

Performance Comparison of Convolutional and Multiclass Neural Network for Learning Style Detection from Facial Images

F.L. Gambo^{1,*}, G.M. Wajiga², L. Shuib³, E.J. Garba², A.A. Abdullahi¹ and D.B. Bisandu⁴

¹Department of Computer Science, Federal University Dutse, Jigawa, Nigeria

²Department of Computer Science, Moddibo, Adama University of Technology, Yola, Adamawa, Nigeria

³Department of Information Systems, Faculty of Computer Science and Information Technology, Universiti Malaya, Malaysia

⁴Department of Computer Science, University of Jos, Nigeria

Abstract

Improving the accuracy of learning style detection models is a primary concern in the area of automatic detection of learning style, which can be achieved either through, attribute/feature selection or classification algorithm. However, the role of facial expression in improving accuracy has not been fully explored in the research domain. On the other hand, deep learning solutions have become a new approach for solving complex problems using Deep Neural networks (DNNs); these DNNs have deep architectures that are capable of decomposing problems into multiple processing layers, enabling and devising multiple mapping of complex problems functions. In this paper, we investigate and compare the performance of Convolutional Neural Network (CNN) and MultiClass Neural Network (MCNN) for classification of learners into VARK learning-style dimensions (i.e Visual, Aural, Reading Kinaesthetic, including Neutral class) based on facial images. The performances of the two networks were evaluated and compared using square mean error MSE for training and accuracy metric for testing. The results show that MCNN offers better and robust classification performance of VARK learning style based on facial images. Finally, this paper has demonstrated a potential of a new method for automatic classification of VARK LS based on Facial Expressions (FEs). Based on the experimental results of the models, this approach can benefit both researchers and users of adaptive e-learning systems to uncover the potential of using FEs as identifier learning styles for recommendations and personalization of learning environments.

Keywords: Learning Style, Artificial Neural Network, Facial Images, VARK learning-Style Model, Deep Learning

Received on 29 June 2021, accepted on 15 October 2021, published on 20 October 2021

Copyright © 2021 F.L. Gambo *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.20-10-2021.171549

*Corresponding author. Email: farouk4142@gmail.com

1. Introduction

Educational technologies have touched almost all aspects of modern learning. They open several opportunities for learners to have easy access to new knowledge through various learning objects, and for instructors to present information in many forms, such as text, pictorial, animation-based, audio presentations, and so on [1]. However, traditional educational computer-based systems,

for example, Massive Open Online Courses (MOOC), Learning Management Systems (LMS), Intelligent Tutoring Systems (ITS), as well as other types of educational systems, suffer from the absence of teachers to recognise the best method of delivery learning. Recently learner's modelling has been employed in these educational technologies to provide adaptivity and personalization of the learning environment. This is due to, that in real life, individual differs in ways and preferences. For example, sometimes learners might have a preference for visual learning material that is in the form of pictures, diagrams,

and graphs. In contrast, other learners might have a preference for audio learning material for assimilating new information. This concept of individual preference of learning is known as the preferred learning style [2, 3]. The application of this concept has now become popular in e-learning practice for improving its efficiency [4, 5]. Learning systems that deliver learning content based on learner's preferences are known as "adaptive learning systems" and their efficiency depends on the efficiency of the learner's model [3]. A learner model is an essential module found in modern adaptive e-learning systems, which represent learners' behaviour for decisions making. These learner models were created through a process called learner modelling [3] which, according to [6] can also be defined as "a process that deals with many cognitive issues such as determining the knowledge level, predicting the student's performance, identifying the misconception, and inferring preferences". Learner models in adaptive systems are either based on learning style and other personality traits that can be monitored from learners' interaction with the computer systems within the electronic learning (e-learning) [7].

Among interrelated Information Technology (IT), psychology and pedagogy studies that gained significant interest recently, learning style is the most useful and is considered among other personality traits in adaptive e-learning systems [7, 8]. Identification of students' learning styles could be of great help to the students in several ways; reducing assimilation time, specific searching of type of materials via the Internet by reducing the amount of time spent, improving learning outcome and many more [9]. Models were developed for the prediction of learning style using various learning behaviours (Predictors) and the classification of several algorithms. However, these approaches only yield an average precision of 77% [9]. This may be as a result of neglecting an important predictor and the use of a simple classification method [4]. This work is an extended study to the existing work "A framework for automatic detection of learning style from facial expressions" by [4] with the belief that the use of facial expressions will provide better accuracy for the prediction of learning style using deep learning approach.

Deep learning is among the outstanding approach in dealing with complex tasks, including pattern recognition, classification, and detection. From what the name entails, the difference between Deep Neural Networks and a shallow/ or network with a single hidden layer is the enormous number of layers network it has; as such, the original input can be transformed more in the deep neural network than the shallow networks. It is through the deep neural network that one can "learn" rigidly than in the shallow networks [10, 11, 21].

In line with the same research idea [4], this paper is set to investigate the emerging performance of the deep learning solution CNN and compare it with a Multiclass neural network for automatic detection of learning style using facial images. The research will help in providing the opportunity for the selection of better architecture to use for the prediction of individual learning styles from facial images.

2. Background

2.1. Overview of Learning Style

Learning style is intended to determine the preference and needs of the individual learner during learning [13]. These learning styles described what type of instruction or learning object a student preferred to internalise new information [12]. People differ in way of learning, and the type of instruction that is most suitable for them can be determined via their learning style. This assumption has gained increasing popularity in adaptive multimedia learning and e-learning in practice [12, 13, 14, 15, 16]. Learning is a composite of cognitive, emotional, and physiological features that serve as an indicator of how learners internalised new concepts and interact with the learning environment [17]. Individuals have unique learning methods. e.g. some like to process information visually (e.g. pictures, diagrams, graphs), while others prefer verbal ways (reading or listening) etc. These preferences are known as preferred learning styles [2].

There are two approaches in the area of learning style detection; traditional and automatic [13, 16]. Traditional detection of learning styles makes use of a questionnaire to detect students learning styles, where each of the learning style dimensions has its unique questions [13]. While the automatic detection makes use of a model developed from learner's records from a different data source in the system [16].

The automatic detection of learning styles can also be categorised into two based on the approaches used: the data-driven and the literature-based approach [14]. But each approach also differs from one another based on the attribute used (Behavioural, personal trait and so on).

The data-driven approach is targeted at constructing a classifier that emulates a learning style instrument and used sample data to develop a model [15]. Here, machine learning classification algorithms are mostly used to model users and produce their learning style/preferences [13, 16]. While the literature-based approach models the user's interaction with the system to get hints about students' LS to estimate the preference. This is achieved by using certain established rules [15, 16]. The automatic approach is said to be more practical and generally accepted regardless of the course domain since the focus lies on the content of the learning object [9, 15]. However, the two approaches rely on the interaction between the user and system which includes reading materials used, online chats, active collaboration discussion forums and online quizzes [9, 16].

