International Journal on Advanced Science Engineering Information Technology

Vol.7 (2017) No. 4-2
ISSN: 2088-5334

Intelligent Learning Model based on Significant Weight of Domain Knowledge Concept for Adaptive E-Learning

Norsham Idris[#], Siti Zaiton Mohd Hashim[#], Ruhaidah Samsudin[#], Nor Bahiah Hj Ahmad[#]

[#]*Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia, Skudai Johor, Malaysia*
E-mail:norsham@utm.my

***Abstract*— In order to support personalized learning, an adaptive learning system should have a capability to provide each student with a suitable learning material regarding his profile. However, the issue of student varieties in acquiring every Domain Knowledge Concept (DKC), and a range of DKC important variations in a particular learning material produced a complex dependency that causes a difficulty in the learning material selection process. Existing rule-based learning material selection approach requires the definition of a huge set adaptation rules. However, this approach usually results in inaccurate and incorrect selection due to the inconsistent, insufficient and confluence of the defined rules. Consequently, the process of learning material selection is hard to be algorithmized, therefore, intelligent methods are applied to handle the complexity challenges. This research proposes a significance weight approach that represents the complex dependency of learning material selection problem to substitute the rules definition in the selection process. In addition, this research proposes an intelligent learning model that combines unsupervised and supervised machine learning techniques to accurately select the learning material for a particular student adaptively. The unsupervised machine learning technique is vital in obtaining a learning material classification and labelling based on the proposed significance weight. Meanwhile, the supervised machine learning technique, the Multilayer Perceptron Artificial Neural Networks conducts the adaptation process that will assign the student to suitable learning materials regarding his performance upon specific DKC. With 98% achievement of classification accuracies, this model can be considered as highly accurate in selecting a correct and suitable learning material based on student's domain knowledge level.**

***Keywords*— adaptive e-learning; domain knowledge concept; learning material selection; supervised machine learning**

I. INTRODUCTION

The advancement of the information and communication technology has changed the process of teaching and learning towards a portable, student-centered and multi-platform environment. The history started with a stand-alone Computer Aided Learning (CAL) and CD-ROM which only provided a set of programmed instructions used for educational purposes. However, such technologies were not able to replace the traditional classroom in which teachers typically consider factors that affect the learning of each student [1]. Therefore, researchers have been studying the adaptation element to be included in CAL system so that the material and the learning session can be personalized to entertain students' heterogeneity [2]. Particularly, Intelligent Tutoring System (ITS) and Adaptive Educational Hypermedia System (AEHS) are then designed under the paradigm of Adaptive Learning System (ALS) in order to cater the issue of personalization [3]. The past decade has seen the rapid development of ALS with a broad range of adaptation and implementation strategies that offered promising solutions through the use of various techniques and approaches.

One major issue in ALS research concerned the mechanism for adaptively selecting a learning material that fits the criteria of students such as the knowledge level, learning style, and goal. Learning material and students profile is a key component in ALS, and both have distinctive characteristics that rather complex. This complexity requires a deliberate construction of adaptation module so that the selection of the learning material really meets the needs of students. In previous research, the selection is carried out through the implementation of adaptation rules. One of the most significant discussions in the rule-based learning material selection is the issue of insufficiency and inconsistency of defined rules that induced incorrect and inaccurate selection [4]. In addition, this approach also involves in high costs due to the high effort of domain experts in defining the adaptation rules. Therefore, several new approaches have been developed to replace the need for adaptation rules in the selection of the learning material in ALS.

Machine learning is one approach that has been utilized to support adaptation in ALS. The capability of machine learning techniques in addressing pattern recognition and prediction problem has caught the interest of researchers in the field of adaptive learning to apply it for adaptation and personalization. Machine learning algorithms can be divided into two types of supervised and unsupervised learning. Usually, supervised learning is used for classification task while unsupervised learning in the clustering task. In most of ALS, supervised learning has been employed for classification of students [5], [6] and prediction of student performance [7], [8], [9]. Whilst clustering in ALS can be divided into clustering of students [10] or clustering of learning materials [11]. To date, there have very few works reported in combining both classification and clustering techniques especially in learning material selection of ALS.

