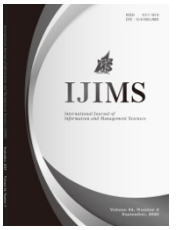
http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information
and
Management
Sciences

International Journal of
Information and Management Sciences

36 (2025), 187-204. DOI:10.6186/IJIMS.202509_36(3).0002



Application of Machine Learning in a Certificate Validation and Alert System (C-VAS)

Azubuike I. Erike[1], *Yusuf Mshelia*[2*], *Anthony N. Isizoh*[3], *Charles O. Ikerionwu*[1], *Ikenna C. Nwandu*[1], *Florence O. Elei*[1]
[1] Federal University of Technology, Owerri. Nigeria
[2] Data Aid, Abuja. Nigeria
[3] Nnamdi Azikiwe University, Awka. Nigeria

Keywords

Certificate Validation, Alert System, Intelligent Machine Learning, Fraud Detection, Neural Network, Authentication

Abstract

For emerging technologies that are driven by digital transactions and online interactions, the threat of sophisticated fraudulent activities poses a significant challenge to institutions and individuals alike. The issue of certificate forgery, a profound offense with far-reaching consequences, has infiltrated reputable organizations, regulatory bodies, and examination authorities. This research presents a comprehensive study aimed at addressing this challenge through the development of an intelligent machine learning-powered Certificate Validation and Alert System (C-VAS). The system leverages cutting-edge technology, data analytics, and machine learning methodologies to enhance fraud detection and notification. Various certificate verification methods, including optical character recognition, blockchain technology and artificial intelligence, are reviewed. The research focuses on machine learning approach that employs neural networks for feature recognition and classification. The model's performance is evaluated through Mean Square Error and regression analysis. Furthermore, an alert management module is integrated to confirm the authenticity of certificates with the issuing institutions. The research concludes by demonstrating the system's responsiveness based on data signal strength, emphasizing the system's reliance on network quality and also the model's high accuracy indicated by a Mean Square Error result of 0.000100mu and a regression score of 0.99373. Ultimately, the research presents a comprehensive solution that combats the grave challenge of certificate forgery and bolsters the integrity of digital credentials.

1. Introduction

In the rapidly evolving technological trend characterized by digital transactions and online interactions, the proliferation of sophisticated fraudulent activities has emerged as a pervasive

*corresponding author. Email: yusuf.mshelia@dataaid.org

threat affecting a wide range of organizations. Paper certificate forgery has remained a challenge even in this era. Certificate forgery, a grave offense subject to legal consequences, has cast its shadow over esteemed organizations such as Tertiary institutions, authoritative bodies like the Joint Admission and Matriculation Board (JAMB), and examination authorities like the West African Examination Council (WAEC). The most basic type of academic dishonesty involves the fabrication or procurement of completely falsified diplomas or certificates (Tamrat, 2022). This criminal act transcends mere manipulation of documents; it distorts the very essence of an individual's identity and ambitions (Noor, 2024). As societies recognize the gravity of this issue, stringent measures have been implemented to criminalize such actions. According to the Criminal Code Act in Australia, the offense of forgery in general stipulates that any individual who counterfeits a document, writing, or seal is deemed to have committed a felony. Unless specified otherwise, this offense can result in a penalty of imprisonment for up to three years, in cases where no alternative punishment is prescribed (Lanham et al., 2006).

In this context, the prevalence of fraudulent activities within digital environments, poses an imminent and substantial threat that demands a comprehensive response. Traditional methods of combating certificate fraud, often reliant on rigid rule-based approaches, have begun to falter in the face of increasingly sophisticated and adaptive fraudulent behaviours. This inadequacy has necessitated the need for innovative and adaptable solutions that can bolster the accuracy, effectiveness, and real-time capabilities of fraud detection procedures. Blockchain solution has been explored, although it is not devoid of constraints especially in the area of online connection (Vinayasree et al., 2024).

This study is motivated by the potential transformative impact that a machine learning-based intelligent certificate verification and alert system could wield in fraud detection and notification, and so proposes an intelligent certificate verification and alert system that leverages on the prowess of machine learning techniques to curb the menace of paper certificate forgery using technology.

The rest of the manuscript is organized as follows: Section two is a brief review of some related works. Section three presents the methodology for the research, development, and empirical evaluation of the proposed intelligent certificate verification and alert system. Section four discusses the results of the study, while section five concludes the paper. This research does not only contribute to the academic discourse surrounding fraud detection but also to the preservation of the integrity and authenticity of credentials within the digital landscape.

1. Review of literatures

Various methods exist for certificate verification. These include manual observation, optical character recognition, artificial intelligence (AI), signature analysis, template matching, scale invariant feature transformation, object fast rotated brief, speed of robust feature techniques, blockchain technology, and so on.

2.1 Manual Methods

The manual method, widely employed, involves visual inspection of signatures, dates, stamps, and seals to determine document authenticity. Although fast, it is prone to overlooking forged documents due to advancements in OCR technology that can replicate originals with high accuracy (Khandpur et al., 2017).