The automatic prediction of learning preference follows a framework that comprises model development and integration of the model into adaptive systems. The development of the model generally starts with choosing the learning styles theory, learning styles attributes selection from data sources, model development, and finally model evaluation for the suitable application into the e-learning framework [12, 18].

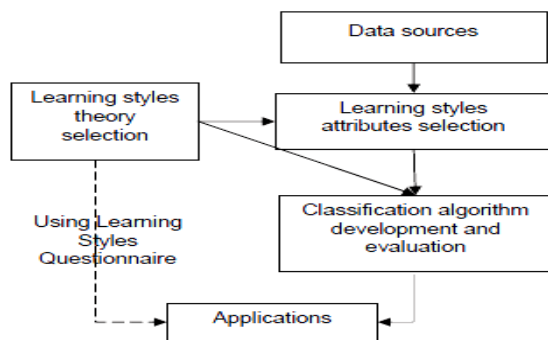


Figure 1. Learning style detection framework [18]

(i) Data Source for Learning Style Prediction

The first step in learning style detection is to identify the data sources that are used to identify the students’ behaviour. The students’ behaviour or preference is vital in the learning process to construct an accurate LS detection model. Different attributes (predictors) from different learning domains namely; cognitive, affective and psychomotor can be used in the prediction where each attribute has the potential toward predicting learning style [7, 18].

(ii) Attributes Selection

This is the second aspect of the detection of student learning styles. The attribute selection from students’ behaviour in a learning process enables the construction of learning style predictive models in e-learning platforms, where different attributes (predictors) from online data sources can be used to develop learning style predictive models [4, 18, 20].

(iii) Learning Style Theory Selection

The third step is the theory selection which plays a vital role in learning style detection. For this research, the VARK Model is selected which is originated from “Gardner’s theory of multiple intelligences”. VARK model has four dimensions which are: the visual dimension for those who prefer internalising new information from video, pictures, or diagrams, the aural dimension for those who prefer internalising new information either from what they heard or spoke, the read/write dimension for those who prefer internalising new information from printed text and the last is the kinesthetic dimension for those who prefer internalising new information through experience and practice [19]. However, the model doesn’t restrict learners to a single dimension but reveals the strength and weaknesses of the student in each of the dimensions. Therefore, one may belong to more than one dimension at a time [20].

2.2. Artificial Neural Network

Artificial Neural Network (ANN) processes information in a similar way that a biological nervous network does; it comprises a huge amount of neurons that are extremely unified which together work in solving complex tasks. The most important aspect of ANNs is their capacity to adapt to certain problems through ‘training’ the network. They work just like the human brain through learning from examples [21]. Just like the human brain, the same set of neurons can be used to solve different problems but with different network settings and training. The neural network can be used to derive meaning from large and complicated data, as well as coming off with patterns and trends that are not be seen by the human expert.

ANN is a collection of united neurons, arrange in layers that can transform given vector input into expected output result. Normally, a function process input to produce an output that can serve as an input to the next layer [22]. The network in ANN is by default in the form of a forward chain; this means that only the preceding layer feeds the succeeding layer but not vice versa. Each unit of input is associated with assigned weight but can change in a process called the training or learning phase, which differs from one problem to another [23].

In ANN, every unit in the network is called a neuron and is represented as a mathematical function (model) similar to a biological neuron [21]. To view the model mathematically,

Let n be the number of inputs with signals $x_1, x_2 \dots x_n$ where each input will be carried along with its weight $w_1, w_2 \dots w_n$. After that, a bias b is applied to produce an output in equation 1.

$$y = f(x) + b \quad (1)$$

Where

$$x = \sum_{i=1}^n x_i w_i + b \quad (2)$$

The symbol f in equation (1) above represents the activation function. The work of the activation function is to apply the certain fixed nonlinear function of an output which can be visualised on a single neuron in figure 2 below.

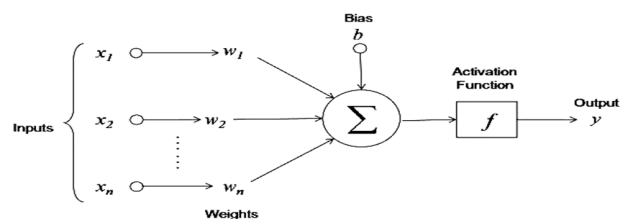


Figure 2. Shows a Single Artificial Neuron [21].

The activation function f is an essential property of a neuron that controls its functioning. One can also refer to the activation function as a step function, which could result in a binary (0 or 1) output of the neuron. The use of other non-linear functions in a small network of neurons can solve complex problems [21]. These activation functions could be sigmoid, tanh, ReLU, e.t.c.

2.2.1. Artificial Neural Network Architectures

There are various ways in which artificial neurons can be arranged and connected. In this section, all the network architectures discuss are feed-forward networks, meaning that neuron in a network is connected in an acyclic graph. This property provides output neurons with the possibility of being input to the next connected neuron. However, there is a possibility if a signal not propagates back in the case of a network that is not cyclically connected, therefore we can say that the network propagated forward in this case [21]. Whenever a network has only one hidden layer, that network is referred to as a "shallow network". While any network with a combination of an input layer, multiple hidden layers and an output layer is referred to as deep learning architecture [24]. This means multilayer feed-forward networks could differ in terms of depth by counting all layers in the architecture (N-layer network) except for the input layer. For example, 1-layer-network consists of just an input layer that is directly connected to an output layer, 2-layer-network consists of just an input layer, one hidden layer and an output layer. Generally, N-layer-network consists of N-1 number of hidden layer and output layer. There are four major types of deep learning architectures, which include deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN) [25, 26]. However, this study only discussed deep learning architectures used in the research.

2.2.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are also called ConvNets are described as deep, feed-forward neural network and which specialised mainly at analysing image data [25]. CNN models are mainly built from three (3) types of hidden layers; convolution layers, pooling layers and fully connected layers. This class of ANN have demonstrated exemplary performance on complex learning problems. CNN models are very powerful and have achieved impressive results in solving a lot of complex problems [21]. As the name implies, the main thing that differentiates CNN from ANN is its architecture that comprised of a large number of layers in the network. Due to the number of layers in the CNN, the original input is transformed many times than shallow networks. This way, the network can 'learn' harder tasks than shallow networks; for example, more complicated features can be extracted in image recognition [17]. The CNN architecture is motivated by the visual cortex of the human brain [27]. CNN consist of layers, namely; convolutional, pooling and fully connected layers.

