# Prediction of Learner's Profile Based on Learning Styles in Adaptive E-learning System

Sucheta V. Kolekar
Manipal University, Manipal, Karnataka, India
sucheta.kolekar@manipal.edu

Radhika M. Pai*
Manipal University, Manipal, Karnataka, India
Radhika.pai@manipal.edu

Manohara Pai M.M.
Manipal University, Manipal, Karnataka, India
mmm.pai@manipal.edu

**Abstract**—The major requirement of e-learning system is to provide a personalized interface with personalized contents which adapts to the learning styles of the learners. This is possible if the learning styles of the learner is known. In this paper, it is proposed to identify the learning styles of the learner, by capturing the learning behavior of the learner in the e-learning portal using Web Log Mining. The learning styles are then mapped to Felder-Silverman Learning Style Model (FSLSM) categories. Each category of the learner is provided with the contents and interface which are apt for that category. Fuzzy C Means (FCM) algorithm is used to cluster the captured learning behavioral data into FSLSM categories. The learning styles of a learner get changed over a period of time hence the system has to adapt to the changes and accordingly provide the necessary interface and contents. For this, the Gravitational Search based Back Propagation Neural Network (GSBPNN) algorithm is used to predict the learning styles of the learner in real-time. This algorithm is a modification of basic Back Propagation Neural Network (BPNN) algorithm which calculates the weights using Gravitation Search Algorithm (GSA). The algorithm is validated on the captured data and compared using various metrics with the basic BPNN algorithm. The result shows that the performance of GSBPNN algorithm is better than BPNN.

**Keywords**—Felder-Silverman Learning Style Model, Fuzzy C Means, Gravitational Search Algorithm, Back Propagation Neural Network Model, Jaccard Index, Xie-Beni Index.

# 1 Introduction

Adaptive E-learning System enhances the efficiency of online education by providing personalized contents and user interfaces. The personalized environment should be adaptive and should change according to the learner's requirements. The learner's requirements can be captured through their usage patterns. These patterns differ among learners who attend the same course.

Typically e-learning of educational system provides the same resources to all learners even though different learners need different information according to their level of knowledge, ways of learning style and preferences. Course sequencing is a technology originated in the area of Intelligent or Adaptive Learning System with the basic aim to provide learner with the most suitable sequence of knowledge to learn, and sequence of learning tasks (examples, exercise, problems, contents etc.) to work with. This sequence of learning task can be defined based on learning behavior of the learner on the portal.

The learning behavior can be captured through the e-learning portal using Web Usage Mining (WUM). WUM is the process of finding out what users are looking for on the E-learning portal [1]. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. WUM also helps find patterns for a particular group of people, or for e-learning users in a particular region. It uses the web logs to extract potential patterns of learning styles. Such kind of patterns is used to study the learner's navigation behavior along with the efficiency of the interface components, which help in adapting and customizing the resource delivery.
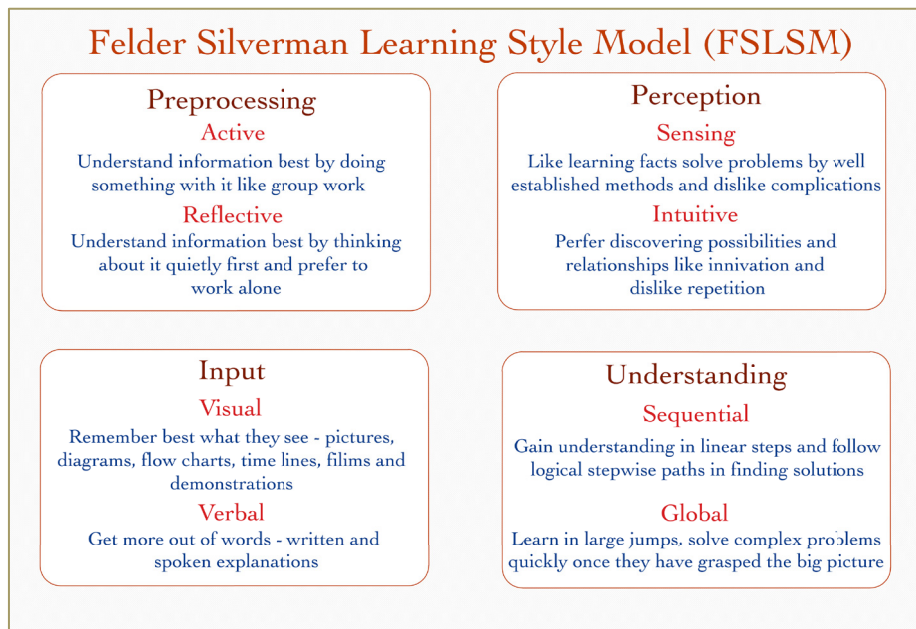


**Fig. 1.** Dimensions and Categories of FSLSM

The captured learning behavior can be mapped to learning styles by using suitable learning style model. The FSLSM is the most popular learning style model available which emphasizes on various categories of learners. The FSLSM defines four dimensions (Pre-processing, Perception, Input and Understanding) and eight categories of learners (Active, Reflective, Sensing, Intuitive, Visual, Verbal, Sequential and Global), which are as shown in Fig. 1.

Depending on the learning style model, the adaptive e-learning system should adopt the mechanism to detect and identify learning styles of new learner. Hence the prediction model needs to be developed to classify a new learner and personalized contents and interface can be provided to the learner on the e-learning portal.

In this paper, the learning styles are captured and identified based on standard FSLSM learning style model. The captured data is mapped to learning style categories of FSLSM, which has been done through FCM Clustering techniques. The mapped data is then used to predict the learning styles using GSBPNN Classification technique [1].

Rest of the paper is organized as follows. Section II discusses related research work in the area of clustering and classification techniques particularly for e-learning. Section III discusses about the learning components of e-learning portal. Section IV describes the methodology used for clustering and classification. Section V presents the experimentation and result analysis and Section VI mentions the concluding remarks.

## 2 Related Works

In literature, there are various techniques that have been used to identify and classify learning styles with various learning style models. Some of the paper are studied thoroughly and some factors are identified which are integrated into the work. The referred literature work cover different aspects of learner's behavior and knowledge depending on the application the authors have considered for their experiments. Abdullah et al. [1] have proposed a new approach to classify students dynamically depending on their learning style. This approach is experimented on 35 students for Data Structure online course created using Moodle. The learning style of each student is identified according to FSLSM by extracting learner's behavior and data from Moodle log. Also, learning styles based on the behavior have been compared with quiz results conducted at the end of the course. But the learning style of a learner changes over a period of time. The e-learning portal considered for study is Moodle and it is static in nature. The authors have not considered the learning components of portal such as pages and files, which are useful to define the learning styles based on FSLSM.

