IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# Evaluation of the academic performance of students' using artificial neural network: Case study of Faculty of Engineering, Nigeria Maritime University, Okerenkoko, Delta State.

***Bolou, D. Bolou[1], Kabiru Ahmed[2], Olateju, O. Abideen[3], Akosubo I. Stevyn[4]***

[1,2,3]Lecturer, Computer Science, Nigeria Maritime University Okerenkoko, Delta State, Nigeria.

[4]Faculty Examination Officer, Faculty of Engineering, Nigeria Maritime University Okerenkoko, Delta State, Nigeria.

[4]Lecturer, Department of Civil Engineering, Nigeria Maritime University Okerenkoko, Delta State, Nigeria.

***Abstract -*** The substantial development in electronic data for university students' academic performance in either supervised or unsupervised learning has resulted in some meaningful information extracted from large volumes of data using diverse data mining techniques. Due to the increase in the rate of poor outcomes, the need to design a system that helps to reduce the menace of students' poor academic performance or having to drop out of school was analysed. The purpose of this study was to develop a system to predict student performance with an Artificial Neutral Network approach to the predictive classification of students in the full range of academic performance (GPA) so as to determine which is more efficient in and in what case one should be preferred over the other, as well as to identify and understand the importance of the variables for each level (low, middle and high) of expected GPA. Artificial Neural networks often need a greater collection of observations to achieve enough predictive ability. The ANN is a suitable model for the prediction of students' academic performance in their final year under conditions of a very complex and great amount of data, in which a large number of variables interact in various complexes. The results attained during this study will allow the identification of the precise influence of every input set of variables on different levels of educational performance (high and low performance), on one hand, and customary processes across all students, on the opposite hand. Additionally, we identified which key factors had an important influence on overall students' performance. Data were collected from the students of the Faculty of Engineering, Nigeria Maritime University, Okerenkoko, Delta State. The study achieved an accuracy of over 92.3 percent, showing Artificial Neural Network's potential effectiveness as a predictive tool for accessing students' academic performance.

***Index Terms -*** **Student performance, University Education, Data mining, Artificial Neural Network**

## I. INTRODUCTION

Predicting student academic performance has long been a very important research topic. Among the problems of the education system, questions concerning admissions into academic institutions remain important (Ting and Man, 2001). The most objective of the admission system is to see the candidates who would likely perform well after being accepted into the university. The standard of admitted students incorporates a great influence on the extent of educational performance, research and training within the institution. The failure to perform an accurate admission decision may lead to an unsuitable student being admitted to the program. Hence, admission officers want to understand more about the educational potential of every student. Accurate predictions help admission officers to differentiate between suitable and unsuitable candidates for an educational program and identify candidates who would likely perform best in the institution (Ayan and Garcia, 2013). The results obtained from the prediction of educational performance could also be used for classifying students, which enables educational managers to supply them additional support, like customized assistance and tutoring resources. The results of this prediction can even be employed by instructors to specify the foremost suitable teaching actions for every group of scholars, and supply them with further assistance tailored to their needs.
Furthermore, the prediction results may help students develop a decent understanding of how well or how poorly they might perform, then develop an acceptable learning strategy. Accurate prediction of student achievement is a way to boost the standard of education and supply better educational services (Romero and Ventura, 2007). Different approaches are applied to predicting student academic performance, including traditional mathematical models and modern data processing techniques.
In these approaches, a collection of mathematical formulas was used to describe the quantitative relationships between outputs and inputs. The prediction is accurate if the error between the anticipated and actual values is within a little range. In machine learning

and scientific discipline, artificial neural networks are a family of statistical learning models inspired by biological neural networks and are used to estimate or approximate functions that may rely on an outsized number of inputs and are generally unknown. Artificial neural networks are generally presented as systems of interconnected neurons that exchange messages with one another. The connections have numeric weights which will be tuned supported experience, making neural nets adaptive to inputs and capable of learning.

The artificial neural network, a soft computing technique, has been successfully applied in several fields of science, like pattern recognition, fault diagnosis, forecasting and prediction. However, as far as we are aware, not much research on predicting student academic performance takes advantage of artificial neural networks. Kanakana and Olanrewaju (2011), utilized a multilayer perception neural network to predict student performance. The research showed that a man-made neural network-based model is in a position to predict student performance within the first semester with high accuracy. Mellamby (1956), observed that universities worldwide are not really satisfied with the methods used for choosing undergraduates. While admission processes in many developed countries have benefited from various advances in information processing and technology, the Nigerian system has yet to require full advantage of those new tools and technology.

Hence, this study takes a scientific approach to tackle the matter of admissions by seeking ways to form the method more practical and efficient. Specifically, the study seeks to explore the chance of using a man-made Neural Network model to predict the performance of a student.

