ISSN No:-2456-2165

Comparative Analysis of Classification Algorithms for Face Matching and Verification in E-Examination

Adetoba, B. T.
Computer Science Dept
Babcock University
Ilishan, Ogun State

Awodele, O.
Computer Science Dept
Babcock University
Ilishan, Ogun State

Alao, O. D.
Computer Science Dept
Babcock University
Ilishan, Ogun State

Nwaocha, V. O.
Computer Science Dept.
National Open University
Nigeria

Abstract:- Classification algorithms have been found to produce a better result for monitoring e-examination in terms of performance and accuracy in detecting examination impersonation in an e-learning environment. This paper presents the results of the comparative analysis of six classification algorithms (Logistic Regression, Multi-Layer Perceptron, Support Vector Machine, Random Forest, Bayes Network and Stochastic Gradient Descent) for face matching and verification in e-examination. This were compared and evaluated based on True Positive Rate (TPR), False Positive Rate (FPR), Precision, Recall, F-measure, Kappa Statistics (KS), Accuracy, Time to build model and Mean Absolute Error, using Waikato Environment for Knowledge Analysis (WEKA) to determine the best fit classifiers for the design of the model. The developed model was tested, using 250 facial images (dataset) acquired from the entire National Diploma students of Computer Science, Yaba College of Technology. The best fit classifiers (LR-based on Logistic Loss Function and SGD-Stochastic Gradient Descent) obtained from the comparison was used for binary classification, image optimization and monitoring. The results from the comparison showed that LR and SGD had leading performances with TPR, FPR, Precision, Recall, F-Measure, KS and Accuracy values of 100%, Mean Absolute Error value of zero. LR had shortest time of 0.01 second and SGD with 0.04 second, based on time taken to build model. LLF of LR and SGD played significance role in producing faster and optimal face recognition results. The technique can be employed by examiners and learning management specialists to conduct malpractice free e-examination.

Keywords:- *Classification Algorithms; E-Examination; Face Matching; Face Verification; E-Learning Environment.*

I. INTRODUCTION

E-examination has been considered a major system for assessing students in different parts of the world; yet, it has been faced with the challenge of security and impersonation. Diverse evolving steps have been taken to ensure a solution to addressing these challenges. In a situation where a stringent policy is missing, examination malpractices are imminent [1]. Cheating have been reported to occur frequently in the education domain [2], [3], [4]. Online examination malpractice could range from looking up answers in printed or handwritten materials availably kept for the purpose of cheating, spying the internet through search engines for possible answers and solution to examination problems, to student impersonation during an online examination [5]. It is also worthy to note that the aforementioned means of cheating in an examination is already been practiced by student. To consider e-examination as valid and reliable to implement, a means of securing the environment and activities becomes crucial for learning management specialists.

For a number of years, authentication approaches using biometrics (facial, fingerprint, signature, etc.) have been employed to keep e-examination systems secure [6]. This has been shown to have effect in the identification of criminals, hence, employable in identifying students and controlling examination malpractices [7]. Biometrics is the science which measures both the cognitive and behavioural properties that are specific to each person for identity verification [6]. The utilization of scientific knowledge is mostly for access control as well as identification. Surveillance cameras are also used for monitoring the behaviour of users during the e-learning process. But despite all efforts taken to reduce examination malpractice, studies have shown that each has its own shortcomings [8]. It becomes essential to prove through practice which machine learning technique is more effective in conducting biometric authentication in an e-examination environment. This will enable researchers make choices considering the metrics used in evaluating them.

This research paper conducts a performance comparison on six classification algorithms (Bayes Network, Support Vector Machine, Multi-Layer Perceptron, Random Forest, Logistic Regression and Stochastic Gradient Descent) to help researchers determine the best fit classifier for face matching and verification in e-examination environment using selected metrics. The metrics considered are accuracy, true positive rate, true negative rate, precision, recall, f-measure, kappa statistics, model building time and mean absolute error.

This paper aim to evaluate and analyzed various classification algorithm in order to choose the best fit techniques that is capable of shunning examination malpractice in e-examination.

II. LITERATURE REVIEW

A number of research works have been conducted in the area of security, face matching and verification using machine learning techniques.

