

MACHINE LEARNING FRAMEWORK FOR EXTRACTION AND RECOGNITION OF NIGERIAN BANKNOTE SERIAL NUMBERS

¹Olatunde T., ²Olabode O., ³Boyinbode O. K., ⁴Akinwonmi A. E., and ⁵Oladoja I. P.



DOI10.51459/jostir.2026.2.2.0262

¹Department of Computing and Information Science, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Ekiti State, Nigeria.

^{2,3,4,5}Department of Computer Science, Federal University of Technology, Akure, Ondo State.

Correspondence

olatunde.taiwo@bouesti.edu.ng
oolabode@futa.edu.ng

History

Received: 12-03-2026

Accepted: 09-06-2026

Published: August, 2026

ABSTRACT

Traceability of physical banknotes is an important issue for the security of financial transactions, fraud detection, forensic auditing and monitoring of the currency. The Serial number on banknotes is unique which makes it difficult to use automatic recognition and extraction techniques because of the complexity of interference from the background, typographic variations, image noise, and degradation of banknotes. As such, this study proposes a machine learning based system for serial number extraction and recognition on Nigerian Naira banknotes. The proposed framework includes adaptive image preprocessing, connected component-based region localization, structured character segmentation, five handcrafted feature extraction and deep learning-based classification. While most of the work reported in the literature works with recognition of currency denomination or generic optical character recognition (OCR), the present study conducts comparison of classical and deep learning machine learning methods for serial number recognition of Nigerian banknotes. A set of 2199 segmented alphanumeric characters collected from eight banknote denominations of Nigeria was used for experimental evaluation. Five feature representation techniques were explored and evaluated with the use of Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) classifiers. Based on the obtained result, the ANN with serialized pixel representation has achieved an accuracy of 99.0% for classification and CNN without feature engineering has achieved the same level of accuracy. Compared to recent works on banknote serial number recognition, accuracies in the range of 94%-98% were reported, whereas the proposed work shows competitive performance and shows its robustness. The results confirm the impact of the deep learning approach in the field of financial document analysis and provide a solid foundation for the development of real-time Nigerian banknote serial number recognition and currency monitoring systems.

Keywords: Serial Number Recognition; Currency Authentication; Deep Learning; Feature Engineering; Nigerian Naira; Computer Vision.



<https://www.futa.edu.ng>



<https://jostir.futa.edu.ng>

1. | Introduction

Fewer and fewer financial systems are currently being digitized, but that hasn't stopped physical currency from being a dominant component in many developing countries' transactional ecosystems. With increased cash flow, the risk of counterfeiting increases,

along with the limitations of tracking stolen cash and forensic auditing difficulties (World Bank, 2024). The serial number is a security feature that is present in paper currency but not being utilized and is in fact crucial. Such alphanumeric marks should be unique and allow tracking of every banknote in financial transactions. Automatic

Serial Number Recognition (SNR) is a specific subset of Optical Character Recognition (OCR) and it comes with unique computational challenges. Banknotes include complex graphic backgrounds, watermark overlays, microtext patterns and security inks, which hinder segmentation processes. Furthermore, variations in scan conditions between polymer denominations and paper denominations are added to the variations due to the denomination differences of the substrates. Recent work has made significant strides in character identification under noisy imaging conditions, particularly due to the development of deep learning techniques, such as convolutional neural networks (CNN) (Liu *et al.*, 2024; Zhang *et al.*, 2023; Li *et al.*, 2021). The literature currently available, however, is mostly focused on the classification of denomination or the general document OCR (Smith, *et al.* 2023), and does not pay much attention to the extraction of serial numbers in African currencies. The denomination of Naira notes contains different information about the serial, so the information is placed and formatted according to the denominations which necessitates specific processing procedures. To fill this gap, this study aimed to create and test a specific machine learning pipeline for the recognition of Nigerian banknote serial numbers. The study compares classical feature engineering methods to ensemble learning models, instead of just relying on end-to-end deep learning. Recent research from 2025 on financial document intelligence highlights the importance of hybrid models that integrate geometric features with deep feature extraction (Rahman *et al.*, 2025). The investigation aims to find out whether it is possible to outperform CNN architectures in terms of performance, in this particular application scenario, using a handcrafted approach, designed to give careful consideration to the features (Chen, X. *et al.* 2022).

1.1 | Research Gap and Motivation

Many researchers have investigated currency recognition and financial document processing using OCR, but most of the current systems are mainly for the classification of the denominations, extraction of generic text or verification of the currency. The extraction and recognition of serial numbers on Nigerian banknotes have received little attention, especially in the presence of varying lighting conditions, image noise, partial occlusion and poor quality bank notes. In addition, existing OCR systems are often very sensitive to hand-designed feature engineering methods that lack the robustness in real-world environment. While it is still limited, recent advances in deep learning have demonstrated impressive performance in character recognition in financial documents; comparison studies between deep neural architectures and classical machine learning models in the African financial document context are still scarce. Furthermore, very little research has been done on the effectiveness of the merging of zoning, geometric, projection-based and serialized pixel features for the recognition of Nigerian bank note serial numbers. There has also been a lack of analysis of model generalization, class imbalance treatment and misclassification behaviour in similar alphanumeric characters in the existing approaches. To address these limitations (Yang *et al.*, 2021), the current work proposes a hybrid feature representation system involving both handcrafted and deep learning to extract and recognize serial numbers of Nigerian banknotes. In order to choose the most reliable method for real-world implementation, the framework compares different models vis-à-vis Naïve Bayes, Decision Tree, XGBoost, Artificial Neural Network (ANN), and Convolutional Neural Network (CNN).

1.2 | Research Contributions

Current OCR research uses generic text extraction or denomination recognition, however this study

develops a dedicated machine learning framework that is used for serial number recognition on the Nigerian banknotes and also evaluates the handcrafted feature descriptors against the deep learning architectures under challenging imaging conditions. The summary of the contributions of this study are as follows: development of a machine learning-based framework for the extraction and recognition of Nigerian banknote serial numbers;

- i. creation of a structured dataset consisting of segmented alphanumeric characters extracted from Nigerian currency notes across multiple denominations;
- ii. comparative evaluation of handcrafted feature extraction methods, including horizontal profiles, vertical profiles, combined projection profiles, zoning plus geometric descriptors, and serialized pixel representations;
- iii. comparative performance analysis of classical machine learning models and deep learning architectures, including Naïve Bayes, Decision Tree, XGBoost, ANN, and CNN;
- iv. investigation of model behaviour under class imbalance and visually ambiguous character conditions such as B/8, S/5, and O/0; and
- v. demonstration that serialized pixel representations combined with ANN and CNN architectures provide superior recognition accuracy and generalization performance for Nigerian banknote OCR applications.