II. Material and Method

The proposed intelligent learning model which is to solve the problem of adaptive selection of learning material, comprehends three major modules, namely Domain Knowledge Module, Domain Experts Module and lastly, Adaptation Module as illustrated in Fig. 1. In addition, the intelligent learning model consists of two major parts, Domain Knowledge Concept Representation (DKCR) and Adaptive Selection of Learning Material.

A. Domain Knowledge Module

Domain Knowledge Module is one of the components of DKCR in the intelligent learning model. This module is to establish the domain knowledge elements that would be structured in order to feed the adaptation and personalization process. In this proposed model, the element of domain knowledge chosen is its DKC. Every learning material provided to students should associate to particular DKC in order to achieve the learning outcome as specified.

B. Domain Experts Module

In the intelligent learning model, Domain Experts Module is a part of DKCR. This module is to constitute the role of the domain experts in this research. In this proposed approach the domain expert roles are to identify the related DKC and to estimate the significance of every DKC in a particular learning material.

C. Data Preprocessing

Data Preprocessing is a task of Representation of DKC in the DKCR in this proposed intelligent learning model. This is to compute the estimated DKC significance weight and to normalize the resulted DKC significance weight.

D. Adaptation Module

The function of this module is as the engine for the adaptation and the selection tasks. For the adaptation and personalization, unsupervised learning technique is used to cluster the learning materials based on the DKC significance weight similarity before being assigned to students. The assignment of learning material to the student is treated as a classification problem. Thus a supervised learning technique is applied.

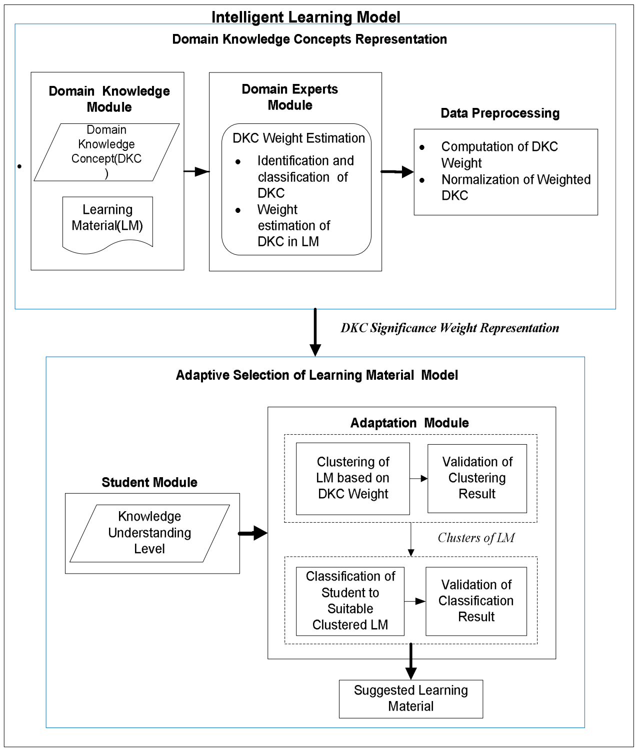

Fig. 1 Intelligent learning model

E. Student Module

The proposed intelligent learning model considers student's knowledge understanding level as adaptation factor to be suited to the learning material. Thus, in this module, the result of student pre-test data is represented in DKC-based format, the same as the representation of the learning material data.

Analysis of DKC is the first task of the DKCR framework. It involves the Domain Knowledge Module and the Domain Expert Module. It is the responsibility of the Domain Knowledge Module to ensure that each knowledge unit called domain knowledge concept of a course are represented correctly and completely. This includes the terms of coverage, sequence, and dependencies between each of the DKC.

