

# Cancer Prediction Using Cascade Generalization and Duo Output Neural Network

Chatree Nilnumpetch  
Department of Computer Science  
Ramkhamhaeng University  
Bangkok, Thailand  
nchatree@ru.ac.th

Somkid Amornsamankul  
Department of Mathematics  
Mahidol University and Centre of  
Excellence in Mathematics  
Bangkok, Thailand  
somkid.amo@mahidol.ac.th

Pawalai Kraipeerapun  
Department of Computer Science  
Ramkhamhaeng University  
Bangkok, Thailand  
pawalai@rumail.ru.ac.th

**Abstract**—This paper proposes the combination of cascade generalization and duo output neural network based on feedforward backpropagation neural networks for cancer prediction. Duo output neural network is a neural network that is created based on two opposite targets in order to predict two opposite results. Cascade generalization is a technique that consists of a set of machines that are sorted together in which the predicted output produced from the previous machine plus the original training input are used for the creation of each machine. In this study, cascade generalization is organized in two levels: the base level and the meta level. In this research, duo output neural network is trained in each level of cascade generalization. Two outputs produced from the base level which are truth output and non-falsity output are averaged. The average result plus the original input are used for training a machine in meta level. The proposed technique is tested using two cancer datasets from UCI machine learning repository and found that our technique provides the best overall results when compared with three individual techniques.

**Index Terms**—Cascade generalization, Duo output neural network, Feedforward neural network, Binary classification, Breast cancer, Cervical cancer.

## I. INTRODUCTION

NEURAL networks are one of the most widely used classification techniques in medicine [1, 2, 3]. This paper proposes a new technique based on neural network used for cancer prediction for small datasets. Breast and cervical cancer are common cancers in women, which develop in the cells of the breasts and cervix, respectively [4, 5]. In this study, two cancer datasets are tested which are breast cancer Coimbra [6] and cervical cancer behavior risk [7]. Both datasets consisted of a small number of instances which are 116 and 72, respectively. The breast cancer Coimbra dataset [6] is very popular, in contrast to cervical cancer behavior risk dataset [7], which has not been tested much in research. For breast cancer Coimbra, several techniques have been used for classification. For example, it was found that extreme learning machine can provide the highest result with 80% accuracy when compared to neural network, k-nearest neighbor and support vector machine (SVM) [8]. The gradient boosting was found to provide the best accuracy which is 74.14% when compared to logistic regression, decision tree, random forest, k-nearest neighbor, SVM and naïve Bayes [9]. However, it was found that the highest accuracy belongs to random forest when compared to logistic regression, decision tree, k-nearest neighbor, naïve Bayes, SVM, and neural network [10]. Instead of using individual technique, bagging is used for the classification [11]. Bagging of decision tree and bagging of k-nearest neighbor were found to give the highest results when compared to bagging of other techniques which are SVM, multilayer perceptron, logistics regression, and random forest. Deep neural network

with two hidden layers was found to provide the highest result with 75.94% accuracy when compared to SVM and decision tree [12]. In [13], deep neural network with four hidden layers was found to give better accuracy results than random forests and logistic regression. In [14], the backpropagation neural network was improved using Nelder Mead and it was found that this technique can provide result with 76.5217% accuracy which is better than the original backpropagation neural network. Polat and Sentürk [15] proposed a three-step hybrid structure in which MAD normalization was used to normalized data in the first step. After that, k-means clustering based feature weighing was used in the second step. AdaBoostM1 was used in the third step for the classification. Their technique provided results with 91.37% accuracy. Yavuz and Eyupoglu [16] used principal component analysis cascaded by median filtering for data transformation and used generalized regression neural network for classification. It was found that their accuracy rate was 0.9773. Chiu *et al.* [17] proposed a novel method consisting of transfer learning, principal component analysis, multilayer perceptron, and SVM. Their technique can give 86.97% accuracy which is the highest accuracy when compared to other six machine learning algorithms. Idris and Ismail [18] proposed a combination of fuzzy-ID3 and FUZZYDBD in order to classify breast cancer. Their technique was found to provide 70.69% accuracy which is better than ID3, SVM, c4.5, naïve Bayes, k-nearest neighbor, and random forest.