2.2 Computational Method

2.2.1 Template Matching and Feature-Based Detection

The advancements in technology have offered great opportunity to engage technological tools in various ways to detect forged documents. For instance, template matching, a high-level machine vision technique, identifies document parts by matching predefined layouts through Gaussian filters (Lee et al., 2012). In yet another approach, signature analysis, a major authentication method, examines signature peaks, inclinations, and content as features for matching documents. Online systems offer enhanced security, measuring various factors like pen inclination and pressure for written documents (Lee et al., 2012). Ke proposed a Speed Up Robust Feature (SURF) which employs wavelet responses and Gaussian weights for key point detection and feature description, facilitating faster matching (Ke & Sukthankar, 2004). Scale Invariant Feature Transform (SIFT) which was proposed by Ye addresses document transformations through keypoint estimation, orientation assignment, and descriptor generation which is necessary for (Ye & Doermann, 2015).

2.2.2 Optical Character Recognition (OCR) and Pre-processing

Researchers have also employed Optical Character Recognition (OCR) approach for fake document detection. This technology converts image-based text into editable format, involving pre-processing and character recognition stages (Neves et al., 2011; Romero et al., 2011). The preprocessing algorithms employed depend on factors like document age, skew, layout, and script type. Another study aimed at comparing the Otsu thresholding algorithm focused on accurately classifying pixels into foreground and background regions. The study performs algorithm development, testing on various images, and evaluation through threshold optimization techniques. While the Gaussian Otsu algorithm triumphs over the Otsu algorithm, particularly in bimodal distribution images, the study misses an in-depth analysis of algorithm performance on intricate histogram patterns, multifaceted modes, and the impact of noise and lighting variations on efficacy (Yousefi, 2011). These have attending gaps.

2.2.3 Text Recognition and Segmentation

In yet another approach, Salvador Espan˜a-Boquera et al.'s work targets the enhancement of offline handwritten text recognition through hybrid Hidden Markov Model (HMM)/Artificial Neural Network (ANN) integration. The task involves recognizing diverse writing styles, tackled by merging HMM's structural attributes and ANN's probability estimation capabilities. Innovative techniques for distortion elimination and text normalization are presented, leading to substantial recognition rate improvements. However, avenues for alternative preprocessing or feature extraction methods remain unexplored. In addition, the computational complexity of hybrid models warrants investigation, as does their adaptability to diverse handwriting datasets and languages (España-Boquera et al., 2011). This work focused more on handwriting

recognition which can further be incorporated into academic certificate validation system. The study by Satadal Saha and team's proposal centres on a Hough Transform-based technique for text segmentation across applications like OCR, business card reader systems, and license plate recognition. The technique showcases promising segmentation accuracy across varied datasets which may be applied to certificate verification (Saha et al., 2010).

Further in text recognition, Krishna Subramanian and team's novel approach to text localization and extraction concentrates on detecting character strokes as the initial step. The method leverages stroke properties to accurately identify strokes, which in turn aid in precise character extraction. The study evaluates the algorithm's performance on a dataset of complex images, revealing strengths in stroke detection while acknowledging limitations with certain font styles. The study underscores the potential of the proposed method for text localization tasks and acknowledges ongoing enhancements and challenges (Subramanian et al., 2007).

2.3 Emerging Technologies

2.3.1 Blockchain-Based Verification

More recent studies focus on the use of blockchain technology. The concept of blockchain is based on the distributed ledger system where each generated digital certificate is integrated into the blockchain with its authentication features (Dinesh Kumar et al., 2020; Fernández-Blanco et al., 2024; Nguyen et al., 2018; Oblikwu & Dekera, 2019). A hash value is first generated for each certificate using a specified hashing algorithm as demonstrated by (Boonkrong, 2024). The fixed length hash value is integrated into the transaction block and is validated by members in the blockchain. Accepting and rejecting a certificate into the chain is done following a consensus protocol.

2.3.2 Artificial Intelligence (AI) Approach

Some recent researches have also explored the use of classification techniques and image recognition techniques in machine learning to approach the issue. Mostly, the use of convolutional neural networks was discussed by (Chen et al., 2018; Nwanze et al., 2023), especially as it concerns image recognition and classification, all with acceptable degrees of accuracy.

These earlier studies lay a strong foundation which can further be explored in academic certificate verification. Meanwhile, although the use of blockchain technology definitely offers a more robust approach to certificate verification, the cost associated with the system involves that an institution be part of the decentralized ledger system. More so, the researches focused only on the domain of the verifying institutions instead of the institution's document (certificate) being verified, while others focus on verifying digital certificates. This research however, aims to develop Intelligent Machine Learning-Powered Certificate Validation and Alert System. This system presents a decentralized approach to certificate verification. This process involves the amalgamation of a verification system coupled with an authentication system which provides validation for each verified certificate. The process is further explained in the next section.

3. Materials and methods

3.1 Data collection

The researchers secured permission from the Exams and Academic Records department at Nnamdi Azikiwe University to gather data that was under their custody for the study. The main data source consisted of 1120 certificates (970 first-degree and 150 postgraduate) from the department spanning 2018 to 2021, augmented by 40 additional certificates from Alumni. Secondary data sources included 25 certificates volunteered by Alumni and 34 certificates replicated with expert assistance using Microsoft Publisher for testing purposes. In total, 1179 samples were collected, acquired through optical character recognition (OCR), and utilized in scanned softcopy form for the research.

