Deep Multi-Classifier Learning for Medical Data Sets

Rosaida Rosly, Mokhairi Makhtar, Mohd Khalid Awang, Hasni Hassan, Ahmad Nazari Mohd Rose

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

ABSTRACT

This paper presents a comparison among the different classifier such as Sequential Minimal Optimization (SMO) , decision tree (J48) , random forests (RFs), Naïve Bayes (NB) and Instance Based for K-Nearest neighbor (IBK) on medical data sets such as Breast Cancer Wisconsin and Hepatitis. Classification accuracy was used in this research based on 10-fold cross validation method. Then, a combination at classification level between these classifiers using deep learning approach was applied to get the highest accuracy and see which the most suitable Deep Multi-classifier Learning (DMCL) approach for the data sets. These medical data sets were taken from the UCI Repository. The results showed that the combination SMO+RF+IBK+NB achieved the highest accuracy for Breast Cancer Wisconsin data set with percentage 96.63%. While for Hepatitis data set, the combination IBk+NB+J48+SMO achieved the highest percentage with 92.50 %. It showed that the proposed method are able to produce the highest prediction accuracy than single and combination of classifier that using majority voting for all these medical data sets.

Keywords : *Classification model, Deep learning, Majority voting, Multi-classifier.*

I. INTRODUCTION

Breast cancer is one of Malaysia's leading murderers of females. According to[1], breast cancer is the second most general cancer in the humanity and, to date, the most common cancer of women with an estimated 1.67 million fresh cancer cases diagnosed in 2012 (25% of all cancers). As per the analysis, one out of eight women in the humanity are at possibility of breast cancer at some end of moment in their life [2].

One of the way to reduce breast cancer mortality rate is timely recognition and effective heal [2]. Due to that, more precise classification of a breast cancer tumor has become a complex problem in the medical application. Nowadays, expert systems and machine learning methods are being widely used in the breast cancer classification case. They offer high classification accuracy and good 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 interpretation correct- ness vary widely, and the surgical biopsy, although reliable, is invasive and high cost.

Besides of breast cancer diseases, hepatitis also is an important public health problem in Malaysia. The word hepatitis comes from the Ancient Greek word hepar (root word hepat) meaning "liver", and the Latin it is meaning inflammation. It means injury to the liver with inflammation of the liver cells. The liver is the largest gland in the human body and the largest internal organ (the largest organ is the skin). It weighs approximately 3 lb (1.36 kg), reddish brown color and separated into four lobes of different sizes and lengths [5]. The three most common causes for hepatitis in Malaysia are hepatitis A, B, and C. Hepatitis A has been a reportable disease in Malaysia since 1988 [6]. About one million people chronically infected with hepatitis B, and 453,700 people were living with Hepatitis C infection in Malaysia in 2009 [6]. The diagnosis of some diseases like hepatitis is very difficult task for a doctor. Due of this cases, it motivates us for suggesting new methods to improve the accuracy of diagnosis and classifier performance, as well as to help doctors and specialists to diagnose hepatitis disease which they are usually determine decision by comparing the current test results of patients with another one who has the same condition.

Nowadays, usage of artificial intelligence in medical disease diagnosis is a new trend and with significant number of applications[7]. Medical data classification is a kind of complex optimization problem and it also needs to provide diagnosis aid accurately. Deep learning is a subfield of machine learning concerned with algorithms inspired by the structure and functions of the brain called Artificial Neural Networks. Artificial Neural Networks (ANN) [12,11] have been used to get high classification accuracy rate. ANN is an artificial representation of the human brain that tries to simulate its learning process. It is an interconnected group of artificial neurons that uses a mathematical model or

computational model for information processing based on connectionist approach to computation [9]. Deep learning is the development of ANN that has more layers. With more layers, it can recognize the process with more complex.

Deep learning means the employ of a deep neural network model [10]. The vital computational unit in a neural network is the neuron. It is a theory stimulated by the study of the human brain. It takes multiple signals as inputs, combines them linearly using weights, and then passes the combined signals through nonlinear operations to produce output signals [10]. 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 are graphical method (multilayer representation)[11] and graphical model (belief network, neural network, hidden markov) [12].

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. Previous researches have showed that combining the predictions from different classifiers can be an effective strategy to improve classification performance[13]. In [14], the fusion of different classifiers to classify the breast cancer dataset achieved the highest accuracy with percentage of 93.98%. Multi-classifier method leads to improved accuracy compared to a single classification or regression model [15].In [16], they proved that multi-classifier approach is better than single classifier for predicting the toxic class of chemical compounds. They applied three performance measures such as Accuracy, False Negative Rate and False Positive Rate which produce better quality prediction models. Using multi-classifier approaches can achieve better accuracy than the single ones.