[9] proposed an approach to secure computer-based tests for tertiary institutions which was stated as an improvement for the protection of examination questions against unauthorized access. The focus of the research was the development of independent and secure computer base testing system with an improved level of integrity, authenticity and confidentiality. The system was modeled using UML use cases, Flowchart and entity relationship (E-R) diagrams. The testing of the system was done locally and shown to be effective in authentication through the use of a Fingerprint Mechanism.

[3] presented a discussion on the effect of examination malpractices in West Africa, proposing technology as the real solution to mitigating them. It was revealed from the study technologies available for the mitigation of malpractices. Some of the solutions stated are facial recognition and biometric system. [10] presented a system to monitor classes live via the use of a face recognition algorithm that is able to process the images of students in a real-time camera-input environment using OpenCV. The captured images and their details are stored in the backend using SQLite. An update that indicates the presence and absence of the student is also done by employing Firebase. A mobile application is used to check the presence and absence of students through a QR code used to authenticate. The detection of faces was done using Haar classifier and local binary pattern histogram has been used to detect the face.

[11] employed support vector machine (SVM) for carrying out face recognition on captured images and Principal Components Analysis (PCA) for the extraction of features on them as well as the reduction of dimensionality. The system made use of KNN for the face classification. After the comparative analysis was done, the combined utilization of SVM and PCA showed a higher accuracy result against other methods. The recognition rate of the approach was up to 92% in the detection on different face databases. [12] made use of machine learning techniques for detecting threats in real time and classifying targets that cross borders in a remote surveillance scenario. Viola-Jones algorithm was used for detecting objects in the videos. Both positive and negative videos were used to conduct the training objects such as humans, vehicles, and handguns. In order to annotate and classify the video in real-life, a threat level classifier and alert warning system were also added. The function of the threat level classifier is to categorize the videos in four-fold—safe, low, medium, and high. The warning is for specifying the type of warning depending on the intrusion type that the system detects. The accuracy for detecting human, vehicle and weapons are above 90%. [13] as a means of achieving higher level of accuracy, focused on the application of an advanced machine learning in face recognition.by using the Caffe and Nvidia DIGITS framework, the creation and training of dataset was done on the GoogleNet (inception) deep learning model. The framework performed better than conventional machine learning techniques in terms of accuracy. [14] explored gender classification from facial images through the use of different learning algorithms. A state-of-art classification of gender using classification methods such as Convolutional Neural Networks (CNN), Dual Tree Complex Wavelet Transform (DTCWT) + a Support Vector Machine (SVM) classifier, and feature extraction techniques such as Principal Component Analysis (PCA), Histograms of Oriented Gradients (HOG) and others with a classifier (SVM, kNN, etc). The comparison of these methods was done on two large datasets - FERET and Adjence. The CNN was proven to perform better than other methods in terms of accuracy considering how large the data are and its deep learning characteristics. [15] proposed a face recognition system for conducting biometric identification. They stated how important it was to computing due to its non-invasiveness. The system operates by first using an algorithm (Haar Classifier Cascade). that detects the faces by extracting them from video frames and generate a face database. The facial images were then filtered and preprocessed to enable them to be employed for recognition. The authors made use of a combination of machine learning algorithms for training the data from the face database. The algorithms tested for classifying the images are: K-Nearest Neighbor, Locally-Weighted Learning, Naive Bayes, Decision Trees, the classifiers are used for classify faces obtained from video frames. The result obtained reveals that the approach is sustainable to analyze large collections of videos in the absence of previous face labels.

Each of these researches have shown that facial recognition using machine learning is a major consideration for security. A number of techniques have been employed, yet, there is need to make researchers decided from comparison which method they should employ through a performance outcome.

III. RESEARCH METHODOLOGY

In comparing the performance of the supervised classification models, Waikato Environment for Knowledge Analysis (WEKA) 3.8 was employed to develop the model with 10-Fold Cross-Validation (10-F-C-V) test mode. The six classification algorithms compared were Bayes Network, Support Vector Machine, Multi-Layer Perceptron, Random Forest, Logistic Regression and Stochastic Gradient Descent. These were chosen due to their vast adoption in the classification phase of facial recognition and verification in e-exam monitoring. Classifier model was done in relation to full training set after which the value of correctly classified instances, sensitivity, specificity, time taken to learn, kappa Statistics and mean absolute error were computed to compare their performance from which two algorithms with leading performances were chosen concerning loss function and image optimization. The metrics considered were given as follows.