Feldman et al. [11] have reviewed the current trends in the field of automatic detection of learning styles. They have presented the analysis result of different techniques and discussed some limitations, implications and research gaps. Authors have described various detection techniques, which are mainly categorized under classification mechanism of Data Mining. Authors have also proposed the user model for the

education system that is used to detect the learning styles. Authors have not implemented the mentioned techniques by considering a learning style model.

Chang et al. [6] have proposed a learning style classification mechanism to classify and identify student's learning styles. The mechanism modifies K-Nearest Neighbor (k-NN) classification algorithm and combines it with Genetic Algorithms (GAs). The experimental results indicate that the classification mechanism can effectively classify and identify student's learning styles. The mechanism has not considered the specific features of learning styles based on a particular learning style model. The selection of learning style model influences the characteristics of mechanism. Authors have not considered the use of learning style model for the improvement of e-learning system with the recommendation mechanism based on learning styles of learners.

Abdullah et al. [2] have discussed about the impact of learning styles in the e-learning environment. The Naive-Bayes Tree (NBTree) classification technique is used to classify the learning style of the learner as per FSLSM. In this approach the learning styles are captured using questionnaire index of FSLSM, which has its own limitations to consider the learner's online usage behavior.

Latham et al. [15] have combined the Multi-Layer Perceptron and Artificial Neural Network (ANN) to classify the learner's data. The learner's data is captured based on FSLSM Questionnaire index and have considered four categories of learners (Active/Reflective and Sequential/Global) for classification. Since the algorithm considered questionnaire data of FSLSM, the identified learning styles cannot be considered as realistic learning styles for adaptive e-learning.

Ahmed and Badii [4] have discussed the importance of integrating the learning styles into Educational Hypermedia Systems (EHS). Authors have also discussed various approaches such as Bayesian Network, Decision Tree, and ANN etc. to classify log data accurately. The algorithms have not been implemented for real time captured learning behavior that can be considered to integrate learning styles in the EHS.

Bernard et al. [5] have proposed an approach that uses ANN to identify student's learning styles. Authors have considered the FSLSM questionnaire approach to identify learning styles of students. Students who had spent more than 5 minutes to fill the questionnaire have been stored in the database. This approach can be applied dynamically to detect the learning styles.

Takacs et al. [19] have proposed a method to make learning more efficient. Authors have tried to adapt the process of learning to learner's individual needs. Learner's characteristics such as learning preferences, sensual types, learning styles and motivation are identified as part of the proposed work. Authors have used questionnaire approach to collect the data. The results of this methodology are compared with clustering algorithms, decision trees and Principal Component Analysis (PCA). Authors have also analyzed that learners cannot be divided into various groups that possess similar properties and it is not possible to deal with every learner individually.

Huong Truong [20] has reviewed various works that has been tried to integrate the learning styles with adaptive e-learning system. He has also discussed about various aspects of learning style theories, real-time learning style predictors, learning style classifications algorithms and adaptive learning system using learning style applications. In learning style theories author has listed Felder and Silverman Model, Kines-

thetic and Kolb's learning style model, Honey and Murnford's Learning style model etc. In real-time learning style predictors, the main sources of data and the corresponding attributes such as log files, user history, background knowledge, intelligent capability, cognitive traits, motivation level etc. have been discussed. In learning style classification the author has explained about rule-based, Bayesian network and hidden Markov Chain.

Hogo [12] has presented the use of different fuzzy clustering techniques such as Fuzzy C-Means (FCM) and kernelled FCM to find the learner's categories and predict their profiles. Fuzzy clustering reflects the learner's behavior which is more than crisp clustering. The author has mentioned three different types of learners such as bad, worker and regular students and not considered any standard learning style model for learning styles of the learners. The usage data of students is captured based on questionnaire approach of FSLSM and not considered real time usage data of learners.

Quang Do [10] have proposed the use of Gravitational Search Algorithm (GSA) in Neural Network and compared with Back Propagation Neural Network for classification. Xu and Zhang [21] combined the GSA with Neural Network and tested on different data sets for classification. The integration of GSA with NN is proved to give better accuracy and can be considered for real time usage data of learners on e-learning portal to classify new learner as per learning style model.

Chellatamilan et al. [7] have discussed about the data mining techniques, which can improve the learning content organization, and learning objects recommendations. Authors have focused on clustering and classification methods to predict the learning styles of the learners. The learning behaviors have been captured on Moodle portal, which is static in nature and does not support dynamic recommendations. Also the approach is implemented for Kolb learning style and can be tried for FSLSM.

Deborah et al. [9] have identified FSLSM as a suitable model for learning style prediction, especially in web environments. The authors have proposed to use Fuzzy rules to handle the uncertainty in the learning style predictions. However this work has not considered all the learner categories of FSLSM. Also the fuzzy rules are defined only on three parameters such as mouse movement, document length and image area.

Markowska-Kaczmar et al. [16] have discussed about providing individual learning path through recommendation system in e-learning. Authors have used clustering and classification for defining the learning profiles but not considered any specific learning style model for learning behaviors.

Based on the studies of related work, it has been identified to consider some important factors such as FSLSM categories of learners, real time captured usage data or learning behavior of learners for specific course, various learning objects in e-learning portal and integration of learning style in the portal to provide adaptive environment.

## 3 Overview of Proposed E-Learning System

In order to develop adaptive e-learning system, the learning behavior of each learner has to be captured on e-learning portal. For capturing the usage data of learn-

ers, an e-learning portal has been developed. The portal is developed using Dot Net Framework and deployed on GodAddy domain space to provide access to all learners on the Internet. The log file option of server is set to W3C extended log file format for the portal, which captures usage details of learners who are accessing the portal. The www.mitelearning.com portal is made available to second year engineering learners for the HTML course. The topics are divided into three categories based on prerequisite concept, main concept and advanced concept. The contents for the topics are made available for the learners through the portal in different file formats such as text (doc/pdf), video (mpeg/mp4) and demo (ppt/pptx). The learners can also go through the exercise modules that have been generated for each topic. The learners can also make use of the facility of different learning components that has been generated as pages to explore more about a specific topic. Portal is also implemented with additional learning components such as Assignments and Forum.