The structure of this paper is organized as follow: Section II explain about the multi-classifier method, followed by deep multi-classifier learning explained in Section III, Experiments in Section IV. Lastly, Section V concludes this paper.

II. MULTI-CLASSIFIER METHOD

Multi-classifier methodology is used to build a predictive model by integrating multiple models. It is well-known that multi-classifier methods can be used for improving performance. Fig.1 depicts the diagram flow of multi-classifier system.

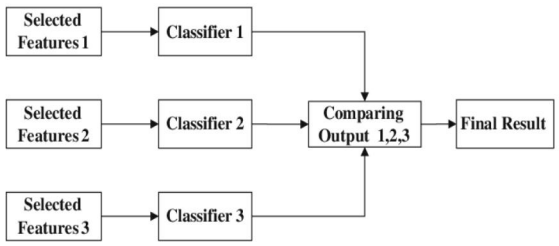

Fig.1. Block Diagram of Multi-classifier System [17]

Researchers from different disciplines such as statistics and artificial intelligence have considered the use of multi-classifier methodology [18]. Multi-classifier methods for supervised machine learning have become popular due to their ability to accurately predict class labels of simple and lightweight "base learners" groups. Researchers from various disciplines such as statistics, pattern recognition and machine learning have seriously explored the use of multi-classifier methodology [19]. Multi-classifier method leads to improved accuracy compared to a single classification or regression model [15]. Tao, Weihua, Haobin, & Zun (2011) also stated that classifiers multi-classifier can effectively improve classification performance than a single classifier. The implementation of multi-classifier mapping techniques showed higher accuracy than any single model, where the yields of numerous models are combined [20]. Multi-classifier models use a combination of several hypotheses, which tend to cancel out overfitting errors [21]. In [22], multi-classifier classifiers were always found to outperform single decision tree classifier in having greater accuracies and smaller predicting errors when applied to a pancreatic cancer proteomic dataset. Other applications of multi-classifier are used in data quality assessment sensor, shellfish farm closure prediction and cause identification, handwriting recognition, benthic habitat mapping, dealing with missing sensor data and algae growth prediction [23,24,25].

III. DEEP MULTI-CLASSIFIER LEARNING (DMCL)

The proposed deep multi-classifier learning model (DMCL) is presented by Fig. 2. There are four processes that involved in this model such as data pre-processing, feature selection, classification by single ones, and combination of classifier based deep learning approach.

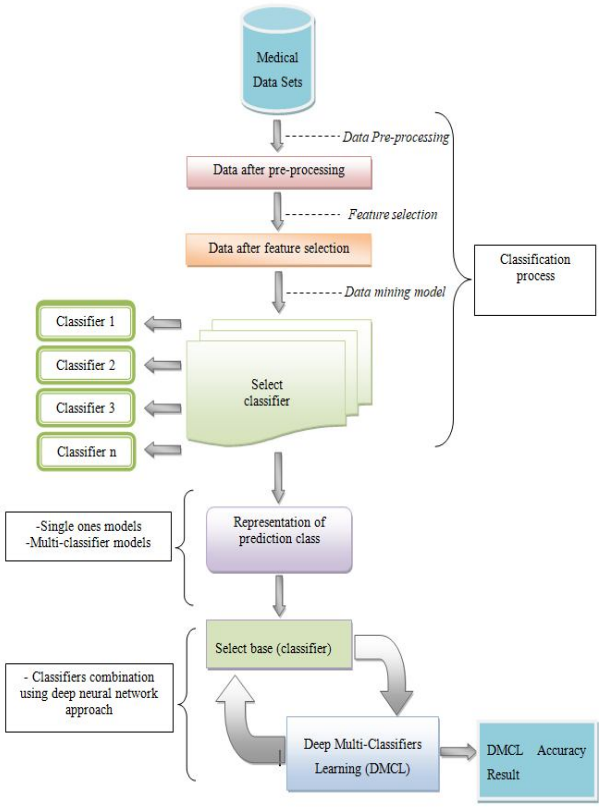

Fig.2.The Proposed of DMCL Models

A. Data-Pre-processing

The classification of data set starts with pre-processing data. Considering the dataset adopted, the pre-processing will focus on manage the missing attributes, the unbalanced data and the number of features used to train the classifier. The data will be prepared and filtered to improve the quality of data and at the same time, clear the noise of data. It started with handling of missing values, discretization of numeric attributes and selection of attribute subsets. In handling the missing values, the instances with missing values will be deleting. In the first step of pre-processing, all instances with missing values were deleted.

