World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:17, No:6, 2023

Breast Cancer Prediction Using Score-Level Fusion of Machine Learning and Deep Learning Models

Sam Khozama, Ali M. Mayya

***Abstract*—**Breast cancer is one of the most common types in women. Early prediction of breast cancer helps physicians detect cancer in its early stages. Big cancer data need a very powerful tool to analyze and extract predictions. Machine learning and deep learning are two of the most efficient tools for predicting cancer based on textual data. In this study, we developed a fusion model of two machine learning and deep learning models. To obtain the final prediction, Long-Short Term Memory (LSTM), ensemble learning with hyper parameters optimization, and score-level fusion is used. Experiments are done on the Breast Cancer Surveillance Consortium (BCSC) dataset after balancing and grouping the class categories. Five different training scenarios are used, and the tests show that the designed fusion model improved the performance by 3.3% compared to the individual models.

***Keywords*—**Machine learning, Deep learning, cancer prediction, breast cancer, LSTM, Score-Level Fusion.

I. Introduction

BREAST cancer is one of the most common forms of cancer among women worldwide, with significant implications for patient outcomes and healthcare systems. Early detection and accurate diagnosis of breast cancer are critical for improving patient outcomes, reducing healthcare costs, and ultimately saving lives. Breast Cancer is classified as one of the most common cancer types [1]. According to the World Health Organization (WHO), cancer is the second leading cause of death [2], [3]. Breast and oral cavity cancers are considered the causes of 25% of deaths around the world [1].

Based on cancer statistics from 2020, breast cancer constitutes 11.7% of all cancer records around the world [4]. From the death side [5], [9], breast cancer was classified as the second deadliest cancer after lung cancer by a percentage of 6.9%.

All these previous facts lead to the importance of the prediction of breast cancer before actual diagnosis. Early prediction can reduce the cancer rate and help physicians predict cancer at its early stages. Fortunately, computer science algorithms have been incrementally developed and enhanced and can be used for the purpose of cancer prediction. Physicians themselves cannot process and analyze all the cancer data since it is huge and very related. Consequently, they need the efficiency of computer science algorithms that can handle large amounts of data in a short time.

Machine learning (ML) is one of the most common computer science fields that is used in cancer prediction [6]. Deep learning (DL), as a sub-field of ML, is a very powerful tool to handle large cancer datasets that are not easy to process using traditional ML technologies [7]. Many previous studies used ML and DL techniques for the purpose of cancer prediction [8], [10]-[15], while others designed ML and DL models for cancer diagnosis [16]-[18].

Khan et al. [19] used three different DL networks: CNN, GoogleNet and ResNet50 for the aim of breast cancer detection. They used a dataset of 8000 images and achieved an accuracy of 97.5%. Other studies in the field of cancer detection were introduced, like [20], which used the CNN network for auto breast cancer detection based on a dataset of mammogram images and achieved 97% accuracy. Zhang et al. [21] also used the ResNet50 network for breast cancer classification. They used four datasets (CLEF-15: 6776, CLEF-16: 10942, ISIC-16: 1279, and SIC-17: 2750). Their models achieved 76.6%, 87.3%, 85.5% and 90.2% accuracy, respectively. Many ML and DL modes were used in cancer prediction systems [22].

Khozama and Mayya [10] studied the effect of weighting the risk factor of breast cancer. They used the well-known BCSC dataset, consisting of 280,660 cases and 12 risk factors. The researchers analyzed the dataset and used some other medical questionnaires in order to build their cancer prediction tool. They constructed a mathematical model to calculate the degree of importance of each risk factor. The decision tree ML model is trained using the original and the weighted version of the dataset. The results indicated an improvement using the weighting methodology of 6.9%. They continued their work and built another breast cancer model using the same dataset [14]. Instead of 0 or 1, the new system provided the breast cancer as a percentage (0-100%). They used Bayesian theory to compute a new range-based cancer score. After that, they used the modified dataset and ensemble learning to build a breast cancer prediction model. Their approach achieved an Area Under Curve (AUC) of 97.95% and a false rejection rate of 1.12%.