Following are the learning components defined for the e-learning portal.

- Announcement: Instructor can convey any important instructions or suggestions through this component. Learner can view by clicking on Announcement link.
- Email: Learner can send mail to instructor or any other learner in order to clarify some doubts or share some materials.
- Discussion Forum: Learner can start discussions on specific topics or can get the doubts solved online with other learners.
- Assignments: Learner can read the Assignment question and upload the completed assignment for assessment.
- References: Learner can see other important links or material by clicking on reference link of each topic.
- Exercise: Learner can solve multiple-choice questions on specific topic after going through the contents.
- Topic List: Learner can view the list of topics, which has been made available in the portal.
- My Account: Learner can modify his/her personal details in the portal.
- Pre/Main/Advanced: Topics have been arranged in the portal in three different levels so that learners can go through any of these levels of the topics based on his/her understanding.

## 3.1 Learning Objects Labeling as per FSLSM

The e-learning portal is a combination of learning components called " pages" and learning contents called "files". Each learning component with the contents in the portal is considered as learning objects. To map learning objects onto FSLSM categories all learning objects need to be labeled depending on the preferences of each learner. Mapping of Learning Objects on to FSLSM's categories are shown in Table 1. These mapped learning objects are considered as feature values for the clustering algorithm to label the learning behaviors.

**Table 1.** Learning objects as per FSLSM

| FSLSM Categories | Learning Objects |
|---|---|
| Active | Videos, PPTs, Demo, Exercise, Assignments |
| Reflective | PDFs, PPTs, Videos, Announcements, References |
| Sensing | Examples, PDFs, Videos, Practical Material |
| Intuitive | PDFs, PPTs, Videos, Forum, Topic, List, References |
| Visual | Images, Charts, Videos, References |
| Verbal | PDFs, Videos, Email, Announcements |
| Sequential | Exercise, References, Assignments, Sequential |
| Global | Topic Lists, References, Exercise, Assignment |

### 3.2 Characteristics of Web Pages as per FSLSM

The learners access various web pages on the e-learning portal. The web pages contain the following different learning objects.

- Learning contents in the form of text, ppt, videos etc.
- Different topic links, forum, email, index of the sequential contents, profile details etc.
- The text contents such as various images, charts, graphs, diagrams, flowcharts, video links etc.
- Assignments with upload and download option for the answers, multiple-choice questions and search option.

The characteristics of the web pages are defined based on the above learning objects that can map to FSLSM categories of learners. The e-learning system captures the usage access of all the learners along with the details of the pages and files accessed by each learner. The captured details of the pages are used to identify the learning styles of learners.

## 4 Methodology

To make traditional e-learning system adaptive, there is a prime necessity to integrate the learning styles into the portal and provide contents to the learners as per learning styles. Two important steps to integrate the learning styles into adaptive e-learning system are mentioned as follows;

- Identify the learning style using captured data (Sequences).
- Modify the e-learning portal as per identified learning styles.

The learning styles can be identified based on standard learning style model such as FSLSM. The captured data can be mapped to learning style categories of FSLSM, which is possible through Clustering and Classification techniques [1]. To perform the clustering and classification of learner's behavior following research questions need be accentuated.

- What are the various parameters/ feature values that need to be considered for each type of learner as per FSLSM categories?
- How to group the captured learner's behavior into FSLSM categories in order to assign labels to each type of learner?
- How to classify the new learner's behavior into defined FSLSM categories of the learners?

The above mentioned research questions can be answered as follows:

- The learning styles of learners are identified based on defined parameters or feature values of FSLSM.
- The grouping can be done using a clustering algorithm and categories of FSLSM can be assigned as a label to each cluster.
- The classification algorithm can be used to classify the new learner's behavior and identify the FSLSM category of the new learner.

For the implementation of clustering and classification algorithms, the learning behaviors have been captured in XML format that has been converted into learning sequences. The various learning components accessed by learners at each session of learner is defined as the learning sequence of that specific learner. The sequences are defined as per time spent on each file and page in each session. The sequence file consists of the captured sequences of all learners and each sequence contains the sequence id, session id, learner id and page id along with time spent on each page and file id along with time spent on each file. Time spent on page or file is considered in minutes.

The example of sequence is as follows.

$$S_1|G_2|u_{24}|p_5|1|p_7|4|p_{10}|12|f_3|10|p_9|3|p_7|19|f_{21}|19|p_{10}|1$$

Where $S_1$ is sequence id, $G_2$ is session id, $u_{24}$ is learner id, $p_5$ is page id, 1 is time spent on page, $p_5$ in minutes, $f_{21}$ is file id and 19 is time spent on file in minutes.

The identified sequences are not labeled and not mapped to learning objects of FSLSM's categories. To classify the new sequence of any learner into any of the eight defined categories of FSLSM, all the existing sequences of sequence file should be labeled or grouped based on the feature values. Hence to identify the learning styles of the learner clustering algorithm is used. To predict the learning style of the new learner classification algorithm is used. The proposed approach uses two algorithms.

- Fuzzy C Means (FCM) algorithm to label or map the sequences into FSLSM categories.
- Gravitational Search based Back Propagation Neural Network (GSBPNN) algorithm. To classify the new sequence of existing or new learner as per FSLSM categories.