These data sets hold integer attributes, so Weka tools cannot support this format. Thus, the shifting of integer attributes to numeric attributes was applied in the discretization task.

B. Feature Selection

The next step, a feature selection process, is implemented on the Breast Cancer Wisconsin dataset. It displayed the number of attributes used to train the classifiers. It tends to increase the classification accuracy by eliminating noise features while irrelevant features are ignored. The benefits of using feature selection are firstly, it enables the machine learning algorithm to train faster, second, it reduces the complexity of a model and makes it easier to interpret, third, it improves the accuracy of a model if the right subset is chosen and lastly it reduces overfitting. By using the classification approach, the different classifiers are used.

C. Classification by Single Classifier

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 popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split. 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, the output of different classifiers will be combined and deep neural network approach will be applied. In ensemble combination classification task, the prediction class result will be combine 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.3 is a description for DMCL.

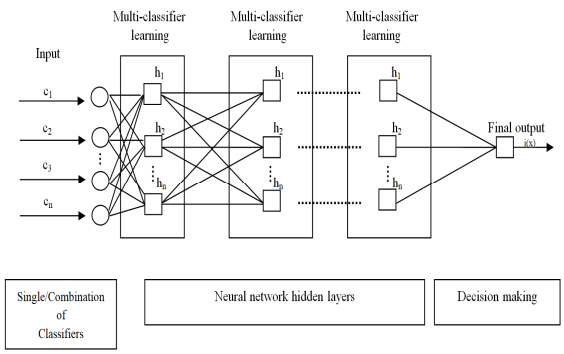

Fig.3. Description for DMCL

To classify the classes, four-layer neural network will be used. The input layer of the network contains five neurons/nodes which represent the attributes (classifiers) of samples in the new data set. In the hidden layers, some experiment with different numbers of nodes in each layer is done for better classification performance. The output layer of the network contains one neuron which output was 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, Breast Cancer Wisconsin and Hepatitis data set will be used. 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

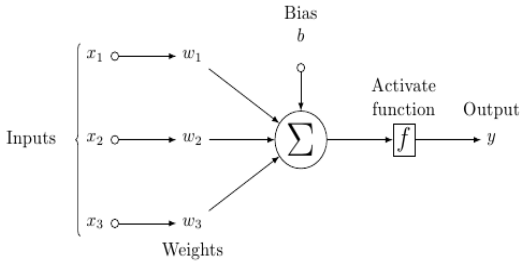

Fig.4. Neural Network Diagram

- Fig.4 shown the example of simple neural network diagram. Models in Keras act as a sequence of layers. The sequential model is created and layers added one at a time. The first thing to get right is to make sure the input layer has the correct number of inputs. This can be particular when building the first layer with the input_dim argument and setting it to 2 for the 2 input variables as an example. There are heuristics that can be use and often the greatest network structure is initiate through a procedure of trial and error experimentation. In this research, a fully-connected network structure with four layers will be applied.
- Fully connected layers are defined using the Dense class. The number of neurons in the layer are specifies as the first argument, 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 weights will be assigning to each mode which is nothing but importance of that node. At the time of initialization, weights should be close to 0 and will randomly initialize weights using uniform function.
- In this situation, the network weights will be set to a small random number generated from a uniform distribution ('uniform'), in this case between 0 and 0.05 because that is the default uniform weight initialization in Keras. Another traditional way would be 'normal' for small random numbers generated from a Gaussian distribution.
- Here the rectifier (relu) activation function and tanh activation function in hidden layer and Sigmoid activation function in output layer as want binary result from output layer are used. 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. Sigmoid are used on the output layer to ensure the network output is between 0 and 1 and easy to map to either a probability of class 1 or snap to a hard classification of either class with a default threshold of 0.5. All together will be portion 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. Table I 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
	2nd layer	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. Three types of activation function are used in this layer which are 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 pick up the best alternative to represent the network for training and making predictions to run on hardware, like as CPU or GPU or even distributed.
- When compiling, some additional properties are specify such as specify the loss function to apply to examine a set of weights, the optimizer used to seek through dissimilar weights for the network and optional metrics to assemble and report during training in searching the greatest set of weights to make predictions for 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 use is 'Adam'. SGD depends on loss thus second parameter is loss. Since out dependent variable is binary, logarithmic loss function called 'binary_crossentropy' will be used. In improving the performance of 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 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

- The model will be evaluate on training dataset using the evaluate() function where pass it the same input and output used to train the model. This will make a prediction for each input and output pair and collect scores, plus 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, the numbers will be see in the diagonal, which mean that all predictions were correct.
  - Precision is a measure of a classifier's exactness. 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.