- Convolutional Layers:

This layer is responsible for feature extraction from input images. A convolution is a form of linear operation used for feature extraction in which a tiny array of numbers called a kernel is applied over the input, which is a tensor of numbers. At each point of the tensor, and element-wise product between each element of the kernel and the input tensor is calculated and added to get the output value at the corresponding position of the output tensor, which is referred to as a feature map [48]. The convolution kernel is reserved as the weight, with its vector denoted as \mathbf{w} , and the pixel vector at the equivalent point of the image which is denoted as \mathbf{x} [49 51]. The result of the convolution of each position is calculated and transformed using equation (2). This technique is repeated with several kernels to produce an arbitrary number of feature maps that represent distinct features of the input tensors; different kernels can thus be thought of as separate feature extractors. The size and number of kernels are two fundamental hyperparameters that determine the convolution operation. The most used kernel is 3x3, although 5x5 or 7x7 can also be used.

- Pooling Layers:

In this layer, the spatial resolution of feature maps is reduced, which serve dimensionality reduction by [52]. After obtaining features via the convolutional layer, the next step is to integrate and classify these features [49]. If the classifier is given all of the features collected using convolution as input, it will have to do a lot of work. Pooling operations are common in convolutional neural networks [50], and the pooling layer is frequently placed behind the convolutional layer. By pooling, the convolutional layer's output feature vectors may be lowered, and the calculation quantity can be reduced while the results are enhanced, making overfitting less likely [49]. Because images are "static," it is easy to achieve this by lowering their dimensions. This indicates that elements that are beneficial in one part of the image are more likely to be beneficial in another. As a result, aggregating statistics of features at multiple locations is a suitable approach for describing huge images. Pooling is a process that integrates each element of the input and then produces a smaller feature map. Earlier studies use average pooling for aggregating all input values, while recent studies use max pooling for maximum aggregating value for receptive fields [25, 26, 28]. See figure 3 for illustration of the two.

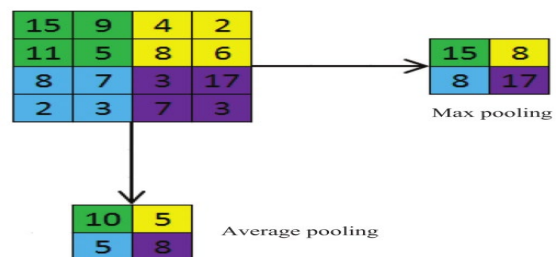


Figure 3. Illustration of Max Vs. Average Pooling [28]

- Connected layers:

This is the highest level in the network responsible for extracting more advanced abstract features [28]. The final convolution or pooling layer's output feature maps are typically flattened, that is, converted into a one-dimensional (1D) array of numbers (or vector), and connected to one or more fully connected layers, also known as dense layers, in which each input is connected to each output by a learnable weight [48]. Once the features extracted by the convolution layers and down-sampled by the pooling layers are formed, they are transferred to the network's final outputs, such as the probabilities for each class in classification tasks, by a subset of fully connected layers. The number of output nodes in the final fully connected layer is usually equal to the number of classes [48]. FC layer is composed of a sigmoidal neuron which sums the outputs of the last preceding convolutional layer while in some recent image classification tasks, softmax function at the last layer of the network which essentially converts the output of the last layer into a probability distribution [27].

3. Related Work

Several models for the detection of learning style were developed from the automatic approach as shown by literature. Models developed from this approach uses machine learning algorithms from either data mining fields or computational intelligence fields to identify LS preference from selected behaviour patterns.

[29] proposed the use of a feed-forward ANN configured with a 3-layer perceptron and trained with backpropagation under supervised learning to detect learning styles. The study used data obtained from 10 behaviour patterns elicited by the student while interacting with the system. For example, the kind of learning content student prefers among others. The study succeeded in building the network but produces only three out of the four FSLSM dimensions as output.

[30] construct a model through accessing various students' learning behaviours such as the number of visits, time spent, and answering questions to a learning object. The research uses certain rules and algorithms with a pattern of activities found with different VAK learning style dimension to build a model that detects learning style. Unfortunately, the average precision for the VAK learning style detection only yielded 52.78% accuracy.

[31] proposed an automatic detection approach for learning styles capable of adapting to the learner's wishes to provide learning objects that suit their learning style. The study combined both data-driven and literature-based methods, specifically through measuring user's time_spent on learning material recommended to each VARKLS dimension/class. The data-driven approach is planned to generate LS from; calculating the difference time_spent and predefined_time_three, result of learners' answer found in example and exercise section and time_pent on outline and content presented by the system. Despite the research opportunity presented by this study, the proposed hybrid model was not evaluated.

Another research by [32] succeeded in proposing a model for reviving prior knowledge from test questions using the Latent Semantic method to overcome the previous method (brainstorming, KWL and cognitive map) that were less effective and dynamic. [33] Further developed and evaluated an internal approach, this approach uses learner's personality trait (prior knowledge) to detecting learning style [34]. The research employed LSI techniques that generate prior knowledge of individuals using single value decomposition then predicts VARK learning style using ANN. a remarkable accuracy of 80% is obtained from the study.

[35] used a modified Back Propagation Neural Network (BPNN) algorithm with the gravitational search algorithm to predict the learning style of learners in real-time. The research captures learning behaviour in an e-learning portal using weblog mining and then maps each of the behaviour to the corresponding FSLSM category using the Fuzzy C Mean (FCM) algorithm. GSBPNN was found to outperform BPNN in the research with an accuracy of 95.93 before the 200th iteration. [36] introduced a different approach that uses ANN to detect student learning styles based on the dimension of the Felder-Silverman LS model. The research called the approach "LSID-ANN" which uses four different neural networks with the same configuration of 3-layered perceptron for all the FSLSM dimensions. Relevant behaviour patterns were used as input to the neural networks similarly [29] also believed it would yield a better result instead of using all attributes from the FSLSM dimension on a single ANN.

Similar research [37] introduces a novel approach for automatic detection of learning style, which harmonises the advantages of rule-based and machine learning techniques from AI. The rule-based method used was extended to consider the different weights of behaviour patterns in the research using a particle swarm optimisation algorithm. Similarly, [9] investigate the use of ANN algorithms train with backpropagation, three optimisation algorithms, namely; genetic, ant colony and particle swarm, to detect learning style with the same behaviour pattern used in [38] research. The investigation showed an improvement in the existing average precision from 67% to 80%. The ANN approach also showed the most promising result when benchmarking with [9].

A successful proposal by [39] proved that other attributes like emotion strongly influence students' learning styles. This research suggests that groups can be established for further studies that involve a relationship between affective factors and learning style. Thus [4] proposed a "framework for automatic detection of learning style from a facial expression using Convolutional Neural Network". The research proposed that through identifying emotional classes that positively correlate with a different dimension of learning style, and effective database can be formed, and classification of learning style can be done on the dataset. The use of the deep learning approach was also proposed in the paper with the belief that CNN can recognise complex patterns on images. A prediction using the same idea was evaluated on the VARK model by [12], the research demonstrated the feasibility of [4] idea through experiment.