The focus of this paper is to improve the cancer prediction technique based on feedforward backpropagation neural network for small datasets. The proposed approach is to use duo output neural network [19] and cascade generalization [20] to classify data.

The remainder of this paper is organized as follows. In Section II, two existing techniques are explained which are duo output neural network and cascade generalization. Section III describes our proposed approach which is the integration of duo output neural network and cascade generalization. In Section IV, two datasets used in this study are illustrated and their experimental results are explained. The conclusion of this paper is explained in Section V.

## II. INDIVIDUAL TECHNIQUES

### A. Duo Output Neural Network

Duo output neural network is a technique that uses two opposite targets for training a neural network in order to solve binary classification or regression problems [19]. In general, a neural network is trained using one target for binary classification. Instead of using only the original target,

duo output neural network also applies the opposite value of the original target for training a neural network. These both targets are called truth target and falsity target, respectively. The falsity target value is computed as 1 minus the original target value. A neural network is trained using both targets and produces two outputs which are truth output and falsity output. Fig.1 shows the model of duo output neural network (DONN) in which a neural network, NN1, is trained using truth target (T) and falsity target (F). Two outputs are produced which are truth output (T1) and falsity output (F1). The classification result will be 1 if T1 is greater than F1, otherwise the classification result will be 0.

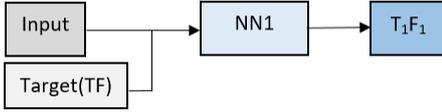


Fig. 1. Duo Output Neural Network (DONN)

### B. Cascade Generalization

Cascade generalization [20] is a technique that consists of several classifiers sequentially connected in which the next classifier is trained based on blending the original input features with the output predicted from the current classifier. Therefore, several levels of classifiers can be created. In our study, two levels named base level and meta level are implemented using a feedforward backpropagation neural network. Fig.2 describes the cascade generalization using neural network (CG-NN). In the base level, the original input features and the original truth target (T) are used to train the feedforward neural network (NN1). The result is the truth output T1. In the meta level, T1 and the original input features are used to train the feedforward neural network (NN2). The truth output T2 is then predicted. The classification result is 1 if T2 is greater than 0.5, otherwise the classification result is 0.

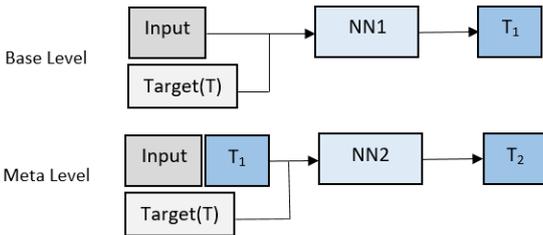


Fig. 2. Cascade Generalization using Neural Network (CG-NN)

## III. CASCADE GENERALIZATION USING DUO OUTPUT NEURAL NETWORK

Fig. 3 illustrates our proposed approach which is cascade generalization using duo output neural network (CG-DONN). In the base level, the duo output neural network (NN1) is trained to produce T1 and F1. The average of the truth output (T1) and the non-falsity output (1-F1) obtained from the base level is used in the meta level. In the meta level, the blending of the original input and the average result are used for training the feedforward neural network (NN2). The truth output (T2) and the falsity output (F2) are produced from the meta level. The classification result will be 1 if T2 is greater than F2, otherwise the classification result will be 0.

