A Multi-Classifier Method based Deep Learning Approach for Breast Cancer

Mokhairi Makhtar, Rosaida Rosly, Mohd Khalid Awang, Mumtazimah Mohamad, Aznida Hayati Zakaria

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

ABSTRACT

Breast cancer is one of the diseases that haunt every woman around the world. It is one of the main killers of women not only in Malaysia, even around the world. Medical diagnosis such as breast cancer is considered a significant but complicated task that needs to be carried out correctly and effectively. In improving the prediction accuracy of breast cancer dataset, this study evaluates the performance of multi-classifier based deep learning approach on datasets. There are five classifiers that are involved like Sequential Minimal Optimization (SMO) , decision tree (J48) , random forests (RFs), Naïve Bayes (NB) and Instance Based for K-Nearest neighbor (IBk). These classifiers will be combined and analyzed using deep learning approach. This strategy utilizes models of deep neural network that is a variant of Neural Network but with big approximation to human brain using an advance system compared to a straightforward neural network. The results of combination different classifiers using deep learning approach indicate the highest accuracy than single classification with 96.63% as a combination SMO+RF+IBK+NB.

Keywords: *Breast Cancer Wisconsin, Deep learning, Multi-classifier methods.*

I. INTRODUCTION

Breast cancer is one of Malaysia's leading murderers of females. According to [1], breast cancer is the second most prevalent cancer in the globe, with an estimated 1.67 million fresh instances of cases diagnosed in 2012 (25% of all cancers). Many women (and men) are still taking care about breast cancer and think it only attacks women who are over 50 years of age. Nevertheless, it was able to attack women as young as 35 years old. According to the study, at some stage in her career, one out of eight females in the globe is at danger of breast cancer [2].

Timely detection and efficient therapy is one of the methods for reducing the mortality rate of breast cancer [2]. As a result, a more precise classification of a tumor with breast cancer has become a difficult medical issue. Today, the issue of classification of breast cancer is widely used by expert systems and machine learning methods. They provide precision and efficient diagnostic capacities for elevated classification. They provide high classification accuracy and effective diagnostic capabilities. Currently, the most used techniques to detect breast cancer in early stages are: mammography (63% to 97% correctness [3]), FNA (Fine Needle Aspiration) with visual interpretation (65% to 98% correctness [4]) and surgical biopsy (approximately 100% correctness). Therefore, mammography and FNA with visual accuracy and surgical biopsy differ extensively, although reliable is invasive and costly.

In this paper, the implementation of multi-classifier based on deep learning approach is proposed. Multi-classifier method is aggregation of predictions of multiple classifiers with the goal of improving accuracy. It is better than single classification accuracy. Previous research has shown that combining distinct classification projections can be an efficient strategy for improving classification efficiency [5]. In [6], the fusion of different classifiers to classify the breast cancer dataset achieved the highest accuracy with percentage of 93.98%. Multi-classifier technique contributes to better precision compared to a single model of classification or regression [7]. In [8], they demonstrated that ensemble strategy is better than a single classifier to predict the chemical compound poisonous class. Three performance measures have been introduced, such as Accuracy, False Negative Rate and False Positive Rate, which producing better prediction models. Using multi-classifier approaches can achieve better accuracy than the single ones.

Deep learning is one of sub-field in machine learning. Deep learning now has a huge stake in every sector and particularly in medical image analysis and is anticipated to hold $300 million in the market for medical imagery by 2021[9]. Thus, it alone will receive more medical imaging investment by 2021 than the entire analytics sector spent in 2016 [9]. It is the most efficient and controlled approach to machine learning. This method utilizes deep neural network models that vary from neural network to human brain using advance mechanism compared to simple neural network.