## II. CONCEPTUAL FRAMEWORK OF ARTIFICIAL NEURAL NETWORKS AND PERFORMANCE

Conceptually, a neural network could be a computational structure consisting of several highly interconnected computational elements, called neurons, perceptrons, or nodes. Each neuron carries out an awfully simple operation on its inputs and transfers the output to a subsequent node or nodes within the constellation (Specht, 1991). Neural networks exhibit polymorphism in structure and parallelism in computation (Mavrovouniotis & Chang, 1992). In general, a synthetic Neural Network consists of an input layer, one or more hidden layers, and an output layer that's akin to a categorical variable quantity (Cascallar et al., 2006; Garson, 1998). All Artificial Neural Networks process data through multiple processing entities which learn and adapt per patterns of inputs presented to them, by constructing a singular mathematical relationship for a given pattern of input file sets on the premise of the match of the explanatory variables to the outcomes for every case (Marshall & English, 2000). Thus, neural networks construct a mathematical relationship by learning the patterns of all inputs from each of the individual cases employed in training the network, while more traditional approaches assume a selected sort of relationship between explanatory and outcome variables and then use a spread of fitting procedures to regulate the values of the parameters within the model. During the training phase, Artificial Neural Networks generate a predicted outcome for every case, and when this prediction is inaccurate the network adjusts the weights of the mathematical relationships among the predictors and with the expected outcome, weights that are represented within the hidden layers of the network. the anticipated output may be a continuous variable with a particular value for every case which has information on the probability of belonging to every one of the specific classifications requested by the developer of the bogus Neural Network. consistent with this architecture, the unreal Neural Network finally recognizes patterns and classifies the cases presented into the requested outcome categories, counting on the target question, and offers the individual probability values for every case. This information is generated by the network through many iterations, gradually changing and adjusting the weights for all the interrelationships between the units after each incorrect prediction. During this training process, the network becomes increasingly accurate in replicating the known outcomes from the test cases. The neural network continues to boost its predictions until one or more of the pre-determined stopping criteria are met. These stopping criteria will be, for example, a minimum level of accuracy, learning rate, persistency, number of iterations, amount of your time, etc. Once trained, the network is tested with the remaining cases within the dataset, which is considered a variety of validation of the network, by observing how the weights within the model, now fixed to those obtained within the training phase, predict classes of outcomes during a new set of information of which outcomes are known to the experimenter but to not the unreal Neural Network system. Afterwards, it may also be applied to predict future cases where the result continues to be unknown (Cascallar et al., 2006). additionally, with complementary techniques in predictive stream analysis, the neural network approach allows us to work out the predictive power of each of the variables involved in the study, providing information about the importance of every input variable (Cascallar et al., 2006; Garson, 1998). Predictive stream analyses (Cascallar & Musso, 2008), based during this case on Artificial Neural Network models, have several strengths:

(a) because these are machine learning algorithms, the assumptions required for traditional statistical predictive models (e.g., ordinary statistical procedure regression) don't seem to be necessary. As such, this system is ready to model nonlinear and complicated relationships among variables. Artificial Neural Networks aim to maximise classification accuracy and go through the information in an interactive process until maximum accuracy is achieved, automatically modelling all interactions among variables;

(b) Artificial Neural Networks are robust, general function estimators. they sometimes perform prediction tasks a minimum of similarly as other techniques and most frequently perform significantly better (Marquez, et. al., 1991);

(c) Artificial Neural Network can handle data of all levels of measurement, continuous or categorical, as inputs and outputs. thanks to the speed of microprocessors in even basic computers, Artificial Neural Networks are more accessible today than once they were originally developed. The Artificial Neural Network learns by examining individual training cases (subjects/students), then generating a prediction for every student, and adjusting the weights whenever it makes an incorrect prediction. This process is repeated again and again, and therefore the network continues to enhance its predictions until one or more of the stopping criteria are met. A minimum level of accuracy is often set because the stopping criterion, although additional stopping criteria is also used further (e.g., number of iterations, amount of processing time). Once trained, the network may be applied, with its structure and parameters, to future cases (validation or holdout sample) for further validation studies and programme implementation (Lippman, 1987). As long because the basic assumptions of the population of persons or events that the substitute Neural Network used for training is constant or varies slightly and/or gradually, it can adapt and improve its pattern recognition algorithms the more data it's exposed to within the implementations. The class of Artificial Neural Network models employed in this research is compared with the more traditional discriminant analysis approach. Both of those methods derive classification rules from samples of classified objects supported known predictors. This general approach is named supervised learning since the outcomes are known and