- Accuracy (ACC): This measures the number of instances properly classified divided by the overall number of instances or the proportion of instances properly classified as in equation (1)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Where TP is the True Position, TN is True Nonposition, FP is Fake Positive and FN is Fake Negative.

- True Positive Rate (TPR): This is often referred to as sensitivity ratio. It measures the number of true instances that are properly classified or abnormal instances as in equation (2).

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \quad (2)$$

- True Negative Rate (TNR): This is also referred to as specificity ratio. It measures the number fake instances properly classified as negative or termed normal instances as in equation (3)

$$True\ Negative\ Rate\ \frac{TN}{TN + FP} \quad (3)$$

True Negative Rate = 1 – Fake Positive Rate (4)

- Precision: the proportion of positive predictions or instances that are actually true

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- Recall: It is used in measuring the success rate of prediction. The actual proportion that were properly identified

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

- F-measure: it measures the harmonic mean between precision as well as recall. F-measure was commonly used in the retrieval of information:

$$F - measure = \frac{2\ x\ Precision\ x\ Recall}{Precision\ + Recall} \quad (7)$$

- Kappa Statistic (KapS): this is a classifier performance measurement that estimates the similarity between the members of an ensemble in multi classifiers systems:

$$Kappa\ Statistic\ (Kaps) = \frac{P\ (O) - P\ (A)}{1 - P\ (A)} \quad (8)$$

where P(O) is the probability of observed among classifiers (accuracy) and P(A) is the probability that agreement among classifiers is due to chance.

Model building time: The time taken to build the classification algorithms in order to build the machine learning model.

Mean Absolute Error (MAE) is the number of predictions that vary from the true probability. P(r, s) is the estimated probability of r class to be in class s taking values [0, 1]:

$$MAE = \frac{\sum_{s=1}^{c} \sum_{r=1}^{m} |f\ (r,\ s) - P(r,\ s)|}{m\ .c} \quad (9)$$

WEKA was chosen to build the model because it is an open source package developed in Java and contains many machine learning algorithms. The chosen algorithms were implemented on Explorer application of WEKA 3.8 for easy and fair comparison of each algorithm. The developed model was tested, using 250 facial images (dataset) acquired from the entire National Diploma students of Computer Science, Yaba College of Technology, Nigeria.

IV. RESULTS AND DISCUSSION

The six binary classification algorithms namely Multi-Layer Perceptron, Support Vector Machine, Random Forest, Bayes Network, Logistic Regression and Stochastic Gradient Descent were built in WEKA 3.8 and evaluated with 10-Fold-Cross-Validation (10-F-C-V) technique for training and testing on the facial recognition, matching and verification dataset. The performance of the six classification algorithms were measured using nine existing performance benchmarks which includes Accuracy (ACCU), True Position (positive) Rate, Fake Positive Rate, True Nonposition (Negative) Rate, Precision, F-Measure, Kappa Statistic (KS), Model building time and Mean Absolute Error (MAE).

A. Comparison Based on Correctly Classified Instances

The study revealed that Support Vector Machine, Multi-Layer Perceptron, Logistic Regression, Random Forest, and Stochastic Gradient Descent had 100% accuracy while Bayes Network, had 98.08% for classification algorithms with 10-F-C-V as the test mode.

Hence from the comparison of the correctly classified instances all the algorithms except Bayes Network performed well, so they could have been used as the classifier. The results can be used to identify the number of instances that are of particular matching group as shown in Figure 1

