FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 3 No. 4, December, 2019, pp 478 –487

PERFORMANCE COMPARISON OF COMPUTER-BASED TEST AND PAPER-PENCIL TEST USING SUPPORT VECTOR MACHINE

[1]Oniyide, M. K, [1]Kolawole, A. O. and [2]Dada, E. G.

[1]Joint Admissions and Matriculation Board (JAMB), National Headquarters, Abuja, Nigeria
[2]Department of Mathematical Sciences, Faculty of Science, University of Maiduguri, Borno State, Nigeria

Corresponding Author's Email: gbengadada@unimaid.edu.ng

ABSTRACT

The Joint Admissions and Matriculation Board (JAMB) have over the years been in the news over the use of Computer Based Test (CBT) mode over the Paper Pencil Test (PPT) mode for its Unified Tertiary Matriculation Examination (UTME). This study examines the two test modes, and also tries to ascertain which particular mode makes the unified tertiary matriculation examination more comfortable for the students in privileged environment and those in the rural areas. Predicting student performance can be useful to the managements in many contexts. The purpose of this research work is to do a performance evaluation of computer-based and paper-based version of Joint Admissions and Matriculation Board (JAMB) test data conducted in the previous year using a robust Support Vector Machine model. This work attempts to determine whether there is any difference in the performance of student when comparing CBT to identical PPT test mode and also investigate the levels of malpractices involved in both test mode. Experimental results demonstrated CBT has better predictive accuracy and root mean square error compared to PPT.

Keywords: Support Vector Machine, Computer Based Test, Paper Pencil Test, Performance, Unified Tertiary Matriculation Examination.

INTRODUCTION

Information and Communication Technology (ICT) has become, within a very short time, one of the basic building blocks of modern society (Bodmann and Robinson, 2004). Many countries now regard understanding it and mastering its basic skills and concepts as very crucial in education. This is because it adds value to the processes of learning and to the organization and administration of learning institutions. It encompasses different types of technologies, which are utilized for capturing, processing and transmitting data and information, using computer facilities. One specific form of Information Communication Technology (ICT) for assessment is the Computer-Based Testing (CBT), also known as Computer-Based Assessment or e-assessment/testing which has been available in various forms for more than four decades (Sorana-Daniela and Lorentz, 2007). Over the years, CBT has grown from its initial focus on certification testing for the IT industry, to a widely accepted delivery model serving elements of virtually every market that was once dominated by Paper-and-Pencil Testing (PPT) (Abubakar and Adebayo, 2014; Alabi, *et al*.2012). It is a method of administering tests in which the responses are electronically recorded, assessed, or both. It is commonly available for several admissions tests throughout the developed countries (Oduntan, Ojuawo and Oduntan, 2015). Computer-based testing or computer-based assessment is a green computing (green IT) strategy used to reduce paper consumption (Peter *et al.*, 2004). Computer-based tests offer several benefits over traditional paper-and-pencil or paper-based tests (Oduntan, Ojuawo and Oduntan, 2015). Technology based assessment provide opportunities to measure complex form of knowledge and reasoning that is not possible to engage and assess through traditional methods (Bodmann and Robinson, 2004). Reducing paper consumption will indirectly reduce greenhouse gases and energy consumption. Minnesota Pollution Control Agency reported that 40 reams of paper are equal to 1.5 acres of pine forest absorbing carbon for a year and each ream of paper is equal to roughly 12 pounds of carbon dioxide not removed from the atmosphere (Minnesota, 2011). Through the paper making process, paper industry represents around 10% of all global greenhouse emissions.

The purpose of this research work is to evaluate the comparability of computer-based and paper-based version of Joint Admissions and Matriculation Board (JAMB) test data conducted in the previous year using a robust Support Vector Machine model. This work attempts to determine whether there is any difference in the performance of student when comparing CBT to identical PPT test mode and also investigate the levels of malpractices involved in both test mode.

RELATED WORK

Several work in literature have been done on the comparability of PPTs and CBTs. Mazzeo and Harvey (1998) provided one of the earliest reviews, which included some 30 comparability studies about a range of tests such as ones focusing on

intelligence, aptitude, personality, and achievement. Their review revealed mixed evidence concerning the comparability of CBTs and PPTs. They found test mode seemed to have no effect on power tests, but a considerable effect on speeded tests. Mead and Drasgow (1993), who used meta-analysis to examine the mode effect on timed power tests and speeded tests, arrived at a similar conclusion. However, in a meta-analysis of ability measure tests performed by Kim (1999), CBTs and PPTs were found to have comparable average scores.

More recently, with the growing interest in CBTs in K-12 education, a number of comparability studies have been conducted focusing on these applications. The Texas Education Agency (TEA) issued a technical report (2008) that reviewed comparability studies across different content areas (Mathematics, English Language Arts including Reading and Writing, Science and Social Studies) in K-12 tests. In each content area, they found discrepancies between the conclusions of some empirical studies—some studies indicate that a CBT is more difficult than a PPT or vice versa while some studies show that CBTs and PPTs are comparable (Kim and Hyunh, 2007). A similar pattern was observed by Paek (2005), although she concluded that "in general, computer and paper versions of traditional multiple-choice tests are comparable across grades and academic content".