relationships are modelled or supervised in keeping with these outcomes (Kohavi & Provost, 1998). But, there are significant differences within the algorithms and procedures for both analyses, like the actual fact that while discriminant analysis assumes linear relationships, neural network analysis doesn't. In terms of comparisons with another common method employed in educational research, simple regression, it's important to notice that although neural networks can address a number of the identical research issues as regression it's inherently a distinct mathematical approach (Detienne et al., 2003). there's another family of predictive systems which are unsupervised (e.g., Kohonen networks), during which the patterns presented to the network don't seem to be related to specific outcomes; it's the neural network itself that derives the commonalities between the predictors, grouping cases into classes on the idea of those similarities. Thus, these analyses are often accustomed explore the information from a unique perspective and learn the grouping of cases supported these predictor commonalities rather than being focused on predictions or individual outcomes (Cascallar et al., 2006; Kyndt et al., 2012, submitted). Neural networks excel within the classification and prediction of outcomes; especially when large data sets are available that are related in nonlinear ways, and where the intercorrelation between variables isn't clearly understood. These properties of Artificial Neural Networks clearly make them particularly suitable for science data where they'll simultaneously consider all variables in a very study (Garson, 1998). Moreover, the assumptions of normality, linearity and completeness that are made by methods like multiple rectilinear regression (Kent, 2009), which are often very difficult to determine for scientific discipline data, aren't made in neural network analysis. Neural networks can work with noisy, incomplete, overlapping, highly nonlinear and non-continuous data because the processing is touching an oversized number of processing entities (Garson, 1998, Kent, 2009). during this regard, it is often said that neural networks are robust and have wide non-parametric applications. there's also evidence that neural models are robust within the statistical sense, and also robust when faced with a low number of grade points (Garson, 1998).

### 2.1. ARTIFICIAL NEURAL NETWORK PROCESSES AND MEASURES TO EVALUATE THE SYSTEM PERFORMANCE

In order to evaluate the performance of the neural network system, there are variety of measures used which offer a method of determining the standard of the solutions offered by the assorted network models tried. the standard measures include the determination of actual numbers and rates for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) outcomes, as products of the factitious Neural Network analysis. additionally, certain summative evaluative algorithms are developed during this field of labor, to assess the general quality of the predictive system.

These overall measures are: Recall, which represents the proportion of correctly identified targets, out of all targets presented within the set, and is represented as: Recall = TP/ (TP + FN); and Precision which represents the proportion of correctly identified targets, out of all identified targets by the system, and is represented as: Precision = TP/ (TP + FP). Two other measures, derived from signal-detection theory (ROC analysis), have also been wont to report the characteristics of the detection sensitivity of the system. one in all them is Sensitivity (similar to Recall: the proportion of correctly identified targets, out of all targets presented within the set), which is expressed as Sensitivity = TP/ (TP + FN). the opposite is Specificity, defined because the proportion of correctly rejected targets from all the targets that ought to be rejected by the system, and which is expressed as Specificity = TN/ (TN + FP). All the normal measures are typically represented in what's called a “confusion matrix” representing all four outcomes.

In addition, the evaluation of Artificial Neural Network performance is additionally disbursed with another summative measure, which is employed to account for the somewhat complementary relationship between Precision and Recall. This measure is defined as F1, and is defined as F1 = (2 * Precision * Recall)/ (Precision + Recall). Such a definitional expression of F1 assumes equal weights for Precision and Recall. This assumption may be modified to favour either Precision or Recall, consistent with the utility and cost/benefit ratio of outcomes favouring either Precision or Recall for any given predictive circumstance.

## III. RESEARCH METHODOLOGY

### 3.1 Area of the Study

The study was conducted in the Faculty of Engineering, Nigeria Maritime University, Okerenkoko, Delta State.

### 3.2 Sample of the Study

The total sample included 232 students in the Faculty of Engineering, of both genders (male 74.6%; female 25.4%), ages between 18 and 25 (*M*age = 20.38, *SD* = 3.78), recently enrolled in the 100 level students in the Faculty of Engineering during the 2020/2021 academic session. In all, 67.8% of the sample was 17 to 20 years old, 24.7% was 21-25 years old, and 7.5% was older than 25 years. An 80% math accuracy criterion was imposed for all participants in the Automated Operation Span (Unsworth et al., 2005). Therefore, they were encouraged to keep their math accuracy at or above 80% at all times. As a consequence of this criterion, 78 participants were excluded from the analyses. The final sample consisted of 198 students.

### 3.3 Instrument for Data and Sources of Data Collection

This computerized task provides a measure for every of the three anatomically defined attentional networks: alerting, orienting, and executive. The ANT is a combination of the cued reaction time (Posner, 1980) and the flanker test (Eriksen & Eriksen, 1974). The participant saw an arrow on the screen that, on some trials, was flanked by two arrows to the left and two arrows to the proper. Participants were asked to see when the central arrow points left or right, by two mouse buttons (left-right). They were instructed to target a centrally located fixation cross throughout the task and to reply as quickly and accurately as possible. During the practice trials, but not during the experimental trials, subjects received feedback from the pc on their speed and accuracy. The practice trials took approximately 2 minutes and every of the three experimental blocks took approximately 5 minutes. The entire experiment took about twenty minutes. The measure for the attention is the average response time regardless of the cues or flankers.