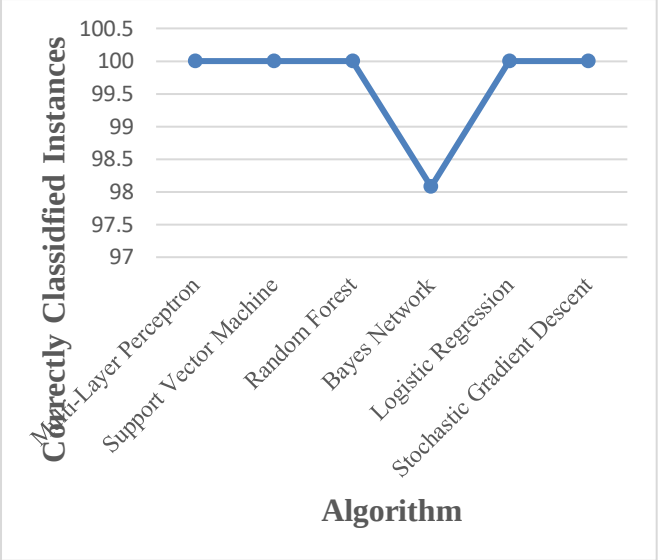

Fig 1:- Comparison based on properly classified instances with 10-F-C-V Method

B. Comparison Based on Model Building Time

The outcome of the time to learn the model shows that both Logistic regression and Bayes Network had the lowest possible running time with vales of 0.001 seconds while Random forest has the highest time. It is worthy of note that, Bayes Network has the lowest possible running time but was still not chosen as the optimal classifier. Therefore, time to learn cannot be considered as just the performance metric of choice. Hence, RF cannot be used as the classifier. The results with 10-F-C-V approach is depicted in Figure 2.

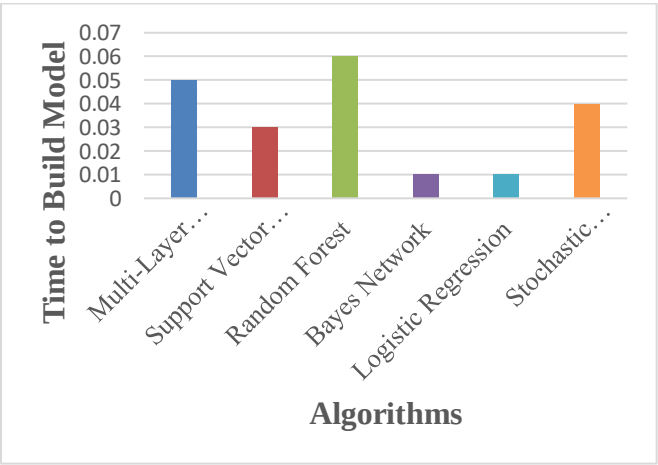

Fig 2:- Comparison Based on Model building time with 10-F-C-V Method

C. Comparison Based on Kappa Statistics

In 10-F-C-V approach, Multi-Layer Perceptron, Random Forest, Logistic Regression, Support Vector Machine and Stochastic Gradient Descent had values of 100%, which shows total similarity among the dataset while Bayes Network, had 95.22%. The performance comparison of the six algorithms based on Kappa Statistics signifies that Multi-Layer Perceptron, Random Forest, Logistic Regression, Support Vector Machine and Stochastic Gradient Descent are the classifiers with better results which shows the inter-agreement between the data items due to its robustness as depicted in Figure 3. However, Bayes network cannot be totally relied upon because it shows dissimilarity of 4.78% among the dataset.

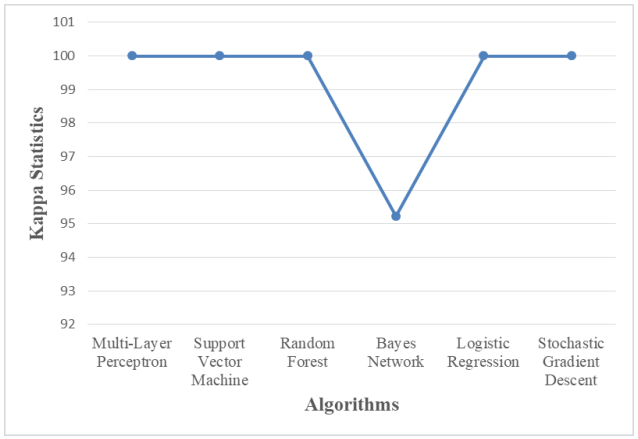

Fig 3:- Comparison Based on Kappa Statistics of 10-F-C-V Method

D. Comparison Based on Sensitivity

In 10-F-C-V method, Multi-Layer Perceptron, Support Vector Machine, Logistic Regression, Random Forest and Stochastic Gradient Descent had the greatest value of 100%, that is 0% insensitivity while Bayes Network, had 98.1%. Hence, from the comparison, the results shows that SVM, MLP, RF, LR and SGD are the classifiers that highly measures the percentage of instances that are properly identified in the dataset while Bayes Network has 1.9%, that is, insensitivity of proportion of instances that are incorrectly identified in the dataset as shown in Figure 4.