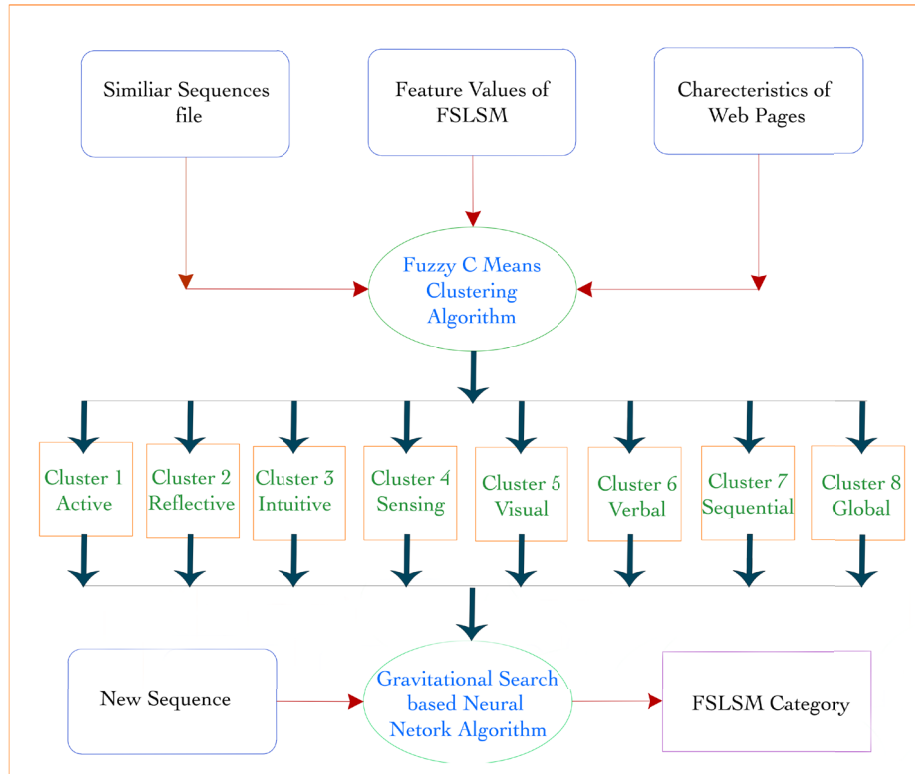The approach of labeling the sequences and classification of new sequence is shown in Fig. 2.

**Fig. 2.** Learning Style Identification Approach

Following subsections describe the technique of clustering and classification algorithms that are used in the proposed system.

### 4.1 Fuzzy C Means (FCM) Algorithm

The FSLSM's feature values and the sequences of learners are given as an input to the FCM algorithm. Based on the feature values the sequences are grouped together [13][3]. The FCM technique is entrusted with the task of assigning membership to each and every sequence values, which are linked to each singular cluster center. If the sequence values are very near to the cluster center, its membership moves further toward the precise cluster center. The sum of all the membership values of each sequence should be equal to one. Further, membership and cluster centers are revised according to the equation after each iteration. After finding the distance value, the value is compared with the threshold value. If the distance value is less than the threshold value then learner's sequence is grouped in that cluster. A sequence can belong to more than one cluster. The step-by-step procedure is shown in Algorithm 1.

In FCM algorithm, the learner's learning style is grouped based on the center value selection. The center values are calculated based on the feature values and assigned to

each cluster. Finally all the sequences are grouped into eight clusters of FSLSM categories and each sequence is assigned the label as per FSLSM.

$initialize\ m = real\ number\ greater\ than\ 1$
$M_{ij} = degree\ of\ membership\ function\ x_i\ in\ cluster\ j$
$F_i = input\ sequence, C_j = center\ of\ the\ cluster, \in\ =\ threshold\ value$
**Step 1**: $Initialize\ the\ membership\ function$
$M = [M_{ij}], M^{(0)}$
**Step 2**: $Calculate\ the\ center\ value\ by\ assigning\ weight\ to\ learning\ objects$
$C^K = C_j\ with\ M^{(K)}\ Where\ K$
$= number\ of\ center\ values\ calculated\ based\ on\ feature\ values$
$$C_j = \frac{\sum_i^n M_{ij}^m F_i}{\sum_{i=1}^n M_{ij}}$$
**Step 3**: $Update\ the\ membership\ function\ M^K\ and\ M^{K+1}$
$$M_{ij} = \frac{1}{\sum_{K=1}^C \left[ \frac{\|F_i - C_j\|}{\|F_i - C_K\|} \right]^{\frac{2}{m-1}}}$$
**Step 4**: $if\ |M^{K+1} - M^K| < \epsilon\ then\ stop, else\ go\ to\ step\ 2$

**Algorithm 1.**   FCM Algorithm for Labeling Sequences

## 4.2   Goodness of Clusters

The clustering algorithm should form high quality clusters with low similarity between the clusters. In order to evaluate the goodness of the clusters the following two indexes are used.

**Jaccard Index:** The Jaccard Index is used to check the similarity between the clusters [8]. The steps of Jaccard Index are shown in Algorithm 2. The FCM algorithm results in the grouping of some sequences into more than one cluster. Hence it is necessary to have the distance between the clusters closer to one based on Jaccard Index algorithm.

$intialize\ Clusters\ of\ Sequences$
**Step 1**: $Find\ unique\ sequences\ of\ each\ of\ the\ cluster$
**Step 2**: $For\ each\ cluster\ calculate\ Jaccard\ Index$
$$Jacc_{index}(C_i, C_j) = \left| \frac{C_i \cap C_j}{C_i \cup C_j} \right|$$
**Step 3**: $If\ Jacc_{index}(C_i, C_j)is\ equal\ to\ 1, the\ exactly\ similiar$
$otherwise\ different$
**Step 4**: $Compute\ the\ average\ Jaccard\ Index\ value\ to\ validate$
$all\ the\ clusters$

**Algorithm 2.**   Jaccard Index Algorithm

**Xie-Beni (XB) Index:** The Xie-Beni (XB) index defines the inter-cluster separation as the minimum square distance between cluster centers, and the intra-cluster compactness as the mean square distance between each data object and its cluster center. To identify the goodness of intra-cluster similarity [8], XB indexing is applied and checked for the similarity of clustered points of FSLSM eight categories. The XB index is calculated using equation 1.

$$X(Z;U,V) = \frac{\sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{M}\|Z_k - V_i\|^2}{C.min_{i\neq j}\left(\|V_i - V_j\|^2\right)} \tag{1}$$

Where X is the XB Index, Z is the center value, U is the member function, V are the center values and μ is the mean value.