Lang et al. [23] predicted oropharyngeal cancer using 3D CNN on a dataset consisting of 675 breast cancer cases. They split the dataset into training (412 cases of the OPC dataset and 263 cases of the HNSCC dataset) and validation (90 cases of the HN PET-CT dataset). For the test, they used 80 cases from the HN1 dataset. The experiments showed that the AUC was 0.81.

In 2022, Ashokkumar et al. [24] predicted the lymph nodes of the breast using the Kohonen self-organizing ANN. They

S.A. Khozama is with the Faculty of Information Technology and Bionics; Pázmány Péter Catholic University, Budapest, Hungary (corresponding author, e-mail: khozama.sam@itk.ppke.hu).

A.M. Mayya is with Tishreen University, Lattakia, Syria (e-mail: ali.mayya@tishreen.edu.sy).

used a dataset of 10,150 images of 850 patients. Their approach achieved 94% accuracy.

Recently, Saleh et al. [25] introduced a DL-based breast cancer prediction model. They used the Recurrent Neural Network (RNN) with five hidden layers and one output layer. Three feature selection models were proposed. The BCWD was used in their study. It had 30 factors (features) and one class (cancer prediction 0 or 1). The results indicated an accuracy of 95.18%.

Previous studies introduced the problem of breast cancer prevention in different ways. In our study, we will use the fusion approach of ML and DL models. For the deep DL part, we suggest using the "Long-Short Term Memory" (LSTM) network, while for the ML part, the ensemble learning supported by hyperparameters optimization will be used. The final cancer prediction model will be obtained using the fusion of both the ML and DL models.

II. Materials and Methods

A. System Description

The proposed breast cancer prediction system steps are illustrated in Fig. 1.

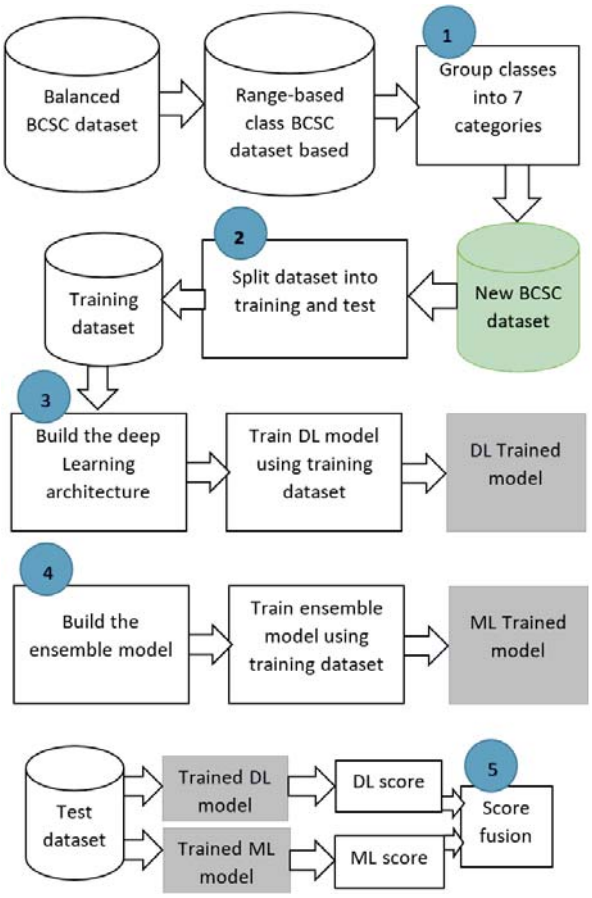


Fig. 1 Proposed methodology

As Fig. 1 shows, there are five main steps in the proposed breast cancer prediction model; where in the first step, the dataset is preprocessed and the target or classes are grouped into specific categories, while in the second step, the data set is split into training and test; for the third step, the DL architecture is constructed and trained using the training dataset, while in the fourth step, the ensemble ML model is built and trained using the same training dataset. In the final step, the DL and ML models' scores are fused using the score-level fusion and the final prediction is computed.