3.2 Data preprocessing

In data preprocessing, various steps are taken to enhance the collected data. Firstly, the Matlab database toolbox is employed to standardize data attributes like size, content, quality, and colour by converting all data into a uniform Matlab.File (.m file) format, facilitating input into an artificial intelligence system for training. Furthermore, there is a strong emphasis on the quality of optical character image acquisition devices due to its substantial impact on the accuracy of document recognition. Utilizing high-definition image acquisition tools, the system captures samples of document certificates, guaranteeing the attainment of optimal image quality. Image binarization simplifies data by converting it to black and white, reducing colour complexity for consistency. Segmentation identifies and isolates important data elements, such as text and signatures, while feature extraction extracts relevant data features, especially related to handwritten text patterns and signatures, for classification purposes, condensing them into a statistical feature vector.

3.3 Data Splitting

The dataset was divided into three in the following ratios: 70:15:15.

_ Training Set: 70% of the data is used for training the model. This is where the model learns patterns and relationships in the data.

_ Validation Set: 15% of the dataset is allocated for validation purposes. This validation set serves the essential role of fine-tuning hyper-parameters and overseeing the model's performance throughout the training process. It proves instrumental in identifying overfitting tendencies and aids in the selection of the optimal version of the model.

_ Testing Set: The remaining 15% is kept as a separate testing set. This set is used to evaluate the model's performance on completely unseen data. This provides an estimate of how well the model will generalize to new data.

A 70:15:15 data split (for training, validation, and testing) was chosen over k-fold cross-validation due to its simplicity, lower computational cost, and suitability for relatively large or stable datasets where variance in performance across folds is minimal. It allows for a clear separation between model training, hyperparameter tuning, and unbiased performance evaluation. Training the algorithm follows an iterative procedure that includes passing variables through the algorithm, contrasting the output with the expected results, then fine-tuning the

weights and biases embedded within the algorithm to potentially enhance accuracy, and rerunning the variables iteratively until the algorithm generates the accurate outcome. The outcome of this rigorous training process is a refined and high precision machine learning model.

3.4 Design Process

ML, often termed predictive analytics, builds and leverages algorithms to learn from data, creating generalized models for accurate predictions (Jordan & Mitchell, 2015). Building an ML application involves data collection, model training, testing, and prediction, refining algorithms with more data (Bell et al., 2022). The process involves three developmental approach as depicted in fig. 1, the block diagram model of the system. The output of the input validation block serves as the input into the model. And the output of the model serves as input to the Alert system.

- Design of the input validation and protocol
- The Alert Management Module

Fig. 1 *Block Diagram Model of the System*

Fig. 2 is the condensed algorithmic flow chart of the certificate fraud detection and alert system. The interlinked process expressed in the block diagram of figure 1 is differentiated using different dashed lines. The first section validates the input document to ensure that it of the correct file type based on a predefined criteria used in developing the user interface. The system rejects any file whose type is not supported. The accepted file are jpg, jpeg, and png.

The second section defines the machine learning model developed in for the system. The development involves the data collection process, data cleaning, etc. The outcome of the model if positive is fed into the alert system. A positive outcome indicates that the model has verified the supplied certificate as original. That notwithstanding, this claim by the model is passed through a second test by issuing a verifying request to the database of issuing institution through an API call. If the system returns the original uploaded certificate alongside the personal details of the owner, then the verification and validation are correct.

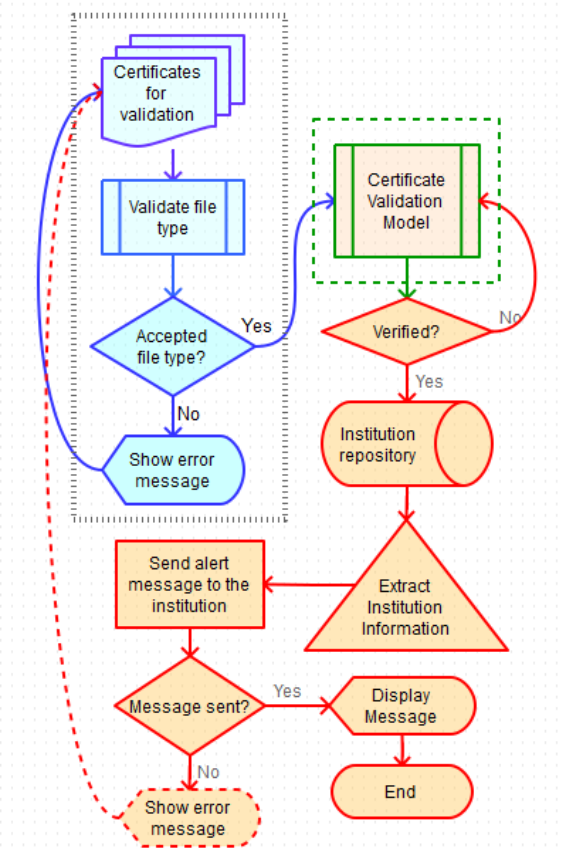


Fig. 2 *Process Flow Diagram of the Certificate Fraud Detection and Alert System Architecture*

3.4.1 Input Validation and Protocol

The input validation and protocol section comes in two parts. The first section is the institution registration section. Since the entire C-VAS is designed to be an open source platform that can be leveraged upon by any institution, any interested institution is meant to follow a registration protocol. The registration protocol requires the institution to provide an API link with which an API call will be made on the institution's server upon successful ML-based verification. The API link is configured to receive the registration number of the certificate submitted for verification.