### 4.3 Gravitational Search based Back Propagation Neural Network (GSBPNN) Algorithm

After labeling the sequences in the FCM algorithm, all the sequences are used to train the classification algorithm. In order to provide the Adaptive User Interface, the new sequence of learner's session need to be classified as per eight classes of FSLSM. To classify the sequences, Gravitational Search based Back Propagation Neural Network algorithm is used. The new sequence of each learner is given as input to neural network. The Gravitational Search is used to improve the accuracy of the basic neural network classification algorithm. GSA is a heuristic optimization algorithm which is based on Newton's law of gravity and interactions between the mass. The algorithm consists of number of agents which are selected randomly or based on specific condition. These agents are considered as communicating objects through the force of gravity. The positions and masses of the objects represent the solution to the specific problem. The iterations of GSA are based on the changing positions. The change in the positions are defined by the fitness function, velocity and acceleration. These parameters change in every iteration of GSA to provide better weight values [21][17]. Neural Network (NN) is used for processing of information that is based on nervous system of biological process. The components of this systems are brain, neurons and connection between them for information process. NN consists of large number of neurons, connected with each other to solve the specific problem. The applications of NN are classification, pattern recognition, image compression, prediction etc. NN works on the learning process which involves the adjustments of weights which are assigned to the connections between the neurons. Neural networks are organized in layers, where layer consists of many interconnected nodes. These interconnected nodes are mimicking the characteristics of biological neurons by processing the input data. The output of each node is called as the activation or node values. The connection between the nodes is represented by weights, which could be positive or negative. These weight values decides how input data are related with the output data. Weight values are determined at the time of training for each iteration through the neural network. Thus, the network learns to identify different classes based on the input data characteristics.

After the training procedure the neural network could be applied for the classification of new data. Using the following steps, classification is performed by the trained neural network.

- Activation of the network input nodes by using data sources that match the data sources used to train the network.
- Forward the data through the neural network i.e. the interconnected nodes.
- Activation of the network output nodes.
- Activation of network output nodes decide the outcome of the classification.

Back Propagation Neural Network (BPNN) learns through the algorithm called Back Propagation. With Back Propagation, the input data is repeatedly presented to the neural network. With each presentation the output of the neural network is compared to the desired output and an error is computed. This error is then fed back (back propagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output. This process is known as training. After training the Neural Network the model is validated and tested [18]. The GSA gives optimized weights which can be used in the process of BPNN. The accuracy of classification algorithm is based on the weights assigned the connections between the neurons in the BPNN. To assign the optimized weights Gravitational Search Algorithm (GSA) is used. The algorithm classifies the learner's sequence into multiple classes i.e eight categories (classes) of FSLSM. The modified GSA with BPNN is given in algorithm 3.

The basic BPNN steps are showed in algorithm 4. The BPNN algorithm is also implemented for same number of sequences and results are compared with modified GSBPNN that is discussed in the next section.

*intialize*

**Step 1**: *Initialization of agents*: *Initialize the positions for N umber of agents*

$X_i = (x_i^d, \dots \dots, x_i^d, \dots \dots, x_i^n \ for \ i = 1,2,3, \dots \dots, N$

$x_i^d$ *is the position of* $i^{th}$ *agent in the* $d^{th}$ *dimensions*

**Step 2**: *Compute the fitness function for each agents*

$fit_j(t)$*represents the fitness value of the* $j^{th}$ *agent in the iteration t*

*best* (t) *and worst* (t) *are given below for maximization or minimization of fitness at iteration t*

*best* $(t) = \min(fit_j(t))$ *and worst* $(t) = \max(fit_j(t))$ *where* $j = 1 \ to \ N$

*or*

*best* $(t) = \max(fit_j(t))$ *and worst* $(t) = \max(fit_j(t))$ *where* $j = 1 \ to \ N$

**Step 3**: *Compute the gravitational cost* $G(t)$

$$= G_0 \left( \frac{Itr}{Total \ Itr} \right) \ where \ G_0 \ is \ constant$$

**Step 4**: *Compute the gravitational mass* $m_k = \dfrac{fit(t) - worst(t)}{best(t) - worst(t)} \ M_i(t)$

$$= \frac{m_k}{\sum_{k=1}^n m_k}$$

**Step 5**: *Caculate the acceleration* $a_i(t) = \frac{F(t)}{M_i(t)}$ *where* $F = total \ force$

**Step 6**: *Update the agent velocity and positions*

$v_i(t + 1) = rand * v_i(t) + a_i(t) \ and \ x_i(t + 1) = x_i(t) + v_i(t + 1)$

*The velocity values are considered as weight values to assign to neurons*

**Step 7**: *Initialize neural network with two nodes for the input layer, three nodes for the hidden layer and eight nodes for the output layer*

**Step 8**: *Assign the velocity values to neurons to reduce the Mean Square Error (MSE)that is* $w_p = v_p$. *Where p is the number of connections between the neurons in the network*

*Activate unit$_i$ in the input layer* $X_i = a_i(m)$ *where* $i$

$$= 1 \ to \ 5 \ input \ parameters \ and \ a_i(m) \ is \ the \ input \ vector$$

**Step 10**: *Activate unit$_j$ in the hidden layer* $Y_j^h = \sum W_{ij}(x) + b_i$ *where* $i$

$$= 1 \ to \ 8 \ (Hidden \ layer)$$

**Step 11**: *Output signal from unit$_j$ in hidden layer* $\Theta_j^h = \dfrac{1}{1 + \exp\left(-X_j^h\right)}$

**Step 12**: *Activate unit$_k$in the output layer*

$Z_k^0 = \sum W_k^j \Theta_j^h \ where \ k = 1 \ to \ 8 \ output \ classes$

**Step 13**: *MSE for the* $k^{th}$ *output* $E = \dfrac{1}{n} \sum (d_k - o_k)^2$

*where* $d_k$ *is the actual output,* $o_k$ *is the experimented output and* $n$ *is the number of patterns in the samples*

**Algorithm 3**.    GSBPNN Algorithm for Classification of Sequences

$intialize$
**Step 1**: $Initialize\ two\ nodes\ for\ the\ inpute\ layer, three\ nodes\ for\ the$
$hidden\ layer\ and\ eight\ nodes\ for\ the\ output\ layer$
**Step 2**: $Activate\ unit_i\ in\ the\ input\ layer\ X_i = a_i(m)\ where\ i$
$= 1,2\ input\ parameters\ and\ a_i(m)\ is\ the\ input\ vector$
**Step 3**: $Activate\ unit_j\ in\ the\ hidden\ layer\ Y_j^h = \sum W_{ij}(x) + b_i\ where\ j$
$= 1\ to\ 3\ (Hidden\ layer)$
**Step 4**: $Output\ signal\ from\ unit_j\ in\ the\ hidden\ layer\ \theta_j^h = \dfrac{1}{1 + \exp(-X_j^h)}$
**Step 5**: $Activate\ unit_k\ in\ the\ output\ layer\ Z_k^0$
$= \sum W_k^j \theta_j^h\ where\ k = 1\ to\ 8\ output\ classes$
**Step 6**: $MSE\ for\ the\ k^{th}\ output\ E = \dfrac{1}{n} \sum (d_k - o_k)^2$
$Where\ d_k\ is\ the\ acutual\ output, and\ o_k\ is\ the\ experimented\ output\ and$
$n\ is\ the\ number\ of\ patterns\ in\ the\ samples$

**Algorithm 4**.    BPNN Algorithm for Classification of Sequences

## 4.4    Validation of Algorithm

The classification algorithm is validated using k-cross fold validation technique. This technique is commonly used for validating the performance of classifier by dividing the data set into k cross folds. For given set of M sequences, one data set is used for training and k-1 data sets are used for testing the classifier. The steps are shown in Algorithm 5.