B. Dataset Preparation

The study suggests using the BCSC dataset. It includes the following risk factors: menopause, age group, race, Hispanic factor, body mass index, age at first birth, number of first relatives with breast cancer, breast procedure, last mammogram before the index mammogram, surgical menopause and current hormone therapy.

The original dataset has 280,660 records, but the problem with this dataset is that it is unbalanced (i.e., the number of "1-class" samples is too small compared to the number of "0-class" samples). After analyzing the dataset, we found that the "0-class" percentage is 96.68%, while the "1-class" is 3.32%. The solution is to re-balance the dataset by using oversampling techniques in which the minor-class samples are repeated until reaching a suitable percentage. The new version of the dataset is oversampled until the "0-class" samples become almost 15% instead of 3.32%.

The next issue with the BCSC dataset is that the cancer prediction within it is a binary classification problem since the target or classes are either cancer or non-cancer. A study [14] resolved this problem by proposing a new range-based cancer score. They updated the BCSC dataset and made the target or class column percentage, including 36 categories.

The balanced range-based cancer score (BCSS) dataset is used in the current study. One modification is applied to the dataset before using it. Since the number of classes is too large (36) classes, making the classification problem very complex, we suggest grouping the classes or the target column of the BCSC dataset into categories. This solution minimized the number of classes from 36 to 7.

After preprocessing of the BCSC dataset is done, the dataset is split into two sub-datasets (training and test) using a 25% percentage for the test set and 75% for training.

C. Building the DL Model

In this step, a specific DL architecture is proposed. Fig. 2 shows this architecture. The first layer of the model is the sequence input layer, which represents the input layer taking the input features of the training samples and forwarding them to the next LSTM layer. The LSTM layer is the main part of the DL model. It consists of 500 neurons, each of which consists of four basic cells: input cell, memory cell, forget cell, update cell, and output cell. The input cell receives the input from the previous cell, while the forget cell decides what information to remember and what to drop, controlling the cell state reset.

The update cell depends on the information received by the input cell, forget cell, and previous hidden LSTM cell h_{t-1}. It produces the output c_t, which represents the current output of the LSTM cell (output at time t). Another output will be

produced, which is the h_t representing the current hidden LSTM output that will be forwarded to the next LSTM layer [26].

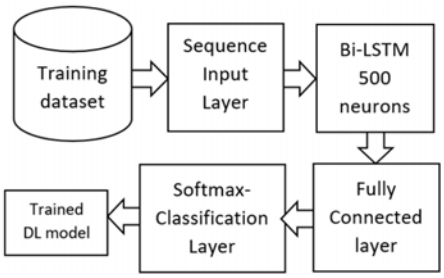


Fig. 2 Proposed DL architecture

D. Building the Ensemble ML Model

Ensemble models are types of ML models that use a combination of multiple ML classifiers in one combination. This methodology works in two different ways to get the final classification. The first approach is boosting, while the second is bagging. In the former, the classification decision is based on an iterative strategy in which the first classifier introduces his decision to the next one that will learn from the first classifier's error and try to minimize the classification error until reaching the final classifier that will produce the final decision with the minimum classification error. In the second approach (bagging), the classifiers work in parallel. The ensemble tries to minimize the prediction variance by generating new samples of the training dataset by repeating the training data and producing sub-datasets to train multiple classifiers, and the final decision is based on the fusion of their scores [27].

In the current study, we suggest using an ensemble of 30 boosting decision trees, and in order to get the best performance, we suggest using the hyperparameters optimization of those 30 decision tree models. Fig. 3 illustrates the architecture of the ML ensemble model.