The second part of the input validation is the user interface depicted in fig. 3 It receives the registration number of the certificate to be verified and the file input.

It is assumed that any subscribing institution must have a database of all their graduated students. By registration and submission of the API to the C-VAS application, the institution authorizes the C-VAS app to remotely send an HTTP request to her database. A positive response from the API confirms the authenticity of the certificate document in question.

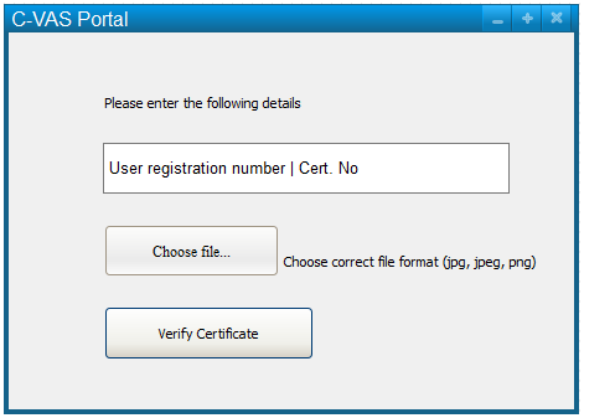


Fig. 3 *A Screenshot of the Developed C-VAS User Interface Showing the Certificate Input Capture Module*

3.5 The ML Model Selection

The ML Model selected for the work is the Artificial Neural Network (ANN). Artificial Neural Network falls under the category of supervised learning, which means that it learns from labeled training data to make predictions or classifications. ANNs are particularly suited for tasks involving complex patterns and relationships in data, such as image recognition, natural language processing, and more.

An ANN consists of interconnected nodes, often referred to as neurons, organized in layers. The three main types of layers are:

1.Input Layer: This layer receives the raw input data, whether it's images, text, or other types of data.

2.Hidden Layers: These layer, positioned between the input and output layers serve as intermediaries. Within each intermediate layer, individual neurons undertake a weighted summation of their inputs, subsequently applying an activation function to this summation before transmitting the outcome to the succeeding layer.

3.Output Layer: This layer provides the final prediction or classification.

The connections between neurons have associated weights that are learned during the training process. The goal of training an ANN is to adjust these weights so that the network can accurately map inputs to their corresponding outputs.

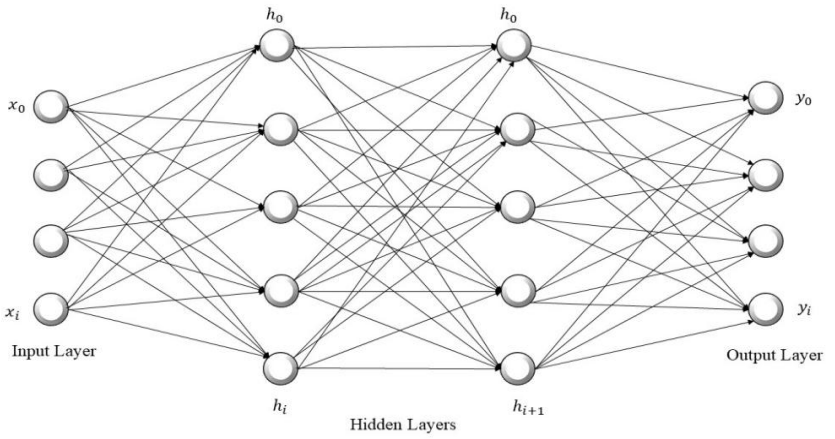


Fig. 4 *Schematic Representation of a Typical Neural Network Architecture*

ANNs are versatile and powerful models that have led to significant advancements in various fields. Hence, the choice of the ML model.

The functioning of the ANN adheres to the data progression illustrated in fig. 5 Neurons equipped with interconnected weights and biases are utilized by the system to identify feature vectors present within the initial document dataset (training dataset). Subsequently, the training algorithm is applied to facilitate data learning for the purpose of classification and precise decision-making.

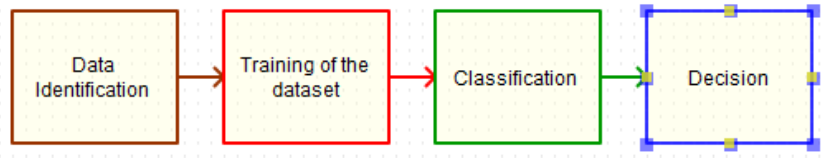


Fig.5 *Block Diagram Illustrating the Machine Learning Workflow*

Note that the process of training an Artificial Neural Network (ANN) involves initializing the network's architecture, including the number of layers and neurons per layer. Then, the network is provided with a labelled training dataset, which consists of input data and their corresponding target outputs. During training, the network adjusts its internal parameters, such as weights and biases, using optimization algorithms like gradient descent. This adjustment process is carried out iteratively with the aim of reducing the disparity between predicted outputs and actual targets. This iterative process persists until the network attains a level of accuracy on the training data that is deemed satisfactory. Once trained, the ANN can make predictions or classifications on new, unseen data.

The adjusted parameter values set for the neural network is depicted in table 1.