$intialize\ M = Number\ of\ sequences\ as\ a\ training\ set$
$k = Number\ of\ folds$
$n_i = Number\ of\ sequences\ classified\ wrong\ in\ fold_i$
**Step 1**: $Arrange\ the\ training\ sequences\ M\ as\ a\ dataset$
**Step 2**: $Divide\ the\ training\ sequences\ into\ k\ parts$
**Step 3**: $For\ each\ i\ where\ i = 1,2, \ldots \ldots \ldots, k$
**Step 3.1**: $Train\ the\ classification\ algorithm\ using\ all\ the\ sequences\ except$
$the\ fold_i$
**Step 3.2**: $Test\ the\ algorithm\ for\ all\ the\ sequences\ of\ fold_i$
**Step 3.3**: $Compute\ n_i: wrongly\ classified\ sequences$
**Step 3.4**: $Compute\ the\ estimate\ E = \dfrac{\sum_{i=1}^{k} n_i}{m}$

**Algorithm 5**.    BPNN Algorithm for Classification of Sequences

### 4.5 Performance Metrics for Classification Techniques

The accuracy of classification algorithms is measured based on the number of correct predictions. To prove the validity of the predictions and robustness of the models, accuracy alone is not sufficient. However there is a requirement to calculate the precision and recall for each class and analyze the performance.

Precision (P): The ratio of the number of relevant sequences identified to the total number of irrelevant and relevant sequences identified. Precision is calculated as in equation 2.

$$P = \frac{TP}{TP+FP} \tag{2}$$

Recall (R): The ratio of the number of relevant sequences identified to the total number of relevant sequences in the sequence file. Recall is calculated as in equation 3.

$$R = \frac{TP}{TP+FN} \tag{3}$$

F-Measure (F1 Score): Since the class distribution is uneven, it is always better to calculate the F1 score. The F1 score provides the weighted average of precision and recall. Hence it considers both False Positive (FP) & False Negative (FN) as in equation 4.

$$F1 = \frac{2(R*P)}{(R+P)} \tag{4}$$

Accuracy (A): The ratio of correctly predicted sequences as in equation 5.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Where

- True Positive ($TP$): Number of sequences that are correctly classified to be positive.
- True Negative ($TN$): Number of sequences that are correctly classified to be negative.
- False Positive ($FP$): Number of sequences that are incorrectly classified as positive.
- False Negative ($FN$): Number of sequences that are incorrectly classified as negative.

# 5    Experimentation and Result Analysis

The www.mitelearning.com portal for HTML course was made available to second year engineering students for duration of two months. Total 108 learners participated in HTML online course and 1235 learning sequences are captured in the web log data. The sequences are pre-processed and are labeled according to eight categories of FSLSM using FCM clustering algorithm. The FSLSM mapped learning objects are considered as feature values for labeling. Also the characteristics of web page are identified based on feature values of FSLSM. These characteristics are used to understand the learning styles of the learner who is accessing a specific web page. Around 1235 sequences are grouped into eight clusters of FSLSM. After clustering the existing sequences, the classifier is trained using GSBPNN algorithm to identify the FSLSM category of the new learner. The structure of experimented neural network is shown in Fig. 3 where all the parameters of each sequence is assigned to each input node i.e. $x_1$ to $x_5$. The parameters of each sequence used for prediction are as follows;

- Filename/Pagename
- Type of file/page
- Time spent on file/page
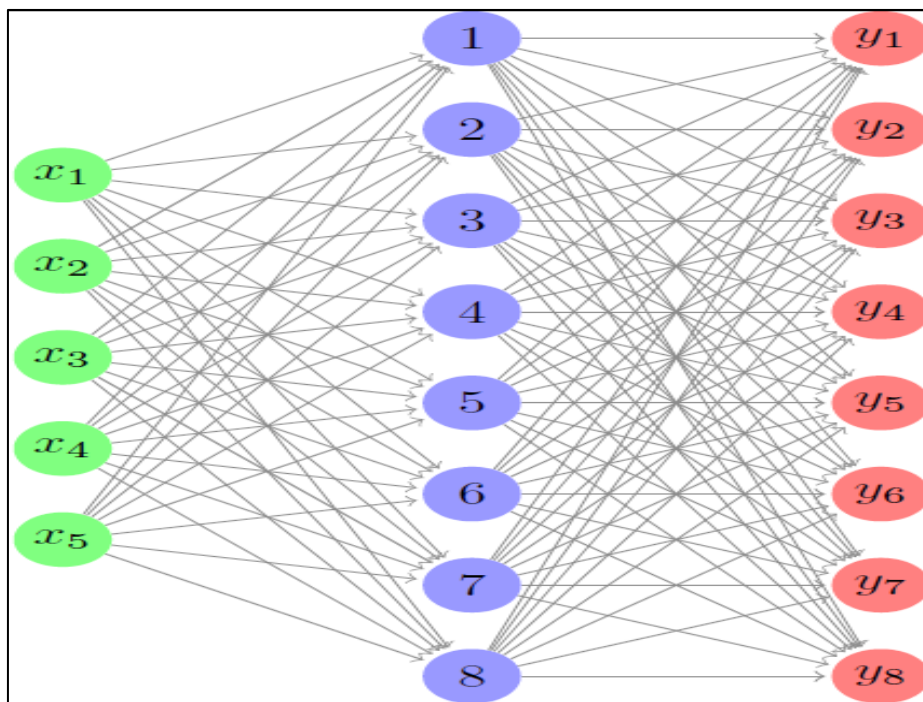- Frequency of file/page
- FSLSM Category



**Fig. 3.** Structure of NN used for GSBPNN

The eight categories of FSLSM are considered as the output nodes for neural network [10]. The number of nodes for hidden layer is calculated as shown in equation 6 where the final value will be truncated to identify the nodes [14].

$$\text{Nodes} = \frac{(Input\ Nodes * Output\ Node) * 2}{3} \qquad (6)$$

To validate the performance of the classification algorithm, k-fold cross validation algorithm is used [3]. The GSBPNN algorithm is compared with basic BPNN using the performance metrics for classification techniques.