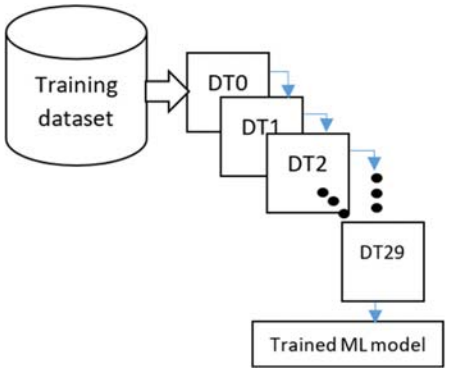


Fig. 3 Proposed ML architecture

E. DL and ML Fusion

After building the ML and DL models, they are combined together using the score-level fusion in which the ML and DL scores are fused, and the final prediction decision will be made.

III. Results and Discussion

In the experimental part, many training and test scenarios are used to evaluate the proposed method.

For the LSTM model, five different test scenarios are proposed. In three of these scenarios, the LSTM architecture is modified by changing the number of neurons and the number of learning epochs, while the last two scenarios are related to the splitting percentage (raining and test percentages).

The performance evaluation metrics are used to evaluate the five scenarios. A true positive rate (TPR) is the percentage of correctly predicted cancer samples among all cancer test samples. A positive predictive rate (PPR) is the ratio of correctly predicted samples per predicted class. The false negative rate (FNR) is the opposite rate of the TPR, while the FDR is the opposite concept of the PPR. Test accuracy is the number of correctly predicted test samples, including cancer and non-cancer samples among all test samples. Table I includes the results of these five test scenarios.

TABLE I
Results of Different Test Scenarios of LSTM Model

Scenario	TPR	FNR	PPR	FDR	Test Accuracy
LSTM (100 iterations, 300 neurons)	65.1148%	34.88%	91.49%	8.509%	86.84%
LSTM (200 iterations, 300 neurons)	94.51%	5.49%	95.55%	5.54%	96.68%
LSTM (200 iterations, 400 neurons)	88.979%	11.02%	96%	4%	95.197%
LSTM (200 iterations, 300 neurons, test Percentage = 30%)	92.05%	7.95%	93.11%	6.89%	91.38%
LSTM (200 iterations, 300 neurons, test Percentage = 35%)	91.7116%	8.28%	96.814%	3.186%	93.85%

Table I proves the fact that the best LSTM architecture is achieved by using 300 neurons and 200 iterations for training. From the splitting point of view, the best case scenario is by using 25% of the dataset samples as a test set. The last two scenarios show that by taking more samples for a test set, the performance decreases.

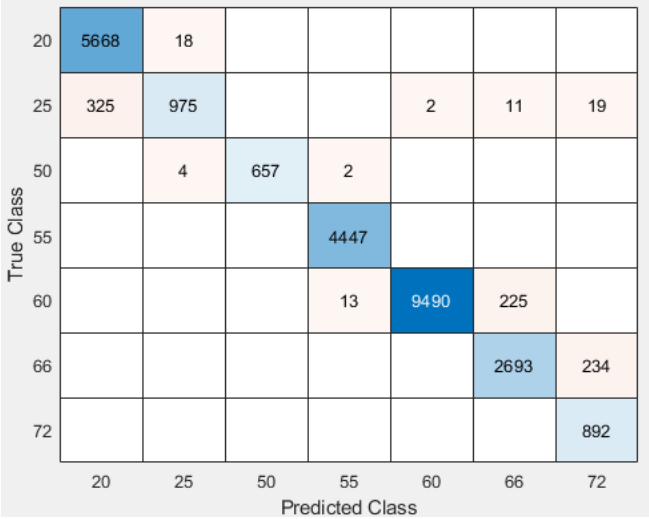
Fig. 4 shows the confusion matrix of the best DL model. It illustrates that the highest FNR error rate is related to class "25" with FNR = 0.26, while the best TPR is related to class "55" with TPR = 100%. However, class "50" has the best PPR value at 100%, while class "72" has the worst FDR value of 0.779.