2. | Related Work

Automatic Serial Number Recognition (ASNR) is a specific type of Optical Character Recognition (OCR) that is employed in banking, financial security systems, document intelligence and forensics. The initial SNR and OCR systems

used template matching and handcrafted feature extraction methods. Even though these methods were developed to be computation efficient, they were very sensitive to noise, distortion, rotation, illumination changes and font variations (He *et al.*, 2016). With the recent development of deep learning and machine learning, many studies have been carried out to improve the recognition accuracy. Convolutional Neural Networks (CNNs) were shown to be successful in text recognition in noisy environments by Liu *et al.* (2024), whereas Zhang *et al.* (2023) reported that hybrid preprocessing methods improve OCR accuracy. Similarly, Khan *et al.* (2023) showed that spatial zoning and geometric features enhance classification robustness, whereas Rahman *et al.* (2025) proposed hybrid models that combine handcrafted and deep features for financial document analysis. The majority of the previous works, however, are related to structured documents and major currency denominations like US Dollar and Euro, and have paid little attention to the recognition of Nigerian banknote serial number.

2.1 | Classical OCR Methods

Due to applications in document processing, finance systems, industrial inspection, automated identification and intelligent transportation, OCR is an important research topic in pattern recognition and computer vision (Subramani *et al.*, 2020). Traditional OCR systems were mainly based on hand-crafted feature extraction and classical machine learning models before the advent of deep learning. Template matching is one of the earliest OCR techniques developed. It compares the input character with a set of character template that are predefined using similarity measures such as Euclidean distance or correlation coefficient. This is accomplished by picking out the template that matches the best. Template matching is simple and easy to compute, but it is not robust to noisy images, variations in scale, skew, rotation or font. However, recent studies have revealed that template-

based systems lack generalization in the degraded document environment (Koponen *et al.*, 2022; Monteiro *et al.*, 2023). Another classical way of OCR is Projection histogram techniques. These methods employ a horizontal and vertical distribution of the pixels for the creation of structural character representations. They are very efficient and have many applications in serial number recognition system. Holila *et al.* (2024) showed their usefulness in structured document recognition task. Using these methods, however, can be tricky to read characters like B, 8, O, 0, or S, 5, especially if the image is poor quality. Edge descriptor methods have also been extensively used due to their ability to describe the contour information, the orientation of the strokes, and the boundary structure. These methods use Sobel, Canny, Prewitt and Laplacian filters to obtain the discriminative features for classification. The findings of Chousangsuntorn *et al.* (2021) indicated that the representations based on the edge can help in the recognition process of a contaminated industrial serial number system with the significant preservation of important stroke information. However, they are still prone to blurring, contour distortion, lighting changes and segmentation mistakes. Another major classical OCR technique is called Connected Component Analysis (CCA). It characterizes clusters of contiguous foreground pixels that may form character regions in binary images. CCA has been successfully applied in structured documents where the separation of characters is well-defined and remains in use during pre-processing and segmentation steps due to its simplicity and interpretability (Monteiro *et al.*, 2023). However, it is not very effective when characters overlap, when characters are close to adjacent symbols, or when characters are set in a complex background, such as a currency note. Classical OCR methods are applicable for embedded and resource limited applications but require a lot of manual feature engineering and are usually not robust against illumination variation, note degradation, motion blur, skew, and noise in real-life

scenarios. These restrictions have spurred the shift to machine learning and deep learning methods that can automatically extract discriminative features from raw image data. In recent years, modern OCR systems have been incorporating increasing use of CNNs, attention mechanisms, and deep feature learning techniques, such as noted by Koponen *et al.* (2022). Similarly, Jang *et al.* (2020) have shown that CNN-based serial number recognition systems perform better than handcrafted feature methods in the low quality banknote recognition task. Despite this progress, there is a lack of research that specifically targets the Nigerian banknote serial numbers and therefore, comparative studies of both classical and modern techniques for serialization are needed.

2.2 | Deep Learning-Based OCR

Deep learning has brought about a revolution in OCR by introducing automatic hierarchical feature learning. Deep learning models learn relevant representations directly from image data, which are more powerful than traditional models that demand manual feature design and are more resistant to blur, deformation, changes in lighting and complex backgrounds. Some of the most popular architectures used for OCR are Convolutional Neural Network (CNNs). CNNs, through their convolution operations, automatically capture what the patterns, contours, textures and spatial relationships are along with the translational invariance. Jang *et al.* (2020) showed that CNNs significantly outperform the other methods in recognizing the serial number of low-quality banknote images and Baek *et al.* (2021) reported that the CNN outperforms the other methods in complex document environments due to its ability to capture fine-grained visual structures. Convolutional Recurrent Neural Network (CRNNs) are CNNs with an additional recurrent layers being added into the recognition pipeline. CNNs are used for feature extraction and recurrent layers are used to model the sequential dependencies

between characters. This combination allows for recognition of variable-length text without any prior segmentation of characters. Shi *et al.* (2021) demonstrated that CRNNs could successfully extract the spatial and contextual information, and thus are widely applicable to serial number recognition and OCR tasks in documents. Because they successfully model long-range sequential dependencies and solve the vanishing gradient issue with conventional recurrent networks, Long Short-Term Memory (LSTM) networks have also been extensively used in OCR. The CNN feature extractors and LSTMs are commonly used in OCR systems. In the field of damaged document recognition, Li *et al.* (2022) demonstrated that a CNN-LSTM architecture can obtain a much higher recognition accuracy by using the spatial feature extraction and sequence modelling methods. They are useful for tasks like serial numbers because they can keep track of the context. In recent years, OCR systems based on transformer technology have become state-of-the-art. Self-attention architectures are able to model global relationships over an entire image. Compared to CNNs and LSTMs, Transformers can attend to multiple image regions at the same time, which boosts the ability to recognize irregular and complex text patterns. The Transformer-based OCR systems have been found to be superior to the traditional CNN and recurrent systems in the document recognition task where the document images are distorted, skewed, or captured at low resolution (Wang *et al.*, 2023). Fang *et al.*, (2024) have demonstrated that the Transformer systems are robust for these scenarios.

However, the deep learning models have their drawbacks: they need large annotated datasets, a lot of computational resources and require longer training time. They can also provide less interpretability than traditional. However, their powerful feature learning and generalization have made deep learning the most prevalent paradigm in the current research of OCR. Such architectures offer promising potential for Nigerian banknote serial number recognition,

as they can be trained to automatically learn the visual patterns of banknote characters and complex stroke structures directly from the banknote images. In this study, the effectiveness of Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) over the traditional machine learning methods is evaluated to find out which one is the best to be used for recognition.