After training the CNN, the result of the experiment showed a performance of 71.03% for Visual learners, 50.90% for Aural learners, 71.01% for Reading learners, and 68.48% for kinaesthetic learners. This paper uses a similar approach with [12] to develop and compare two Neural networks that model cognitive-affective features for learning style detection.

[40] proposed a deep learning approach, the approach was proposed for learning style classification on the large number of behaviour in an online environment called DBNLS. The approach employed a deep belief network where the multi-layered RBM of the DBN model was used for the extraction of relevant features for the identification of learning style, and the BP layer of the DBN fine-tune the network and prediction of the learning style. The researcher trained and evaluated the model on the dataset that was labelled based on the FSLSM Index of Learning Style (ILS) questionnaire from the weblog. The model achieved an accuracy of 84% for Act/Ref class, 81% for Sen/Int, 89% for Vis/Vrb, 69% for Seq/Glo, and 79% for Soc/Asoc, which outperformed both BP and BN when compared.

[41] proposed an approach called "deep multi-target prediction" that applied deep neural networks on different class/dimensions of FSLSM, referred to as "target". The research investigated a 3-layered network with 2 hidden layers and an output layer in improving the accuracy of automatic identification of learning style. This new method identifies features/descriptors from previous literature related to a specific dimension of FSLSM, dataset constructed from the raw information collected in the massive open online course (MOOC) and result obtained from students' filled ILS questionnaire. Finally, a remarkable result was achieved through training the network with a various number of hidden layers and neurons. The final result shows that the best model achieved with 26 neurons and a 3-layered network (2-hidden layer and output layer) where 85%, 76%, 75%, and 80% accuracy recorded for active/reflective, sensing/intuitive, visual/verbal and sequential/global respectively.

[42] proposed a new mechanism of automatic detection of learning style based on EEG features. The study considered the Felder-Silverman model's processing dimension because there is a behavioural difference between reflective and active learners. The mechanism first labelled learners according to their actual learning style from the result of the questionnaire they filled, RAPM was used to stimulate learning style difference by asking the subjects to think logically based on certain associated rules to answer questions. Questions asked were simple to avoid cognitive loads, and the goal is to stimulate brain processing to enable data collection. Data were collected from a total of 504 experiments using wires EEG called Emotiv EPOC+. The EEG data collected were labelled using the actual learner's LS and divided into a ratio of 80:20 for training and testing. The study first trained the classification model using SVM with backpropagation and later used a one-dimensional convolutional neural network to improve the existing EEG classification model. Finally, the mechanism demonstrated another significant and efficient learning style recognition with an accuracy of 71.2%. However, the mechanism only demonstrated on processing dimension of the Felder-Silverman model.

4. Methodology

The methodology in this study involves stages used for the development of the two networks for learning style detection using facial images. To model the learners' facial images for the learning style detection, this research focuses on CNN architectural design, which is a variant of [26] network and compare its performance with Multiclass ANN. This enables us to determine a better performance either through working on the feature extraction and classification separately (Multiclass ANN) or simultaneously (CNN) [43]. The steps involve data collection, pre-processing, model development and evaluation, as seen in Figure 4.

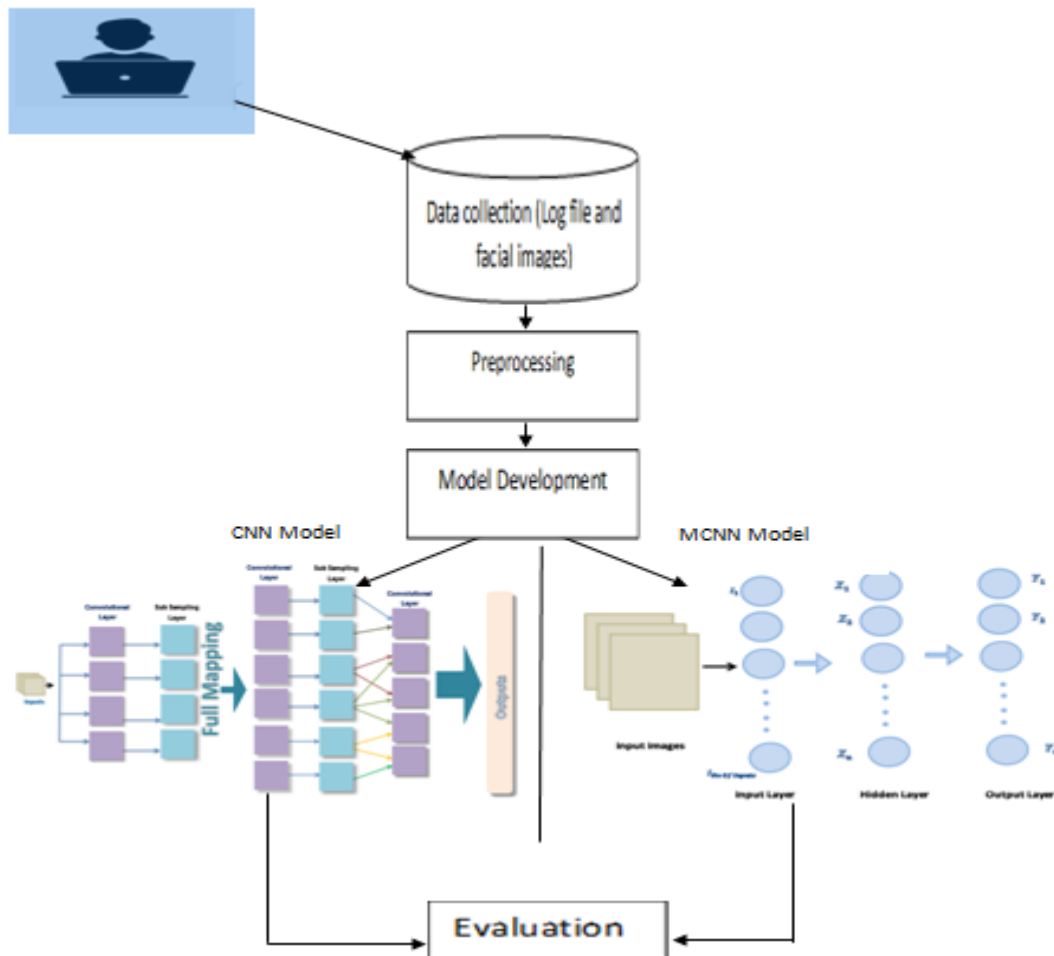


Figure 4. Framework for the proposed model comparison

4.1. Data Collection

The facial images and student behaviours were collected from learners interaction with our developed learning system specifically designed to contain learning object that corresponds to each of the four dimensions of VARK LS Model. The data gathering tool was developed using C# programming language and MySQL database. Images captured were taken at fixed intervals from students learning process (that is one image per second) to enable us to capture different facial expressions elicited by learners during their learning processes.

