T_A}{T_A + F_A} \tag{5}$$

$$R_{EC} = \frac{T_A}{T_A + F_B} \tag{6}$$

$$F - 1S = \frac{2 \times T_A}{2 \times T_A + F_B + F_B} \tag{7}$$

$$S_{PE} = \frac{T_B}{T_B + F_A} \tag{8}$$

$$F_{NR} = \frac{F_B}{T_A + F_B} \tag{9}$$

$$F_{PR} = \frac{F_A}{T_B + F_A} \tag{10}$$

$$M_{CC} = \frac{(T_A + T_B) - (F_A + F_B)}{\sqrt{(T_A + F_A)(T_A + F_B)(T_B + F_A)(T_B + F_B)}} \tag{11}$$

When datasets contain several instances, employing the DL method, ANN alone (which we termed "approach-1") is reasonable. Furthermore, several experiments are conducted on ANN models that use weighted averaging, soft voting, and hard voting ensemble approaches (also known as "approach-2," "approach-3," and "approach-4"). The detailed performance evaluations for various proposed hybrid approaches are provided in Table 2. In contrast, the findings in % on the accuracy, recall, specificity, precision, FPR and FNR, MCR, and MCC are shown in Figures 2 to 8, respectively. According to reports based on the performance measurements, the model "approach-4", i.e., ANN with hard voting classifiers, as seen in Table 2 and Figure 2, surpasses all other recommended approaches with 97.47% accuracy and is deemed to be the proposed DL approach. In order to justify the novelty of the work, we have compared the results obtained using this suggested work with some of the state-of-the-art works considered earlier, as seen in Table 3. It can be noticed that this suggested work outperforms some other works employed in the literature in this work in terms of various performance parameters considered for the comparison.