The rest of this paper is arranged as follows: Section II discusses the multi-classifier method, followed by deep learning explained in Section III and Deep Multi-classifier Learning (DCML) method is shown in section IV. The experimental results are shown in section V and lastly, Section VI concludes this paper

II. MULTI-CLASSIFIER METHOD

The primary aim of multi-classifiers is to combine a set of models, each of which solves the same initial job, so that a outcome can be achieved with more precise and reliable estimates or choices than a single model can achieve [10]. By incorporating various models to produce a final forecast, it creates a predictive model and the more we include the better it performs. The development of various base classifier models and their combination are two primary stages of the ensemble or multi-classifier [11]. Multi-classification techniques are well known to be used to improve predictive efficiency. In [12], classification tasks are conducted on seven publicly accessible cancerous microarray information and the classification/prediction efficiency of these techniques is compared. The outcome was that ensemble models (bagged and boosted decision trees) in this classification assignment often perform better than single decision trees. The output of the ensemble classifiers is better than any other ensemble techniques or single classifiers using Soft Set ensembles selection and optimization [13]. In multi-classifiers fusion classification task, multiple classifiers will be combining in order to get the best accuracy [14].

III. DEEP LEARNING

The use of artificial intelligence in the diagnosis of medical diseases is now a new trend and has many applications [15]. Classification of medical data is a complicated issue of optimization and it also needs to provide diagnosis aid accurately. Deep learning is a machine learning subfield that focuses on algorithms inspired by the brain's structure and functions called Artificial Neural Networks. Artificial Neural Networks (ANN) [12,11] have been used to get high classification accuracy rate. ANN is an artificial human brain representation that attempts to simulate its learning process. It is an interconnected group of artificial neurons that utilizes a mathematical model or computational model to process data based on connectionist computational strategy [17]. Deep learning is the development of ANN that has more layers. With more layers, it can recognize the process with more complexes.

The word deep learning involves the use of a model of deep neural network [9]. The basic computational unit in a neural network is the neuron, a concept inspired by the study of the human brain that takes multiple signals as inputs, combines them linearly using weights and then transmits the combined signals to generate output signals through nonlinear operations [9]. It is a machine learning method that learns (features and tasks) directly from data. This data can include images, text or sounds. The features of deep learning are covers the main architecture (fully connected, convolutional, and recurrent), flexible and fast prototyping and compile to run on GPU. The approach that commonly applied to implement deep learning is graphical method (multilayer representation) and graphical model (belief network, neural network, hidden Markov).

The advantages of deep learning are best-in-class performance, scales effectively with data, no need for feature engineering and adaptable and transferable[18]. In addition, the characteristics can be learned automatically through deep learning using a general-purpose procedure. Usually this operation is performed by a multilayer stack of simple neural networks with nonlinear input−output mappings, involving deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent or recursive neural networks (RNNs) and other deep networks with more than one hidden layer and more neurons in each layer[19].

IV. DEEP MULTI-CLASSIFIERS LEARNING METHOD (DMCL)

A. Data-Pre-processing

The classification of data set starts with pre-processing data. The pre-processing will concentrate on managing missing features, unbalanced data and the amount of characteristics used to train the classifier, taking into account the information set taken. The data will be ready and filtered to enhance data quality and at the same time, clear the noise of data. It began with a focus on managing missing values, discretizing numerical attributes, and selecting subsets of characteristics. In handling the missing values, the removal of instances with missing values was applied. In the first step of pre-processing, all instances with missing values were removed.

These data sets contain integer attributes, so this format cannot be supported by Weka instruments. Thus, in the discretization task, the change of integer attributes to numeric attributes was applied.

B. Feature Selection

The next step, a feature selection process, is implemented on the Breast Cancer Wisconsin dataset. It showed how many attributes classifiers were trained. It tends to increase the classification accuracy by

eliminating noise features while irrelevant features are ignored. The advantages of using feature selection are firstly, it allows the machine learning algorithm to train more quickly, secondly, it decreases the complexity of a model and makes it simpler to interpret, thirdly, it increases the accurcay of a model if the correct subset is selected and finally it decreases overfitting [20]. By using the classification approach, the different classifiers are used.

C. Single Classification

Firstly, classification is performed using single base classifier. The classifier model is built using 10-fold cross validation method. It will divide into 10 folds. It is a common technique because it is easy to comprehend and usually outcomes in a less biased or less optimistic estimation of model skill than other techniques, such as a simple train/test split [21,25]. Then, the training data is load into Weka. Using supplied test set through Weka, the testing data will be loaded which it predicts the class of a set of instances and the result of accuracy prediction for each folds are achieved. Then, this model is used for measurement for the next stage.