Table 1. *Values Set for the Neural Network Parameters*

Parameters	Values
The network hidden layers	10
Train epoch values	16
No. delayed reference input	2
Max epoch values	30
Maximum feature output	3.1
No. delayed output	2
Maximum interval per sec	2
Number of non-hidden layers	10
Maximum reference value	0.7
Minimum reference value	-0.7
No. delayed feature output	2

The ANN, set up with numerous neurons featuring weighted connections and bias functions, underwent a feed-forward process using the feature vectors extracted from the training dataset. At the start of the process, the features are first identified and the variables selected. Then, the weight and threshold are initialized. The feature vector from the training set are set. If the values are not accepted, the features are further trained. The training ends when the output features have been deemed acceptable, otherwise, the iterative process continues.

The number of hidden layers used was meant to ensure that the model adequately learns complex non-linear patterns and hierarchical features from data. In other to provide a balance between overfitting and under fitting, the number of train epochs was adjusted to 16. The number of delayed reference inputs was set to allow the model to consider previous input states for better prediction accuracy while the max epoch values was set to provide a safety range to prevent unnecessary training that could lead to overfitting the model. To provide numerical stability and consistency with known domain values, the maximum feature output range was set at 3.1. The summation of these features was subjected to a non-linear activation function, specifically the Tansig function, which maps the values to a statistical range between 0 and 1. This step introduces non-linearity and removes negative values from the feature vectors. The adjusted values are then iteratively refined through the back propagation algorithm, as illustrated in Fig. 6 This process facilitates learning from the feature vectors, leading to the development of a benchmark classification model. The inclusion of two delayed outputs and two delayed feature outputs was to enable the ANN to capture temporal dependencies, needed to enhance prediction accuracy in dynamic systems. Limiting the interval to two updates per second was to balance temporal resolution with computational efficiency, while the use of ten non-hidden layers helps manage input-output complexity. The reference value limits of 0.7 and -0.7 support data normalization, ensuring model stability and effective learning.

Note, the Tansig function, also known as the hyperbolic tangent sigmoid function, is a mathematical function commonly used in artificial neural networks as an activation function. It introduces nonlinearity to the output of neurons. The Tansig function transforms input values into a range between -1 and 1, allowing the network to model complex relationships in the data. Mathematically, the Tansig function is defined as:

$$f(x) = \frac{2}{1+e^{-2x}} - 1 \tag{1}$$

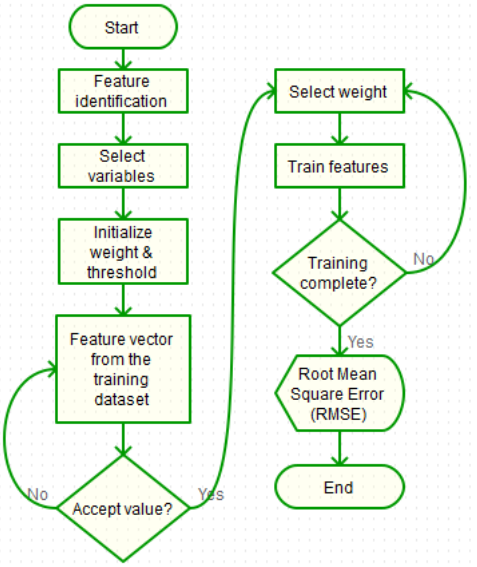


Fig. 6 *Flowchart Illustrating the Training Process of the Artificial Neural Network (ANN)*

In the context of neural networks, this function is often used to introduce nonlinearities in the network's computations and enable the modelling of more intricate patterns in the data. It's particularly useful for tasks where data relationships are not linear.

3.6 The Alert Management Module

From the system flow chart of figure 2, the alert management module is only triggered as a second check if the ML model verifies an uploaded certificate to be original. The server of the institution is at this time requested via an HTTP request to search and send back the data of the requested certificate under verification to the requesting server. This serves as a confirmatory test whether the certificate under verification has been issued by institution. This process is depicted in fig. 7.

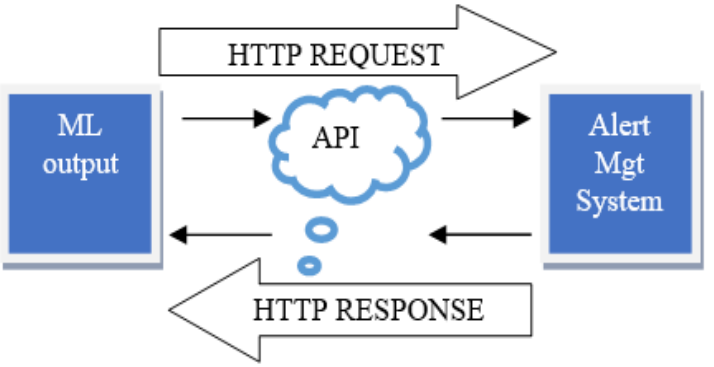


Fig. 7 *Data Flow from ML Model to Alert Management System*

The integration of the alert management system follows a dual combat technique proposed by (Erike, 2024; Erike et al., 2015). The system accepts the certificate number via the API and uses it to perform a database search. If a record is returned, the system returns the data via an HTTP response. An alert signal is sent to the institution's certificate fraud management office if no such record exists in the institution's repository.