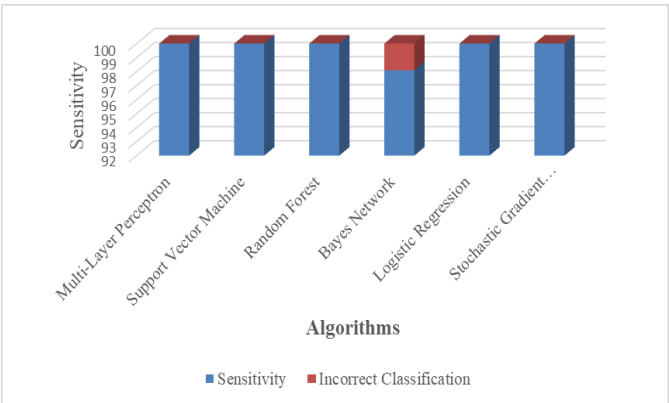

Fig 4:- Comparison Based on Sensitivity in 10-F-C-V Method

E. Comparison Based on Specificity

In 10-F-C-V method, Multi-Layer Perceptron, Support Vector Machine, Logistic Regression, Random Forest and Stochastic Gradient Descent had the greatest value of 100%, that while Bayes Network, had 99.3%. That is, all the negatives are identified with MLP, SVM, RF, LR and SGD except BN. Hence, from the comparison of the six algorithms on the basis of Specificity, the results shows that MLP, SVM, RF, LR and SGD were the classifiers that has higher values of negative instances identified in the dataset while Bayes Network has 0.7% of proportion of negatives that are not identified in the dataset as depicted in Figure 5.

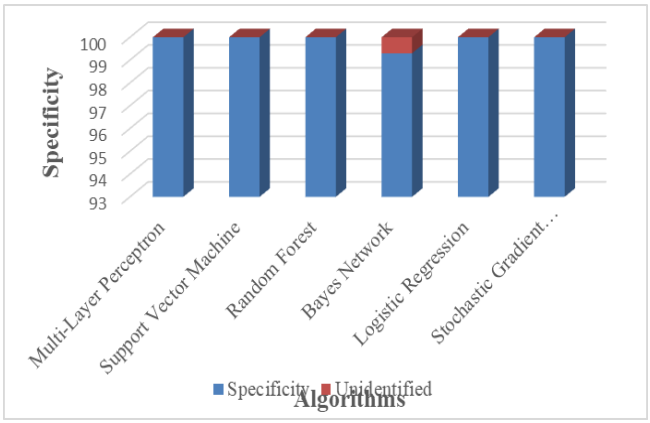

Fig 5:- Comparison Based on Specificity in 10-F-C-V Method

F. Comparison Based on Mean Absolute Error

In 10-F-C-V method, SVM, LR and SGD had the greatest values of 0.000 (zero), that is, the predictions did not deviate from the true probability while Multi-Layer Perceptron, Random Forest and Bayes Network had values of 0.0103, 0.0108 and 0.0201 respectively, that signifies the degree of deviation of predictions from the true probability to be 1.03%, 1.08% and 2.01% for Multi-Layer Perceptron, Random Forest and Bayes Network accordingly as depicted in Figure 6.

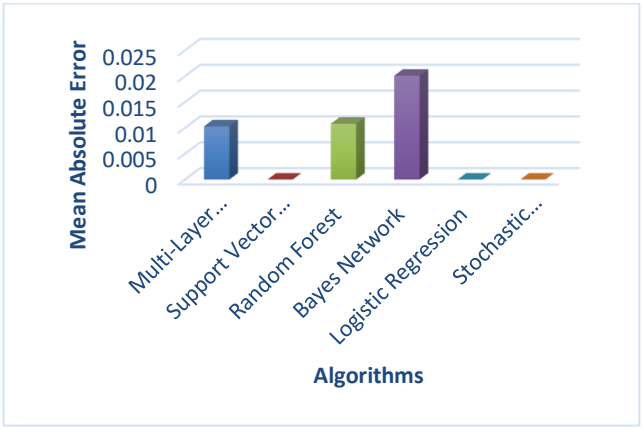

Fig 6:- Comparison Based on Mean Absolute Error in 10-F-C-V Method

G. Comparison Based on Mean Absolute Error, Precision and Time taken to build

In 10-F-C-V method, Multi-Layer Perceptron, Random Forest and Bayes Network had values have errors as depicted in Figures 7 which makes them unacceptable as the best classifiers even though Multi-Layer Perceptron and Random Forest have higher precisions. This is evident from the fact that the number of fault-prone modules classified that are in fact fault-prone are higher in Multi-Layer Perceptron and Random Forest compared to that of Bayes Network, Logistic Regression, Support Vector Machine and Stochastic Gradient Descent.