This trend toward CBT and PPT comparability has also been echoed in studies that use meta-analysis to examine the mode effect for K-12 populations. Wang,*et al.* (2007) conducted two meta-analysis studies on the K-12 student math and reading achievements, respectively. The results indicate that administration mode has no statistically significant effect on student math or reading achievement scores. Likewise, Kingston (2009) synthesized 81 comparability studies in K-12 multiple-choice tests performed between 1997 and 2007 and found that the estimated effect size was small across all the studies. Although the majority of recent comparability studies have indicated that CBT and PPT are comparable across delivery medium, the results are not unanimous. The inconsistency in the findings is not surprising, given that these comparability studies involve a wide range of variations in content areas, participants, data collection designs, and item format.

In their empirical study, Olsen *et al.* (1986) compared paper administered, computer-administered, and computer- adaptive tests by giving third- and sixth grade students' mathematics applications achievement tests. This study found no significant difference between paper-administered and computer-administered tests, and equivalences among the three test administrations in terms of score rank order, means, dispersions, and distribution shapes. Mazzeo and Harvey (2008) pointed out that computer-based test graphics may affect test scores and consequently their equivalence with paper-and-pencil versions, and that test with reading passages may be more difficult when given on computers. Bunderson *et al.* (1992) suggested performance on some item types such as paragraph comprehension are likely to be slower if presented by computer, while some types such as coding speed items are likely to be faster.

In reviewing all above-mentioned studies, Bugbee (1996) concluded that the use of computers indeed affects testing. However, computer-based and paper-and-pencil tests can be equivalent provided the test developers take responsibility for showing that they can be administered to evaluate the performance of students the same way. Bugbee (1996) stated that the barriers to the use of computer-based testing are inadequate test preparation and failure to grasp the unique requirements for implementing and maintaining computer tests. In other words, Bugbee (1996) reminded us that some factors such as the design, development, administration and user characteristics must be taken into consideration when computers are used for testing.

As computer-assisted instruction (CAI) has grown in popularity, computer-based testing has become more and more appropriate for assessing students' CAI learning achievement. As Bugbee (1996) states, if what is being tested is done on or learned from a computer, then it is more appropriate to assess it by computer. Thus, computers are used as the sole vehicles for distributing tests, not only as alternatives to paper-and pencil testing. Alessi and Trollip (1991), in their classic book on computer-based instruction, devoted a chapter to the design, development, and use of computer-based testing. They pointed out that the two main ways of incorporating computers into the testing process are for constructing or administering tests. When constructing tests, test developers use computers' word processing abilities to write test items and use their storage capacities to bank and later retrieve test items. Jacob and Chase (1992) pointed out that computers can present test materials paper-and-pencil test cannot, for example, 3-D diagrams in computer graphics, motion effects, rotating geometric forms, animated trajectories of rapidly moving objects, and plants seen from different angles. Shavelson *et al.* (2002) further suggest using computer simulations for hands-on performance assessment. Computer-based testing can also be designed to provide test-takers with immediate feedback and scoring.

However, Wise and Plake (1989) found that immediate feedback may contribute to students' test anxiety. Bernt *et al.* (1990) also pointed out that general computer-test anxiety may influence test takers. They considered that, although anxiety tends to be a random variable among people, it must be identified and dealt with. Jacob and Chase (1992) also suggested discontinuing item-by item feedback until further research has been done on the computer-test-anxiety issue.

In addition to the traditional multiple-choice, fill in the blank, and short essay type questions, Rasmussen *et al.* (1997) suggested Web-based instruction include participation in group discussions and portfolio development to evaluate students' progress. Khan (2007) also suggested Web-based instruction designers have facilities that allow students to submit comments about courseware design and delivery.

Although many researchers, e.g., Rasmussen *et al.* (1997) and Ravitz (1997) considered testing and evaluation to be of utmost importance in Web-based instruction and suggested some design strategies and techniques, few usable systems have been developed and no empirical data collected to explore the feasibility of computer assisted testing and evaluation on the Web. The search for creative and effective tools and methods for conducting testing and evaluation in such a complicated technology-dependent learning environment represents a challenge for system designers and instructional designers.

Support Vector Machine

A Support Vector Machine is a complex nonlinear modeling technique based on a model of a classifier. A SVM is used to predict outputs (dependent variables) from a set of inputs (independent variables) by taking linear combinations of the inputs and then making nonlinear transformations of the linear combinations using activation function (Vapnik, 1998). It can be shown theoretically that such combinations and transformations can approximate virtually any type of response function. Thus, SVM use large numbers of parameters to approximate any model. SVM are often applied to predict future outcome based on prior experience. For example, a SVM application could be used to predict who will respond to a direct mailing. Support vector machine are becoming very popular with data mining practitioners, particularly in medical research, finance and marketing. This is because they have proven their predictive power through comparison with other statistical techniques using real data sets. The example of a simple feed forward SVM model processing is shown in Figure 1.