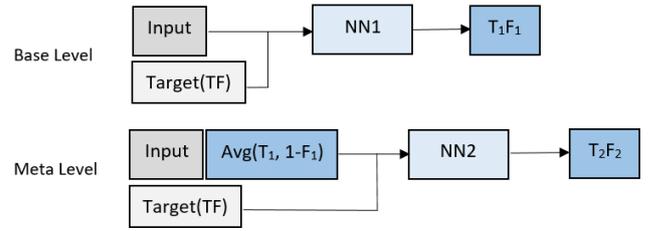


Fig. 3. Cascade Generalization using Duo Output Neural Network (CG-DONN)

## IV. EXPERIMENTAL METHODOLOGY AND RESULTS

### A. Dataset

Two cancer datasets which are breast cancer Coimbra [6] and cervical cancer behavior risk [7] from UCI machine learning repository [21] are tested on our proposed technique. Breast cancer Coimbra contains nine input features, which are anthropometric data gathered from blood analysis. These features are used to classify 116 people into two groups, which are patient and healthy people. Cervical cancer behavior risk consists of eighteen input features gathered from behavior, intention, attitude, norm, perception, motivation, social support, and empowerment. Seventy-two instances are divided into two classes, which are “has cervical cancer” and “no cervical cancer”.

### B. Experiment and Results

Four techniques shown in Table I are compared in this paper. The fourth technique is our proposed technique. Feedforward backpropagation neural network is applied to each technique. For all neural networks, the number of nodes in the hidden layer is set to double the number of input features. Other parameters are set as default except the initial weight which is set to be different. Each dataset is randomly divided into three folds so that three-fold cross validation is applied.

TABLE I. FOUR TECHNIQUES USED IN THIS PAPER (TECHNIQUE 4 IS THE PROPOSED APPROACH)

No.	Technique
1	Neural network (NN)
2	Cascade generalization using neural network (CG-NN)
3	Duo output neural network (DONN)
4	Cascade generalization using duo output neural network (CG-DONN)

For each round of three-fold cross validation, fifteen sets of machines are built so that we can see the average and the best results. In the first technique, fifteen feedforward backpropagation neural networks are trained in each round. In order to control the environment, these three sets of neural networks are used in the second technique.

In the second technique, all neural networks created in the first technique are used in the base level. In each round of meta level, fifteen neural networks are created using the blending of original input features and output predicted from each neural network obtained from the base level.

In the third technique, the environment is also controlled by using the same parameter values of the first technique except the number of targets and outputs which is set to two. Therefore, fifteen duo output feedforward backpropagation neural networks are created.

In the fourth technique, all duo output neural networks created in the third technique are used in the base level of cascade generalization. In each round of meta level, fifteen duo output neural networks are created based on the blending of original input features and the average of truth and non-falsity values obtained from each machine in the base level.

For each round of three-fold cross validation, the best and the average of fifteen results are computed. For breast cancer Coimbra dataset, Table II and Fig.4 show the best results whereas Table III and Fig.5 show the average results obtained from each fold of each technique.

TABLE II. THE BEST PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR BREAST CANCER COIMBRA DATASET

Fold	Technique (% correct)			
	NN	CG-NN	DONN	CG-DONN
1	53.85	51.28	82.05	82.05
2	46.15	56.41	74.36	82.05
3	39.47	39.47	81.58	76.32

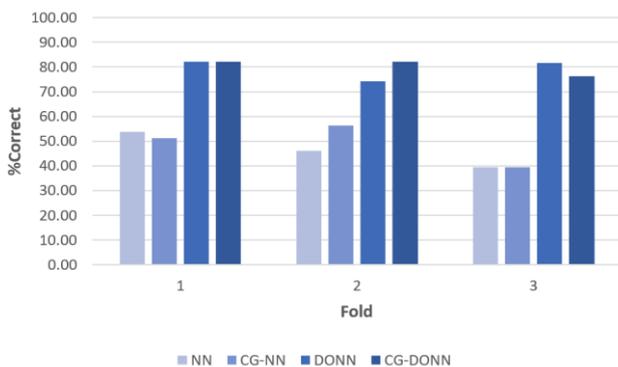


Fig. 4. The best result obtained from each fold of each technique for breast cancer Coimbra dataset