4.1.1. Mapping Facial Images onto VARK Model

To model learners' facial images for learning style detection, a new dataset (facial images) would have to be labelled by mapping the facial images onto different VARK learning styles model classes. For this reason, this research used some learner's traces as a reference to label the facial images for the VARK LS classification. Although many datasets were developed and used for automatic learning style detection, the strategies used in labelling datasets are

still vague. Some researcher does not even provide a clear method they used in labelling their dataset [41]. However, our dataset needs to be labelled based on VARK learning theory model to enable a supervised training algorithm for the classification task. Therefore, this research used the same rules used [30] to labelled our dataset. The approach uses two types of measures (count and time(s)) for calculating the number of visits towards VARK content and the learner's time spent in visiting each VARK content (see Table 1) in the Appendix.

Firstly, the rule determined the predicted time for each learning object and the actual time a user spent on each learning objects to calculate the ratio time for each learning object RTSL.

$$RTSL_j = \frac{\sum_i^n Time_spent}{\sum Predicted_time_j} \quad (3)$$

Where j = (v, a, r, k) to represent each of the VARK learning object and i= (1,2,3,...n) is the number of visits to the learning objects.

Step1: Calculate ($Predicted_Time_j$) for $j = (v, a, r, k)$ represents each of the VARK learning object.

Step2: Calculate $RTSL_j$ using equation (3) for $j = (v, a, r, k)$ represent each of the VARK learning object.

Step3: $LS == Neutral$

$IF(RTSL_v > RTSL_a)$ then $LS == Visual$
Else

$IF(RTSL_a > RTSL_r)$ then $LS == Auditory$
Else

$IF(RTSL_r > RTSL_k)$ then $LS == Reading$

Else $LS == Kinesthetic$

4.2. Data Pre-processing

After labelling the datasets, the data is imbalanced across the VARK classes. Thus data pre-processing is needed to balance the dataset to avoid the problem of overfitting some classes over others. According to [44], if there is a priori knowledge of a class imbalance, one straightforward method to reduce its impact on model training is to select a training set sample to have roughly equal event rates during the initial data collection [44]. However, if prior sampling is not possible, down-sampling and up-sampling of data reduce the effect of imbalance during model training. Down-sampling techniques reduce the number of some classes to the minority class instance, while Up-sampling is any technique that simulates or imputes additional data points to improve balance across classes. In this study, we used the Down-sampling technique to reduce the number of images from VARK classes to the minority class.

4.3. Model Development

The study proposed two potential classes of feed forward Neural Networks (NN), namely Multi-Class Neural Networks (MCNN) and Convolutional Neural Networks (CNN). The first proposed network is a Multiclass feed-forward neural network which consists of the input layer, N-Hidden layer, and output layer. While CNN harnesses the three convolutional layers, two pooling layers and the output layer.

4.3.1. Proposed Multiclass Neural Network

The first proposed network is a Multiclass feed forward neural network which consists of the input layer, N-Hidden layer, and output layer. Figure 3 presents the structure of a variant of the network consisting of I_n the number of neurons in the input layer determined by the size of the input image, X_n denotes the number of hidden neurons determined through parameter variation, and the number of a layer can be optimised accordingly. Each neuron manipulates its input using the equation and 1 & 2, while Y_n represents the number of neurons in the output layer, which is determined by the number of classes (VARK) in the problem space.

4.3.2. Proposed Convolutional Neural Network

The proposed network is also a class of feed forward deep neural networks (LeNet) that has different structure from the first networks mentioned above. The network couples three convolutional layers, two pooling layers and output layer, where each of them is fed into an activation function. All images were resized to 32X32 to reduce the time complexity. The first layer C1 takes the Images with 5X5 learnable filters and produce six (6) sets of 28X28 feature maps. P1 then subjected feature map pooling over a 2X2 window, resulting in six sets of 14X14 feature maps which C2 takes and produce new sixteen feature maps of 10X10. These sixteen C2 feature maps are also over a 2X2 window, resulting in sixteen sets of 5X5 feature maps. The result of C2 with 5X5 learnable filters convolved for the third time with no stride and produce one hundred and twenty 5X5 sized filters that are fully connected to each feature maps. Finally, the output of the fully connected layer would then have to be fed into the performance metric to determine the error [12, 45].

- Convolutional layer: -

This is the first and core building block of CNN, which uses convolution operation * on the input image and learnable filters, also called a kernel.

For each of the two-convolution layers, say C1 and C2 in the network (L), the input images are mapped with a specified filter to produce a set of feature maps using the two operations, respectively.

$$C1_{i,j}^k = \sigma \left(\sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^k * I_{i+m,j+n} + b^k \right) \quad (4)$$

Where C1 is the representation of the first convolutional layer feature maps, k is filter number, while (m, n) and (i, j) are the indices of kth filter and output, respectively.

$$C3_{i,j}^k = \sigma \left(\sum_{d=0}^5 \sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^{k,d} * P2_{i+m,j+n}^k + b^k \right) \quad (5)$$

Where C3 represent sixteen output feature maps, k is filter number, while (m, n) and (i, j) are the indices of kth filter and output respectively, d is the index of the number of channel in the input

$$C5_{i,j}^k = \sigma \left(\sum_{d=0}^{15} \sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^{k,d} * P2_{i+m,j+n}^k + b^k \right) \quad (6)$$

Where C5 represent one hundred and twenty (120) output feature maps of with size, (m, n) represent the indices of kth filter and indices of output are the same size with input, k representing the filter number.

- Sub Sampling Pooling Layer

In CNN, what follows each convolution and activation function before the final convolution is the pooling layer which works as dimensionality reduction. This pooling layer proceeds with every feature map from the convolution layer and remove less important data but conserves the detected feature in smaller representation as specified by the architecture. For each of the two pooling layers, say P2 and P4 in the network (L), the input feature map from the convolution layer are reduced using filter. Equation 4 and 5 below shows how $P2_{(i,j)}^k$ and $P4_{(i,j)}^k$ are to be calculated

$$P2_{(i,j)}^k = \max \begin{pmatrix} C1_{(2i,2j)}^k & C1_{(2i,2j+1)}^k \\ C1_{(2i+1,2j)}^k & C1_{(2i+1,2j+1)}^k \end{pmatrix} \quad (7)$$

$$P4_{(i,j)}^k = \max \begin{pmatrix} C3_{(2i,2j)}^k & C3_{(2i,2j+1)}^k \\ C3_{(2i+1,2j)}^k & C3_{(2i+1,2j+1)}^k \end{pmatrix} \quad (8)$$

Where (i, j) is the indices output filter feature map, while k represents feature map index

- Fully Connected Layer

LeNet contained 5 neurons for five output classes, this layer can be seen mathematically in the equation below.

$$F6^k = \sum_{i=1}^{120} W_i^k * C5^i \quad (9)$$

The output of this layer is fed into softmax classifier which output the class score using the equation below

$$Z^k = \text{soft max}(F6^k) \quad (10)$$

4.3.3. Implementation

For the implementation, MATLAB (2014a) tool was used for training and testing of a classification model based on facial images for VARK LS detection. Due to the high computational requirement of CNN, a GeForce GTX 1050 system, 64-bit Operating System, core i7 768CUDA with 8125 MB total available graphics memory was used. All experiments run on the divided data set which contains 5 distinct classes for both the training and testing. Several parameter variations were applied in a quest for finding optimal performance of the two networks, which includes no of iteration, training function, learning rate et al.