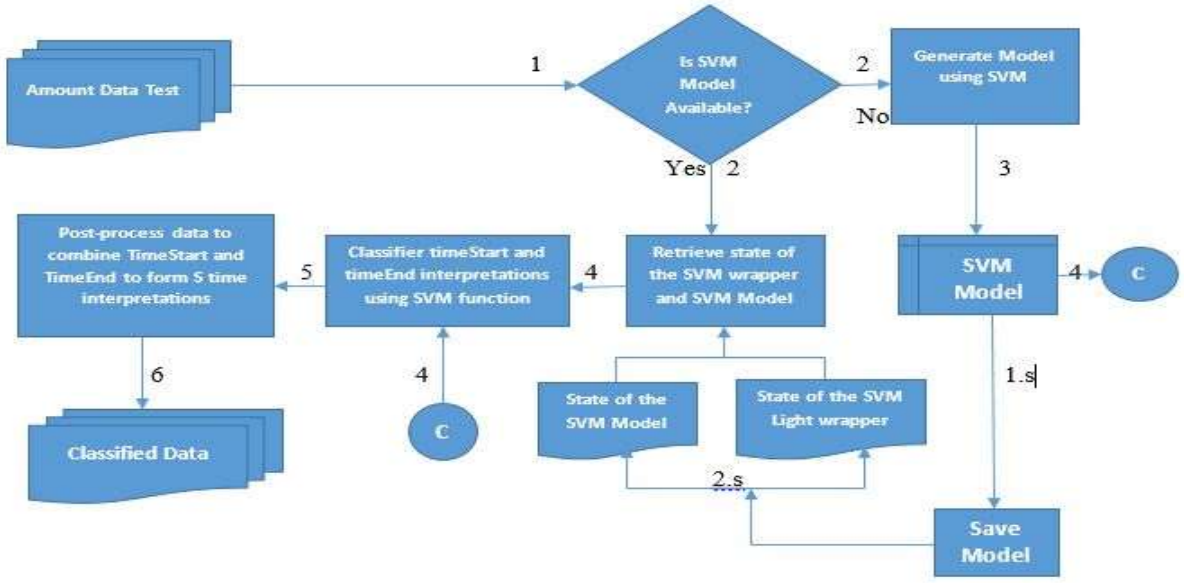

Fig. 1: Simple Feed Forward SVM Model Processing (Source: http://svmlight.joachims.org)

A Support Vector Machine (SVM) is a supervised machine-learning method that can be used for binary regression: associating a non-labeled sample to one class or another (Gunn, 1998). Given a set of labeled training data (features) the SVM finds the best plane that separates the two classes. If the data is not linearly separable, then it is possible to translate the data into a higher-dimensional space to find a separator, referred to as the *kernel-trick*. Once the plane has been found, the SVM can associate new non-label data to a specific class. Furthermore SVMs allow the inspection of weight functions, making it possible to identify which features are most responsible for regression. The SVM can be used for predicting that will greatly enhance the performance and reliability of computer-Based testing (CBT) in the future.

The architecture of the ensemble of Support Vector Machine regression is a collection of m SVM regression, each trained on a balanced subsample of the training data (approximately equal number of positive and negative instances obtained by sampling with replacement from the entire training data). The ensemble of SVM regression is trained and evaluated on the original distribution of the glycosylation data. The prediction of the ensemble of SVMs is computed from the predictions of the individual SVM regression. Figure 2 shows the architecture of the ensemble of SVM regression.

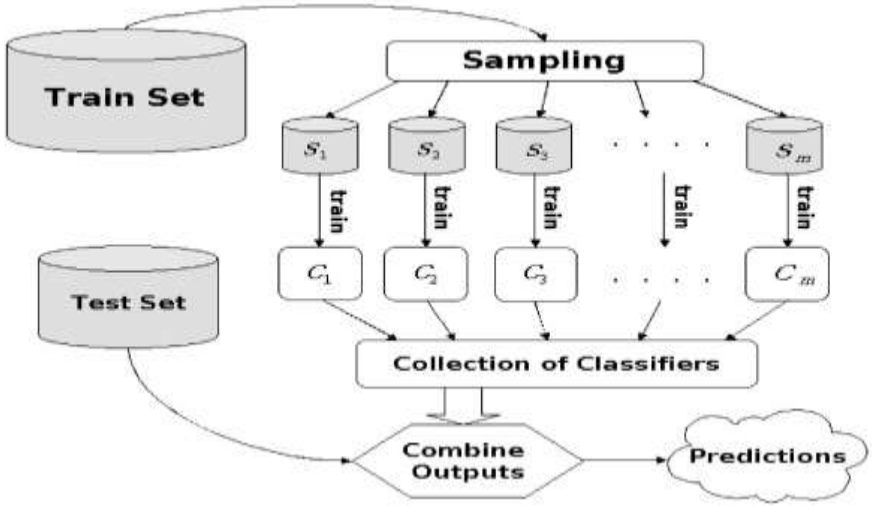

Fig. 2: Architecture of the ensemble of Support Vector Machine regression a collection of mSVM. (Cornelia, *et al*., 2007)

METHODOLOGY

Research Design

The research design used in conducting the study was descriptive assessment. The target population for the study was the totality of candidates who did JAMB examination seeking admissions into various tertiary institutions in Nigeria for both CBT and PPT, which runs into hundreds of thousands. Thus, both quantitative and qualitative instruments were employed to reveal more valid and reliable results and consequently reach more solid conclusions. From the two groups, valid data samples were collected from CBT and PPT respectively. Table 1 illustrates the design of this study in summary:

Table 1: The order of test taking across the two groups

Group	Prompt
A	Paper and Pencil Test
B	Computer-Based Test

Research Participants

The study population consists of hundreds of thousands candidate who were enrolled by JAMB for paper-pencil testing from 2008 to 2014 and computer based testing from 2013 to 2016 for both UME and UTME seeking admissions to the tertiary institution either the public or private university.