For the ensemble learning ML model, the minimum classification error of the boosted decision tree models is 1.1%. The confusion matrix with TPR, FNR, PPR, and FDR is illustrated in Fig. 6. It shows that the best TPR is related to the "50". The results also indicate that the best PPR corresponds with class "55".

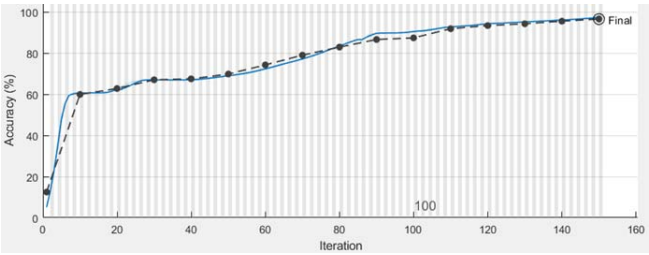
These two results are the same as those obtained by the DL model.

The final test scenario is the fusion scenario, in which the scores of the DL and ML models are fused together. Table II includes a detailed comparison between the cases of individual models and the fusion model. As illustrated in Table II, the fused model has a better performance than the individual ones. The accuracy is increased by 1.08% and 3.3% compared to the ML and DL individual models, respectively. The TPR is increased by 1.66% and 5.46%, while the PPR is improved by

2.01% and 5.44% compared to the DL and ML individual models.



(a) Confusion matrix of the best LSTM scenario



(b) Training progress

Fig. 4 Results of LSTM training

Results prove the fact that fusion of the DL and ML models improves the performance significantly.

TABLE II
COMPARING THE RESULTS OF INDIVIDUAL MODELS AND THE FUSION MODEL

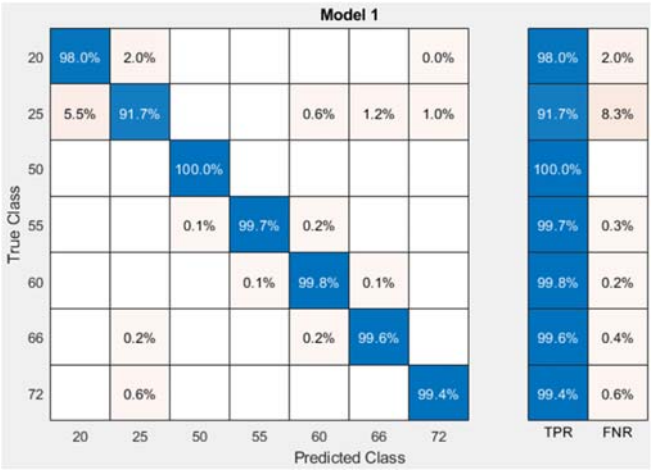
Model	TPR	FNR	PPR	FDR	Test Accuracy
LSTM	94.51%	5.49%	94.55%	5.45%	96.68%
Ensemble ML	98.31%	1.69%	97.98%	2.02%	98.9%
Fusion	99.9795%	0.0205%	99.9922%	0.0078%	99.98%

IV. CONCLUSION

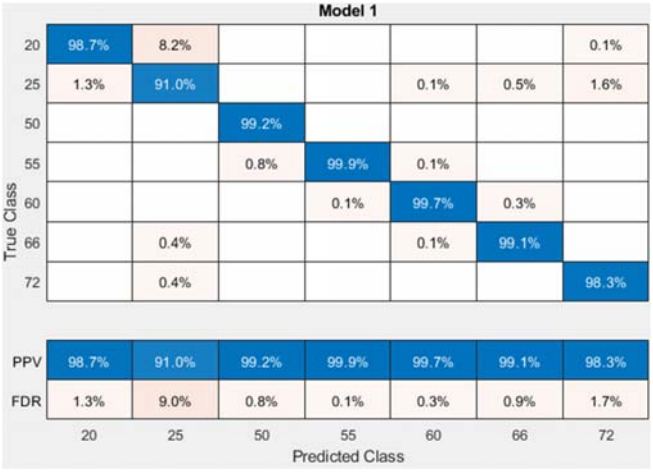
In this study, a range-based breast cancer prediction system is proposed. The proposed system includes a fusion of two different models. The first model is the LSTM DL model, while the second one is an ensemble of boosted decision trees.