4. Results and discussion

4.1 The ANN Model Performance

The implemented ANN model, utilizing the training dataset, is constructed using the neural network toolbox in Matlab with the parameters specified in table 1. The assessment of the algorithm's performance was conducted using the neural training tool, as detailed in the implementation section. This tool encompasses vital elements for evaluating the efficiency of the learning process and measuring performance, as illustrated in Fig. 8.

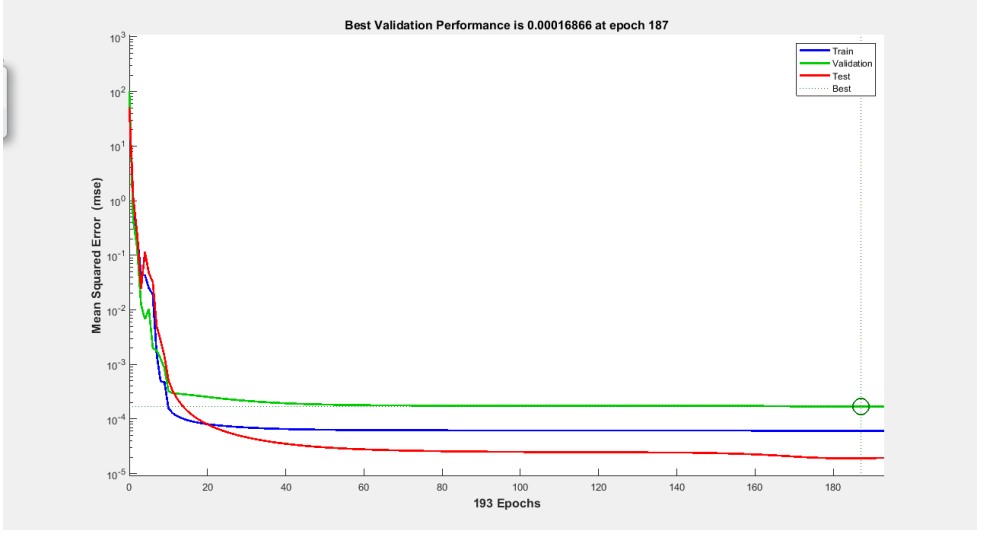


Fig. 8 *The Mean Square Error (MSE) Performance*

The purpose of this tool is to quantify the degree of error encountered during the system's training, with the ultimate objective of attaining a Mean Square Error (MSE) value that approaches zero. The findings unveiled that during the 187th epoch, the MSE performance attained a commendable value of 0.00100Mu, coupled with a significant validation value of 0.0016866. A total of 180 certificates were used for the testing shared in the ratio of 1:1 for genuine and counterfeit certificates respectively. Table 2 shows the confusion matrix obtained from feeding the model with know data labels over a period of time.

Table 2. *Confusion Matrix for Certificate Forgery Detection Using ANN Model*

	Predicted Genuine	Predicted Fake
Actual Genuine	TP = 72	FN = 3
Actual Fake	FP = 2	TN = 73

The Precion, Recall and the F1-Score is computed as follows:

$$Precision = \frac{TP}{(TP+FP)} = \frac{72}{(72+2)} = 0.9729$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{72}{(72+3)} = 0.96$$

$$F1\ Score = 2 \times \frac{Precision*Recall}{Precision+Recall} = 2 \times \frac{0.9729 \times 0.96}{0.9729+0.96} \approx 0.97$$

These outcomes show that there is a minimal training error, hence affirming the algorithm's effective acquisition of knowledge from the supplied training data.

Subsequently, the system's regression performance is presented in Fig. 9. Regression analysis is employed to assess the overall training performance. The objective was to attain a regression value around one. The outcome highlighted that the mean regression value, denoted as R = 0.99373, was used to evaluate the multiset's performance across training, testing, and validation. This result underscores the system's reliability and robust capability in distinguishing between genuine and counterfeit documents when deployed for verification.

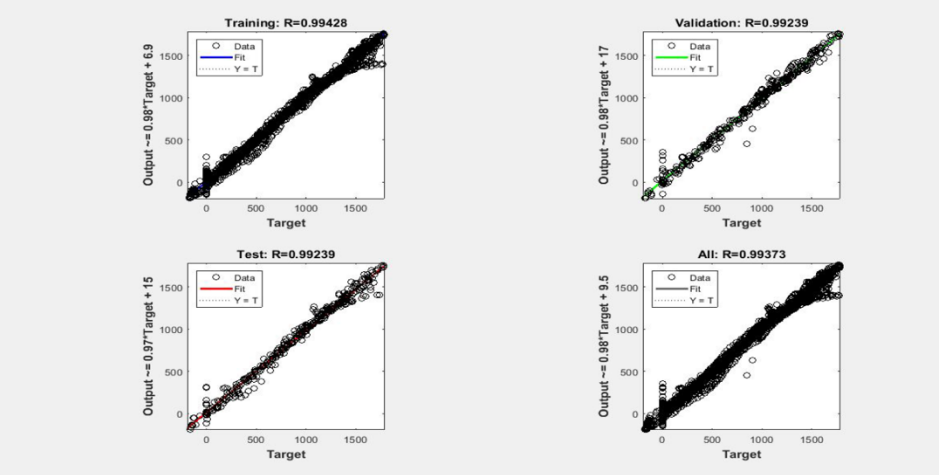


Fig. 9 *Regression Result*

The screenshots depicting the preprocessing step results, which encompass binarization, segmentation, and classification, have intentionally been excluded due to concerns regarding data confidentiality. This decision stems from the utilization of authentic student certificates in the research, necessitating precautions to prevent any breach of data privacy.