To analyse the effect of the three attentional networks, a collection of cognitive subtractions described by Fan et al. (2002) were used. The efficiency of the three attentional networks is assessed by measuring how response times are influenced by alerting cues, spatial cues, and flankers (Fan et al., 2002). The alerting effect was calculated by subtracting the mean reaction time of the double-cue conditions from the mean time interval of the no-cue conditions. For the orienting effect, the mean time interval of the spatial

cue conditions (up and down) was subtracted from the mean reaction time of the centre cue condition. Finally, the effect of the executive control (conflict effect) was calculated by subtracting the mean reaction time of all congruent flanking conditions, summed across cue types, from the mean reaction time of incongruent flanking conditions (Fan et al. 2002). The test-retest reliability of the overall response times (in this study used as a measurement of general attention), calculated by Fan et al. (2002) equaled 0.87. The test-retest reliability of the subtractions is less good. The executive control is that the most reliable (r = 0.77), followed by the orienting network (r = 0.61). The alerting network showed to be the lesser reliable (r = 0.52) (Fan et al. 2002).

**3.4 Data Analysis Techniques**

A computer-administered version of the Ospan instrument (Unsworth et al., 2005) that measures memory capacity was employed for the study. The responses were collected via the press of a button. First, participants receive practice and secondly, the participants perform the particular experiment. The practise sessions are further counteracted into three sections. the primary practice could be a simple letter-span task. They see letters appear on the screen one at a time and hence, altogether experimental conditions, letters remain on-screen for 800 milliseconds. Then, participants must recall these letters within the same order they saw them from a 4 x 3 matrix of letters (F, H, J, K, L, N, P, Q, R, S, T, and Y) presented to them. Recall consists of clicking the box next to the acceptable letters; the recall phase is untimed. After each recall, the pc provides feedback about the number of letters correctly recalled. Next, participants practice the maths portion of the experiment. Participants first see a math operation (e.g. (1*2) + 1 =?). Once the participant knows the solution they click the mouse to advance to the following screen. Participants then see variety (e.g. "3") and are required to click if the quantity is the correct solution by clicking on "True" or "False." After each operation participants are given feedback. the maths practice serves to familiarize participants with the mathematics portion of the experiment, similarly on calculating how long it takes a given person to resolve the mathematics problems, establishing a personal baseline. Thus, it attempts to account for individual differences within the time it takes to resolve math problems. this can be then used as an individualized limit for the maths portion of the experimental session. the ultimate practice has participants perform both the letter recall and math portions together, even as they're going to knock off the experimental block. The participants first are presented with a math operation, and after they click the push indicating that they need to solve it, they see the letter to be recalled. If the participants take longer to resolve the maths operations than their average time plus 2.5 SD, the program automatically moves on and counts that trial as a slip. This serves to stop participants from rehearsing the letters once they should be solving the operations. Participants complete three practise trials, each of set size 2. After the participant completes all of the practice sessions, the program moves them on to the 000 trials. the 000 trials accommodate 3 sets of every set size, with the set sizes starting from 3 to 7 letters. This makes for a complete of 75 letters and 75 math problems. Subjects are instructed to stay their math accuracy at or above 85% in any respect times. During recall, a percentage in red is presented within the upper right-hand corner. Subjects are instructed to stay a careful watch on the proportion so as to stay it above 85%. This study reports absolutely the Ospan score (the sum of all perfectly recalled sets) that's interpreted because the measure of overall remembering capacity, and one latency score (operations). The task takes approximately 20–25 minutes to finish (Unsworth et al., 2005). This measure of memory capacity encompasses a high correlation with other measures of remembering and general intelligence, like Ospan and Raven Progressive Matrices. additionally, AOSPAN incorporates a good test-retest reliability (r = 0.83) and an adequate internal consistency (α = 0.78) (Unsworth et al., 2005).

**3.5 Analyses Procedure**

The ANN model used was a backpropagation multilayer perceptron neural network, that is, a multilayer network composed of nonlinear units, which computes its activation level by summing all the weighted activations it receives and which then transforms its activation into a response via a nonlinear transfer function, which establishes a relationship between the inputs and the weights they are assigned (Musso, et. al., 2013).

During the training phase, these systems evaluate the effect of the weight patterns on the precision of their classification of outputs, and then, through backpropagation, they adjust those weights in a recursive fashion until they maximize the precision of the resulting classifications. ANN parameters and variable groupings, as well as all other network architecture parameters, were adjusted to maximize predictive precision and total accuracy. Confusion matrices have been determined for each ANN, as well as ROC analyses for the evaluation of sensitivity and specificity parameters. Parameters such as learning rate (they control rate the size at of weight which and bias changes the ANN during learning), momentum (adds a fraction of the previous weight update to the current one, and is used to prevent the system from converging to a local minimum), number of hidden layers, stopping rules (when the network should stop-fitting "learning" the current sample), activation to functions avoid (which over define the output of a node given an input or set of inputs to that node or unit), and the number of nodes were specified and varied in the model construction phase in order to maximize the overall performance of the network model.