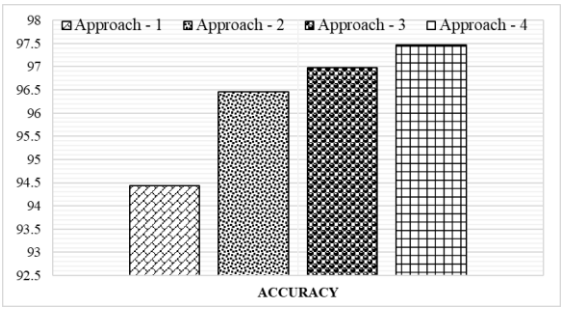

Figure 2. Accuracy comparison among different approaches

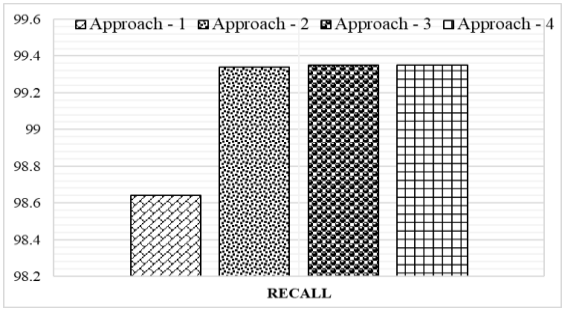

Figure 3. Recall comparison among different approaches

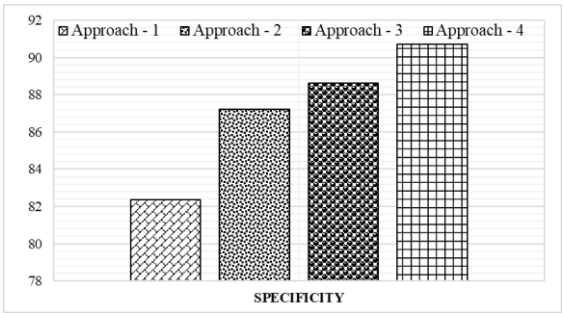

Figure 4. Specificity comparison among different approaches

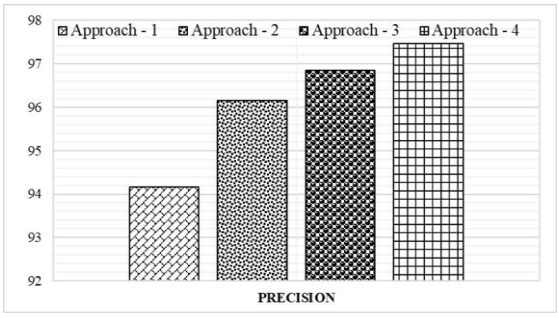

Figure 5. Precision comparison among different approaches

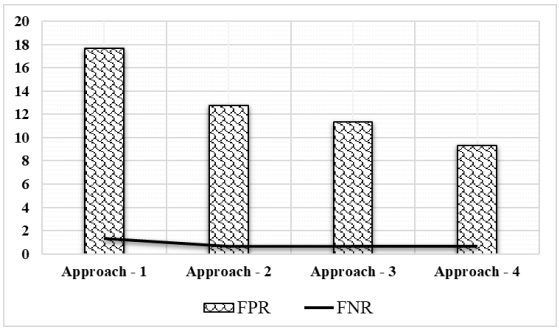

Figure 6. FPR and FNR comparison among different approaches

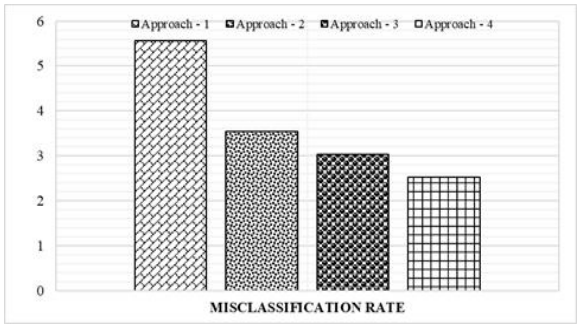

Figure 7. Misclassification rate (MCR) comparison among different approaches

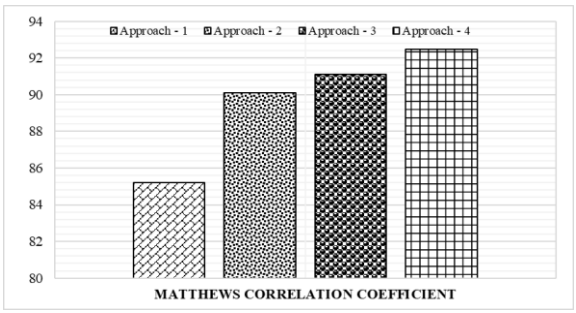

Figure 8. Matthews correlation coefficient (MCC) comparison among different approaches

Table 2. Performance evaluation for different approaches

Proposed approaches	Accuracy in (%)	MCR	PRE	REC	F1-S	SPE	FNR	FPR	MCC
Approach-1	94.44	5.56	94.16	98.64	96.35	82.35	1.36	17.65	85.19
Approach-2	96.46	3.54	96.15	99.34	97.72	87.23	0.66	12.77	90.1
Approach-3	96.97	3.03	96.84	99.35	98.08	88.64	0.65	11.36	91.11
Approach-4	**97.47**	2.53	97.47	99.35	98.4	90.7	0.65	9.302	92.48

Table 3. Performance comparison of proposed work in contrast to the existing literature

Work	Methodologies	Dataset(s)	Comparison parameters in (%)			
			ACC	PRE	REC	F1S
Menéndez *et al.* [12]	SOM, SVM, and MARS	Breast cancer dataset	80.28	-	99.00	-
Janghel *et al.* [13]	ANN, MLP, RBFN, LVQ, CL	Breast cancer dataset	95.82	-	98.53	-
Utomo *et al.* [15]	ANN with ELM	WBC dataset	96.40	-	94.80	-
Nasser and Abu-Naser [17]	ANN	Lung cancer dataset	96.67			
Muhammad *et al.* [18]	ANN	PLCO datasets			87.30	
Alshayeji *et al.* [20]	Shallow ANN	WBCD and WDBC dataset	99.85	99.69	100.00	99.84
Zhao *et al.* [21]	ANN	GEO and TCGA datasets	88.20	-	-	-
Proposed work	ANN with EL approaches	Leukemia dataset	97.47	97.47	99.35	98.40