D. Deep Multi-Classifier Method (DMCL)

In this stage, we use a deep neural network approach to combine the output of various classifiers. In our ensemble combination classification task, we combine the prediction class result from the highest 2 classifiers in accuracy and then the 3rd, then 4th and so on until the accuracy decreases then stops.

In deep learning program, classifiers act as attributes class or input nodes. It will be started with combining two input nodes (classifiers) which got the highest accuracy on single classification, followed by choosing the combination that achieves the highest accuracy and then combine with other input nodes. From two input nodes, three input nodes are derived and so on. The input nodes in this stage refer to the number of classifiers used in combination process. Repeat the same process until the latest level of combination and pick the highest accuracy. Then, the combination of input nodes will be analyzed using deep neural network program. Shown in Fig.I is a description for deep multi-classifier learning (DMCL).

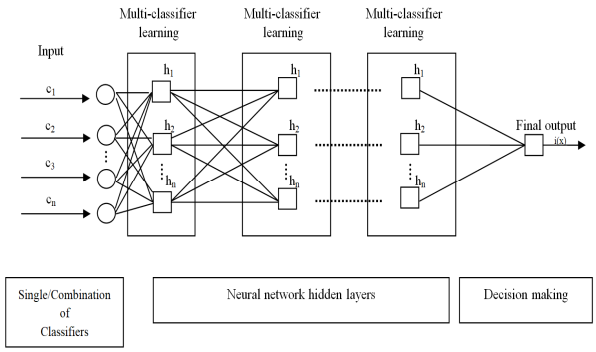

Fig.I. Description for DMCL

To classify the classes, we are using a neural network of four layers. The network's input layer includes five neurons/nodes representing the sample attributes (classifiers) in the new data set. In the hidden layers, we are experimenting for better classification results with different node numbers in each layer. The network output layer includes one neuron, the output being 0 or 1.

The step in implement deep learning neural network in Keras (framework for deep learning) are as below:

i. Load data

- In this research, we are going to use Breast Cancer Wisconsin and Hepatitis data set. This is a standard machine learning dataset from the UCI Machine Learning repository. It describes patient medical record data for Breast Cancer Wisconsin and Hepatitis.
- As such, it is a binary classification problem (onset of Breast Cancer Wisconsin as 1 or not as 0). The input variables that describe each patient are numerical and nominal. The class input (benign, malignant) is in nominal then converted to numerical input to makes it easy to use directly with neural networks that expect numerical input and output values.

ii. Define model

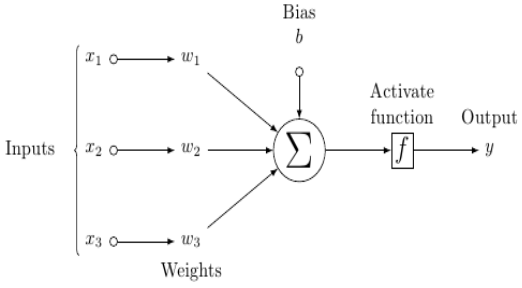

Fig.II. Neural Network Diagram

- Fig. II shown the example of simple neural network diagram. Models in Keras are described as a sequence of layers. The model is developed sequentially and layers are added one at a time. The first thing to get correct is to guarantee the input layer has the right number of inputs. This can be indicated with the input dim argument when generating the first layer and setting it to 2 for the 2 input variables as an example. We can use heuristics and often find the best network structure through a trial and error experimentation method. In this research, we will use a fully-connected network structure with four layers.
- Using the Dense class, Fully connected layers are described. As the first argument, we specify the amount of neurons in the layer, the initialization method as the second argument as init and specify the activation function using the activation argument. Init is the initialization of Stochastic Gradient Decent. In Neural Network we assign weights to each mode which is nothing but importance of that node. At the time of initialization, weights should be close to 0 and we will randomly initialize weights using uniform function.
- In this case, we initialize network weights to a small random number generated from a uniform distribution ('uniform'), in this case between 0 and 0.05 because in Keras this is the default uniform weight initialization. Another traditional option for small random numbers produced from a Gaussian distribution would be 'normal'.
- In our hidden layer we use rectifier (relu) activation function and tanh activation function and in our output layer we use Sigmoid activation function as we want binary result from output layer. It used to be the case that sigmoid and tanh activation functions were preferred for all layers. These days, better performance is achieved using the relu activation function[22]. We use a sigmoid on the output layer to guarantee that our network output is between 0 and 1 and can be easily mapped to either a probability of class 1 or snap to a hard classification of either class with a default threshold of 0.5. We piece it all together by adding each layer.