4.3.4. Network Training

The two networks were trained using backpropagation algorithm; each training begins with assigning random weight to 32x32 size inputs vector (image) that fed into the network K. The network sum the weights of the input vector and activation function determine the output Y_K . At each iteration, parameters that contribute to loss function is recorded to check the difference with expected output Y_{DK} is kept minimal while training. If the network error is not minimised, then gradient backpropagation is used for

updating the weight, which entails a backward minimise the error (MSE) calculated using equation (12). The process is repeated until the network optimises its loss function.

4.4. Model Evaluation

The performances of the two networks for VARK LS during training and testing were evaluated in terms of MSE and Accuracy. MSE measures how well the networks were trained for the detection of learning style from the dataset (see equation 12). In contrast, Accuracy is used to measure the ratio of true positive and true negative to the sum of all true positive, true negative, false positive and false negative (see equation 11). Samples are partitioned into 5-folds in this method to avoid both over-fitting and under-fitting [27]. After training the two models, the Accuracy metric was used to test the classification of VARK LS models since it can measure both CNN and MCNN [28].

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$E = \frac{1}{n} \left(target(i) - output(i) \right)^2 \quad (12)$$

5. Result and Discussion

Results presented in this section were based on optimal performance achieved by the networks. The proposed CNN achieved optimal results with 3 convolutional and 2 pooling layers to extract features from the facial images, hyperbolic tangent and sigmoid (tansig) activation functions were used for all convolutional layers and pooling layers respectively in the designed CNN except the last layer (output), which used softmax activation function. While the proposed MCNN used achieved optimal performance with fifty hidden neurons, one hundred PCA for feature and selection. Finally, the two networks were trained using gradient descent backpropagation algorithm for updating weight, and sigmoid activation function for 200 epochs.

A termination or network goal is set at 0.001 to avoid overfitting on the network so that a class of new instance that is not exact with trained instances can be predicted. The two tables (Table 2 and Table 3) represent the errors calculated during the training of the two networks. The errors are presented at an interval of 25 epochs.

Table 2 from Appendix shows that training error (MSE) with MCNN begins with a high value at 25 epochs. All five classes recorded an error with the range of 33 to 40, but the error reduced as the number of iteration increases. It can be observed that even though the reading class was not having the least error in the first 25 epochs, but it recorded the least error in the last 200 epochs. This show the inconsistency in comparing the error at different epoch.

It can be seen from Table 3 above that the MCNN converge faster based on MSE in all VARK LS classes. Also, considering the time comparison, the training CNN model varies within 6 to 8 hours depending on the size of the dataset. While MCNN varies within 17 to 20 seconds.

One of the likely reason is that MCNN takes the matrix form of the image and store it in one vector, whereas CNN used a matrix for computation.

Moreover, it can be observed that the CNN network achieved different performance level with an increase in sample size as shown in Table 4, three out of the five classes recorded a better increase in performance; visual class increased from 50% to 71.03%, reading class from 52 to 71.01%, neutral class from 33.10% to 60.835. While little it affects the aural and kinesthetic class. From Table 2, it can be seen that all the 5 classes recorded above 90% performance with 99.5% accuracy for the Reading class as the highest while 97.3% for the visual class as the lowest. It is also found that MCNN recorded an average accuracy of 99.8%, while CNN recorded 64.45%. Comparing the performance results from Table 4 in particular, it is seen that the reading class with MCNN has the highest performance with an accuracy of 99.5% for all classes, while the aural class with CNN records the lowest performance with an accuracy of 50.90%.

Although CNN is outperformed by MCNN, we conducted two experiments using different sample sizes with the same experimental setup. The experiment used the initial setup of our experiments. The first experiment used 864 samples each for a class, while the second experiment used 5000 images for each class

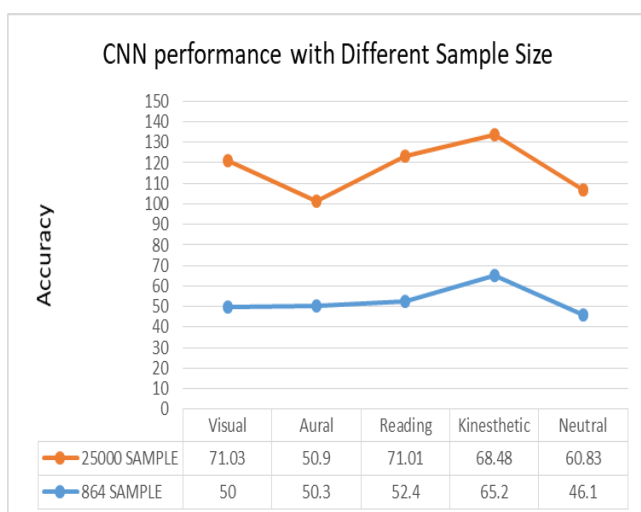


Figure 5. Produced by Table 10

From Table 5, we can see that when the CNN network is trained with 864 images for each VARK class, the performance is poor. However, when trained with 5000 images each for the class, the performance of the network significantly improved.

Similarly, when analysing the result from the illustration in Figure 2 we can see that on average of all the classes, CNN gets better when trained with 5000 images. This implies that the larger the training set, the higher the accuracy of the CNN network. The findings are similar to [47] that the performance of classification with CNN is proportional to the size of the training set, which means that

the performance of our CNN model could further improve when more images are used for the training.

The performance achieved by the two networks in this study is, however, limited to the data collected from learners that used our learning system for learning “Emotional Intelligence” as a course. To address this limitation, other technical courses like mathematic could be considered for data collection. Another limitation is the challenge in comparing findings with other studies; most studies obtained results from a different dataset. Probably, this may be due to privacy and other personal concerns. This study only compared and presented the finding of CNN and MCNN used for the VARK LS detection.

To make a comparison with other recent related studies, evaluating the power of different attributes based on a common dataset is still an open issue [12]. However, the performance of our model verify the effectiveness of our approach, the best model built recorded an average accuracy of 98.7% for the VARK LS classes. The developed model record the best performance (accuracy) compare to the study in[53] that yield average accuracy of 85% through modelling blood pressure with DT. It also outperforms the recent studies in [41, 42] that model student behavioral pattern using ANN with an average of 77.7% and DBN with an average accuracy of 80.4 for FLSM respectively.