Identification of related DKC is the step in the analysis part of the Domain Expert Module in the process of DKCR. This step implemented according to the procedure used by [12]. Domain experts have been asked to identify related concepts from the domain knowledge. The domain experts involved are three (3) lecturers from the Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia who had more than 3 semesters experience in teaching the course. A study on the current course syllabus used in the faculty that contemporized with the Association for Computing Machinery (ACM) Curriculum (Computer Science Curriculum 2008) also has been conducted along with investigation on some popular textbooks which is used by the faculty.

The following step of the DKC analysis is to classify the DKC regarding its complexity levels. There are three levels of DKC complexity that have been identified, namely Prerequisite, Basic, and Advanced. The prerequisite concept is a concept which is classified as a requirement for understanding Basic concepts of the topic Array. The basic level is a concept which is considered as the main concept that should be learned by all students for the Array topic. Advance is the concept that is considered tough and complex. Determination of the complexity levels is suggested by the domain experts. For the purpose of data preparation, the DKC order in the list is ranked or sorted according to the complexity level classification. Table 1 shows the list of 25 DKC that have been identified with the associated level of complexity [13].

Meanwhile, this task involves the preparation of learning materials that cover all of the identified DKC. In the area of education, learning materials for teaching and learning activities can be in the type of forum, textual notes, animation, simulation, game, on-line question, Powerpoint slides, hypertext and etc. [14]. In this research, the type of learning material that we focused is a test or quiz question which is normally used for student assessment. For the purpose of this study, the selection of 120 test questions is varied and cover from Prerequisite, Basic to Advance DKC complexities.

In order to evaluate the significance of DKC in our test question collection, we have conducted a survey among the domain experts. Each of them is given a set of test questions which covers the topic of Array and the prerequisites of it. For every test question, they were required to rate how significant or important the listed 25 DKC to one particular question. This Likert scale type of concept relevance evaluation is rated from 0 to 5 (Not relevant to Strongly Relevant) [13].

TABLE I
LIST OF DOMAIN KNOWLEDGE CONCEPTS

ID	**CONCEPTS**	**COMPLEXITY LEVEL**
C1	Primitive data type	Prerequisite
C2	Declaration of primitive data type	Prerequisite
C3	For loop	Prerequisite
C4	Assignment statement	Prerequisite
C5	Identifier	Prerequisite
C6	Element indexing	Prerequisite
C7	Passing parameter by value	Prerequisite
C8	Passing parameter by reference	Prerequisite
C9	Nested for loop	Prerequisite
C10	Passing one element	Prerequisite
C11	Returning value	Prerequisite
C12	Class and object	Basic
C13	Object declaration	Basic
C14	Array declaration	Basic
C15	Method definition	Basic
C16	Assign value using nested for loop	Basic
C17	Declare, create and initialize value using array initializer	Basic
C18	Access array element using indexed variable	Basic
C19	Create object in each/particular index of the array	Basic
C20	Assign value using nested for loop	Basic
C21	Passing entire array	Advance
C22	Returning array	Advance
C23	Create array of type class	Advance
C24	Declare array reference variable	Advance
C25	Copy contents from one array to another	Advance

To date, various methods have been developed and introduced to measure the difficulty and complexity of test question and learning materials. Among the methods that have been reported are based on Item Response Theory [15], Learner Feedback [12], Expert Rating [11] and Latent Semantic Analysis [16]. However, far too little attention has been paid to the significance of DKC in learning the material for the purpose of personalization. There is no research so far in the field of adaptive learning that takes this significance factor into account despite its reliability in supporting adaptation.

Consequently, in this study, the expert rating approach was chosen in estimating the significance of DKC in a particular learning material in order to encapsulate the human domain expert point of view or decision. After the domain expert rating of the learning materials is collected, next is to compute and transform the rating into a DKC significance weight. For that purpose, the formula as depicted in (1) which is adopted from [11] was employed.

$$D_{exp} = \sum_{n=0}^{5} \frac{C_n}{C_{experts}} Rel_n \qquad (1)$$

D_{exp} refers to the experts' decision of the concept relevancy level for each question. C_n denotes the number of an expert who chose a significance level of Rel_n. The significance level of the concept is classified into 6 points (from Not relevant to Strongly relevant), depicted as Rel_0, Rel_1, Rel_2, Rel_3, Rel_4, Rel_5. $C_{experts}$ denotes the number of domain expert [13].