Neuron applies activation function to weighted sum (summation of Wi * Xi where w is weight, X is input variable and i is suffix of W and X). The closer the activation function value to 1 the more activated is the neuron and the more neuron passes the signal. Which activation function should be used is critical task. Below are descriptions of deep neural network architecture layer for this research

Table I Description of deep neural network architecture layer

Num. of input nodes (classifiers)	Layer	Neurons /nodes	Activation function
2	1st layer	8	Relu
	2ndlayer	6	Relu
	3rd layer	4	Tanh
	4th layer	1	Sigmoid
3	1st layer	8	Relu
	2ndlayer	6	Relu
	3rd layer	4	Tanh
	4th layer	1	Sigmoid
4	1st layer	8	Relu
	2ndlayer	6	Relu
	3rd layer	4	Tanh
	4th layer	1	Sigmoid

Table I shown the details number of neuron/nodes and activation function that are used for each layer. This layer uses three kinds of activation function: rectified linear unit (relu), tanh and sigmoid.

iii. Compile model

- Compiling the model uses the efficient numerical libraries under the covers (the so-called backend) TensorFlow. The backend automatically selects the best way to train the network and predict how to operate on your hardware, such as CPU or GPU or, even distributed.
- When compiling, certain additional properties are specified such as specifying the loss function to be used to evaluate a set of weights, the optimizer used to search the network by different weights and the optional metrics that we would like to collect and report during training to find the best weights to predict this problem.
- In this case, first argument is Optimizer to find optimal set of weights which will optimize weights in turn making out neural network more powerful. This algorithm is Stochastic Gradient descent (SGD). Among several types of SGD algorithm the one which we used is **'**Adam**'**. SGD depends on loss thus our second parameter is loss. Since out dependent variable is binary, we will have to use logarithmic loss function called 'binary crossentropy'. We want to improve performance of our neural network based on accuracy so add metrics as accuracy.

iv. Fit model

- Now it is time to execute the model on some data. The model and loaded data are train using

the fit () function on the model.

- Batch size is used to specify the number of observation after which you want to update weight. Epoch is nothing but the total number of iterations. Choosing the value of batch size and epoch is trial and error there is no specific rule for that. Then the test set result is predicting. The prediction result will give the probability of class distribution. The probability will be converted into binary 0 and 1.

v. Evaluate model

- Using the evaluate () function on the model, we evaluate the training dataset model and pass it on to the same input and output used for model training. This will produce a prediction and collect scores for each input and output pair, including the average loss and accuracy.
- Apart from calculate the accuracy, some basic classification evaluation techniques are done in this research such as:
  - The confusion matrix, which breakdown of predictions into a table showing correct predictions and the types of incorrect predictions made. Ideally, you will only see numbers in the diagonal, which mean that all your predictions were correct.
  - Precision is a measure of the accuracy of a classifier. The higher the precision, the more accurate the classifier.
  - Recall or known as sensitivity (true positive rate) is a measure of a classifier's completeness. The higher the recall, the more cases the classifier covers.

The F1 Score or F-score is a weighted average of precision and recall.

V. EXPERIMENTAL RESULT

The dataset is based on the Breast Cancer Wisconsin taken from the UCI Repository of Machine Learning Databases. It comprises of 699 instances. There are two classes which are benign and malignant where 458 instances of benign and 241 instances of malignant. The removal of instances with missing values was implemented in managing the missing values. From 699 instances, only 683 instances are involved. Then, the attribute selection is applied on them. From 10 attributes, only 9 attributes are selected which are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, and Mitoses. The classifier model is built using 10-fold cross validation.