4. CONCLUSION

The proposed model is evaluated over the Leukemia dataset. The ANN deep learning technique has been applied to obtain the initial prediction to the normalized dataset. Then, different ensemble techniques, including the weighted average, hard voting, and soft voting techniques, are applied to the initial prediction of base learner model ANN with 50 number epochs. The results are then compared with each other to obtain the best result. The empirical analysis shows that the ANN with hard voting, "approach-4," has a 97.47% accuracy level. However, the ANN with weighted average and ANN with soft voting shows an accuracy level of 96.46% and 96.97%, respectively, whereas the ANN model shows a 94.44% accuracy level. Furthermore, this work can be extended with other ML and DL approaches on other various microarray datasets.

REFERENCES

[1] M. Saminathan, M. Ramachandran, A. Kumar, K. Rajkumar, A. Khanna, and P. Singh, "A study on specific learning algorithms pertaining to classify lung cancer disease," *Expert Systems*, vol. 39, no. 3, Aug. 2022, doi: 10.1111/exsy.12797.

[2] H. A. Miller, V. H. van Berkel, and H. B. Frieboes, "Lung cancer survival prediction and biomarker identification with an ensemble machine learning analysis of tumor core biopsy metabolomic data," *Metabolomics*, vol. 18, no. 8, Jul. 2022, doi: 10.1007/s11306-022-01918-3.

[3] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," *Neural Computing and Applications*, vol. 34, no. 12, pp. 9579–9592, Apr. 2022, doi: 10.1007/s00521-020-04842-6.

[4] J. A. Alzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, and C. Thaventhiran, "Boosted neural network ensemble classification for lung cancer disease diagnosis," *Applied Soft Computing Journal*, vol. 80, pp. 579–591, Jul. 2019, doi: 10.1016/j.asoc.2019.04.031.

[5] S. Doppalapudi, R. G. Qiu, and Y. Badr, "Lung cancer survival period prediction and understanding: Deep learning approaches," *International Journal of Medical Informatics*, vol. 148, p. 104371, Apr. 2021, doi: 10.1016/j.ijmedinf.2020.104371.

[6] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, p. 106885, Jan. 2021, doi: 10.1016/j.asoc.2020.106885.

[7] M. Mamun, A. Farjana, M. Al Mamun, and M. S. Ahammed, "Lung cancer prediction model using ensemble learning techniques and a systematic review analysis," in *2022 IEEE World AI IoT Congress, AIIoT 2022*, Jun. 2022, pp. 187–193, doi: 10.1109/AIIoT54504.2022.9817326.

[8] M. Mohammed, H. Mwambi, I. B. Mboya, M. K. Elbashir, and B. Omolo, "A stacking ensemble deep learning approach to cancer type classification based on TCGA data," *Scientific Reports*, vol. 11, no. 1, Aug. 2021, doi: 10.1038/s41598-021-95128-x.

[9] S. Gupta, M. K. Gupta, and R. Kumar, "A novel multi-neural ensemble approach for cancer diagnosis," *Applied Artificial Intelligence*, vol. 36, no. 1, Dec. 2022, doi: 10.1080/08839514.2021.2018182.

[10] M. Masud, N. Sikder, A. Al Nahid, A. K. Bairagi, and M. A. Alzain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–21, Jan. 2021, doi: 10.3390/s21030748.

[11] Y. Xie *et al.*, "Early lung cancer diagnostic biomarker discovery by machine learning methods," *Translational Oncology*, vol. 14, no. 1, p. 100907, Jan. 2021, doi: 10.1016/j.tranon.2020.100907.

[12] L. Á. Menéndez, F. J. C. Juez, F. S. Lasheras, and J. A. Á. Riesgo, "Artificial neural networks applied to cancer detection in a breast screening programme," *Mathematical and Computer Modelling*, vol. 52, no. 7–8, pp. 983–991, 2010, doi: 10.1016/j.mcm.2010.03.019.