The Kappa or Cohen's kappa is the classification accuracy normalized by the imbalance of the classes in the data.

IV. EXPERIMENTS

The data set is based on the data taken from the UCI Repository of Machine Learning Databases. Table II shows the total of instances after removal of instances with missing values and selected features for each Medical Data Sets.

Table II Features for each Medical Data Sets

Data Set	Total of Instances	Selected Features
Breast Cancer Wisconsin	683	Clump Thickness Uniformity of Cell Size Uniformity of Cell Shape Marginal Adhesion Single Epithelial Cell Size Bare Nuclei Bland Chromatin Normal Nucleoli Mitoses Class
Hepatitis	80	Sex Anorexia Ascites Liver Big Albumin Protime

To evaluate the proposed model, a number of experiments were performed. Table III shows the comparison of accuracies for the five classifiers used based on a 10-fold cross validation as a test method

namely SMO, J48, RF, NB, and IBk. Based on the results of the experiment as indicated in Table III, it demonstrates the prediction accuracy of the proposed single classification on different two Medical Data Sets which are Breast Cancer Wisconsin and Hepatitis. SMO classifier achieved the highest prediction accuracy in Breast Cancer Wisconsin (96.05%). While for Hepatitis, IBk achieved the highest prediction accuracy with 85.00%.

Table III Comparative Study of Single Classifier for Medical Data Sets

Data Set /Classifier	Breast Cancer Wisconsin	Hepatitis
SMO	**96.05**	81.25
J48	93.86	76.25
RF	95.77	83.75
NB	95.61	81.25
IBk	94.89	**85.00**

In experimental results for DMCL method, in BCW data set, SMO as base input nodes was chosen because it has shown the highest accuracy (Refer Table III). While Hepatitis data set, IBk as base input nodes was chosen because it has shown the highest accuracy (Refer Table III). For BCW data set that use SMO as base input nodes have shown that the combination of SMO+RF+IBK+NB is superior to the other classifiers as the combination of input nodes that achieved the highest accuracy with percentage of 96.63%.

The obtained results are shown in Table IV. It shows the result of combination two input nodes, three input nodes and four input nodes that using deep learning approach. Apart of accuracy, the dataset also evaluate using the precision, recall, F1_Score, and Cohen kappa score. From the test it indicated that combination SMO+RF+IBK+NB achieved the highest percentage with 96.63%.

Table IV Deep Multi-classifier Learning Prediction for BCW Data Set

Input nodes	Classifiers	Precision (%)	Recall (%)	F1_Score (%)	Cohen Kappa Score (%)	Accuracy (%)
2	SMO+RF	0.93	0.97	0.95	0.92	96.49
	SMO+NB	0.93	0.95	0.94	0.91	96.05
	SMO+IBk	0.92	0.97	0.95	0.92	96.34
	SMO+J48	0.93	0.95	0.94	0.91	96.05
3	SMO+RF+IBk	0.93	0.97	0.95	0.92	96.49
	SMO+RF+NB	0.93	0.97	0.95	0.92	96.49
	SMO+RF+J48	0.93	0.97	0.95	0.92	96.49
4	**SMO+RF+IBk+NB**	**0.93**	**0.97**	**0.95**	**0.93**	**96.63**
	SMO+RF+IBk+J4	0.93	0.97	0.95	0.92	96.49

For experimental results in Hepatitis data set that use IBk as base input nodes have shown that the combination of IBk+NB+J48+SMO is superior to the other classifiers as the combination of input nodes that achieved the highest accuracy with percentage of 92.50%.

The obtained results are shown in Table 6.3. It shows the result of combination two input nodes, three input nodes and four input nodes that are using deep learning approach. Apart of accuracy, the dataset also evaluate using the precision, recall, F1_Score, and Cohen kappa score. From the test it indicated that combination IBk+NB+J48+SMO achieved the highest percentage with 92.50 %.