Research Questions

The purpose of this study was to examine comparability of student scores obtained from computerized and paper-and-pencil test scores. This study attempts to answer the question: Is there any statistically significant difference between performance on computer-based tests and their paper-based ones?

Research Modelling Framework

This section describes the general framework of the method proposed in this research. It first introduces the data collection methods followed by feature selection, data resampling, classification, and weighted voting methods. Figure 3 shows the block diagram of the proposed method.

Data Collection

Students who were enrolled in for Unified Matriculation Examination (UME) from 2007-2009 and Unified Tertiary Matriculation Examination (UTME) from 2010-2016 participated in this study.

Data Analysis

The accuracy of PPT UME and CBT UTME is known by comparing the reliability using the classical test theory approach and comparing the value of information function of both tests directly and through its relative efficiency. The reliabilities are estimated by computing the reliability using support vector machine model. One comparison was performed for analysis. A comparison was made between the candidates who took part in the paper and pencil and the candidates who took the computer-based test. Data were stored and sorted on Microsoft Office 2013 Excel spreadsheet software

SVM Algorithm

The proposed model for the prediction of student performance who partook in CBT and PPT was implemented using support vector machine analyzer tool. The acquired candidates' dataset from Joint Admissions and Matriculation Board (JAMB), Nigeria was analyzed and pre-processed into the required format using spreadsheet statistical tool. The entire dataset has a dimension of 6x12 (representing the dataset candidates' records with each having 6 attributes) and it was partitioned into training and testing sets. The training process begins by selecting the training dataset using a working set selection strategy based on kernel function conditions of optimality. The process waits for the SVM model to be available, before the model was trained, updated and then released to unblock waiting threads. As the training continues, kernel values were cached and the workload of kernel computations decreases. This procedure was repeated until training stopped and optimum accuracy with training error rate was computed. The trained model was tested by dividing the testing dataset into small sizes and the outputs of the model on each of them were computed. Also, the performance of the model was further evaluated using SMV regression.

The SVM algorithm was used on the training and testing data collected from the previous process. Moreover, the sigma and constant values are used for the regression. The data are collected from the database. The training records are used for testing. The resulted attributes from the genetic algorithm are used. The kernel function (radial basis), sigma (0.5) and constant input data (class = 1or 2 or 3 and the other class = –1). The SVM regression function for regressive data get the arrangement details. The analysis produced the results and find the accuracy for the overall function.

RESULT AND DISCUSSION

Data source

The Psychometric data of students were elicited by means of direct link with educational body database of student's record performance from Joint Admissions and Matriculation Board (JAMB) in Nigeria. The dataset comprises of information about student records' percentage that did UTME from 2007 to 2014 and 2013 to 2016 for Paper-Pencil Test (PPT) and Computer-Based Test (CBT) respectively. The data sources provided the yearly information about total candidate's demographics, the percentage of male and female candidates and the percentage of malpractices recorded during the examination in which six significant attributes of candidates were extracted for experimentation in the study. Table 2 presents the dataset characteristics, account, and classification of PPT attributes into various categories while table 3 shows the description of candidates dataset for CBT.

Table 2: Training data of PPT candidates' record

YEARS	TOTAL CANDIDATES	% MALES	% FEMALES	TOTAL CENTERS	% MALPRACTICES
2007	969231	57.42	42.58	2785	3.80
2008	1076273	57.11	42.89	3045	3.96
2009	1242599	56.81	43.19	2568	3.21
2010	1375671	55.93	44.07	2810	3.63
2011	1493611	55.78	44.22	2872	1.20
2012	1503933	55.80	44.20	3055	3.79
2013	1629105	55.95	44.04	1047	0.78
2014	1012990	57.04	42.96	1865	0.98

Table 3: Predicted Testing data set with few records

YEARS	TOTAL CANDIDATES	% MALES	% FEMALES	TOTAL CENTERS	% MALPRACTICES
2007	969231	57.42	42.58	2785	3.80
2008	1076273	57.11	42.89	3045	3.96
2009	1242599	56.81	43.19	2568	3.21
2010	1375671	55.93	44.07	2810	3.63
2011	1493611	55.78	44.22	2872	1.20
2012	1503933	55.80	44.20	3055	3.79
2013	1629105	55.95	44.04	1047	0.78
2014	1012990	57.04	42.96	1865	0.98
2015	1205648	55.34	44.66	2145	

Table 4: Training data of CBT candidates' record

YEARS	TOTAL CANDIDATES	% MALES	% FEMALES	TOTAL CENTRES	% MALPRACTICES
2013	916996	59.39	40.61	1092	0.02
2014	619182	55.85	44.15	3371	0.47
2015	1475600	56.01	43.99	8495	0.18
2016	1592462	56.22	43.78	7590	0.37