6. Conclusion

The development of the model aimed to help instructors detect learners’ learning styles to enable them to provide preferred learning material. The findings in Table 2 shows that the MCNN outperformed the CNN on average performance for all the five classes of the VARK LS model. This is in line with a recent study in [46] that a shallow network works best for small datasets. For the performance of the CNN, Table 2 shows that the accuracy increases as the sample size used in the experiment increases. These findings suggest that CNN needs a large sample size for it to perform better. It also validates the [10] recent suggestions that CNN performs better with a huge dataset. Although many studies recommend that CNN is more suitable for image recognition related problems, findings from this study showed that MCNN could outperform CNN in some instances. This new method for automatic LS detection is expected to be among the recent contribution in the area. For future work, the study suggests that CNN performance can be improved through; 1) Use of more sample images from learners, perhaps by requiring learners to spend more time with the computer learning system. 2) Considering other LS models like FLSM that have more classes.

Appendix . Table of Results.

Table 1. Summary of VARK learning style dimensions with corresponding LO used and parameters used

VARK learning style dimensions	Content and Measure used for labelling	
	Learning content	Measure & Parameter
Visual	Video	VContent_visit (Count)
Auditory	Recorded note	VTime_spent (Time)
Reading	Power Point slide	AContent_visit (Count)
Kinesthetic	Hands on exercise	ATime_spent (Time)

Table 2. Results of Training MCNN with Back propagation Algorithm

Epoch	Results of Training Error (MSE) of Multiclass Neural Network			
	Visual	Aural	Reading	Kinesthetic
25	33.99058	37.22782	38.36802	39.25908
50	5.5436	6.40112	6.39668	5.70094
75	1.452212	1.60692	1.63912	1.3476
100	0.442564	0.481448	0.548604	0.421634
125	0.189914	0.197118	0.2575584	0.179568
150	0.120626	0.1168206	0.153169	0.1023646

Table 3. Results of Training CNN with Back propagation Algorithm

Epoch	Results of Training Error (MSE) of Convolutional Neural Network			
	Visual	Aural	Reading	Kinesthetic
25	0.172	0.1133	0.1733	0.172
50	0.1704	0.1165	0.1705	0.1704
75	0.1641	0.08701	0.1649	0.1642
100	0.1643	0.09983	0.1642	0.1643
125	0.1612	0.08444	0.161	0.1612
150	0.1618	0.114	0.1617	0.1618

Table 4. Shows performance comparison of individual class of VARK classification with CNN and MCN

VARK Learning Style	Performance of CNN on Two Different Samples Size		
	CNN	MCNN	Total In %
Visual	50.00%	97.30%	100
Aural	50.03%	97.90%	100
Reading	52.40%	99.50%	100
Kinesthetic	65.20%	99.40%	100
Neutral	46.10%	97.40%	100

Table 5. Results of performance of CNN with two different sample size

VARK Learning Style	Performance of CNN on Two Different Samples Size		
	864 Sample	5000 Sample	Total In %
Visual	50.00%	71.03%	100
Aural	50.03%	50.90%	100
Reading	52.40%	71.01%	100
Kinesthetic	65.20%	68.48%	100
Neutral	46.10%	60.83%	100

Acknowledgements.

We acknowledge TETFUND for its support, Bufferedky Plus Computer Technology for developing the learning system that serves as a data gathering tool and the participants who volunteered to use the learning system.

References

- [1] Moser, S and Zumbach, J. Exploring the development and impact of Learning Style: An empirical investigation based on explicit and implicit measure, Computers and Education 2018.
- [2] John, A., Shahzadi G., and Khan, K. Student preferred learning style and academic performance, Sci. Int. (Lahore); 2016. 28 (2016): 337-341.
- [3] El aissaoui, O., and Oughdir, L (2020). A literature Review on Student Modeling purposes. In: Ezziyyani, M. Advance Intelligent Systems for sustainable Development; July 08-11; Marrakech Morocco. Computational Intelligence and Complexity: Springer; 2020.
- [4] Gambo, F.L., Wajiga, G.M., and Garba, E.J. A conceptual framework for detection of learning style from facial expressions using convolutional neural network, In Oluwade, B. The role of Computing in the Evolution and Development