### 5.1 Clustered Sequences based on FCM Algorithm

1235 sequences are clustered using FCM algorithm and labeled with FSLSM categories of learners. Center values for each cluster are computed based on the feature values i.e. based on learning objects to FSLSM categories. The result of clustering is shown in Table 2.

**Table 2.** Results of FCM Algorithm

| Clusters | Number of Sequences |
|----------|---------------------|
| Active | 169 |
| Reflective | 144 |
| Sensing | 138 |
| Intuitive | 162 |
| Visual | 256 |
| Verbal | 133 |
| Sequential | 144 |
| Global | 152 |

Since some of the sequences belong to more than one cluster, the total number of sequences in cluster is more than the total number of sequences in the output set. This is because one learner can belong to more than one type of categories as per FSLSM spectrum. It is analyzed that, the learner can be active and at the same time interested in different types of visual contents on the portal to understand the specific topic [3][8].

The goodness of clusters is computed by applying Jaccard Index Algorithm. The average Jaccard Index computed for eight clusters is 0.72. The Jaccard Index value is not exactly 1, as some sequences are clustered into more than one cluster. The validity of FCM is also evaluated by calculating XB index value for all the clusters. The calculated index value is 0.02. The minimum index value defines the best partitions of the clusters.

## 5.2    Classification of Sequences based on GSBPNN Algorithm

The GSBPNN algorithm is used to classify the new sequence as per the eight categories of FSLSM. Once the model is trained based on sequences, which are labeled as per FCM, the new sequence can be classified and learning styles of the learner identified. The algorithm is executed for various number of Iterations (I) and performance metrics are obtained. These metrics are compared with the metrics of BPNN, as shown in Table 3.

According to values defined in the Table 3, the accuracy of GSBPNN classification algorithm is 95.93% for 200 iterations. As the number of iteration increases, the accuracy also increases but algorithm takes more time for execution.

**Table 3.**  Comparison of Classification Algorithms

| I | GSBPNN | | | | BPNN | | | |
|---|---|---|---|---|---|---|---|---|
| | A | P | R | F1 | A | P | R | F1 |
| 60 | 87.16 | 88.09 | 90.30 | 88.89 | 75.13 | 76.38 | 79.43 | 77.20 |
| 100 | 89.45 | 90.03 | 92.18 | 90.87 | 77.09 | 78.18 | 81.02 | 79.03 |
| 200 | 95.93 | 96.33 | 99.05 | 97.67 | 78.83 | 79.90 | 83.10 | 80.12 |

The proposed system emphasizes mainly on two important aspects of AES. One is dynamic way of identifying the learning styles by capturing the web logs in real time and second is consideration of learning style model to categorize the learner. In AES, the need for identifying the learning styles is mainly for adapting the contents and user interface on the portal. In this scenario, the system should be able to identify as well as adapt the learning styles and to the changing learning patterns. The proposed system captures and identifies the learning styles of the learners dynamically as the learner progresses on the portal for the particular course. Also the learning styles are identified based on standard FSLSM learning style model which has eight categories of learning styles. Any specific learner fall into any one of the eight categories of the FSLSM initially and category changes over the period of time. The dynamically changed learning styles are considered in order to provide adaptive user interface and contents on the developed portal.

Some of the existing systems discussed in literature survey are static in nature where the learning styles have been captured once and contents are provided accordingly. Also some of the other systems categorize the learners into general categories such as beginner, advanced, expert and so on.  Such type of systems do not focus on learning abilities. The proposed system provides the learning styles of the learner during the learning process. The system is tested in real time environment by offering the online course and the result of prediction is validated.

## 5.3    Time Complexity of Clustering and Classification Algorithms

Two algorithms have been used for experimentation, where FCM is used once to label all the sequences and GSBPNN algorithm have been implemented to train the sequences with iterations. Thus the time complexity of both the algorithms are com-

pared separately. Following are the parameters and complexity for FCM and GSBPNN:

- The time complexity of FCM algorithm is $O(ndc^2i)$

Where $n$ is the number of data points, c is the number of clusters, d is the number of dimensions and i is the number of iterations.

- The time complexity of GSBPNN algorithm is $O(vXYZ_i)$

Where $v$ is the number of weights to be assigned to all the nodes, $X$ is the number of input nodes, $Y$ is the number of hidden nodes, $Z$ is the number of output nodes and $i$ is the number of iterations.

## 6      Conclusion

To enhance the efficiency of online course, Adaptive E- learning Systems (AES) is a promising research area. A necessary requirement in this area is to identify the learning styles of the learners as per standard Learning Style Model, which is useful to provide adaptive learning experiences. In this context the work describes a methodology to automatically detect and identify the learning styles of learners using Web Log Analysis approach.

The captured web usage data is pre-processed and converted into XML format to identify the unique sequences of each learner as per their sessions. These sequences are mapped to eight categories of Felder-Silverman Learning Style Model (FSLSM) based on the learning objects defined for each category. The mapping has been achieved by applying Fuzzy C Means (FCM) algorithm on the sequences and clustered them into eight groups of FSLSM which are defined as Active, Reflective, Sensing, Intuitive, Visual, Verbal, Sequential and Global. Some sequences are assigned multiple labels according to the featured values of FSLSM learning objects.

The output of FCM algorithm is checked for goodness of the clusters. The sequences are further used for classification using Gravitational Search based Back Propagation Neural Network (GSBPNN) model. Applying k-cross fold mechanism the classification algorithm is validated. The model is executed for number of iterations and compared for various metrics with basic BPNN algorithm. The result shows that the GSBPNN algorithm performs better than BPNN algorithm.

The future work will focus on to provide the adaptive contents and interface to the new learner based on the learning style of the learner, which is identified through GSBPNN algorithm.