Table 5: Predicted testing data of CBT candidates' record

YEARS	TOTAL CANDIDATES	% MALES	% FEMALES	TOTAL CENTRES	% MALPRACTICES
2013	916996	59.39	40.61	1092	0.02
2014	619182	55.85	44.15	3371	0.47
2015	1475600	56.01	43.99	8495	0.18
2016	1592462	56.22	43.78	7590	0.37
2017	1698742	57.67	42.33	8532	

After the training and cross Validation, the model was tested with the test data set and the following results were obtained. This involves the input variable data that is supplied to the model without the output variable results. The output from the model is then compared with the actual variable data. The comparison is summarized in table 8. The descriptive output results obtained from PPT data sets are depicted in figures 3 and 4. The output results obtained from training CBT data sets are as shown in figures 5 and 6.

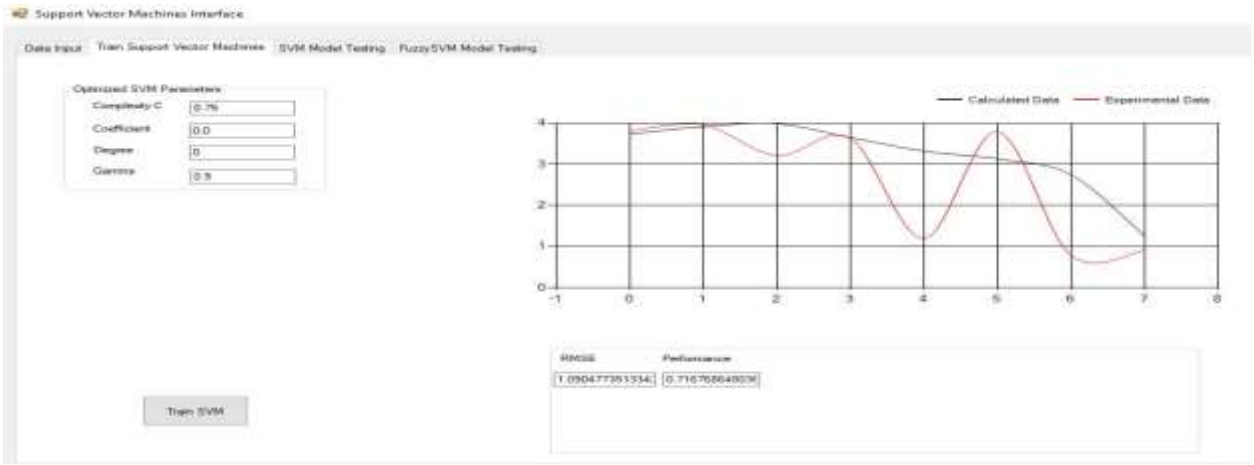

Fig. 3: The Training PPT data set from Support Vector Machine Model

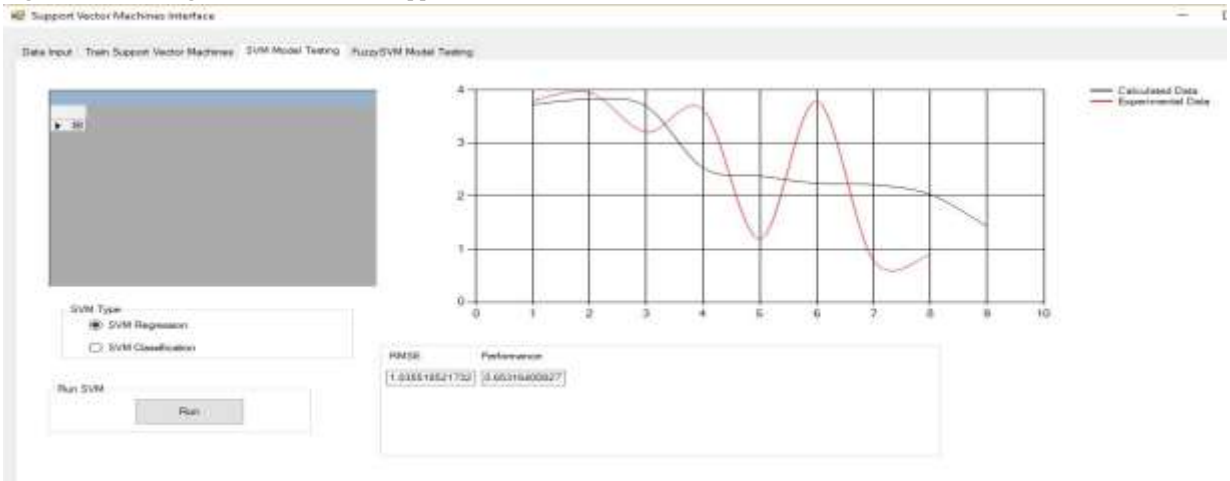

Fig. 4: The Testing PPT data set from Support Vector Machine Model

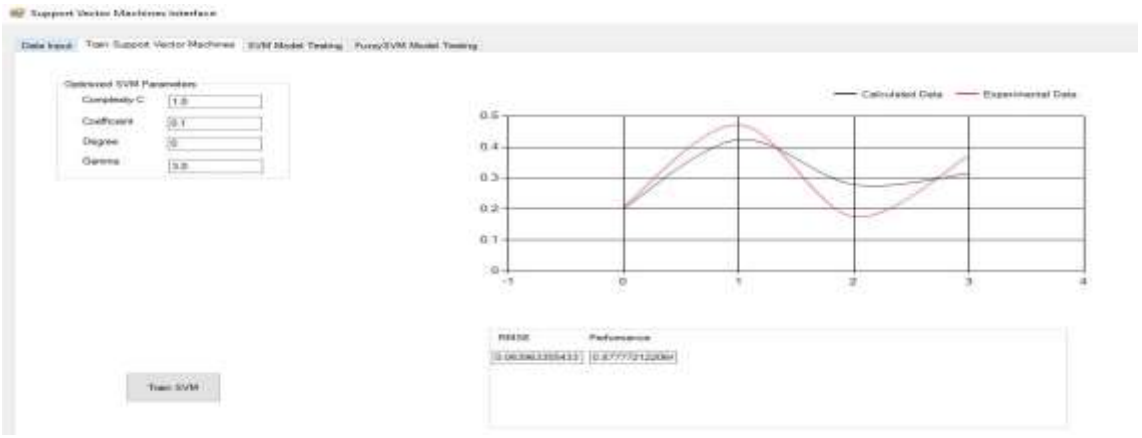

Fig. 5: The Training CBT data set from Support Vector Machine Model

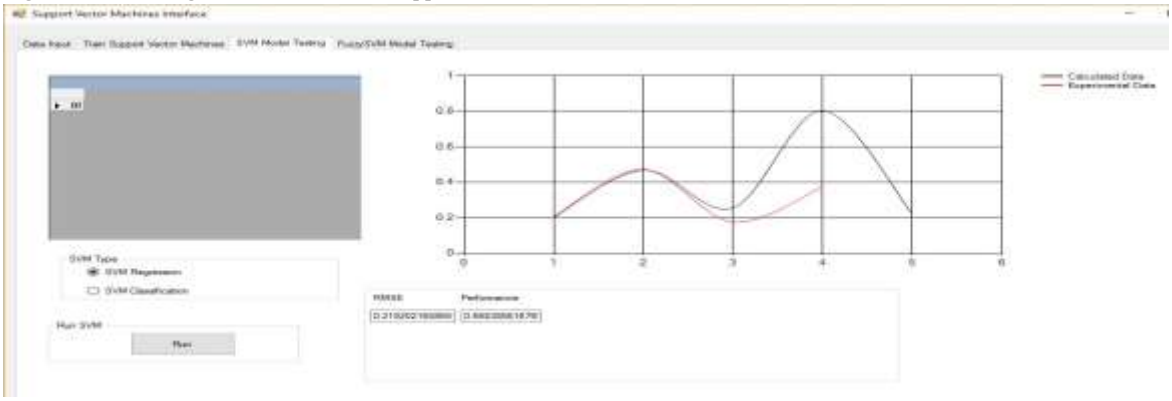

Fig. 6: The Testing CBT data set from Support Vector Machine Model

Performance Evaluation and Comparison of the results