In the first step, a balanced range-based BCSC dataset is used and preprocessed in order to group the classes into seven different categories. The new dataset is split into training and testing by using a percentage of 25% for the test set.

The DL model consists of a sequence input layer, one LSTM layer, one fully-connected layer, and one softmax-classification layer. The ML model, on the other hand, consists of 30 decision tree classifiers trained using the hyperparameters optimization and boosted learning approach.



(a) The TPR and FNR



(b) The PPR and FDR

Fig. 5 Confusion matrix of the best LSTM scenario

The final prediction is obtained using the fusion of ML and DL scores. Five different training scenarios, including different LSTM cells and different training epochs, are performed. The experiments are also applied using the individual modes and the fusion one in order to measure the effect of the fusion approach. The results indicate an improvement in performance. Future work can focus on using other different breast cancer datasets and studying the effect of using different risk factors on the performance of range-based breast cancer prediction.

ACKNOWLEDGMENT

Data collection and sharing was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium (HHSN261201100031C), available at: http://www.bcsc-research.org/.

This article publication was funded by PPCU supported by NKFIH, financed under Thematic Excellence Programme (TUDFO/51757-1/2019-ITM).

REFERENCES

[1] G. Chugh, S. Kumar and N. Singh, "Survey on machine learning and deep

learning applications in breast cancer diagnosis," Cognitive Computation, vol. 13, no. 6, pp. 1451-1470, 2021.
[2] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro and A. Liotta, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, vol. 223, p. 106951, 2022
[3] World Health Organization, WHO position paper on mammography screening, Geneva, Switzerland: WHO Library Cataloguing-in-Publication Data, 2014.
[4] world bladder cancer, "GLOBOCAN 2020: Bladder cancer 10th most commonly diagnosed worldwide," World Bladder Cancer, Lyon, France, 2020.
[5] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. J. DMV and 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, 2020.
[6] S. Sarumathi, M. Vaishnavi, S. Geetha, P. Ranjetha, " Comparative Analysis of Machine Learning Tools: A Review", International Journal of Computer and Information Engineering, Vol. 15, No. 6, 2021.
[7] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo and G.-Z. Yang, "Deep Learning for Health Informatics," IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 4-21, 2017.
[8] P. Ferroni, F. Zanzotto, S. Riondino, N. Scarpato, F. Guadagni, & M, Roselli, "Breast cancer prognosis using a machine learning approach", Cancers, Vol. 11, no. 3, 2018.
[9] A. Ahmad and A. M. Mayya, "A new tool to predict lung cancer based on risk factors," Heliyon , vol. 6, no. 2, p. e03402, 2020.
[10] S. Khozama, A. Mayya, "Study the Effect of the Risk Factors in the Estimation of the Breast Cancer Risk Score Using Machine Learning", Asian Pacific Journal of Cancer Prevention, Vol. 22, no.11, pp.3543-3551, 2021.
[11] C. Dalmiglio, L. Brilli, M. Campanile, C. Ciuoli, A. Cartocci and M. G. Castagna, "CONUT score: a new tool for predicting prognosis in patients with advanced thyroid cancer treated with TKI," Cancers, vol. 14, no. 3, 2022.
[12] C. Huang, Q. Su, Z. Ding, W. Zeng and Z. Zhou, "A novel clinical tool to predict cancer-specific survival in patients with primary pelvic sarcomas: A large population-based retrospective cohort study," Cancer Medicine, 2022.
[13] S. Nagalpara and B. M. Patel, "A Deep Learning Strategy for Predicting Liver Cancer Using Convolutional Neural Network Algorithm," Indian Journal of Computer Science, vol. 7, no. 3, 2022.