**3.6 Architecture of the Neural Networks**

According to the objectives of this research, three different neural networks (ANN) were developed as predictive systems for the GPA of the scholars during this study. ANN1 was developed to maximise the predictive classification of rock bottom 33% of scholars, which might be scoring the bottom average GPA at the top of the tutorial year. ANN2 was developed to maximise the predictive classification of the best 33% of scholars, which might be scoring the best GPA. ANN3 was developed to predict the classification of scholars into the three levels of expected GPA at the identical time. the info set was partitioned into a training set and a testing set for every ANN, and for every network, training and testing samples were chosen indiscriminately by the software, from the available set of cases. One suggested criterion is that the quantity of coaching inputs (cases) should be a minimum of 10 times the amount of input and middle layer neurons within the network (Garson, 1998). Similarly, it's suggested that about 2/3 (or 3/4) of the cases within the available data set be used for the training innovate order to incorporate a collection of cases representing most of the patterns expected to be present within the data (patterns represented by the vector for every case). The remaining 1/3 or 1/4 of the information is employed for the testing phase of the network. the precise architecture of every of the three neural networks developed is as follows:

$ANN_1$ - (Maximizing the prediction for the Low 33% performance group): All cognitive variables, learning strategies, and background variables were introduced within the analysis. They were used for the event of the vector-matrix containing all predictor variables for every student. The resulting network contained all the input predictors, with a complete of 18 input units (Reaction

Time Operation, latent period Math, interval Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/ Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, lyceum from which the coed graduated, the best level of education completed by father, and also the highest level of education completed by mother). The model built contained one hidden layer, with 15 units. The output layer contained a variable quantity with two units (categories equivalent to "belongs to lowest 33% or below the network, a consistent method for the rescaling of the scale-dependent variables was used. The hidden layer had a hyperbolic tangent activation function which is that the most typical activation function used for neural networks due to its greater numeric range (from -1 to 1) and therefore the shape of its graph. The output layer utilized a softmax activation function that's useful predominantly within the output layer of a clustering system, converting a raw value into a posterior probability. The output layer used the cross-entropy error function during which the error signal related to the output layer is directly proportional to the difference between the specified and actual output values. This function accelerates the backpropagation algorithm and it provides good overall network performance with relatively short stagnation periods (Nasr, Badr, & Joun, 2002). The training was dispensed with the web method 0.4, and momentum adequate to 0.9. The optimization algorithm was gradient descent (which takes steps proportional to the negative of the approximate gradient of the function at the present point), and therefore the minimum relative change in training error was 0.0001.

$ANN_2$ - (Maximizing the prediction for the High 33% performance group): All cognitive, learning strategies, and background variables were introduced within the analysis. They were used for the event of the vector-matrix containing all predictor variables for every student. The resulting network contained all the input predictors, with a complete of 18 units (Reaction Time Operation, latent period Math, latent period Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, the very best level of education completed by father, and also the highest level of education completed by mother). The model built contained one hidden layer, with nine units, and an output layer with two units (categories correspondence 67%"). In terms of the, the same architecture method for his or her scaling of the size network-dependent variables was used. The hidden layer had a hyperbolic tangent activation function. The output layer utilized a softmax activation function. Cross-entropy was chosen because the error function. The dataset was partitioned into a training set and a testing set. Th an initial learning rate of 0.5, and momentum adequate to 0.7. The optimization algorithm was gradient descent, and also the minimum relative change in training error was 0.0001.

$ANN_3$ - (Maximizing the simultaneous prediction for all the performance groups: Low 33% - Middle 33% - High 33%, simultaneously): All cognitive, learning strategies and background variables were introduced within the analysis. They were used for the event of the vector-matrix containing all predictor variables for every student. The resulting network contained all the input predictors, with a complete of 19 input units (Reaction Time Operation, interval Math, response time Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/ Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, school, the very best level of education completed by father, and therefore the highest level of education completed by mother, Ln of Attention Total RT). The model built contained one hidden layer, with 20 units, and one output layer with three units (categories comparable to "belongs to the middle", "low 33%", or "below" performance groups). In terms of the architecture of the network, a uniform method for the rescaling of scale-dependent variables was used. The hidden layer and also the output layer both had a hyperbolic tangent activation function. a regular method for the rescaling of covariates was used. The Sum of squares was chosen because the error function. The dataset was partitioned into the training set and testing set. The training was applied with an initial online learning rate of 0methodol.4, and momentum capable 0.8. The optimization algorithm was gradient descent, and therefore the minimum relative change in training error was 0.0001.