[13] R. R. Janghel, A. Shukla, R. Tiwari, and R. Kala, "Breast cancer diagnosis using artificial neural network models," in *Proceedings - 3rd International Conference on Information Sciences and Interaction Sciences, ICIS 2010*, Jun. 2010, pp. 89–94, doi: 10.1109/ICICIS.2010.5534716.

[14] A. Oustimov and V. Vu, "Artificial neural networks in the cancer genomics frontier," *Translational Cancer Research*, vol. 3, no. 3, pp. 191–201, 2014.

[15] C. Prasetyo, A. Kardiana, and R. Yuliwulandari, "Breast cancer diagnosis using artificial neural networks with extreme learning techniques," *International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 7, 2014, doi: 10.14569/ijarai.2014.030703.

[16] L. Bertolaccini, P. Solli, A. Pardolesi, and A. Pasini, "An overview of the use of artificial neural networks in lung cancer research," *Journal of Thoracic Disease*, vol. 9, no. 4, pp. 924–931, Apr. 2017, doi: 10.21037/jtd.2017.03.157.

[17] I. Nasser and S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 3, pp. 17–23, 2019.

[18] W. Muhammad *et al.*, "Pancreatic cancer prediction through an artificial neural network," *Frontiers in Artificial Intelligence*, vol. 2, May 2019, doi: 10.3389/frai.2019.00002.

[19] W. B. Mao *et al.*, "Application of artificial neural networks in detection and diagnosis of gastrointestinal and liver tumors," *World Journal of Clinical Cases*, vol. 8, no. 18, pp. 3971–3977, Sep. 2020, doi: 10.12998/wjcc.v8.i18.3971.

[20] M. H. Alshayeji, H. Ellethy, S. Abed, and R. Gupta, "Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach," *Biomedical Signal Processing and Control*, vol. 71, p. 103141, Jan. 2022, doi: 10.1016/j.bspc.2021.103141.

[21] D. Zhao *et al.*, "Diagnosis and prediction of endometrial carcinoma using machine learning and artificial neural networks based on public databases," *Genes*, vol. 13, no. 6, p. 935, May 2022, doi: 10.3390/genes13060935.

[22] Z. Zhu, Y. S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognition*, vol. 40, no. 11, pp. 3236–3248, Nov. 2007, doi: 10.1016/j.patcog.2007.02.007.

[23] A. A. Abd 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.

[24] 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.

[25] 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.
[26] 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.

BIOGRAPHIES OF AUTHORS

Manmath Nath Das is currently pursuing Ph.D. as a research scholar at Shiksha 'O' Anusandhan University, Bhubaneswar, Odisha. Mr. Das completed his M.Tech. in CSE in 2011 from BPUT, Odisha. His research interest includes machine learning, deep learning, and cloud computing. He can be contacted at email: manmathnath.das@gmail.com.

Niranjan Panda is working as an Associate Professor in the CSE Department, Siksha 'O' Anusandhan University, Bhubaneswar, India since 2011. He was awarded his Ph.D. degree from Siksha 'O' Anusandhan University, Bhubaneswar, India, in 2019. He got his M.Tech. Degree from KIIT University, Bhubaneswar, India, and B.Tech. Degree from Utkal University, Bhubaneswar, India, in 2010 and 2005, respectively. His research interests include ad hoc networks, computer security, intelligent systems, and machine learning. He can be contacted at email: niranjanpanda@soa.ac.in.

Rasmita Rautray received her BE degree and M.Tech. degree in CSE from Utkal University, Bhubaneswar, India, in 2003 and 2006, respectively, and the Ph.D. degree in CSE from Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India, in 2017. She is currently working as an Associate Professor in the Department of CSE, FET (ITER), Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India. Her research interests include data mining, soft computing, and information retrieval. She has contributed over 30 research papers in many National and International journals and conferences. She can be contacted at email: rashmitaroutray@soa.ac.in.