Table V Deep Multi-classifier Learning Prediction for Hepatitis Data Set

Input nodes	Classifiers	Precision (%)	Recall (%)	F1_Score (%)	Cohen Kappa Score (%)	Accuracy (%)
2	IBk+RF	0.59	0.77	0.67	0.59	87.50
	IBk+SMO	0.78	0.54	0.64	0.58	90.00
	IBk+NB	0.80	0.62	0.70	0.65	91.25
	IBk+J48	0.59	0.77	0.67	0.59	87.50
3	IBk+NB+SMO	0.80	0.62	0.70	0.65	91.25
	IBk+NB+RF	0.80	0.62	0.70	0.65	91.25
	IBk+NB+J48	0.80	0.62	0.70	0.65	91.25
4	**IBk+NB+J48+SMO**	**0.89**	**0.62**	**0.73**	**0.69**	**92.50**
	IBk+NB+J48+RF	0.80	0.62	0.70	0.65	91.25

Table VI shows the final result from all experimental method in this study. Based on these results, it can be concluded that the accuracy for all data sets improved and outperforms DMCL methods than single ones and multi-classifier method that using majority voting as combination method. This scenario shows that DMCL method is able to gain higher accuracy for the data sets.

Table VI Accuracy Comparison Medical Data Sets by Different of Method (Highest Accuracy)

Data Set	Single Ones (%)	Majority Voting (%)	DMCL (%)
Breast Cancer Wisconsin	96.05	96.49	96.63
Hepatitis	85.00	86.25	92.50

V. CONCLUSION

Based on classification method, the previous work has demonstrated that single classification and multi-classifier that using majority voting as improving the classification performance. The proposed method is evaluated using two medical data sets taken from the UCI repository. The results showed that the proposed method is able to produce the highest prediction accuracy than single and combination of classifier that using majority voting for all these medical data sets. In the future the DMCL methods can be applied to enhance multi class classifiers and a method like as pruning a neural network can be applying to the future research in determine an optimal structure for a neural network.

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. [Online]. 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] A. . Nancy.P, Sudha.V, "*Analysis of feature selection and classification algorithms on hepatitis data*," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 6, no. 1, 2017.

[6] R. Raihan, "*Hepatitis in Malaysia: Past, Present, and Future*," Euroasian Journal of Hepato-Gastroenterology, vol. 6, no. 1, pp. 52–55, 2016.

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

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

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

[10] 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.

[11] S. Min, B. Lee, and S. Yoon, "*Deep learning in bioinformatics*," 2016.

[12] P. Doetsch, "*Optimization of Hidden Markov Models and Neural Networks*," Rheinisch-Westf¨alische Technische Hochschule Aachen, 2011.

[13] 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.

combination rule are not enough performed well in improving the prediction accuracy of medical applications. The results of this work have proven that proposed DMCL method

[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*," in Proceedings of the International Conference on Telecommunication, Electronic and Computer Engineering (ICTEC '15), 2015.

[15] 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.

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

[17] F. Keyvanfard, M. A. Shoorehdeli, M. Teshnehlab, K. Nie, and M. Y. Su, "*Specificity enhancement in classification of breast MRI lesion based on multi-classifier*," Neural Computing and Applications, vol. 22, no. SUPPL.1, pp. 35–45, 2013.

[18] L. Rokach, "*Ensemble-based classifiers*," Artificial Intelligence Review, 2010.

[19] L. Rokach, "*Taxonomy for characterizing ensemble methods in classification tasks : A review and annotated bibliography*," Elsevier, 2009.

[20] L. Rokach, "*Ensemble methods for classification*," Data Mining and Knowledge Discovery Handbook, pp. 957–980, 2005.

[21] M. Elish, T. Helmy, and M. Hussain, "*Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation*," Mathematical Problems in, vol. 2013, 2013.

[22] G. Ge and G. W. Wong, "*Classification of premalignant pancreatic cancer mass-spectrometry data using decision tree ensembles.*," BMC bioinformatics, vol. 9, p. 275, 2008.

[23] [23] A. Rahman and S. Tasnim, "*Ensemble classifiers and their applications : A review*," International Journal of Computer Trends and Technology (IJCTT), vol. 10, no. 1, pp. 31–35, 2014.

[24] 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.

[25] Naveed, Q. N., Qureshi, M. R. N., Tairan, N., Mohammad, A., Shaikh, A., Alsayed, A. O., ... & Alotaibi, F. M. (2020). "*Evaluating critical success factors in implementing E-learning system using multi-criteria decision-making*". *Plos one*, *15*(5), e0231465.