In the present study, the performance comparison of SVM classifier for CBT and PBT malpractices predictions was done. The findings of the study suggested that there was statistically significant difference between malpractice performance levels across the two modes. The outcome suggests that in establishing the CBT, the test items must be clearly stated and avoidance of complexity, that is, complex items should be excluded in the test, as this would help reduce the level of malpractices in test participants.

Table 6: PPT prediction performances of examination malpractices

Input	Original Data (%)	Predicted Data (%)
2007	3.80301496753612	3.73008054091463
2008	3.95578073592852	3.82582203719557
2009	3.20932175223061	3.69390754509885
2010	3.62804769454324	2.53383899048745
2011	1.19582675810502	2.38411104950000
2012	3.78919805603042	2.24298253375010
2013	0.78405013795918	2.21791364451157
2014	0.90642553233496	2.03127998702028
2015		1.43074549400000

The results presented in table 6 and figure 7 show that if JAMB conducted the UTME in 2015, the level of malpractices will be higher as predicted than that of 2014 UTME as shown in the results (1.43% and 0.91%) respectively. With these outcomes, the continuation of PPT shows that the level of malpractices such as leakage of examination papers, use of machineries of all sorts by candidates, bride taking by examination officials, impersonation, use of unauthorized gadgets, and so on would be increased. By comparing table 6, the results demonstrated that CBT conducted in 2015 is 0.175% as original data achieve the objectives of

ensuring 100 percent elimination of all forms of examination malpractice that had been a major challenge in the conduct of public examinations in the country.