The software used was SPSS v.19 –Neural Network Module, for the event and analysis of all predictive models during this study. Two development phases of the predictive system were carried out: training of the network and testing of the network developed. During the training phase, several models were attempted, and a number of other modifications of the neural network parameters were explored, such as: learning persistence, learning rate, momentum, and other criteria. These tests continued until achieving desired levels of classification, maximizing the advantages of the model chosen. In these analyses both precision and recall, as outcome measures of the network, got equal weight. There was no need to trim the quantity of predictor inputs within the three models. The validation procedure used was the leave-one-out methodology.

### 3.7 Discriminant Analyses

Discriminant Analyses (DA) were carried out using the same data and the same categories of GPA used in the Neural Networks Analyses. DA1 was performed to discriminate between the students belonging to the lowest 33% of GPA and contrast them against those, not in that category. DA2 was focused on identifying students in the highest 33% of academic performance versus those, not in that group, and DA3 was calculated to discriminate the students belonging to each one of the three levels of GPA performance. In order to give every variable, the opportunity to contribute significantly to the prediction, a stepwise discriminant analysis was calculated for each category including all independent variables (Musso, et. al., 2012). Furthermore, we calculated three discriminant analyses, one for each category including the independent variables of the maximised neural networks of each category.

## IV. RESULTS AND DISCUSSION

The research study performed on undergraduate students, which represent one of the leading respondents, is in a group of faculties of technology and engineering sciences of Lagos state university. The faculty of Sciences is a higher education institution that deals with education, scientific research and consultancy through the development of knowledge and skills in management, information systems and technology with the aim to enable future professionals to develop the potential of commerce and society. This research includes the first five generations of students, with over 2000 graduated students, who have been registered in accordance with the principles of higher education implemented by the Nigerian law, and had completed their studies by November 2013. In this time frame, more than 60% of enrolled students have graduated, achieving an overall Grade point average (GPA) of 3.5 on an average time of study of almost 4 years. As an input data (predictors), 15 variables were used and represent data correlated to students'

personal characteristics (students' gender), high school information (high school GPA and high school type), admission data (entrance examination points) and the first-year examination grades (individual grades at 11 courses examinations of the first year of studies). On the other side, as an output from neural networks, we used a GPA at the end of their studies. The ratio between the training data and the testing data was 70:30. Model evaluation is performed on so far non-used data Performances of developed models are measured by Absolute average error, Standard deviation and Linear correlation. Results and performance of developed ANN are presented in Table 4.1. The best values for each criterion of comparison are shown in bold font.

**4.1 Results of Descriptive Statics of Study Variables**

Table 4.1: Results of developed ANNs (6 algorithms)

| *Algorithm* | *Absolute average error* | *Standard Deviation* | *Linear correlation* |
|---|---|---|---|
| *Quick* | *0.233* | *0.295* | *0.893* |
| *Dynamic* | *0.239* | *0.300* | *0.889* |
| *Multiple* | *0.234* | *0.294* | *0.892* |
| *Prune* | *0.238* | *0.297* | *0.892* |
| *RBFN* | *0.259* | *0.324* | *0.873* |
| *Exhaustive Prune* | **0.231** | **0.293** | **0.894** |

Table 4.1, shows the comparative analysis of the results, on the test sample, showed that the developed ANN gives acceptable results. An absolute average error of prediction in all the networks is from 0.231 to 0.259 and the linear correlation coefficient is always over 87%. Exhaustive Prune stands out as the best algorithm by all criteria, giving the smallest absolute average error (0.231, the smallest standard deviation (0.293), and at the same time, the largest linear correlation coefficient (89.4%). A developed multilayer neural network with the Exhaustive Prune method consists of an input layer with 15 variables, an output layer with one variable and two hidden layers with 30 and 20 neurons. Results with the best ANN indicate that almost 90% of systems will correctly predict the final student GPA, which represents more than good precision in prediction, especially if we know that GPA will be round on one or two decimals, then the standard deviation will have even less influence on the final result. The usage of ANN gives one more advantage compared to some other data mining methods and techniques, which is represented in the calculation of the relative importance of the input variables for predicting an output variable. In order to identify the major influenced factors of students' success, we used this advantage of ANN. The importance of the attributes of the developed ANN model is given in Table 2. It can be seen that the 6 firstly ranked attributes within the developed model are considered the most important (relative value over 0.1). By doing further analysis, we can conclude that grades on the exam in the first semester of studies have a great influence on the overall success of studies, as well as data from high school (High school type and High school GPA), where both groups of variables achieved importance over 0.1. On the other side, as it can be seen from Table 4.2, the gender of the students and entrance examination points are not of the greatest importance for predicting students' success. Also, some grades from the first-year exams, especially on subjects from the second semester had a small influence on it, such as Engineering Drawing, Engineer-in-Society and Physics.