TABLE III. THE AVERAGE PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR BREAST CANCER COIMBRA DATASET

Fold	Technique (% correct)			
	NN	CG-NN	DONN	CG-DONN
1	49.05	48.90	65.82	68.21
2	46.15	47.18	63.08	73.67
3	39.47	39.47	67.03	72.29

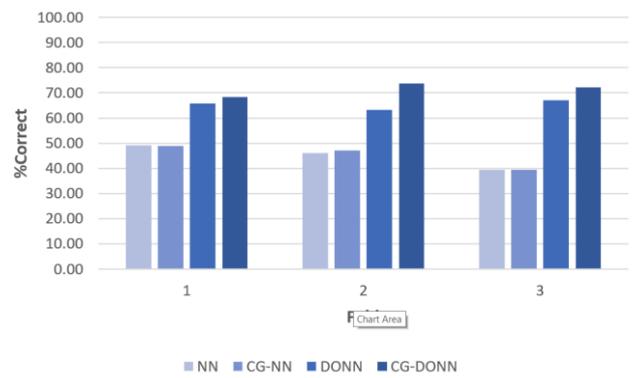


Fig. 5. The average result obtained from each fold of each technique for breast cancer Coimbra dataset

The average of best results for breast cancer Coimbra dataset are shown in Table IV and Fig.6 whereas the overall average are shown in Table V and Fig.7. It can be seen that CG-DONN can give the highest correct percentage.

TABLE IV. THE AVERAGE OF BEST PERCENT CORRECT VALUES FOR BREAST CANCER COIMBRA DATASET

Technique (% correct)			
NN	CG-NN	DONN	CG-DONN
46.49	49.06	79.33	80.14

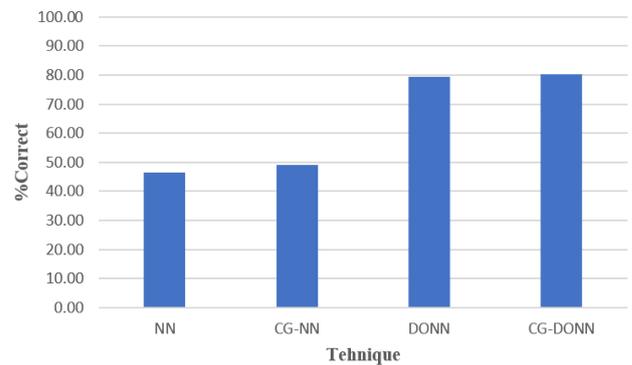


Fig. 6. The average of best results for breast cancer Coimbra dataset

TABLE V. THE OVERALL AVERAGE PERCENT CORRECT VALUES FOR BREAST CANCER COIMBRA DATASET

Technique (% correct)			
NN	CG-NN	DONN	CG-DONN
44.89	45.18	65.31	71.39

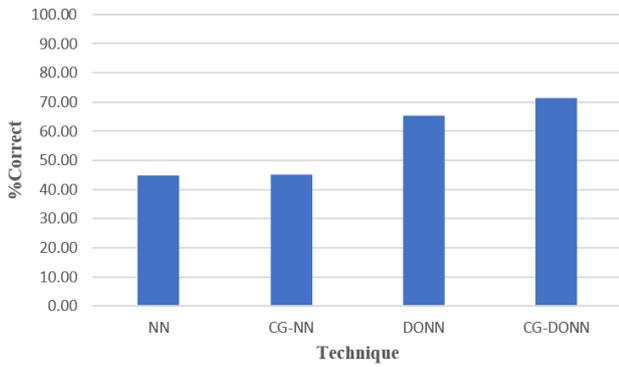


Fig. 7. The overall average for breast cancer Coimbra dataset

Table VI shows the percentage increase in the percent correct of the overall average values from the existing techniques to our proposed technique CG-DONN for breast cancer Coimbra dataset.