For dataset, five different classifiers such as Sequential Minimal Optimization (SMO) , decision tree (J48) , random forests (RFs), Naïve Bayes (NB) and Instance Based for K-Nearest neighbor (IBk) are chosen . To evaluate the proposed model, the experiments were performed.

Table II Accuracy for single classification

Classifiers	Accuracy (%)	Error rate (%)
SMO	96.05	3.95
J48	93.86	6.14
RF	95.77	4.23
NB	95.61	4.39
IBK	94.89	5.11

The percentage of single classifier accuracy and error rate are noted for this dataset with different classifiers. We create our heterogeneous multi-classifiers by selecting five different classifiers which are listed in Table II. SMO was found as first accuracy with the 96.05% followed by RF, NB, IBK and J48.

Table III Deep Multi-Classifier Learning Prediction for Breast Cancer Wisconsin

Input nodes	Classifiers	Precision	Recall	F1_Score
2	SMO+RF	0.93	0.97	0.95
	SMO+NB	0.93	0.95	0.94
	SMO+IBK	0.92	0.97	0.95
	SMO+J48	0.93	0.95	0.94
3	SMO+RF+IBK	0.93	0.97	0.95
	SMO+RF+NB	0.93	0.97	0.95
	SMO+RF+J48	0.93	0.97	0.95
4	**SMO+RF+IBK+NB**	**0.93**	**0.97**	**0.95**
	SMO+RF+IBK+J48	0.93	0.97	0.95

Table IV Accuracy Result for Deep Multi-Classifier Learning Prediction in Breast Cancer Wisconsin

Input nodes	Classifiers	Accuracy (%)	Error Rate (%)
2	SMO+RF	96.49	3.51
	SMO+NB	96.05	3.95
	SMO+IBK	96.34	3.66
	SMO+J48	96.05	3.95
3	SMO+RF+IBK	96.49	3.51
	SMO+RF+NB	96.49	3.51
	SMO+RF+J48	96.49	3.51
4	**SMO+RF+IBK+NB**	**96.63**	**3.37**
	SMO+RF+IBK+J48	96.49	3.51

The obtained results are shown in Table III and Table IV. It shows the result of combination two input nodes, three input nodes and four input nodes that are using deep learning approach. In addition to accuracy, the dataset also can evaluate using the area under the curve (AUC) which is a performance metric for binary classification issues [23]. The AUC is the capacity of a

model to discriminate between positive and negative classes. An area of 1.0 reflects a model that perfectly produced all predictions. An area of 0.5 is as good as a random model [23]. The Precision-Recall curve is using due to imbalances data we used. The area under the PR curve used to measure the relationship between a classifier's accuracy, recall, and efficiency [24]. From the test we indicated that combination of SMO+RF+IBK+NB achieved the highest percentage with 96.63%.

ACKNOWLEDGEMENT

This research was supported by Research Management and Innovation Centre, Universiti Sultan Zainal Abidin.

REFERENCES

[1] W. H. Organization, "*GLOBOCAN 2012:Estimated cancer incidence, mortality and prevalence worldwide in 2012*," International Agency for Research on Cancer, 2012. Available: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. [Accessed: 03-Aug-2018].

[2] L. Dora, S. Agrawal, R. Panda, and A. Abraham, "*Optimal breast cancer classification using Gauss–Newton representation based algorithm*," Expert Systems with Applications, vol. 85, pp. 134–145, 2017.

[3] J. G. Elmore, K. Armstrong, C. D. Lehman, and S. W. Fletcher, "*Screening for breast cancer*," Journal of the American Medical Association, vol. 293, no. 10, pp. 1245–1256, 2005.

[4] R. W. M. Giard and J. Hermans, "*The value of aspiration cytologic examination of the breast a statistical review of the medical literature*," Cancer, vol. 69, no. 8, pp. 2104–2110, 1992.

[5] W. Tao, L. Weihua, S. Haobin, and L. Zun, "*Software defect prediction based on classifiers ensemble,"* Journal of Information & Computational Science 8, vol. 16, pp. 4241–4254, 2011.

[6] R. Rosly, M. Makhtar, M. K. Awang, M. N. A Rahman, and M. M. Deris, "*Multi-classifier models to improve accuracy of water quality application*," in Proceedings of the International Conference on Telecommunication, Electronic and Computer Engineering (ICTEC '15), 2015.