4.2 The Alert Management Model Performance

The system's web interface was designed using HTML and CSS, while the backend and API functionalities were constructed using PHP. The data retrieval aspect of the system functions seamlessly at a 100% success rate. Nonetheless, the system's responsiveness varies due to the strength of the data signal accessible during usage. The system underwent testing on an HP Spectre 360, with its specifications outlined in section three. The internet source was through the MTN Nigeria 4G Mifi. The time-response table along with the graph depicting the system's performance is illustrated in table 2 and fig. 10 respectively.

Table 2 *Time Response of the Alert Management System*

Test case	Response Time(s)
1	492
2	522
3	402
4	392
5	442
6	23445
7	399
8	433
9	1217
10	45416
11	399

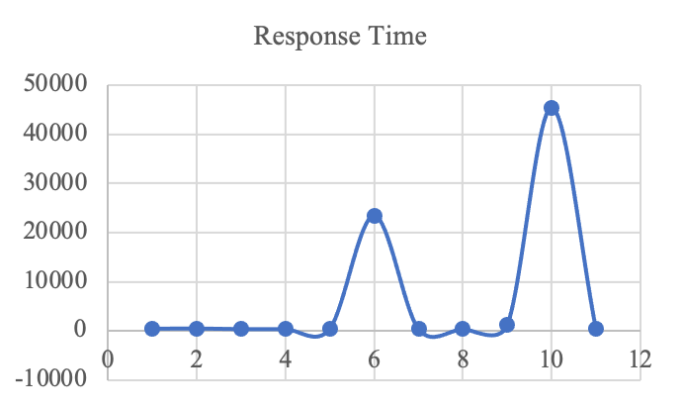


Fig. 10 *Time Response Graph of the Alert Management System*

The data in the table and graph indicate that responses from the Alert Management System were occasionally nearly instantaneous, while in other instances, there was a noticeable delay. This reiterates the key point that the system's response time is contingent on the strength of the available data signal at any given moment.

5. Conclusion.

In conclusion, the research presents a significant stride towards combating certificate forgery and ensuring credential authenticity in the digital landscape. The Intelligent Machine Learning-Powered Certificate Validation and Alert System (C-VAS) presents a holistic approach that integrates advanced technologies and methods to effectively identify fraudulent certificates. The employed artificial neural network demonstrates commendable performance in feature recognition and classification, signifying its suitability for real-world deployment. Additionally, the inclusion of an alert management module acts as a secondary safeguard, enhancing verification accuracy. The research highlights the importance of data signal strength in system responsiveness, underscoring the need for robust network infrastructure.

Following these research findings, several recommendations are put forward. Firstly, the research recommends that institutions should be encouraged to adopt the C-VAS, thereby making a more robust certificate verification process. There is another need for a strong collaboration between certificate awarding institutions and the C-VAS to streamline verification efforts and ensure real-time fraud detection. Secondly, further research is encouraged to refine the system's performance, potentially exploring hybrid models or incorporating other advanced AI techniques which may include the original owner of the certificate in the verification process. Lastly, government and other stakeholders should prioritize network infrastructure development to ensure consistent and reliable system responsiveness.

The robustness of the alert management module can be enhanced to mitigate the limitations that is clearly caused by network variability. Future work could explore the implementation of a caching mechanism or a fail-safe queuing system—such as a message buffer using Redis—that temporarily stores unsent alerts in a FIFO order. To prevent network congestion, the system should incorporate periodic retry logic with exponential back-off for reliable and efficient alert delivery.

In summary, the Intelligent Machine Learning-Powered Certificate Validation and Alert System proves to be a transformative solution in the fight against certificate forgery. Its integration into existing verification processes can lead to enhanced accuracy, reduced fraud incidents, and the preservation of the integrity of awarded certificates and the consequent awarding institutions.

References

Bell, J. (2022). What is machine learning?. *Machine Learning and the City: Applications in Architecture and Urban Design*, 207-216. https://doi.org/10.1002/9781119815075.ch18

Boonkrong, S. (2024). Design of an academic document forgery detection system. *International Journal of Information Technology (Singapore)*, Advance online publication. https://doi.org/10.1007/s41870-024-02006-6

Chen, C., Diao, W., Zeng, Y., Guo, S., & Hu, C. (2018). DRL-GenCert: Deep learning-based automated testing of certificate verification in SSL/TLS implementations. In *2018 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 48–58). https://doi.org/10.1109/ICSME.2018.00014

Dinesh Kumar, K., Senthil, P., & Manoj Kumar, D. S. (2020). Educational certificate verification system using blockchain. *International Journal of Scientific and Technology Research, 9*(3), 82–85.