The computed DKC significance weight (D_{exp}) data are within the range of [0...5]. Therefore a scaling method has to be implemented for the data set to normalized them into [0...1] range of data. This is to ensure that the data is compatible with machine learning process in the next stage.

III. RESULTS AND DISCUSSION

This research aims to study the ability to utilize both supervised and unsupervised machine learning techniques in order to select suitable categories of learning materials (Prerequisite, Basic and Advanced) for a particular group of students (Beginner, Intermediate, and Advanced). The learning material data which is originated from domain experts' estimation of DKC significance weight produce a set of data which describes the characteristics of each learning material through the significance weight of 25 DKC that have been computed. These unlabelled and unclassified data need to be clustered using the unsupervised algorithm in order to obtain a grouping of similar learning material based on DKC weight.

A. Clustering of Learning Material

The data set has a total of 3000 DKC significance weight data and a dimension of 120 rows and 25 columns. The rows represent 120 learning materials whilst the columns represent 25 DKC. The experiment is implemented using MATLAB tool for Self-Organizing Map (SOM) clustering. SOM has been widely used in many areas as its ability in clustering of high dimensional data [17].

Cluster number determination is made based on the number of student categories. In this study, students are categorized based on their understanding level upon each DKC as being collected during the pre-test earlier. In this study, three (3) categories of student performance are being considered for the classification namely Beginner, Intermediate and Advanced. Therefore, the learning material data will be grouped into 3 clusters to satisfy the requirement of each category of students.

For experiment using SOM, a network has been created with 25 input neurons and has been trained with different dimension sizes of the map (2 x 2, 3 X 3 and 5 X 5) in order to get the best clustering result. From the experiments, using our data, the best result was obtained when the size of the map dimension is 5 x 5. The Euclidean similarity measure is used in this experiment. Table 2 presents the confusion matrix for SOM test result.

TABLE II
CONFUSION MATRIX FOR OVERALL SOM CLUSTERING

		Predicted by SOM			**Accuracy**
		Prerequisite	**Basic**	**Advanced**	
Actual (Domain Expert Decision)	**Prerequisite**	78 (TP)	7 (FN)	0 (FN)	0.9176
	Basic	4 (FP)	12 (TP)	0 (FN)	0.750
	Advanced	0 (FP)	2 (FP)	15 (TP)	0.8823

Table 2 explains that TP (true positives): is the number of correct predictions that an instance is positive; TN (true negatives): is the number of correct predictions that an instance is negative; FP (false positives): is the number of incorrect predictions that an instance is positive; FN (false negatives): is the number of incorrect of predictions that an instance negative. The accuracy of the classification can be defined by the following formula:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2)$$

B. *Learning Material Selection for Student Using Classification Technique*

The proposed approach is based on the intelligent mechanism to model the decision from domain experts in the selection of learning materials. The neural network is chosen considering its ability to learn from data and predict the class for new input. A classifier from neural network paradigm is built to select learning materials according to the characteristics of its class. Compared to rule-based approach, this approach has the advantage that it can facilitate the work should be done by domain experts, especially in the definition of adaptation rules.

In this approach, the selection of learning materials do not require the definition of rules, but machine learning will determine the connection between the characteristics of the learning material with the characteristics of the students. Thereby, the selection of learning materials is expected to be the same as the selection made by the domain experts. Therefore, the testing data for validating the classifier are labeled by the domain expert prior to the experiment. Based on the pre-test result which is presented in each DKC, the domain expert will identify whether the student is Beginner, Intermediate or Excellent. The manual classification performed by domain experts will produce a set of data for benchmarking the classifier model. The input data of the model is clusters of learning materials and the knowledge understanding level of students. Outputs of the model are groups of learning materials that are assigned to students. Fig. 2 illustrates the framework of the classification task.