Table 4.2: Relative importance of the input variables

| *Input variable* | *Relative importance* |
|---|---|
| *Mathematics 1* | *0.1223* |
| *Basics of information-communicational technologies* | *0.1103* |
| *High school type* | *0.1072* |
| *Basics of organization* | *0.1034* |
| *High school GPA* | *0.1024* |
| *Introduction into the informational systems* | *0.1004* |
| *Production systems* | *0.0602* |
| *Mathematics 2* | *0.0513* |
| *English language 1* | *0.0502* |
| *Entrance examination points* | *0.0426* |
| *English language 2* | *0.0411* |
| *Engineering Drawing* | *0.0402* |
| *Student gender* | *0.0343* |
| *Engineer-in-Society* | *0.0215* |
| *Physics* | *0.0126* |

Table 4.2 represents results and performance of developed ANN by using same performance measurements. After selection of 6 the most important input variables, we repeated process of predicting GPA only with that 6 variables, in order to examine potential performance of prediction model with less input variables.

Table 4.3: Results of developed ANNs (6 algorithms with 6 input variables)

| *Algorithm* | *Absolute Average Error* | *Standard Deviation* | *Linear Correlation* |
|---|---|---|---|
| *Quick* | *0.259* | *0.323* | *0.886* |
| *Dynamic* | *0.262* | *0.330* | *0.881* |
| *Multiple* | *0.255* | *0.319* | *0.889* |
| *Prune* | *0.259* | *0.323* | *0.886* |
| *RBFN* | *0.294* | *0.367* | *0.849* |
| *Exhaustive Prune* | **0.253** | **0.317** | **0.890** |

Table 4.3 represent the developed models with reduced number of inputs which gives satisfactory results. Furthermore, Exhaustive Prune stands out as the best algorithm by all criteria, giving the smallest absolute average error (0.253, the smallest standard deviation (0.317), and at the same time, the largest linear correlation coefficient (89%). A developed multilayer neural network with Exhaustive Prune method consists of input layer with 6 variables, output layer with one variable and two hidden layers with 24 and 18 neurons. Comparing two developed ANN with Exhaustive Prune method it can be concluded that there exists no significance, so by occasion it can be also used model developed with 6 input variables

## V. CONCLUSION

The model for prediction of students' success is developed to enable exam officers to predict students that have the potential for learning advanced courses and also students who need additional education so as to enhance their knowledge. Using information about students after their first year of studies as input variables, a developed model of multilayer neural network has the flexibility to predict the success of students at the end of their studies. Development of such a model gives an opportunity of recognizing which aspects of the academic plan and program should be improved so as to induce students to figure harder and improve their knowledge in certain scientific branches. this method should even be useful to students so they may adapt their learning habits, their work and grades so they might achieve the general success they need.

Future developments and research, firstly in terms of advanced use of the concepts EDM, will include integration of an outsized number of input variables, like people who are directly associated with studies and socio-economic and demographic indicators, their comparative analysis, and secondly in way of model development for predicting study success supported variables that were analyzed in this research.

Furthermore, this predictive systems approach facilitates and maximizes the identification of these factors (or predictors) of the educational processes which participate in varying degrees within the modelling of various levels of performance in academic outcomes in education. If we are able to identify specific profiles of scholars, specializing in the foremost important variables, this opens major possibilities for the development of assessment procedures and also the planning of pre-emptive interventions. provided that this system allows for the accurate prediction of actual academic performance a minimum of one year beforehand of it actually being measured (GPA), it's implications for the appliance of those methods in educational research and within the implementation of warning diagnostic programmes in early settings. These results also inform cognitive theory and help within the development of improved automated tutoring and learning systems. Although a number of the variables involved, like the academic level of the oldsters, are impossible to change in their effects on academic performance at the time of the assessment, they are doing inform policy and indicate the burden that a lot of social and environmental factors influence future academic performance. This methodological and conceptual approach allows us to contemplate an outsized number of variables simultaneously and choose those which are most relevant and permit a greater degree of intervention to enhance student performance, including early intervention programmes for college kids in need of special support. The capacity to very accurately classify expected student performance, which is additionally what tests try and do, without the performance sampling problems with traditional testing, and employing a much broader spectrum of all factors influencing a student's performance. In fact, it also represents a more valid approach to educational assessment thanks to its overall accuracy and also the breadth of the constructs considered to classify the expected performance. Traditional assessments aren't sufficient for more complex assessments or for assessment systems that shall serve multiple direct and indirect purposes, in complex educational situations (Mislevy, 2013; Mislevy, et. al., 2003) during this respect, this new approach allows for the conceptualization and development of recent modes of assessment which could facilitate breaking aloof from traditional styles of testing while at the identical time improving the standard of the assessment process (Segers, et. al., 2003).

Finally, the employment of ANN along with other methods like cluster analyses and Kohonen networks could contribute to the study of the precise patterns of these variables which influence the educational process for every level of performance. In fact, a serious observation resulting from the information during this study is that variables contribute to the prediction in relatively small proportions, and it's the joint effect of the many contributing variables that might cause significant changes in performance. In other words, there's no magic but rather the buildup of effects from of these various sources that produce significant changes in outcomes. These results provide an insight into learning questions from a special perspective and one that has important implications for educational policy and education at large.