2.3 | Financial Document Recognition

One of the primary use cases of OCR and machine learning is the recognition of financial documents, which is vital for financial institutions, payment systems, and commerce to automate financial transactions and operations. The systems automatically extract, confirm and analyse information from banknotes, cheque, invoices, receipts, and transaction records. Bank note authentication and recognition is one of its important applications. Old systems relied on manually engineered features like colour histograms, texture features and geometric patterns. Recent studies, however, have used Deep learning techniques to enhance performance. Ahmad *et al.* (2022) showed that CNN-based banknote authentication systems are able to detect counterfeits with a higher accuracy due to their ability of learning intricate visual features straight from the banknote image. Likewise, Albahli *et al.* (2023) found that the accuracy of currency classification in various denominations and imaging conditions was enhanced. Another important app is Cheque OCR. Scalped cheque processing software reads the cheque scanning documents to extract the account number, amount, signature and routing information. CNN-LSTM architectures proved to be crucial in handwritten cheque recognition by learning both spatial and sequential features, which is a vital part in modern banking operations as demonstrated by Sharma and Gupta (2021). Invoices are a critical asset today because of the digitization of financial processes, which has made invoice OCR important. These systems retrieve invoice

numbers, vendor data, transaction quantities, tax information and payment data. In order to solve the above problems, Xu *et al.* (2021) proposed LayoutLVv2 to fuse three modal data (visual, textual, and layout) to facilitate intelligent document understanding and enhance the performance of invoice data extraction. Another research showed that for financial document OCR, Transformer-based methods work better than regular OCR pipelines (Liu *et al.*, 2023). Financial document recognition is a specialized type of financial document recognition that primarily targets the identification of a unique alphanumeric identifier from financial instruments and currency notes. These identifiers are used for banknote authentication, fraud detection, inventory management, forensic investigations and currency tracking. In degraded bank note images, CNN-based serial number recognition (SNR) frameworks have been shown to be more effective than conventional OCR approaches by Jang *et al.* (2020). Similarly, Chousangstorn *et al.* (2021) reported high classification accuracy in contaminated and partially degraded environment by deep learning approaches. Although significant work has been done in OCR or financial document intelligence, limited research has focused on the recognition of

Nigerian banknotes' serial number. The existing studies are mainly dedicated to either denomination classification or counterfeit detection, generic OCR application, or foreign currency systems. Moreover, there is hardly any research comparing the handcrafted feature extraction techniques with deep learning architectures in the African Financial document classes. This gap emphasizes the need for developing a localized machine learning framework that can extract and recognize the serial numbers of Nigerian banknotes with varying image quality.

3 | Materials and Methods

The research methodology was structured into sequential stages as follows: dataset creation, image preprocessing, region-of-interest (ROI) detection, Character Segmentation, Feature Extraction, Model training, character recognition, and evaluation. The architecture is represented in Figure 1.

3.1 | Data Acquisition and Preparation

Eight (8) denominations of Nigerian Naira (₦5- ₦1000) were digitized using a flatbed scanning

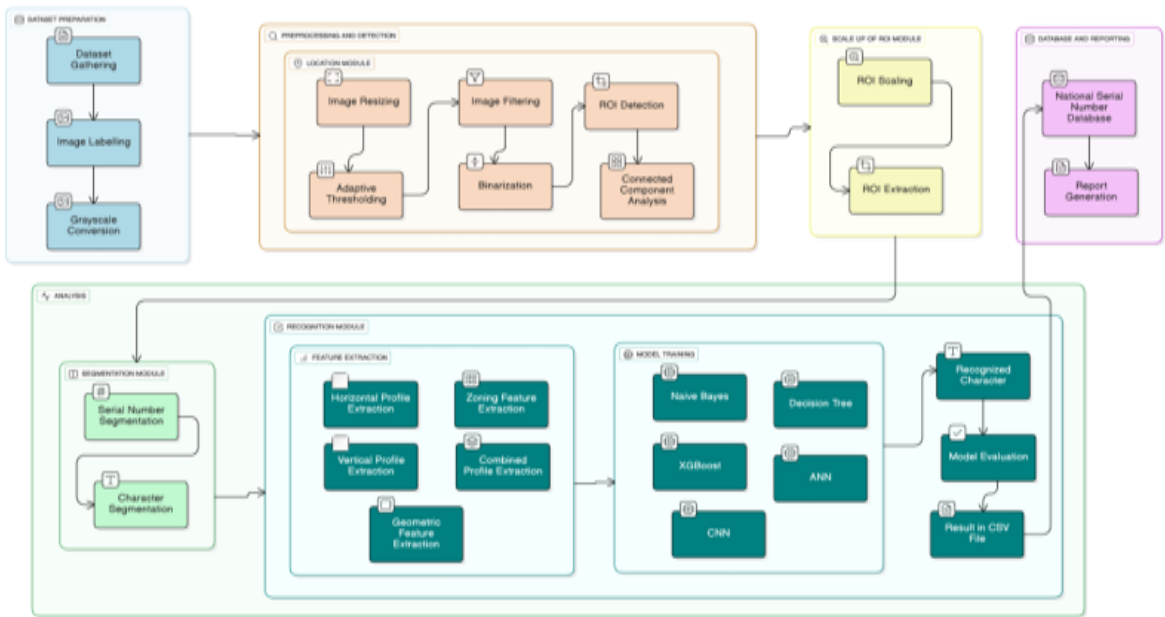


Figure 1 | System Architecture for Extraction and Recognition of Nigerian Banknote Serial Number



Figure 2 | Samples of scanned currency notes

device under consistent illumination settings. Serial number regions were manually annotated to ensure precise ground-truth labelling. Character-level segmentation produced 2,199 isolated alphanumeric samples. To maintain computational efficiency and uniformity, each segmented character was resized to a 10x20 binary image representation.

3.2 | Preprocessing

Colour images were converted to grayscale using a weighted luminance transformation. Adaptive thresholding was subsequently applied to mitigate illumination heterogeneity. A localized mean filter (10x10 kernel) was selected after empirical testing

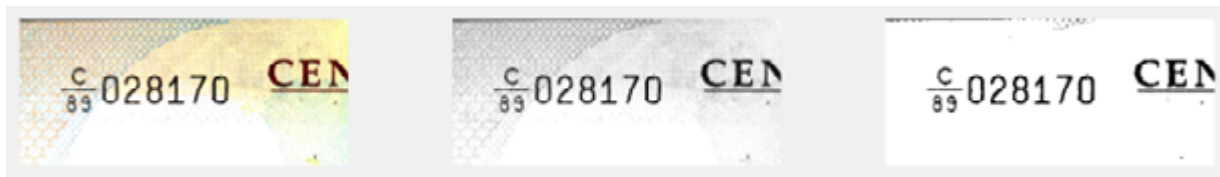


Figure 3 | Gray Scale Image

to preserve character edges while suppressing background artifacts.

3.3 | Feature Representation

Five different representations of features were explored:

6. This is useful in identifying characters that have horizontal strokes e.g. distinguishing between 'E' and 'F', where the bottom stroke is present in 'E' but not 'F'.