[7] I. Partalas, G. Tsoumakas, E. V Hatzikos, and I. Vlahavas, "*Greedy regression ensemble selection : Theory and an application to water quality prediction*," Information Sciences Journal, vol. 178, pp. 3867–3879, 2008.

[8] M. Makhtar, Y. Longzhi, D. Neagu, and M. Ridley, "*Optimisation of classifier ensemble for predictive toxicology applications*," Computer Modelling and Simulation (UKSim), 2012. Available: http://www.academia.edu/4750498/Optimisation_of_Classifier_Ensemble_for_Predictive_Toxicology_Applications. [Accessed: 15-Apr-2015].

[9] M. I. Razzak, S. Naz, and A. Zaib, "*Deep learning for medical image processing: Overview, challenges and future*," CoRR, vol. 1704.06825, pp. 1–30, 2017.

[10] P. K. Srimani and M. S. Koti, "*Medical diagnosis using ensemble classifiers - A novel machine-learning approach,"* Journal of Advanced Computing, pp. 9–27, 2013.

[11] M. K. Awang, M. Makhtar, M. N. A. Rahman, and M. M. Deris, "*A new soft set-based pruning algorithm for ensemble method,"* Journal of Theoretical and Applied Information Technology, vol. 88, no. 3, pp. 384–391, 2016.

[12] A. C. Tan and D. Gilbert, "*Ensemble machine learning on gene expression data for cancer classification*," Applied bioinformatics, vol. 2, pp. 1–10, 2003.

[13] M. Awang, M. Makhtar, and M. Rahman, "*Ensemble Selection and Optimization Based on Soft Set Theory for Customer Churn Classification*," Journal of Theoretical, no. October, 2016.

[14] R. Rosly, M. Makhtar, M. K. Awang, M. N. A. Rahman, and M. M. Deris, "*Multi-classifier models to improve accuracy of water quality application*," ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 5, p. 3209, 2016.

[15] E. Alickovic and A. Subasi, "*Data mining techniques for medical data classification*," The International Arab Conference on Information Technology (ACIT) 2011, no. May, 2011.

[16] R. Setiono and H. Liu, "*Symbolic representation of neural networks*," Computer, vol. 29, no. 3. pp. 71–77, 1996.

[17] L. da Fontoura Costa and G. Travieso, "*Fundamentals of neural networks*," Neurocomputing, vol. 10, no. 2, pp. 205–207, 1996.

[18] G. Seif, "*Deep learning vs classical machine learning – Towards data science*," 2018. Available: https://towardsdatascience.com/deep-learning-vs-classical-machine-learning-9a42c6d48aa. [Accessed: 12-Aug-2018].

[19] Y. Xu, Z. Dai, F. Chen, S. Gao, J. Pei, and L. Lai, "*Deep learning for drug-induced liver injury*," Journal of Chemical Information and Modeling, vol. 55, no. 10, pp. 2085–2093, 2015.

[20] G. Chandrashekar and F. Sahin, "*A survey on feature selection methods*," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[21] W. Wu, R. J. May, H. R. Maier, and G. C. Dandy, "*A benchmarking approach for comparing data splitting methods for modeling water resources parameters using artificial neural networks*," Water Resources Research, vol. 49, no. October, pp. 7598–7614, 2013.

[22] G. Zhao, Z. Zhang, P. Tang, and J. Wang, "*Rethinking ReLU to Train Better CNNs*," 2018.

[23] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "*Evaluation Measures for Models Assessment over Imbalanced Data Sets*," Journal of Information Engineering and Applications, vol. 3, no. 10, pp. 27–39, 2013.

[24] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "*A deep learning-based multi-model ensemble method for cancer prediction*," Computer Methods and Programs in Biomedicine, vol. 153, pp. 1–9, 2017

[25] Naveed, Q. N., Qureshi, M. R. N. M., Shaikh, A., Alsayed, A. O., Sanober, S., & Mohiuddin, K. (2019). "*Evaluating and ranking cloud-based e-learning critical success factors (CSFs) using combinatorial approach*"**.** *IEEE Access*, *7*, 157145-157157.