## REFERENCES

**[1]** Almond, R., Mislevy, R.J. and Steinberg, L.S., 2003. On the structure of educational assessments. Measurement: Interdisciplinary research and perspectives, 1(1), pp.3-62**.**

**[2]** Ayán, M.N.R. and García, M.T.C., 2008. Prediction of university students' academic achievement by linear and logistic models. *The Spanish journal of psychology*, *11*(1), pp.275-288.

**[3]** Cascallar, E. and Musso, M., 2008. Classificatory stream analysis in the prediction of expected reading readiness: Understanding student performance. *International Journal of Psychology*, *43*(3-4), pp.242-242.

**[4]** Cascallar, E., Boekaerts, M. and Costigan, T., 2006. Assessment in the evaluation of self-regulation as a process. *Educational Psychology Review*, *18*(3), pp.297-306.

**[5]** Detienne, K.B., Detienne, D.H. and Joshi, S.A., 2003. Neural networks as statistical tools for business researchers. *Organizational Research Methods*, *6*(2), pp.236-265.

**[6]** Eriksen, B.A. and Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & psychophysics*, *16*(1), pp.143-149.

**[7]** Fan, J., McCandliss, B.D., Sommer, T., Raz, A. and Posner, M.I., 2002. Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience*, *14*(3), pp.340-347.

**[8]** Garson, G.D., 1998. *Neural networks: An introductory guide for social scientists*. London: Sage Publications Ltd.

**[9]** Isljamovic, S., & Suknovic, M., 2014. Predicting Students 'academic Performance Using Artificial Neural Network: A Case Study from Faculty of Organizational Sciences. *The Eurasia Proceedings of Educational and Social Sciences*, *1*, 68-72.

**[10]** Kanakana, G. and Olanrewaju, A., 2011. Predicting student performance in engineering education using an artificial neural network at Tshwane University of Technology. In *Proceedings of the ISEM*. South Africa: Stellenbosch. pp. 1–7.

**[11]** Kent, R., 2009. Rethinking data analysis-part two: Some alternatives to frequentist approaches. *International Journal of Market Research*, *51*(2), pp.1-16.

**[12]** Provost, F. and Kohavi, R., 1998. Guest editors' introduction: On applied research in machine learning. *Machine learning*, *30*(2), pp.127-132.

**[13]** Kyndt, E., Cascallar, E. and Dochy, F., 2012. Individual differences in working memory capacity and attention, and their relationship with students' approaches to learning. *Higher Education*, *64*(3), pp.285-297.

**[14]** Kyndt, E., Musso, M., Cascallar, E. and Dochy, F., 2015. Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. In *Methodological challenges in research on student learning* (pp. 55-76). Antwerp: Garant.

**[15]** Lippmann, R., 1987. An introduction to computing with neural nets. *IEEE ASSP Magazine*, *4*(2), pp.4-22.

**[16]** Marquez, L., Hill, T., Worthley, R. and Remus, W., 1991, January. Neural network models as an alternative to regression. In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences* (Vol. 4, pp. 129-135). IEEE.

**[17]** Marshall, D.B. and English, D.J., 2000. Neural network modeling of risk assessment in child protective services. *Psychological Methods*, *5*(1), p.102-124.

**[18]** Mavrovouniotis, M.L. and Chang, S., 1992. Hierarchical neural networks. *Computers & chemical engineering*, *16*(4), pp.347-369.

**[19]** Mellanby, K., 1958. The birth of Nigeria's university. Methuen.

**[20]** Mislevy, R. J., 2013. Measurement is a Necessary but not Sufficient Frame for Assessment. *Measurement*, *11*, 47–49, 2013

Musso, M., Kyndt, E., Cascallar, E. and Dochy, F., 2012. Predicting mathematical performance: The effect of cognitive processes and self-regulation factors. *Education Research International*, *2012*.

**[21]** Musso, M.F., Kyndt, E., Cascallar, E.C. and Dochy, F., 2013. Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1*(1), pp.42-71.

**[22]** Posner, M.I., 1980. Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), pp.3-25.

**[23]** Romero, C. and Ventura, S., 2007. Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, *33*(1), pp.135-146.

**[24]** Segers, M., Dochy, F. and Cascallar, E. eds., 2006. *Optimising new modes of assessment: In search of qualities and standards*, Vol. 1. Springer Science & Business Media.

**[25]** Specht, D., 1991. A general regression neural network. *IEEE transactions on neural networks*, *2*(6), 568-576.

**[26]** Ting, S.M.R. and Man, R., 2001. Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education*, *17*(1), pp.75-80.

**[27]** Unsworth, N., Heitz, R.P., Schrock, J.C. and Engle, R.W., 2005. An automated version of the operation span task. *Behavior research methods*, *37*(3), pp.498-505.