7. Column-wise projection density (vertical profiles): Vertical profile will measure the intensity of a pixel across the columns of an image, instead. This can be useful for determining characters that

stand out from the rest in the vertical axis like 'L' vs 'T' and '1' vs '7'.

8. Combined projection vectors: These combine both views into a single feature vector. This hybrid method provides a more comprehensive picture of the character's form in terms of both its vertical and horizontal distributions. When characters have similar silhouettes in one direction and varying silhouettes in the other direction, it can be quite useful, e.g. 'Z' and '2'; they look alike vertically but different horizontally. The model is more enriched with geometry understanding by merging both profiles. The two features extractions in combination led to a set of 30 independent variables and 1 class target.

9. **Zoning plus geometric features:** This is an advanced and spatially-aware approach is using zonal density descriptors with geometric shape metrics. The image is subdivided into smaller areas (in this example 4x4 blocks) and the density of the image is computed on each area. This will help to capture local patterns and stroke placement, which are important to differentiate characters that are close in meaning. Contour features are based on the shape of the character itself and complement zoning. Geometric descriptors such as area, perimeter, aspect ratio etc. can be derived from the image using contour detection techniques. The combination of zoning and geometric features offers a combination of both texture and shape analysis, and is therefore very effective in pattern recognition of complex or stylized characters. The zoning plus contour feature extraction model yielded a set of 20 independent features and a target class.

10. **Direct Pixel serialization (flattened intensity vector):** This is the most straightforward method of feature extraction. It is a technique of converting a two-dimensional image into one dimensional array by reading the values of the pixels in a specific order (generally row-wise or column-wise). For instance, a picture of size 10 x 20 is now a 200-element vector. This approach maintains the integrity of the pixel intensity and spatial arrangement while being vulnerable to noise, distortions, and misalignments. Serialization is easy and quick however, lacks abstraction that is required for good recognition in a wide range of fonts. With the Serialization feature extraction model, a total of 200 features were obtained, and the target class was added to make 201 features. This was done by flattening the picture data set of 10 x 20 character-images. The final result is a csv file with 2,199 entries, without any missing values. The independent variables are binary (0 or 1) and the dependent variable is 34 valued (representing each digit and each alphabet except for O and I); that is, each one has not been observed in the data set.

3.4 | Classification Models

Four (4) supervised classifiers were implemented to evaluate feature separability:

1. **Naïve Bayes (probabilistic baseline):** It uses Bayes theorem for finding the probability of the event. Building statistical models of classes from training datasets is how the probabilistic classifier works. A class is represented by the class c , while characteristics determined separately are denoted by x , as indicated in Equation (1)

$$P(c|x) = \frac{p(x|c)p(c)}{p(x)} \quad (1)$$

Where:

$P(c|x)$ = the posterior probability of class c , given x ,
 $p(c)$ = the prior probability of the class,
 $p(x|c)$ = the likelihood which is the probability of predictor, given class and
 $p(x)$ = the prior probability of predictor.

2. **Decision Tree (non-linear rule-based model):** The feature space is recursively partitioned into disjoint regions to make predictions. Its effectiveness in feature extraction and pattern recognition stems from algorithms that use specific mathematical criteria to select the most informative features for splitting the data at each node. A decision tree is modelled as a function $y(x)$ that maps an input vector x to a predicted output y as represented in Equation (2)

$$y(x) = \sum_{m=1}^T c_m I(x \in R_m) \quad (2)$$

Where:

T = number of terminal nodes (leaves) in the tree
 R_m = the m -th region (leaf) of the feature space
 c_m = the prediction value assigned to region R_m .
 $I(x \in R_m)$ = indicator function that equals 1 if the input x belongs to the region R_m , and 0 otherwise.

3. **XGBoost (gradient boosting ensemble):** XGBoost is an ensemble learning technique that combines multiple weak decision trees in a boosting framework to improve classification accuracy for serial number recognition. The model is represented by Equation (3) to (7).

Prediction Function:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (3)$$

Where:

$\hat{y}_i$ = Predicted output for the i^{th} training sample.

x_i = Feature vector of the i^{th} sample.

K = Total number of decision trees in the ensemble.

f_k = The K^{th} regression tree.

F = Space of all possible regression trees

$\sum_{k=1}^K$ = Summation of predictions from all trees.

Equation (3) indicates the final prediction obtained by summing the outputs of multiple decision trees. Each tree contributes a prediction value, and the ensemble combines these contributions to produce a more accurate and robust prediction.

Objective Function

The optimized objective function by XGBoost is defined as:

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

Where:

L = Overall objective (loss) function.

N = Total number of training samples.

$l(y_i, \hat{y}_i)$ = Loss function measuring the difference between the actual and predicted outputs.

y_i = Actual (ground truth) value of the i^{th} sample.

$\hat{y}_i$ = Predicted value of the i^{th} sample.

$\Omega(f_k)$ = Regularization term associated with the k^{th} tree.

K = Total number of trees.

The objective function consists of two components:

i. Training Loss

$$\sum_{i=1}^N l(y_i, \hat{y}_i) \quad (5)$$

which evaluates how well model fits the training data.

ii. Regularization Term

$$\sum_{k=1}^K \Omega(f_k)$$

Which controls model complexity and prevents overfitting. The regularization component of XGBoost is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (6)$$

Where:

$\Omega(f)$ = Complexity penalty of a decision tree.

γ = Penalty coefficient for adding a leaf node

T = Number of leaf nodes in the tree

λ = L2 regularization parameter

w = Vector of leaf weights (scores assigned to terminal nodes)

$\|w\|^2$ = Squared Euclidean norm of the leaf-weight vector, given by:

$$\|w\|^2 = \sum_{j=1}^T w_j^2 \quad (7)$$

w_j = Weight associated with the j^{th} leaf node.

The regularization term penalizes excessively complex trees by:

i. Increasing the cost when additional leaf nodes are created (γT).

ii. Penalizing large leaf weights through L2 regularization ($\frac{1}{2} \lambda \|w\|^2$)

This mechanism helps XGBoost achieve strong performance generalization and reduces the occurrence of overfitting.

Furthermore, XGBoost uses an ensemble of decision trees in a stage-wise manner, where the final prediction is obtained by aggregating the outputs of multiple regression trees as shown in Equation (3). The learning process optimizes the objective function in Equation (4), which combines prediction loss and model complexity penalties. To improve generalization and prevent overfitting, XGBoost incorporates the regularization term shown in Equation (6), which penalizes both the number of leaf nodes and the magnitude of leaf weights. This regularized optimization framework enables the model to achieve high prediction accuracy while maintaining robustness and computational efficiency for Nigerian banknote serial number recognition.

4. **Artificial Neural Network (multi-layer perceptron):** ANN models nonlinear relationship between input features and serial number classes using interconnected layers of neurons. The network learns complex nonlinear relationships between the extracted feature vectors and the corresponding alphanumeric character classes through a supervised learning process.