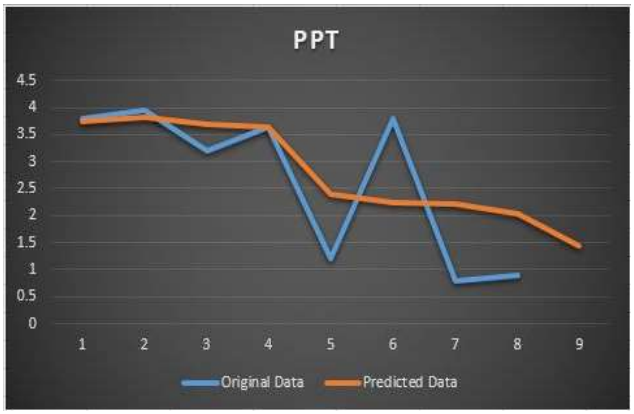

Fig**. 7:** Graph of PPT between the original and predicted data of malpractice levels

Table 7: CBT prediction performances of examination malpractices

Input	Original Data (%)	Predicted Data (%)
2013	0.2072064212179	0.2018940867636
2014	0.4720744466085	0.4689730628324
2015	0.1751152073732	0.25581892992919
2016	0.3725677598586	0.80343597521002
2017		0.22232137164502

From table 7 and figure 8, the results indicated that computer-based testing (CBT) can curb examination malpractices and enhanced safety as the level of malpractices decline from 0.373% in 2016 to 0.222% in 2017 as predicted. The configuration of the level of model is a good indicator for proper comprehension of CBT.

Fig**. 8:** Graph of CBT between the original and predicted data of malpractice levels

Table 8: Comparison of PPT and CBT test mode

Test Mode	Training Dataset		Testing Dataset	
	RMSE	PERFORMANCE	RMSE	PERFORMANCE
PPT	1.09048	71.7%	1.03552	65.3%
CBT	0.06396	87.8%	0.21920	66.8%

These results are supported by the comparison between the values of the information function obtained on PPT and that of CBT. First, the test mode analysis showed that there is a difference in students' performance on paper based tests and computer based tests. This difference becomes quite obvious when looking at the root mean square errors (RMSE) of paper based and computer based tests for both training and testing dataset. Table 8 shows a summary of all tests' root mean square errors (RMSE) and performance. Second, by running both test mode, there is a high significant correlation between the performances of the CBT having 87.8% and 66.8% respectively

compare to that of both PPT that have 71.7% and 65.3% respectively.

Analysis of the results indicate that both CBT and PPT of the SVM regressive are reliable for experimental research and no testing effects were found for test scores of the two testing modes. In other words, the SVM score is stable and consistent over time. It shows that a participant who sits for the CBT and PPT would most probably yield a similar post test score. However, testing effects occurred in testing time and testing motivation for the PPT mode. The finding is consistent with previous research results which state that the PPT mode has internal and external validity problems. Compared to the PPT mode, the results showed that the CBT mode was more stable and consistent in terms of internal and external validity because no testing effects were found in intrinsic and extrinsic testing motivation, except for the work avoidance dimension. They reported that students performed faster in CBTs because they did not have to spend time writing down their responses. Besides that, as a whole, there was a treatment effects on testing motivation. The results showed that the CBT has effectively increased intrinsic and extrinsic motivation of the test takers in challenge, curiosity, self-efficacy, involvement, joyfulness, comprehension and social dimensions. Regarding examination paper leakage and other examination malpractices which have be-deviled the assessment systems in Nigeria, 100% of the respondents were of the opinion that CBT can deliver a malpractice (especially impersonation) free examination due to the deployment of full biometric authentication system and also eliminate the practice of miracle centers in other examinations similar to the UTME. However, there is the issue of comparability between the testing modes that needs to be resolved. CBT is more than a decade old. Standardization of testing procedures, lower costs, time savings, improved scoring accuracy, immediate feedback, adaptability are some benefits of CBT but still its comparability to PPT is controversial. The mode of item presentation is a basic driving factor in the comparability of test modes. It is the cognitive workload that can be affected with different presentation formats. Another issue is the comfort of the examinee with the testing medium: students may be more familiar with scrolling and clicking a mouse or tapping on a touch sensitive screen instead of using the pencil as a primary writing tool. In this way they enjoy CBT and prefer them to PPT.

CONCLUSION

This study is an attempt to use SVM algorithms with feature extraction and selection technique for predicting student performance and comparing the performance of candidates who did the JAMB using SVM with the real time results. In summary, the CBT mode is more reliable in terms of internal and external validity and no testing effect on test score was found in CBT mode. The CBT mode reduced testing time and increased testing motivation of the participants. An advantage of the CBT is increasing testing motivation would increase response rates. This study shows that factors such as test-mode effect, characteristics of test takers, features of computer-based testing systems and the test content are possible confounding factors which affect the performances of students using any of the test modes. Once these parameters are controlled on CBT, similar test performance can be reached with the PPT. Accordingly, once the CBT environments are decided, the possible confounding factors such as personal characteristics of test takers, the features of computer-based testing systems and the test content need to be controlled.

Conventional model for predicting academic performance of student are limited in their performance accuracy and generalization, but the suggested methodology can deal with uncertainty in its determinants. The model presented proves to have prospective application in educational system. Future work in the same area will include exploring additional determinants which affect academic performance of a student in an institute. Such determinant factors can be analyzed for their significance in contribution to successful prediction.

REFERENCES

Abubakar, S. A. and Adebayo, O. F. (2014). Using Computer Based Test Method for the Conduct of Examination in Nigeria: Prospects, Challenges and Strategies. Mediterranean Journal of Social Sciences Vol 5 No 2, pp. 47-55.