Erike, A. I. (2024). Dual combat technique-based cyber systems protection against password attacks. *Nigerian Journal of Technology (NIJOTECH), 43*(4), 706–715. https://doi.org/10.4314/njt.v43i4.11

Erike, A. I., Inyiama, H. C., & Nwalozie, G. C. (2015). Securing enterprise information using dual combat technique. *International Journal of Computer Science and Telecommunications, 6*(8), 12–18. http://www.ijcst.org

España-Boquera, S., Castro-Bleda, M. J., Gorbe-Moya, J., & Zamora-Martinez, F. (2011). Improving offline handwritten text recognition with hybrid HMM/ANN models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(4), 767–779. https://doi.org/10.1109/TPAMI.2010.141

Fernández-Blanco, G., Froiz-Míguez, I., Fraga-Lamas, P., & Fernández-Caramés, T. M. (2024). A blockchain-based system for preventing academic forgery: Design and practical evaluation for CPU-based and low-power computers. In *2024 6th International Conference on Blockchain Computing and Applications (BCCA)* (pp. 406–413). https://doi.org/10.1109/BCCA62388.2024.10844456

Ke, Y., & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2004). https://10.1109/CVPR.2004.1315206

Khandpur, R. P., Ji, T., Jan, S., Wang, G., Lu, C. T., & Ramakrishnan, N. (2017). Crowdsourcing cybersecurity: Cyber attack detection using social media. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM)* (pp. 1049–1057). https://doi.org/10.1145/3132847.3132866

Lanham, D., Wood, D., Bartal, B., & Evans, R. (2006). *Criminal laws in Australia*. Federation Press.

Lee, H., Verma, B., Li, M., & Rahman, A. (2012). Machine learning techniques in handwriting recognition: Problems and solutions. In *Machine Learning Algorithms for Problem Solving in Computational Applications: Intelligent Techniques* (pp. 12–29). IGI Global.

Neves, R. F. P., Lopes Filho, A. N. G., Mello, C. A. B., & Zanchettin, C. (2011). A SVM-based off-line handwritten digit recognizer. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 510–515). https://doi.org/10.1109/ICSMC.2011.6083734

Nguyen, D.-H., Nguyen-Duc, D.-N., Huynh-Toung, N., & Pham, H.-A. (2018). CVSS: A blockchainized certificate verifying support system. In *Proceedings of the 2nd International Conference on Blockchain Technology and Applications* (pp. 436–442). https://doi.org/10.1145/3287921.3287968

Noor, Z. Z. (2024). Proof of the legal power of electronic certificates against criminal acts of forgery. *Journal of Law, Politics and Humanities (JLPH), 4*(6), 2571–2575.

Nwanze, D. E., Okechukwu, O. P., & Nnaji, C. H. (2023). Document verification system for fraud detection using machine learning technique. *International Journal of Computing, Science and New Technologies (IJCSNT), 1*(1), 1–9.

Oblikwu, P., & Dekera, K. (2019). A generic certificate verification system for Nigerian universities. *International Journal of Computer Science and Mobile Computing, 8*(11), 137–148.

Romero, V., Serrano, N., Toselli, A. H., Sánchez, J. A., & Vidal, E. (2011). Handwritten text recognition for historical documents. In *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage* (pp. 90–96).

Saha, S., Basu, S., Nasipuri, M., & Basu, D. K. (2010). A Hough transform-based technique for text segmentation. *arXiv preprint arXiv:1001.0021.*

Sánchez, J. A., Bosch, V., Romero, V., Depuydt, K., & De Does, J. (2014). Handwritten text recognition for historical documents in the Transcriptorium project. In *Proceedings of the 1st International Conference on Digital Access to Textual Cultural Heritage* (pp. 111–117). https://10.1145/2595188.2595193

Subramanian, K., Natarajan, P., Decerbo, M., & Castañòn, D. (2007). Character-stroke detection for text-localization and extraction. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 33–37). https://doi.org/10.1109/ICDAR.2007.4378671

Tamrat, W. (2022). Academic credential fraud: In search of lasting solutions. In *Higher Education in Ethiopia* (pp. 125–127). Brill. https://10.1163/9789004513488_033

Vinayasree, P., Narsimhulu, P., Sriganesh, P., & Ganesh, P. P. S. (2024). Online blockchain-based certificate generation and validation. *International Journal of Engineering Innovations and Management Strategies, 1*, 1–7.

Ye, Q., & Doermann, D. (2015). Text detection and recognition in imagery: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(7), 1480–1500. https://doi.org/10.1109/TPAMI.2014.2366765

Yousefi, J. (2011). Image binarization using Otsu thresholding algorithm. *University of Guelph, Ontario, Canada.* https://doi.org/10.13140/RG.2.1.4758.9284

Azubuike I. Erike
Federal University of Technology, Owerri. Nigeria
E-mail address: azubuike.erike@futo.edu.ng
Major area(s): Software Engineering

Yusuf Mshelia
Data Aid, Abuja. Nigeria
E-mail address: yusuf.mshelia@dataaid.org
Major area(s): Data Analysis

Anthony N. Isizoh
Nnamdi Azikiwe University, Awka. Nigeria
E-mail address: an.isizoh@unizik.edu.ng
Major area(s): Electronic and Computer Engineering

Charles O. Ikerionwu
Federal University of Technology, Owerri. Nigeria
E-mail address: charles.ikerionwu@futo.edu.ng
Major area(s): Software Engineering

Ikenna C. Nwandu
Federal University of Technology, Owerri. Nigeria
E-mail address: ikenna.nwandu@futo.edu.ng
Major area(s): Software Engineering

Florence O. Elei
Federal University of Technology, Owerri. Nigeria
E-mail address: florence.elei@futo.edu.ng
Major area(s): Software Engineering

(Received March 2024; accepted June 2025)