Extracted features serve as input to a multilayer perceptron.

Input Layer

Let the extracted feature vector be represented as:

$$X = [x_1, x_2, x_3, \dots, x_n] \quad (8)$$

Where:

X = Input feature vector.

x_i = The i^{th} extracted feature.

n = Total number of input features.

Hidden Layer Computation

The weighted sum entering the hidden neuron j is computed as:

$$z_j = \sum_{i=1}^n w_{ij} x_i + b_j \quad (9)$$

Where:

z_j = Net input to hidden neuron j .

w_{ij} = Weight connecting input neuron i to hidden neuron j .

x_i = Input feature value.

b_j = Bias term associated with hidden neuron j .

n = Number of input neurons.

Hidden Layer Activation

The output of hidden neuron j is:

$$H_j = f(z_j) \quad (10)$$

Where:

h_j = Output of hidden neuron j .

$f(.)$ = Activation function.

In this study, the Rectified Linear Unit (ReLU) activation function is used:

$$f(z) = \max(0, z)$$

Where:

Max $(0, z)$ returns z if $z > 0$, otherwise 0.

Output Layer Computation

The weighted sum entering output neuron k is:

$$o_k = \sum_{j=1}^m v_{jk} h_j + c_k \quad (11)$$

Where:

o_k = Net input to output neuron k .

v_{jk} = Weight connecting hidden neuron j to output neuron k .

h_j = Output of hidden neuron j .

c_k = Bias associated with output neuron k .

m = Number of hidden neurons.

Softmax Output Function

For multi-class character recognition, the Softmax function is used:

$$\hat{y}_k = \frac{e^{0_k}}{\sum_{r=1}^C e^{0_r}} \quad (12)$$

Where:

$\hat{y}_k$ = Predicted probability of class k.

0_k = Output score of class k.

C = Total number of character classes.

e = Euler's exponential constant.

Classification Decision

The predicted character class is obtained as:

$$\hat{C} = \underset{k}{\operatorname{arg\,max}} (\hat{y}_k) \quad (13)$$

Where:

$\hat{C}$ = Predicted character class.

$\hat{y}_k$ = Predicted probability of class k.

arg max = Function that returns the class with the highest probability.

Loss Function

The network parameters are optimized using the categorical cross-entropy loss function:

$$L = - \sum_{k=1}^C y_k \log(\hat{y}_k) \quad (14)$$

Where:

L = Classification loss.

y_k = Actual class label.

$\hat{y}_k$ = Predicted probability.

C = Number of output classes

ANN Feature Representation

The complete feature vector supplied to the ANN can be represented as:

$$F = [P_H, P_V, F_{ZG}, F_{SP}] \quad (15)$$

Where:

F = Complete feature vector.

P_H = Horizontal Profile Features.

P_V = Vertical Profile Features.

F_{ZG} = Zoning and Geometry Features.

F_{SP} = Serialized Pixel Features.

An input layer, one or more hidden layers, and an output layer make up the Multi-Layer Perception (MLP) architecture of the artificial neural network used in this investigation. Using weighted connections and activation functions, the network learns nonlinear decision limits after receiving extracted character attributes as input. Backpropagation and categorical cross-entropy loss minimization are used to optimize the network parameters during training. By learning intricate feature interactions that conventional machine learning classifiers would miss, the ANN allows for efficient identification of visually alphanumeric letters found in Nigerian banknote serial numbers.

3.5 | Convolutional Neural Network Architecture and Training Configuration

A convolutional Neural Network (CNN) architecture was implemented in the final phase of the character classification process to perform automatic feature extraction and classification directly from raw character images without reliance on handcrafted descriptors. The CNN model was designed to capture localized spatial pattern, stroke structures, contour variations, and character-specific geometric representations within the banknote serial number dataset.

The dataset used was partitioned into training, validation, and testing subsets using a 60:20:20 ration to ensure consistency with previous handcrafted feature experiments and to maintain balanced evaluation conditions across all models.

The implemented CNN architecture consisted of the following layers:

- i. Input Layer: 10 x 20 grayscale character images
- ii. Convolutional Layer 1: 32 filters of size 3 x 3 with ReLU activation
- iii. Max Pooling Layer 1: 2 x 2 pooling operation
- iv. Convolutional Layer 2: 64 filters of size 3 x 3 with ReLU activation
- v. Max Pooling Layer 2: 2 x 2 pooling operation
- vi. Fully Connected Dense Layer: 128 neurons with ReLU activation
- vii. Dropout Layer: 0.5 dropout rate for overfitting mitigation
- viii. Output Layer: Softmax activation for multi-class classification
- ix. Optimizer: Adam

- x. Learning Rate: 0.001
- xi. Batch Size: 32
- xii. Epochs: 50
- xiii. Activation: ReLU
- xiv. Loss Function: Categorical cross entropy

The model adopted Adam optimizer for training with a learning rate of 0.001 and a categorical cross-entropy loss function. A batch size of 32 and a maximum of 50 epochs were employed during training. Early stopping was implemented to prevent overfitting by monitoring validation loss convergence.

The training data was augmented with some transformations such as rotation, brightness adjustment, scaling, injection of gaussian noise, and a small translation, to make the data more robust and generalizable under real world imaging conditions. The training validation accuracy curves presented in Figure 4 showed that the training and validation losses stayed close to each other with very little divergence, which suggested convergence without over fitting.

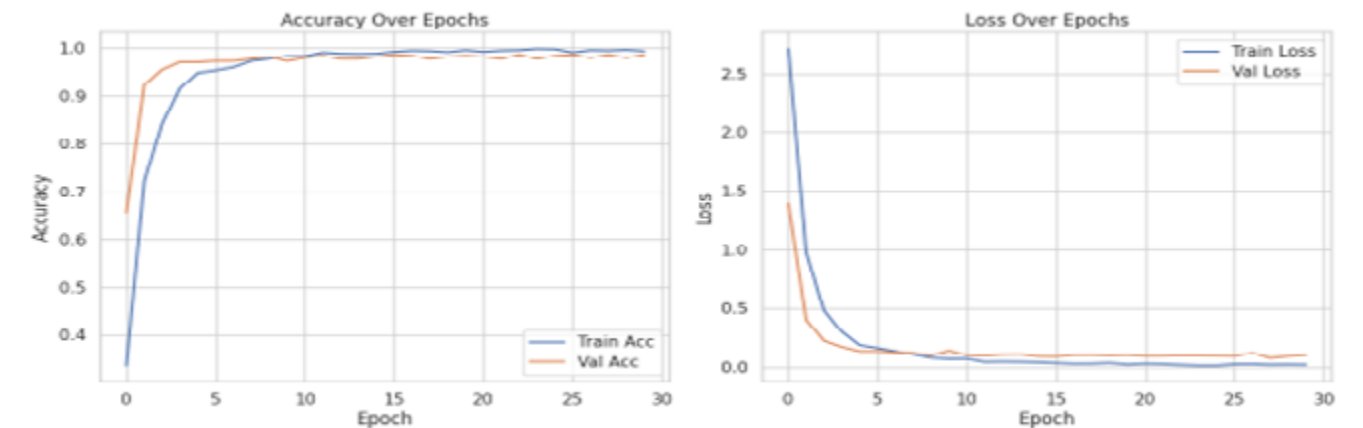


Figure 4 | Training Accuracy and Loss graph for the CNN model. Performance evaluation metrics using Accuracy, Precision, Recall, and F1-score to ensure balanced assessment was used.