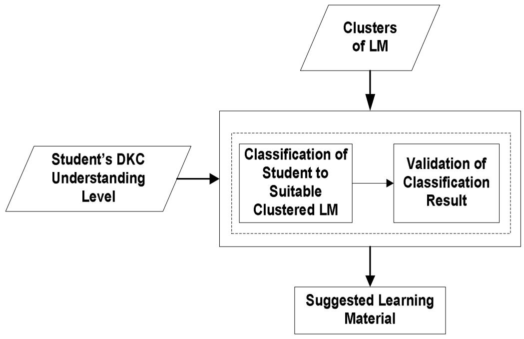

Fig. 2 Classification framework

A common multi-layer perceptron neural network architecture (MLP ANN) consists of three (3) layers or groups of units as follows; input, hidden and output. In this study, the input layer represents the 25 DKCs, where the input vector is a set of values {0…1}. After the network is constructed, next is to train the network using learning material data set as discussed previously. In order to choose the best classifier model, experiments using three different training algorithms are performed.

The training algorithms that are tested namely Scaled Conjugate Gradient (SCG), Levenberg-Marquardt (LM) and Resilient Backpropagation (RP). For each of the algorithm, training is executed with three different sizes of the hidden layer. During the training of the network, the output layer is assigned to 3 classes of learning material that have been identified in previous stage namely Prerequisite, Basic and Advanced. Table 5.15 presents the result of the network training. The MLP ANN model with Levenberg-Marquardt (LM) training algorithm with 50 hidden nodes size is selected to be used in the testing experiment based on the accuracy result of different types of training algorithm. Performance is measured by mean squared error (MSE) value. The training will stop when MSE is obtained, that is the minimum error between targeted and predicted by the model.

TABLE III
ACCURACY RATE OF DIFFERENT TRAINING ALGORITHM

Training Algorithm	**Hidden Layer Size**					
	13		25		50	
	Acc (%)	MSE	Acc (%)	MSE	Acc (%)	MSE
SCG	96.8	0.013	94.6	0.024	96.4	0.015
LM	**95.6**	**0.014**	**97.2**	**0.008**	***97.5***	**0.012**
RP	97.0	0.015	97.1	0.011	93.8	0.062

Data used in this experiment is the average score of 60 student's pre-test results. The pre-test consists of questions that cover all 25 DKCs. Since one DKC may be represented by more than one pre-test question, an average of marks that student gained from each DKC is taken as data for this experiment. For classifier validating purpose, each student data is classified by the domain expert and labelled as Beginner, Intermediate or Advanced based on his pre-test marks. Table 4 depicts the distribution of student data used in this experiment.

TABLE IV
DISTRIBUTION OF STUDENT DATA FOR TESTING

Class	**No. of instances**	**Percentage (%)**
Beginner	21	35
Intermediate	20	33
Excellent	19	32
TOTAL	**60**	**100**

Table 5 presents the performance of the classifier model that classify student into three classes based on their level of domain knowledge understanding. From the table, we can see that the proposed classifier model performed excellently with 98% accuracy in terms of correctly classified instances for the classes. The Kappa statistic value is high, 0.9712 reveals the similarity of classifier decision and the domain expert's decision. Error rate produced by the classifier model is considered low.

TABLE V
CLASSIFICATION RESULT

Performance Measures	**Result**
Correctly Classified Instances (%)	98
Kappa statistic	0.9712
Mean absolute error	0.0245
Root mean squared error	0.115
Relative absolute error (%)	5.5476
Root relative squared error (%)	24.445

Outputs from the Adaptation Module of the proposed intelligent learning model are three categories of students namely Beginner, Intermediate and Advanced. The learning material selection scheme for different categories of the student is shown in Table 6. Based on the table, a Beginner type of student is assigned a learning material from Prerequisite class, the Intermediate student is assigned with learning material of Basic type, whilst the Advanced student is given learning material which contains more Advanced DKC. This learning material selection scheme is adopted from previous works.