Alabi, A. T., Isaa, A. O., and Oyekunle R. A. (2012). The Use of Computer Based Testing Method for the Conduct of Examinations at the University of Ilorin, International Journal of Learning & Development, ISSN 2164-4063 2012, Vol. 2, No. 3.

Alessi, S. M. andTrollip, S. R. (1991). Computer-Based Instruction: Methods and Development. – Englewood Cliffs, NJ: Prentice-Hall, pp. 205-243.

Bernt, F.M., Bugbee, A.C. and Arceo R.D. (1990). Factors influencing student resistance to computer administered testing, J. Res. Comput. Educ. Vol. 22. – № 3. – P. 265-275.

Bugbee Jr., A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. Journal of Research on Computing in Education, 28, pp. 282–299.

Blanzieri, E. and A. Bryl, 2007. Evaluation of the highest probability SVM nearest neighbor classifier with variable relative error cost. Proceedings of the 4th Conference on Email and Anti- Spam, Aug. 2-3, Mountain View, California, USA., pp. 5-9.

Bodmann, S. M. and Robinson, D. H. (2004). Speed and Performance Differences among Computer Based and Paper-

Pencil Tests. Journal of Educational Computing Research, 31(1), 51 – 60.

Cortes, C. andVapnik, V. (1995). Support-vector network. Machine Learning, 20, 1-25.

Cornelia C, Jivko S, Adrian S and Vasant H. (2007). Glycosylation site prediction using ensembles of support vector machine classifiers.

Gunn, S. R. (1998). Support vector machines for classification and regression. Retrieved from the Support Vector Machines website: http://www.svms.org/tutorials/ Gunn1998.pdf

Peter, C., Bill, I. and David, S. (2004). Using computers for Assessment in medicine. British Med. J., 329(7466):606-609.

Sorana-Daniela, B., and Lorentz, J. (2007). Computer-based testing on physical chemistry topic: A case study. Inter. J. Educ. Develop. using Info. Com. Technol., 3(1): 94-95.

Oduntan, O.E., Ojuawo, O.O. and Oduntan, E.A. (2015). A Comparative Analysis of Student Performance in Paper Pencil Test (PPT) and Computer Based Test (CBT) Examination System. Research Journal of Educational Studies and Review Vol. 1 (1), pp. 24-29.

Jacobs, L. C., and Chase, C. I. (2011). Developing and Using Tests Effectively: A Guide for Faculty. – San Minnesota Pollution Control Agency, http://156.98.19.245/paper/, 2011.

Kim, D. H., and Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of Algebra and Biology assessments. Journal of Technology, Learning, and Assessment, 6(4). Available from http://www.jtla.org.

Kim, J. P. (1999). Meta-analysis of equivalence of computerized and P&P tests on ability measures. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association. Chicago, IL.

Kingston, N. M. (2009). Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. Applied Measurement in Education, 22(1), 22-37.

Malik Yousef, Segun Jung, Louise C Showe and Michael K Showe (2007). Recursive Cluster Elimination (RCE) for classification and feature selection from gene expression data. BMC Bioinformatics, 8:144 doi:10.1186/1471-2105-8-144

Mead, A. D., and Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114, 449-458.

Mazzeo J., and Harvey A. L. (1988). The Equivalence of Scores from Automated and Conventional Educational and Psychological Tests. – College Entrance Examination Board, New York, College Board Rep. 88-8.

Olsen J. B., Maynes D. D., Slawson D., and Ho K. (1986). Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement. – San Francisco, CA,1986.

Paek, P. (2005). Recent trends in comparability studies (PEM Research Report 05-05). Available from http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf.

Rasmussen K., Northrup P., and Lee R. (1997).Implementing Web-based instruction // Web-Based Instruction. NJ: Educational Technology, pp. 341-346.

Ravitz, J. (1997).Evaluating learning networks: A special challenge for Web-based instruction. – NJ: Educational Technology, 1997. – P. 361-368.

Shavelson R. J., Baxter G. P., and Pine J. (1992).Performance assessments: Political rhetoric and measurement reality // Educ. Res. – 1992. – Vol. 21. – №. 4. – P. 22-27.

Vapnik, V. N. (1998). Statistical Learning Theory; Wiley: New York, NY, USA.

Wang, S., Jiao, H., Young, M. J., Brooks, T. E., and Olson, J. (2007). A meta-analysis of testing mode effects in Grade K–12 mathematics tests. Educational and Psychological Measurement, 67, 219-238.

Wise S. L. andPlake, B. S. (1989). Research on the effects of administering tests via computers, Educ.Meas.: Issues Practice, vol. 8, No. 3, pp. 5-10.

Yousef, M, Jung, S, Showe L.C., and Showe, M. (2007). Recursive cluster elimination (RCE) for classification and feature selection from gene expression data. The Wistar Institute, Systems Biology Division, Philadelphia, PA 19104, USA. BMC Bioinformatics.

Zhang, X., Lu, X., Shi, Q., Xu, X. Q., Leung, H.C., Harris, L., Iglehart, J., Miron, A., Liu, J., Wong. W. (2006). Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinformatics, 7(1):197