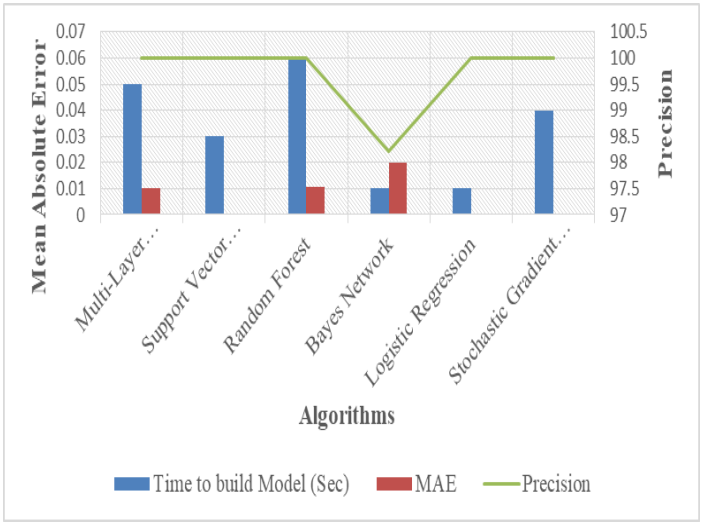

Fig 7:- Comparison Based on Mean Absolute Error, Time taken to build and Precision in 10-F-C-V Method

H. Summary of Algorithms Performance compared

The performance of the six algorithms namely MLP, SVM, RF, BN, LR and SGD were compared using Accuracy, TPR, FPR, TNR, Precision, F-Measure, KS, Model building time and MAE performance metrics. Logistic Regression performed excellently well in relation to all the metrics considered, followed by SVM and then SGD. However, due to computationally expensive nature of SVM, problem of selection of the right kernel function makes SGD the second choice. Furthermore, SGD has the ability to combine the determination of the loss function based on hinge loss using SVM with squared loss in relation to linear regression for facial image optimization. This incorporates in SGD the advantage of over fitting problem in SVM that is not as much as other algorithms as well as ability to handle complex non-linear data points. This in turn brings about better accuracy for facial image optimization. Table 1 shows the summary of the six algorithms and nine performance metrics used in the binary classification of the facial image capturing, verification and matching during the examination monitoring process.

ISSN No:-2456-2165

MLA	Accuracy (%)	TPR (%)	FPR (%)	TNR (%)	Precision (%)	F-Measure (%)	KS (%)	Model building time (Sec)	MAE
MLP	100	100	0	100	100	100	100	0.05	0.0103
SVM	100	100	0	100	100	100	100	0.03	0.0000
RF	100	100	0	100	100	100	100	0.06	0.0108
BN	98.08	98.1	0.7	99.3	98.22	98.1	95.22	0.01	0.0201
LR	100	100	0	100	100	100	100	0.01	0.0000
SGD	100	100	0	100	100	100	100	0.04	0.0000

Table 1:- Summary of the Model Performance in 10-F-C-V Method

Machine Learning Algorithm = MLA, True Position Rate =TPR, Fake Positive Rate = FPR, Kappa Statistics = KS, Mean Absolute Error = MAE, Logistic Regression = LR, Support Vector Machine = SVM, Multi-Layer Perceptron = MLP, Random Forest= RF, Bayes Network = BN, Stochastic Gradient Descent = SGD