TABLE VI
LEARNING MATERIAL SELECTION SCHEME

Students	**Appropriate Type of Learning Material**
Beginner	Prerequisite
Intermediate	Basic
Advanced	Advanced

Therefore, students that have been classified into three classes by the proposed model are assigned to the type of learning material automatically based on the allocation in Table 6.

In order to validate the accuracy of the learning material selection by the proposed approach, an experiment is conducted to compare the selection result with the rule-based selection approach. The learning material selection rules employed for this experiment are depicted in Fig. 3. These rules are extracted from the decision of domain experts on the learning material assignment to students.

For every learning material
Calculate means and for each complexity level
Compare means value for each complexity level

IF (0.938≤ means_Advance≥ 0.982) AND
IF (0.578≤ means_Basic≥ 0.933) AND
IF (0.009≤ means_Prerequisite≥ 0.867)
THEN
Assign LM to ADVANCED-STUDENT

IF (0.00≤ means_Advance≥ 0.067) AND
IF (0.524≤ means_Basic≥ 0.908) AND
IF (0.349≤ means_Prerequisite≥ 0.486)
THEN
Assign LM to INTERMEDIATE-STUDENT

IF (0.017≤ means_Advance≥ 0.121) AND
IF (0.020≤ means_Basic≥ 0.153) AND
IF (0.133≤ means_Prerequisite≥ 0.489)
THEN
Assign LM to BEGINNER-STUDENT

Fig. 3 Rules for learning material classification

The rule-based selection experiment is performed using the learning material and student knowledge understanding level data that have been used by the proposed model. Table 7 shows the comparison of learning material selection result obtained by the proposed approach and rule-based approach. As presented in the previous section, the proposed approach has correctly selected learning materials for 59 students which represent 98% of accuracy. While rule-based approach achieved 95% accuracy where 57 students are assigned a correct type of learning materials.

TABLE VII
COMPARISON WITH RULE-BASED SELECTION

Learning Material Selection Approach	**Accuracy (Correct Selection)**	**Number of students with correct selection (out of 60 students)**
Proposed approach	98 %	59
Rule-based	95%	57

Based on the experimental results, the proposed adaptive selection of learning material model has successfully selected 98% correct learning materials for a particular group of students. From 60 students, only 1 are being misclassified or given a different type of learning material other than suggested by a domain expert. The performance of the proposed model slightly better than the rule-based selection approach as presented in Table 7.

IV. Conclusion

In this research, learning materials is personalized for every student through an adaptive selection that suited to his knowledge understanding level of each specified DKC. Towards the end, the student will be assigned a group of learning material that appropriate to his knowledge understanding. Most of previous researches solved this problem by defining adaptation rules which were tedious, highly cost and complicated. This study however, treated the student to learning material assignment task as supervised classifications that can be done without prior define of adaptation rules.

The first problem to encounter is to find good quality groups of learning materials with similar characteristics. In this case, the characteristics are the significance weight of domain knowledge concept. One learning material may consist of many DKC and each of the concepts has different levels of significance. For the domain knowledge 'Array in Java', 25 domain knowledge concepts are identified by the domain experts, i.e the lecturers. They are also required to rate the significance weight of the concepts for each learning material. Clustering techniques are then applied towards the learning material data to simplify the assignment (classification) process by providing labeled instances. Next, for the purpose of personalization of learning material selection, a supervised classification is performed by constructing and training of ANN model that could select suitable learning material for each student.

It was hypothesized that the representation of DKC significance weight and employment of machine learning techniques in a learning material selection model could provide a better accuracy of selection. The finding of this study is consistent and supports the hypothesis of this research. The proposed approach has achieved 98% accuracy in selecting a correct type of learning material for a particular student. This achievement overcomes the result of rule- based selection using the same dataset (95% of accuracy) as well as improving of previous work on non-rule based selection with ANN MLP technique (97% of accuracy).

Acknowledgment

This research was funded by UTM Research University Project grant, Q.J130000.2628.12J35.