TABLE VI. THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR BREAST CANCER COIMBRA DATASET

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
59.02	57.99	9.31

For cervical cancer behavior risk dataset, the best and the average of fifteen results are computed. Table VII and Fig.8 show the best results whereas Table VIII and Fig.9 show the average results obtained from each fold of each technique.

TABLE VII. THE BEST PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	62.50	79.17	100.00	95.83
2	83.33	75.00	95.83	95.83
3	75.00	87.50	91.67	95.83

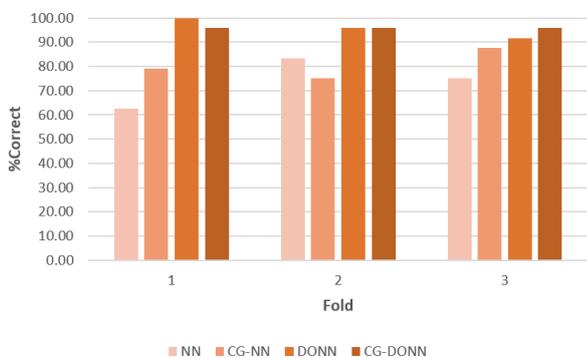


Fig. 8. The best result obtained from each fold of each technique for cervical cancer behavior risk dataset

TABLE VIII. THE AVERAGE PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	42.50	30.54	92.50	86.67
2	45.29	39.46	92.21	91.67
3	53.04	57.21	78.33	85.29

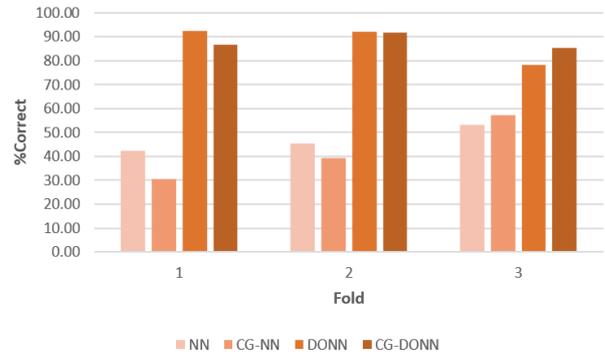


Fig. 9. The average result obtained from each fold of each technique for cervical cancer behavior risk dataset

The average of best results for cervical cancer behavior risk dataset are shown in Table IX and Fig.10 whereas the overall average are shown in Table X and Fig.11.

TABLE IX. THE AVERAGE OF BEST PERCENT CORRECT VALUES FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
73.61	80.56	95.83	95.83

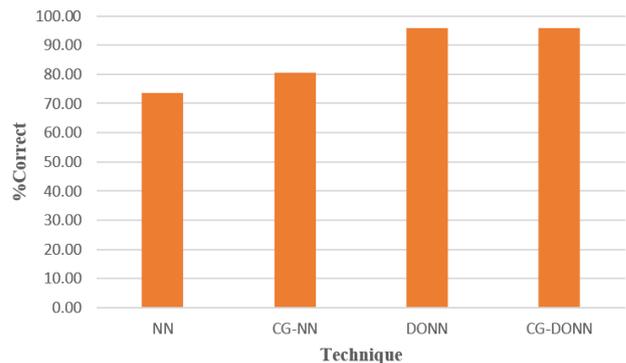


Fig. 10. The average of best results for cervical cancer behavior risk dataset

TABLE X. THE OVERALL AVERAGE PERCENT CORRECT VALUES FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
46.94	42.40	87.68	87.88