- of Emergent and Alternative Technologies; October 14-17; Zaria Nigeria. 2nd International conference of the IEEE Nigeria Computer Chapter: IEEE Xplore; 2019. <https://ieeexplore.ieee.org/document/8949656>.
- [5] Gambo, F.L., Hamada, M. Garko, A.B. Abba, A.H and Yahaya, A.S. Learning Information theory anywhere at any time on smartphone, Dutse Journal of Science and Applied Science; 2015. 1 (1): 82-89.
 - [6] Greer, J. E. & McCalla G. (Eds.). Student modeling: The key to individualised knowledge-based instruction. Berlin: Springer Verlag; 1994. Formal approaches to student modeling. [295–348]
 - [7] Normadhi, N.B.A. Shuib, L. Nasir, H.N.M Bimba, A. Idris, N. and Balakrishnan, V. Identification of personal traits in adaptive learning environment: Systematic literature review, Computers & Education; 2018. DOI: 10.1016/j.compedu.2018.11.005.
 - [8] Thalmann, S. Adaptation criteria for the personalised delivery of learning materials: A multi-stage empirical investigation. Australasian Journal of Educational Technology; 2014. 30(1)
 - [9] Bernard, J. Chang, T.W., Popescu E and Graf, S. Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithm, Expert Systems with Application; 2017. 7 594-108.
 - [10] Jauro, F., Chiroma, H., Gital, A.Y. et al., (2020). Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend, Applied Soft Computing Journal. 1-63
 - [11] Abdullahi A.A., Gambo, F.L. and Adam, J.M. Hardware architecture and neural computing, Dutse Journal of Pure and Applied Sciences; 2017. 3(1) 332-337.
 - [12] Gambo, F.L. Wajiga, G.M. Garba, E.J and Abdullahi, A.A. Prediction of learning style using facial expressions with convolutional neural network. International Journal of Information Processing and Communication; 2020. 9 (special issue), 389-399.
 - [13] Feldman, J. Monteserin, A. and Amandi A., Automatic detection of learning style: State of the art, Springer Science Business Media Dordrecht; 2015. 157-186. DOI: 10.1007/s10332-014-9422-6
 - [14] Nam, V. A Method for detection of learning styles in learning management systems, UPB Scientific Bulletin, Series C: Electrical Engineering; 2013. 75, 3-12.
 - [15] Ahmad, N., Tasir, Z. Kasim, J. and Sahat H, Automatic detection of learning style in learning management systems by using literature based method, Social and Behavioral Science; 2013. 103, 181-189
 - [16] Graf, S. Adaptivity in learning management systems focusing on learning style, Vienna University of Technology; 2007.
 - [17] Keefe, J.W. Learning style: An overview. NASSP's Student learning styles: Diagnosing and proscribing programs Reston, VA. National Association of Secondary School Principals; 1979. 1-17
 - [18] Truong, H.M. Integrating learning style and adaptive e-learning: Current developments, problems and opportunities, Computer in Human Behavior; 2015. 55, 1185-1193.
 - [19] Fleming, N. and Baume D, Learning style again: VARKing up the right tree!!, Education Developemnet; 2006. 7, 4-7.
 - [20] Shuib, L. M Learning style based information seeking tool, PhD Thesis. University of Malaya; 2013.
 - [21] Sordedragar, R Automatic malaria diagnosis using convolutional neural network in an on-field setting: The analysis of low quality smartphone based microscopic images, Delft University of Technology Unpublished Master's Thesis; 2018.
 - [22] Aggarwal, C.C. Data Mining. New York: Springer International Publishing Switzerland; 2015.
 - [23] Han, J. and Kamber, M. Data Mining: Concepts and Techniques. [Second edition]. San Francisco: Morgan Kaufmann Publishers; 2006.
 - [24] Singaravel, S., Suykens, J., & Geyera, P. Deep-learning neural-network architectures and methods: Using component-based models in building design energy prediction. Advanced Engineering Informatics; 2018. 38, 81-90
 - [25] Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems; 2012. 1097-1105
 - [26] Lecun Y. Connections in perspective: Zurich, Switzerland: Elsevier; 1989. Generalisation and network design strategies. 117–155.
 - [27] Hubel, D. H., & Wiesel, T. N. Receptive Fields of Single Neurons in the Cat's Striate Cortex. Journal of Physiology; 1959. 148(1), 574–59.
 - [28] Rawat, W., & Wang, Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Computation; 2017. 29, 2352-2449.
 - [29] Villaberde, J.E, Godoy D, and Amandi, A, Learning style recognition in e-learning environment with feed-forward neural networks, Journal of Computer Assisted Learning; 22, 197-206.
 - [30] Hamtini, T. A Proposed Dynamic Technique for Detecting Learning Style using Literature Based Approach, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies; 2015.
 - [31] Hasibuan, M.S. Nugroho L.E, Santosa, P.L. and Kusumawardani S.S, A proposed model for detecting learning style based on agent learning, International Journal of Emerging Technologies in Learning; 2016. DOI:10.3991/ijet.v11i10.5781.
 - [32] Hasibuan, M.S. and Nugroho, L.E. Detecting Learning Style using Hybrid Model, IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (2016).
 - [33] Hasibuan, M.S. Nugroho, L.E and. Santosa, P.L Learning style model detection based on prior knowledge in e-learning systems, Proceedings ICIC (2017).
 - [34] Hasibuan, M.S. Nugroho, L.E and. Santosa, P.L Model detecting learning style with artificial neural network, Journal of R+Technology and Science Education (2019), 85-95.
 - [35] Kolekar, S. V. Pai, R. M. & Pai, M.M. Prediction of learner's profile based on learning styles in adaptive e-learning system". International Journal of Emerging Technologies in Learning (iJET); 2017. 12, 31–51.
 - [36] Bernard, J. Chang T.W, Popescu, E. and Graf, S. Using artificial neural networks to identify learning styles, International Conference on Artificial Intelligence in Education; 2015. 541-544.
 - [37] Bernard, S Clément, F and Kaufmann, L. Rules trump desires in pre-schoolers' predictions of group behaviour. Social Development; 2016. 453–337, DOI:10.1111/sode.12150.
 - [38] Graf, S., Kinshuk, and T.C. Liu, Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach, Educational Technology & Society; 2009. 12, 3–14.
 - [39] Habibi R., Setyohadi, D.B and Santoso, K.I. Student learning style and emotional intelligence tendencies based on twitter, 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE); 2017, 239-217.
 - [40] Zhang, H., Huang, T., Liu, S., Yin, H., Li, J., Yang, H., & Xia, Y., (2020). A learning style classification approach

- based on deep belief network for large-scale online education. *Journal of Cloud Computing*; 2020. 9, 1-26.
- [41] Gomedede, E. De Barros, R.M and Mendes, L.D.S Use of deep multi-target prediction to identify learning styles. *Applied Science*; 2020. 1-10
- [42] Zhang, B. Chai, C. Yin, Z. Shi, Y. Design and Implementation of an EEG-Based Learning-Style Recognition Mechanism. *Brain Sci.*; 2021. <https://doi.org/10.3390/brainsci11050613>.
- [43] Huang, F. Chen, Y Lv, S. S.and Wang X, Facial Expression Recognition: A Survey, *Symmetry*; 2019. 1-11
- [44] Kuhn, M & Johnson, K. *Applied Predictive Modeling*. New York, NY: Springer New York; 2013. doi:10.1007/978-1-4614-6849-3. ISBN 9781461468486.
- [45] Wani, M.A., Bhat, F.A. Afzal, S and. Khan, A.I Advance Deep Learning”, Springer; 2019.
- [46] Muhammad, A. Aseere, A.M Chiroma H., Sha, H. Gital, A.Y. Hashem I.A.T, Deep learnig application in smart cities: recent development, taxonomy, challenges and research prospect, *Artificial Intelligence Review*; 2020. 1-27, DOI; 10.1007/s10332-019-09770-z
- [47] Yu, H., Samuels, D.C., Zhao, Y., & Guo, Y. Architectures and accuracy of artificial neural network for disease classification from omics data. *BMC Genomics*, Article; 2019. 10.1186/s12864-019-5546-z.
- [48] Yamashita, R., Nishio, M., Do, R.K.G. & Togashi, K.. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*; 2018. 9, 611–629. <https://doi.org/10.1007/s13244-018-0639-9>.
- [49] Li, B. Facial expression recognition via transfer learning. *EAI Endorsed Transactions on e-Learning*; 2021. 21(21) 1-14 DOI: 10.4108/eai.8-4-2021.169180.
- [50] Sharma, S. Mehra, R. & Kumar, S. Optimised CNN in conjunction with efficient pooling strategy for the multi-classification of breast cancer. *IET Image Processing* [Article; Early Access]; 2021. doi: 10.1049/ipr2.12074.
- [51] Muhammad, K. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, *Multimedia Tools and Applications*; 2019. 78, 3613-3632.
- [52] Wang, S. H. DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification, *ACM Trans. Multimedia Comput. Commun. Appl.*; 2010. 16, 60.
- [53] Dutsinma, L. I. F., & Temdee, P. VARK learning style classification using decision tree with physiological signals. *Wireless Personal Communications*; 2020 .1–22