3.6 | Model Training Enhancement

To improve reliability and generalization, the following enhancements were incorporated:

Cross-Validation and Overfitting Mitigation

To achieve statistical reliability and minimize overfitting bias, all machine learning models were

evaluated using 5-fold cross-validation. The reported performance metrics show the average results obtained across the validation folds in Table 1.

Table 1 | Results of Performance Metrics of Cross-Validation

Model	Accuracy (%)	Std. Dev
Naïve Bayes	82.4	±9
Decision Tree	88.2	±1.4
XGBoost	96.1	±0.8
ANN	99.0	±0.3
CNN	99.1	±0.2

Several regularizations, and generalization, strategies have been used in the model development. These encompassed dropout regularizations, data augmentation, monitoring the validation loss, and early stopping mechanisms. Results of these techniques were used to enhance the reliability of models under different image distortions, illumination and character shapes. CNN and ANN models had good generalization behaviour as shown in Figure 4 where the training and validation accuracy curves are close. Moreover, high accuracy across the folds of the cross validation was not due to memorization or overfitting of the model, given the low variation between folds. Although the deep learning models perform well, it is important to consider potential drop-off in performance in extreme real-world scenarios like extreme note

degradation, motion blur, folded notes, and low illumination in future deployments.

4. | Results and Discussion

The feature sets based on projections showed moderately good discriminative performance, with the highest accuracy of 82% to 84%. These descriptors were easy to compute but sensitive to slight distortion and thinness of the characters. The horizontally and vertically projected images demonstrated a significantly better performance (96%), indicating that the two projections would complement each other in spatial encoding.

Zoning features with geometric descriptors also achieved 96% accuracy using XGBoost classification.

However, serialized pixel representation yielded the strongest performance (99%) when paired with an ANN classifier.

The improvement indicates that preserving complete spatial configuration allows neural networks to internalize subtle pixel-level variations. Figure 6 shows the comparative analysis of classifiers based on different feature representation, where it is



Figure 5 | Sample data from the Horizontal profile-based features

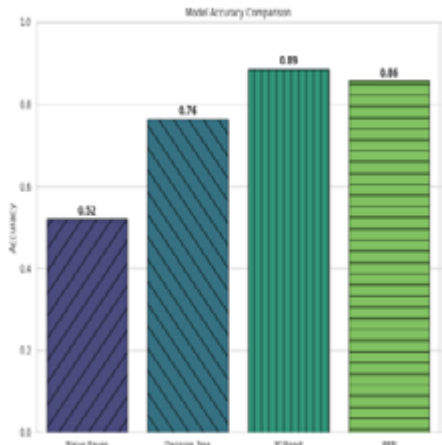


Figure 6 | Graph of Accuracies of Models with Horizontal Profile-Based Features

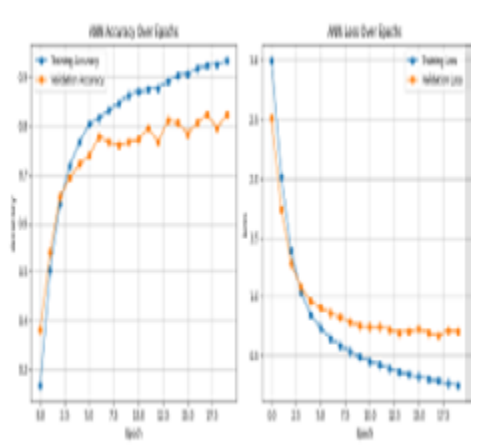


Figure 7 | Training Accuracy and Loss Graphs for Horizontal dataset



Figure 8 | Sample data from the vertical profile-based features

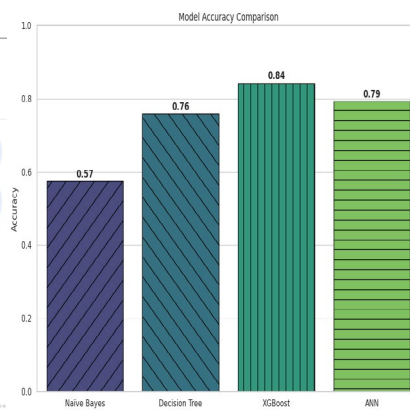


Figure 9 | Graph of Accuracies of Models with Vertical Profile-Based Features.

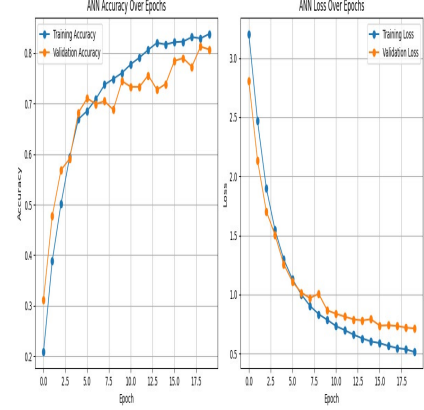


Figure 10 | Training Accuracy and Loss Graphs for Horizontal dataset.



Figure 11 | Sample data from the combined profile-based features.

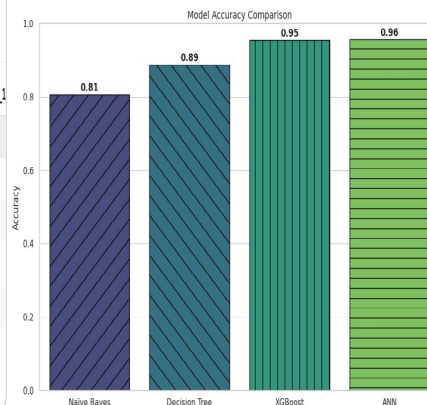


Figure 12 | Graph of Accuracies of Models with Combined Profile-Based Features.

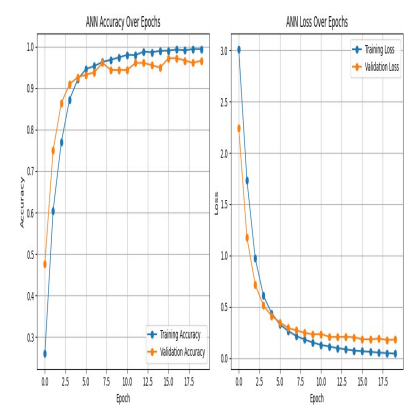


Figure 13 | Training Accuracy and Loss Graphs for Combined dataset.