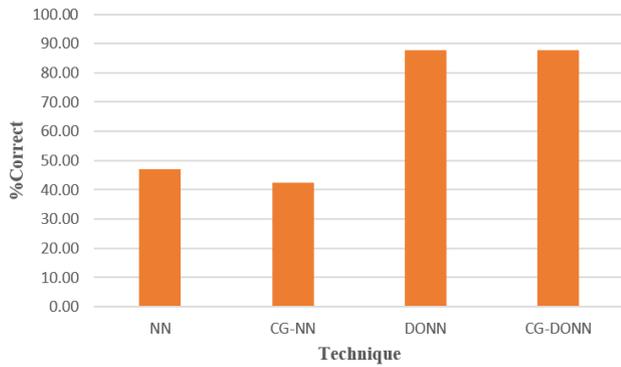


Fig. 11. The overall average for cervical cancer behavior risk dataset

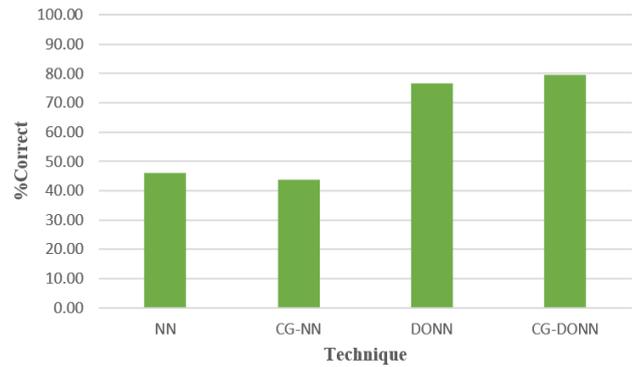


Fig. 12. The average results of both cancer datasets

Table XI shows the percentage increase in the percent correct of the overall average values from the existing techniques to our proposed technique CG-DONN for cervical cancer behavior risk dataset.

TABLE XI. THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
87.19	107.24	0.22

From Table II, it can be seen that the best accuracy of breast cancer Coimbra dataset obtained from NN, CG-NN, DONN, and CG-DONN are 53.85, 56.41, 82.05, and 82.05, respectively. DONN and CG-DONN give the same highest accuracy. However, the technique that gives the highest average of best results shown in Table IV is CG-DONN which is 80.14.

From Table VII, the best accuracy of cervical cancer behavior risk dataset obtained from NN, CG-NN, DONN, and CG-DONN are 83.33, 87.50, 100, and 95.83, respectively. It is found that DONN provides the highest accuracy. However, the highest average of best results shown in Table IX is 95.83 which belongs to both DONN and CG-DONN.

From the overall average percent correct shown in Table V and X, it can be noticed that CG-DONN provides the maximum overall average for breast cancer Coimbra and cervical cancer behavior risk datasets which are 71.39 and 87.88, respectively. It is also found that DONN provides better average results than CG-NN and NN.

From both datasets, the average percent correct of both datasets can be computed as shown in Table XII and Fig.12. Table XIII shows the percentage increase in the percent correct from the existing techniques to the proposed technique CG-DONN for both cancer datasets. It can be seen that CG-DONN provides the highest accuracy.

TABLE XII. THE AVERAGE PERCENT CORRECT VALUES OF BOTH DATASETS

Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
45.92	43.79	76.49	79.63

TABLE XIII. THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR BOTH CANCER DATASETS

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
73.42	81.83	4.10

V. CONCLUSION

CG-DONN is a combination of duo output neural network and cascade generalization. It is found that CG-DONN can improve learning accuracy of feedforward neural networks since it applies various kinds of learning data and utilizes multiple classifiers that are connected systematically. The proposed technique is compared to three existing individual techniques which are NN, CG-NN, and DONN. Two cancer datasets are experimented based on three-fold cross validation. Fifteen machines are run in each round of cross validation for each technique. It is found that CG-DONN is able to produce the average results better than DONN, CG-NN, and NN.