## 7      References

[1] Manal Abdulaziz Abdullah. Learning style classification based on student's behavior in moodle learning management system. Transactions on Machine Learning and Artificial Intelligence, 3(1):28, 2015.

[2] Manal Abdullah, Wafaa H Daffa, Reem M Bashmail, Mona Alzahrani, and Malak Sadik. The impact of learning styles on learner's performance in e-learning environment. International Journal of Advanced Computer Science and Applications, 6(9):24–31, 2015.

[3] Oluwatoyin Catherine Agbonifo. Fuzzy c-means clustering model for identification of students' learning preferences in online environment. Int J of Computer Application and Information Technology, 4(1):15–21, 2013.

[4] Ahmed Al-Azawei and Atta Badii. State of the art of learning styles-based adaptive educational hypermedia systems (ls-baehss). International Journal of Computer Science & Information Technology, 6(3):1–10, 2014. https://doi.org/10.5121/ijcsit.2014.6301

[5] Jason Bernard, Ting-Wen Chang, Elvira Popescu, and Sabine Graf. Using artificial neural networks to identify learning styles. In Artificial intelligence in education, pages 541–544. Springer, 2015. https://doi.org/10.1007/978-3-319-19773-9_57

[6] Yi-Chun Chang, Wen-Yan Kao, Chih-Ping Chu, and Chiung-Hui Chiu. A learning style classification mechanism for e-learning. Computers & Education, 53(2):273–285, 2009. https://doi.org/10.1016/j.compedu.2009.02.008

[7] T Chellatamilan, M Ravichandran, RM Suresh, and G Kulanthaivel. Effect of mining educational data to improve adaptation of learning in e-learning system. In International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011), pages 922–927. IET, 2011.

[8] Zakrzewska Danuta. Using clustering technique for student's grouping in intelligent e-learning systems. In Intelligent Systems for Knowledge Management, pages 23–37. Springer, 2008.

[9] L Jegatha Deborah, R Sathiyaseelan, S Audithan, and P Vijayakumar. Fuzzy-logic based learning style prediction in e-learning using web interface information. Sadhana, 40(2):379–394, 2015. https://doi.org/10.1007/s12046-015-0334-1

[10] Quang Hung Do. A hybrid gravitational search algorithm and back-propagation for training feedforward neural networks. In Knowledge and Systems Engineering, pages 381–392. Springer, 2015.

[11] Juan Feldman, Ariel Monteserin, and Anala Amandi. Automatic detection of learning styles: state of the art. Artificial Intelligence Review, 44(2):157–186, 2015. https://doi.org/10.1007/s10462-014-9422-6

[12] Mofreh A Hogo. Evaluation of e-learners behaviour using different fuzzy clustering models: a comparative study. International Journal of Computer Science and E-education, 7(2):131–140, 2010.

[13] Manish Joshi, Ravindra Vaidya, and Pawan Lingras. Automatic determination of learning styles. In Proc. 2nd international Conference on Education and Management Technology, IACSIT Press, Singapore, volume 13, pages 121–132, 2011.

[14] Saurabh Karsoliya. Approximating number of hidden layer neurons in multiple hidden layer bpnn architecture. International Journal of Engineering Trends and Technology, 3(6):713–717, 2012.

[15] Annabel Latham, Keeley Crockett, and David Mclean. Profiling student learning styles with multilayer perceptron neural networks. In Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, pages 2510–2515. IEEE, 2013. https://doi.org/10.1109/smc.2013.428

[16] Urszula Markowska-Kaczmar, Halina Kwasnicka, and Mariusz Paradowski. Intelligent techniques in personalization of learning in e-learning systems. In Computational Intelligence for Technology Enhanced Learning, pages 1–23. Springer, 2010. https://doi.org/10.1007/978-3-642-11224-9_1

[17] Ebru Özpolat and Gözde B Akar. Automatic detection of learning styles for an e-learning system. Computers & Education, 53(2):355–367, 2009. https://doi.org/10.1016/j.comp edu.2009.02.018

[18] Norlina Mohd Sabri, Mazidah Puteh, and Mohamad Rusop Mahmood. A review of gravitational search algorithm. Int J Advance Soft Comput Appl, 5(3):1–39, 2013.

[19] Ondrei Takács, Jana Šarmanová, and Katerina Kostolányová. Analysis of learning styles for adaptive e-learning. In Digital Information Processing and Communications, pages 368–376. Springer, 2011. https://doi.org/10.1007/978-3-642-22389-1_33

[20] Huong May Truong. Integrating learning styles and adaptive e-learning system: current developments, problems and opportunities. Computers in Human Behavior, 55:1185–1193, 2016. https://doi.org/10.1016/j.chb.2015.02.014

[21] Bao-Chang Xu and Ying-Ying Zhang. An improved gravitational search algorithm for dynamic neural network identification. International Journal of Automation and Computing, 11(4):434–440, 2014. https://doi.org/10.1007/s11633-014-0810-9

## 8 Authors

**Sucheta V. Kolekar** is currently working as an Assistant Professor in the Department of Information and Communication Technology, MIT, Manipal University, Manipal, Karnataka, India. She has an experience of eight years in the field of teaching and research. She is pursuing her PhD in the area of Adaptive E-Learning. Her primary research interests include e-learning, web usage mining and cloud computing. She has published 10 papers in national and international journals/conference proceedings.

**Radhika M. Pai\*** (corresponding author) is currently a Professor in the Department of Information and Communication Technology at Manipal Institute of Technology (MIT), Manipal University, Manipal Karnataka, India. She is having a rich experience of 24 years in industry, research and teaching. She holds a PhD from National Institute of Technology, Surathkal, Karnataka, India. Her primary research interests include database management systems, data mining, operating systems, wireless sensor networks and big data analytics. She has published more than 30 papers in national and international journals/conference proceedings.

**Manohara Pai M.M.** is currently a Professor and Associate Director of Research and Consultancy, Manipal Institute of Technology (MIT), Manipal University, Manipal, Karnataka, India. He is having a rich experience of 25 years in industry, research and teaching. He holds a PhD from the University of Mysore. His primary research interests include wireless sensor networks, intelligent transportation system, internet of things, big data analytics and cloud computing. He holds six patents and has published more than 60 papers in national and international level. He has published two books, guided three PhDs and more than 65 master thesis. He is a Visiting Professor to ESIGELEC-IRSEEM, University of Rouen, France.