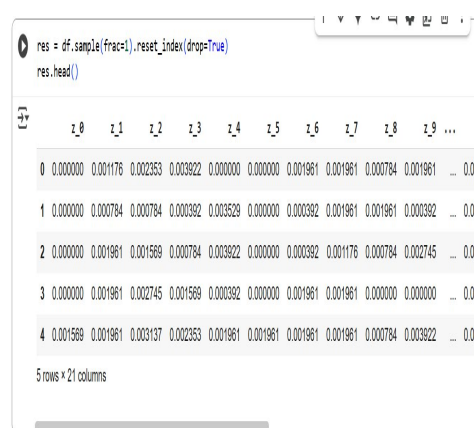


Figure 14 | Sample data from the zoning plus geometric profile-based features.

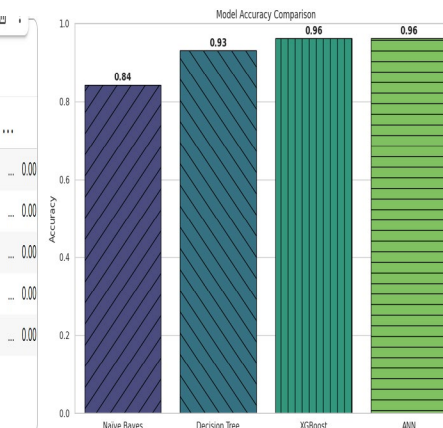


Figure 15 | Graph of Accuracies of Models with Zone + Geometric Features.

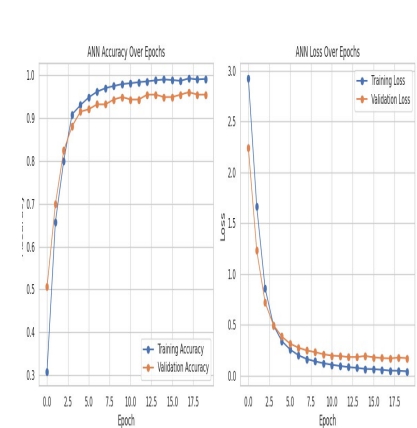


Figure 16 | Training Accuracy and Loss for Zoning plus Geometric dataset.

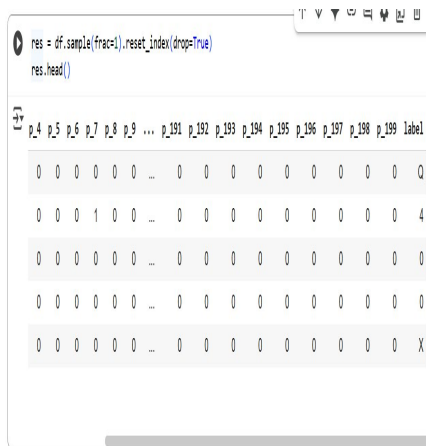


Figure 17 | Sample data from Serialized profile-based features.

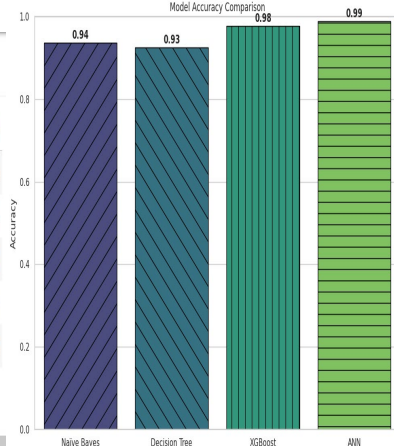


Figure 18 | Comparison of the model performance on Serialized dataset.

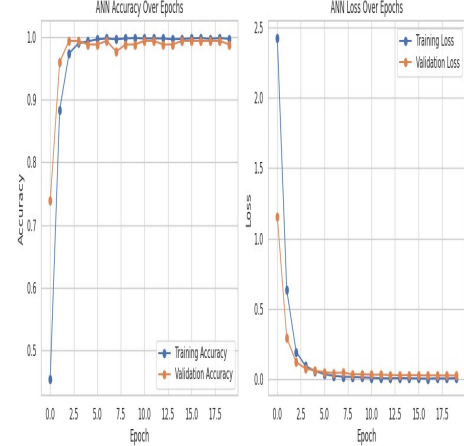


Figure 19 | Training Accuracy and Loss Graphs for Serialized dataset

observed that classification using serialized and deep learning approaches outperform other classifiers.

Without any additional feature engineering, the

CNN model had a comparable 99% accuracy rate. The training accuracy and loss graph of the CNN model is shown in Figure 4. It is consistent with recent results on the ability of convolutional models

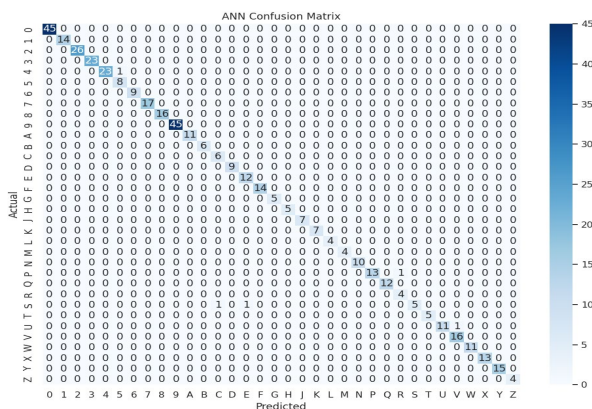


Figure 20 | Confusion Matrix for ANN with Serialized Dataset

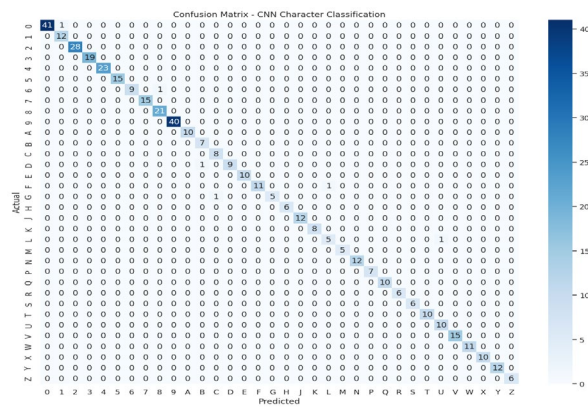


Figure 21 | Confusion Matrix for CNN Model

to automatically extract hierarchical features for noisy visual data (Liu *et al.*, 2024).