V. SUMMARY AND CONCLUSION

This study aimed at developing an E-learning Authentication and Monitoring model for mitigating examination fraud based on the algorithms that has the minimum loss function among the best algorithms with the optimal performance. Evaluation of performance was carried out on six classification algorithms namely: Multi-Layer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), Bayes Network (BN), Stochastic Gradient Descent (SGD) and Logistic Regression (LR); out of which two algorithms were chosen. Logistic Regression was chosen for image classification with emphasis on loss function and Stochastic Gradient Descent (SGD) as the other for training the model in order to achieve highest facial image optimization. The performance of the six algorithms were tested based on the existing benchmark which includes True Negative Rate (TNR), True Position Rate (TPR), Fake Positive Rate (FPR), Accuracy (ACCU), Precision, *F*-Measure, Kappa Statistic (KS), Model building time and Mean Absolute Error (MAE) and comparison were made on their performances. The results of the comparison showed that Logistic Regression and Stochastic Gradient Descent had Mean Absolute Error (MAE) of zero values, that is, they did not record any error, hence, their consideration as optimal algorithms. Therefore, the study concludes that Logistic Regression can be used of binary image classification and Stochastic Gradient Descent for image optimization for training logistic loss function model for continuous facial image detection, recognition, verification, monitoring and matching processes and researches. This outcome would be used in developing a monitoring system in biometric recognition in an e-examination environment.

ACKNOWLEDGMENT

We like to acknowledge and appreciate all lecturers and Postgraduate students of Computer Science Department, Babcock University for their moral support and encouragement during this research work. Thank you and God bless.

REFERENCES

[1]. C. G. King, R.W. Guyette and C. Piotrowski, “Online Exams and Cheating: An Empirical Analysis of Business Students' Views”, Journal of Educators Online, 6(1), 1-11, 2009.

[2]. G. R. Watson, and J. Sottile, “Cheating in the Digital Age: Do Students Cheat More in Online Courses”, Online Journal of Distance Learning Administration, 13(1), 1556-3847, 2010

[3]. E. M. Onyema, A. U. Eucheria, N. A. David, A.I. Omar, and Q. N.Alsayed, “The Role of Technology in Mitigation of Examination Malpractices” in West Africa. sexual abuse, 7(10), 2019.

[4]. G. O. Jimmy, “Gsm technology and e-cheating in the Nigerian higher Institutions: a case of Akwa Ibom state polytechnic, Ikot osurua in Ikot Ekpene, Nigeria.” Journal of Research in Education and Society; 3(1), 2012.

[5]. C. O. Onyibe, U. U. Uma, and E. Ibina, “Examination malpractice in Nigeria: Causes and effects on National Development” Journal of Education and Practice, Vol. 6 No. 26, 2015

[6]. W. A. Al-Hamdani, “Secure e-learning and cryptography”, In K. Sullivan, P. Czigler & J. Sullivan Hellgren (Eds.), Cases on professional distance education degree programs and practices: successes, challenges and issues. Hershey, PA: information Science, 2014.

[7]. Q. Gao, “Biometric Authentication to Prevent e-Cheating”, International Journal of Instructional Technology and Distance Learning, 2012

[8]. O. Adebayo and S. Abdulhamid, “E- Exams System for Nigerian Universities with Emphasis on Security and Result Integrity”, International Journal of the Computer, the Internet and Management (IJCIM), 18(2), 2014.

[9]. O. Sarjiyus, “Securing Computer Based Testing (CBT) System for Tertiary Institutions in Nigeria”, Volume 3(3) 1-16, 2019

[10]. A. Krishna, and H. Tuli, “Live Class Monitoring Using Machine Learning”, In Advances in Computing and Intelligent Systems (pp. 385-389). Springer, Singapore, 2020.

[11]. P. Rakshit,, R. Basu, S. Paul, S. Bhattacharyya, J. Mistri, and I. Nath, “Face Detection using Support Vector Mechine with PCA”, Available at SSRN 3515989, 2020

[12]. A. Goyal, S. B. Anandamurthy, P. Dash, S. Acharya, D. Bathla, D. Hicks, and P. Ranjan, "Automatic Border Surveillance Using Machine Learning in Remote Video Surveillance Systems", In Emerging Trends in Electrical, Communications, and Information Technologies (pp. 751-760). Springer, Singapore, 2020.
[13]. R. Anand, T. Shanthi, M. S. Nithish, and S. Lakshman, "Face Recognition and Classification Using GoogleNET Architecture", In Soft Computing for Problem Solving (pp. 261-269). Springer, Singapore, 2020.
[14]. J. Lemley, S. Abdul-Wahid, D. Banik, and R. Andonie, "Comparison of Recent Machine Learning Techniques for Gender Recognition from Facial Images", In MAICS (pp. 97-102), 2016.
[15]. E. G. Amaro, M. A. Nuño-Maganda and M. Morales-Sandoval, "Evaluation of machine learning techniques for face detection and recognition", In CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers (pp. 213-218). IEEE February, 2012.