REFERENCES

- [1] T. E. Idriss, A. Idri, I. Abnane and Z. Bakkoury, "Predicting Blood Glucose using an LSTM Neural Network," 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), 2019, pp. 35-41.
- [2] H. Benbrahim, H. Hachimi and A. Amine, "Deep Convolutional Neural Network with TensorFlow and Keras to Classify Skin Cancer Images," Scalable Computing: Practice and Experience, vol. 21, no. 3, pp. 379-389, 2020.
- [3] Y. Dai, B. Xu, S. Yan and J. Xu, "Study of cardiac arrhythmia classification based on Convolutional Neural Network," Computer Science and Information Systems. vol. 17. no. 2, pp. 445-458, 2020.
- [4] A. Birsen, I. A. Aydin, S. Rızalar, H. Oz, D. Meral, "Breast and Cervical Cancer Knowledge and Awareness among University Students," Asian Pacific journal of cancer prevention: APJCP, vol. 16, pp. 1719-1724, 2015.
- [5] R. Agrawal, "Predictive Analysis Of Breast Cancer Using Machine Learning Techniques", ing. Solidar, vol. 15, no. 3, pp. 1-23, Sep. 2019.
- [6] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seça, F. Caramelo, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer," BMC Cancer, vol. 18, 2018.
- [7] Sobar, R. Machmud, A. Wijaya, "Behavior Determinant Based Cervical Cancer Early Detection with Machine Learning Algorithm," Advanced Science Letters, vol. 22, pp. 3120-3123, 2016.

- [8] M. F. Aslan, Y. Celik, K. Sabanci, and A. Durdu, "Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data", *IJISAE*, vol. 6, no. 4, pp. 289-293, Dec. 2018.
- [9] Y. Austria, M. Goh, L. Jr, J. Lalata, J. Goh, H. Vicente, "Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset," *International journal of simulation: systems, science & technology*, vol 20, 2019.
- [10] G. Ullah, HaiYan, "Comparative performance analysis of machine learning models for breast cancer diagnosis," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 10, no. 1, 2020.
- [11] Naveen, R. K. Sharma, A. Ramachandran Nair, "Efficient Breast Cancer Prediction Using Ensemble Machine Learning Models," In *Proceedings of the 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, 2019, pp. 100-104.
- [12] S. Poorani, P. Balasubramanie, "Deep Neural Network Classifier in Breast Cancer Prediction," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, pp. 2106-2109, 2019.
- [13] A. Karaci, "Predicting Breast Cancer with Deep Neural Networks," In: Hemanth D., Kose U. (eds) *Artificial Intelligence and Applied Mathematics in Engineering Problems. ICAIAME 2019. Lecture Notes on Data Engineering and Communications Technologies*, vol. 43, pp. 996-1003, 2020.
- [14] E., J. Kusuma, G. F. Shidik, R. A. Pramunendar, "Optimization of Neural Network using Nelder Mead in Breast Cancer Classification," *International Journal of Intelligent Engineering & Systems*, vol. 13 , no. 6, pp. 330-337, 2020.
- [15] K. Polat and U. Sentürk, "A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018, pp. 1-4.
- [16] E. Yavuz and C. Eyupoglu, "An effective approach for breast cancer diagnosis based on routine blood analysis features," *Medical & Biological Engineering & Computing*, vol. 58, pp. 1583–1601, 2020.
- [17] H. -J. Chiu, T. -H. S. Li and P. -H. Kuo, "Breast Cancer–Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine," in *IEEE Access*, vol. 8, pp. 204309-204324, 2020.
- [18] N. F. Idris and M. A. Ismail, "Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method: automatic fuzzy database definition," *PeerJ Computer Science*, vol. 7, e427, 2021.
- [19] P. Kraipeerapun, S. Amornsamankul, C.C. Fung, S. Nakkrasae, "Applying Duo Output Neural Networks to Solve Single Output Regression Problem," In: Leung C.S., Lee M., Chan J.H. (eds) *Neural Information Processing. ICONIP 2009. Lecture Notes in Computer Science*, vol. 5863, pp. 554-561, 2009.
- [20] J. Gama, P. Brazdil, "Cascade Generalization," *Machine Learning*, vol. 41, no. 3, pp. 315–343, 2000.
- [21] D. Dua, C. Graff, *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, 2019.