References

[1] S. Bhattacharya and S. Nath (2016). International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No2 DOI: 10.9781/ijimai.2016.4212

[2] Fatiha Elghibari,Rachid Elouahbi and Fatima El Khoukhi,(2017)."An Automatic Updating Process to Control The E-learning Courseware," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, no. 2, pp. 546-551. [Online]. Available: http://dx.doi.org/10.18517/ijaseit.7.2.1871.

[3] Dziuban, Charles; Moskal, Patsy; Johnson, Constance; and Evans, Duncan (2017) "Adaptive Learning: A Tale of Two Contexts,"Current Issues in Emerging eLearning" Vol. 4: Iss. 1, Article 3.

[4] Karampiperis, P., Lin, T. and Sampson, D. G. (2006). Adaptive cognitive-based selection of learning objects. *Innovations in Education and Teaching International*, *43*(2), 121–135. doi:10.1080/14703290600650392

[5] Oancea, B., Dragoescu, R. and Ciucu, S. (2013). Predicting students ' results in higher education using neural networks. In International Conference on Applied Information and Communication Technology (pp. 190–193).

[6] López, M. I., Luna, J. M., Romero, C. and Ventura, S. (2012). Classification Via Clustering For Predicting Final Marks Based On Student Participation In Forums. In *In Proceedings Of The 5th International Conference On Educational Data mining (pp. 148–151).* (pp. 148–151).

[7] P. A. Khodke, M. G. Tingane, A. P. Bhagat, S. P. Chaudhari and M. S. Ali, "Neuro Fuzzy intelligent e-Learning systems," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, 2016, pp. 1-7.doi: 10.1109/GET.2016.7916766

[8] Mashiloane, L. and Mchunu, M. (2013). Mining for Marks : A Comparison of Classification Algorithms when Predicting Academic Performance to Identify " Students at Risk ," 541–552.

[9] Mishra, T., Kumar, D. and Gupta, S. (2014). Mining Students' Data for Prediction Performance. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 255–262. doi:10.1109/ACCT.2014.105

[10] Moucary, C. El. and Khair, M. (2011). Improving Student ' s Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education, *II*(Iii).

[11] Lu, F., Li, X., Liu, Q., Yang, Z., Tan, G. and He, T. (2007). Research on Personalized E-Learning System Using Fuzzy Set Based Clustering Algorithm. In *ICCS 2007, Part III, LNCS, Springer-Verlag Berlin Heidelberg* (pp. 587–590)

[12] Kim, S. B., Yang, K. M. and Kim, C. M. (2006). A Diagnostic Model Using a Clustering Scheme, 278–287.

[13] N. Idris N. Yusof S. Z. M. Hashim (2013). A Model For Adaptive Selection Of Learning Material In An IntelLigent Learning System Using Combination Of Supervised And Unsupervised Machine Learning Techniques. *Proceeding Int. Conf. Artif. Intell. Comput. Sci. ICT AICS2013 ISBN 978-967-11768-3-2*.

[14] Nor Bahiah Hj Ahmad. (2012). Granular Mining Approach For Identifying Student's Learning Style In E-Learning Environment. PhD Tesis, Universiti Teknologi Malaysia.

[15] Baylari, A. and Montazer, G. A. (2009). Design a Personalized E-Learning System Based On Item Response Theory And Artificial Neural Network Approach. *Expert Systems with Applications*, *36*(4), 8013–8021. doi:10.1016/j.eswa.2008.10.080

[16] Chang, T.-H., Sung, Y.-T. and Lee, Y.-T. (2013). Evaluating the Difficulty of Concepts on Domain Knowledge Using Latent Semantic Analysis. *2013 International Conference on Asian Language Processing*, 193–196. doi:10.1109/IALP.2013.58

[17] Lee, L. C., Liong, C. Y., & Jemain, A. A. (2016). Applying Fourier-Transform Infrared Spectroscopy and Self-Organizing Maps for Forensic Classification of White-Copy Papers. *International Journal on Advanced Science, Engineering and Information Technology*, *6*(6), 1033-1039.