[14] S. Khozama, A. Mayya, "A new range-based breast cancer prediction model using the Bayes' theorem and Ensemble learning", Information Technology and Control Journal, 2022, to be published.
[15] P. Gupta, and S. Garg, "Breast cancer prediction using varying parameters of machine learning models", Procedia Computer Science, vol. 171, pp. 593-601, 2020.
[16] M. Saii and A. Mayya, "Lung Detection and Segmentation Using Marker Watershed and Laplacian Filtering", Journal of Biomedical Engineering and Clinical Science, vol. 1, no.2, pp. 29-42, 2015.
[17] Y. Benhammou, B. Achchab, F. Herrera and S. Tabik, "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights.," Neurocomputing, vol. 375, no. 2020, pp. 9-24, 2020.
[18] Y. A. Haşim, E. H, İ. T and K. S., "Detection of breast cancer via deep convolution neural networks using MRI images," Multimed Tools Appl., vol. 79, no. 21, pp. 15555-15573, 2019.
[19] C. CAO, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo and Z. Xie, "Deep learning and its applications in biomedicine," Genomics, proteomics & bioinformatics, vol. 16, no. 1, pp. 17-32, 2018.
[20] D. Selvathi and A. Aarthy Poornila, "Deep learning techniques for breast cancer detection using medical image analysis," in Biologically rationalized computing techniques for image processing applications., Cham, 2018.
[21] S. KHAN, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo and Z. Xie, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," Pattern Recognition Letters, vol. 125, pp. 1-6, 2019.
[22] P. Ferroni, F.M. Zanzotto., S. Riondino, N. Scarpato, F. Guadagni, and M. Roselli, "Breast cancer prognosis using a machine learning approach", Cancers, vol. 11, no.3, pp.328, 2019.
[23] D. M. Lang, J. C. Peeken, S. E. Combs, J. J. Wilkens and S. Bartzsch, "Deep Learning Based HPV Status Prediction for Oropharyngeal Cancer Patients," Cancers, vol. 13, no. 786, pp. 1-11, 2021.
[24] N. Ashokkumar, S. Meera, P. Anandan, M. Yaswanth, B. Murthy, T. A. Alahmadi, S. A. Alharbi, S. S. Raghavan and A. Jayadhas, "Deep Learning Mechanism for Predicting the Axillary Lymph Node Metastasis in Patients with Primary Breast Cancer," BioMed Research International, vol. 2022, pp. 1-14, 2022.
[25] H. Saleh, H. Alyami, and W. Alosaimi, "Predicting Breast Cancer Based on Optimized Deep Learning Approach", Computational Intelligence and Neuroscience, special issue, 2022.
[26] K. Greff, R. Srivastava K., J. Koutník, B. Steunebrink and J. Schmidhuber, "LSTM: A search space odyssey", IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pp.2222-2232, 2016.
[27] O. Sagi, and L. Rokach, 'Ensemble learning: A survey", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Rokach, vol. 8, no. 4, 2018.

Sam Khozama received his MSc in Computer Science Engineering from Pázmány Péter Catholic University, Hungary in 2018. He also received his BSc from Informatics Engineering Collage, Tishreen university, Syria. At present, he is finalizing his PhD at Pázmány Péter Catholic University, Hungary. He published 3 scientific papers in hardware acceleration, image processing, cancer prevention, and machine learning.

Ali M. Mayya received his PhD in computer engineering (pattern recognition specialty) from Tishreen University, Syria in 2018. He also received his MSc from computer and automatic control engineering department, Tishreen university, Syria. He also received his B.Tech degree in computer engineering form computer and automatic control engineering department, Tishreen university, Syria. At present, he works as a professor in the department of computer engineering, Tishreen university, Syria. He also a professor at Al Andalus university for medical science. He published more than 20 scientific papers in many fields including image processing, cancer prevention, machine learning, deep learning, pattern recognition and computer engineering.