Especially, the CNN showed good performance in separating visually similar character pairs including B/8, 0/9, and S/5, which are often confused by a rule-based system. Although the high accuracy (99%) was obtained, some considerations have to be made on possible overfitting caused by the medium data size (2199) samples. The use of cross-validation and augmentation helps reduce this risk, however. This consistency of ANN and CNN results also indicates

the robustness of the results. The findings indicate that although engineered features are competitive, deep learning architectures have the benefits of scalability and adaptability, such as working with partial notes and varying light conditions in real world scenarios.

4.1 Grad-CAM Visualizations for Nigerian Banknote Serial Number Recognition

The discriminative image regions used by the Convolutional Neural Network (CNN) for character

classification are shown in the following Grad-CAM visualization. The highlighted activation regions show that the CNN mainly paid attention to the intersections of the strokes, boundaries, and structured actions in the local regions

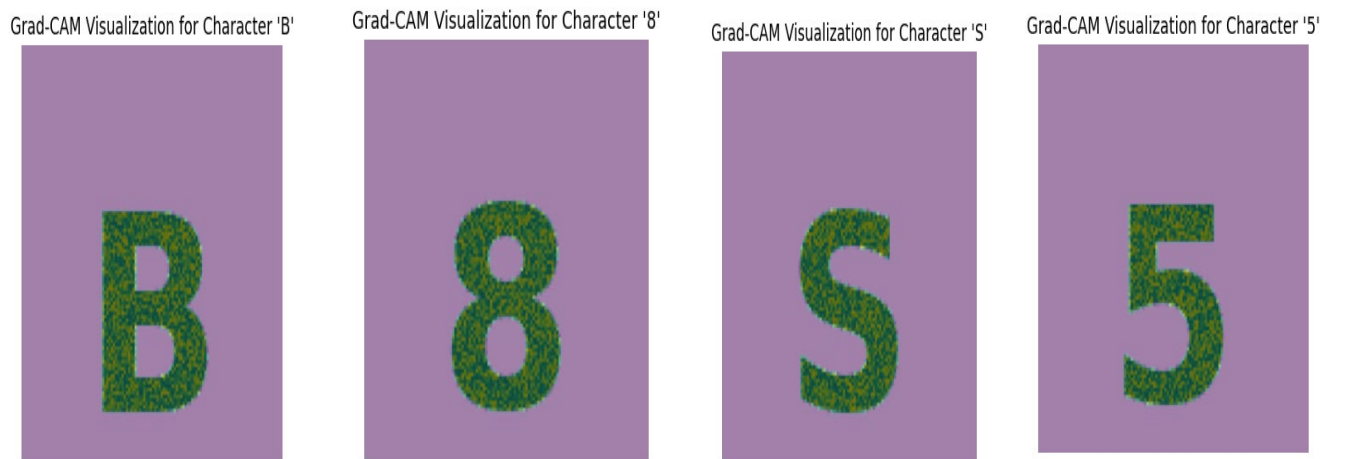


Figure 22 | Grad-CAM Activation Map for character ‘B’

Figure 23 | Grad-CAM activation for Map character ‘8’

Figure 23 | Grad-CAM activation for Map character ‘S’

Figure 23 | Grad-CAM activation for Map character ‘5’

The colour highlighted areas represent the areas that have the most significant influence for CNN-based classification.

‘5’) had overlapping activation regions, which is why there was some misclassification between behaviours seen when the characters were tested during the experiment.

4.2 | Discussion of Grad-CAM

During the character classification task, the CNN model focused on the small local structures and contour intersections at the stroke-level during the Grfad-CAM analysis. Characters that looked similarly geometric (e.g. ‘B’ and ‘8’ or ‘S’ and

4.3 | Error and Misclassification Analysis

The proposed models, although had high classification accuracy, some characters present in similar images still had misclassification behaviour from time to time. Some of the more frequently

Table 2 | Summary of Handcrafted Profile Features

Classifier	Horizontal Features		Vertical Features		Combined Features		Zoning & Geometric Features		Serialized Features	
	Accuracy	Macro F1-Score	Accuracy	Macro F1-Score	Accuracy	Macro F1-Score	Accuracy	Macro F1-Score	Accuracy	Macro F1-Score
Naïve Bayes	52%	0.49	57%	0.54	78%	0.76	84%	0.82	94%	0.92
Decision Tree	76%	0.72	75%	0.72	91%	0.89	91%	0.89	93%	0.91
XGBoost	89%	0.86	84%	0.81	96%	0.95	96%	0.95	97%	0.96
ANN	86%	0.84	82%	0.79	96%	0.94	95%	0.93	99%	0.98

swapped letters are B with 8, S with 5, O with 0 and 1 with I. Low-resolution character representations, contour overlap and structural similarity were the two main factors that caused these ambiguities.

The Naive Bayes classifier was the most sensitive to such ambiguities, because of its conditional feature independence assumption. The Decision Tree models were able to make better discriminations thanks to hierarchic threshold partitioning but were still not good at handling highly correlated geometric patterns. The XGBoost and ANN models showed strong robustness capabilities by capturing the interaction between features and complex spatial relationships, which is not linear. The CNN model outperformed other models as it captured the features in the raw images at the stroke level and identified contour hierarchy, which is necessary for identifying strokes. The misclassification analysis also showed that the misrecognition errors were mainly due to degraded notes, partial character occlusion, uneven illumination, and segmentation artifacts. A few other very small classes (M, Q, Z) showed lower recall because of the class imbalance in the dataset. In general, the errors indicate that deep learning models are better suited for complicated spatial variations and noisy imaging than handcrafted feature-based models.

5 | Limitation

Although the proposed framework has a promising recognition performance, there are some concerns. First, the datasets were acquired mostly under controlled scanning and pre-processing conditions that are not exactly what occurs in real-world situations where images are acquired in a mobile environment where they are subject to motion blur, uneven illumination, folded notes, and significant degradation of the currency. Secondly, the number

of samples is still relatively small compared to a large-scale industrial OCR system, which can lead to the model not being able to generalize well in new environments. Third, some character classes had imbalances that could have affected performance in recalling infrequently used alphanumeric characters. Furthermore, it was limited to Nigerian banknotes, and might need to be adapted for deployment on banknotes issued by other countries using different characters and serial number formats. Lastly, deep learning models have demonstrated better performance, but they also demand more computational power than traditional machine learning models, which could impact real-world systems with low-power embedded systems or mobile devices.

6 | Conclusion

In this study, a structured machine learning framework for serial number recognition on banknotes from Nigeria was developed and tested on Nigerian banknotes. Comparative analysis shows that the accuracy of recognition is high in both cases of using a serialised pixel representation with neural classifiers and convolutional neural networks. The handcrafted feature engineering shows generalization potential, and so does deep learning, but they are different. Handcrafted feature engineering has interpretable intermediate representations, as does deep learning, but in different ways. The framework provides a basis to embed automated serial number recognition into financial monitoring frameworks. The operational deployment feasibility of future work should be expanded by the development of real-time mobile integration of capture and object detection-based localization models like YOLOv8 (Redmon et al., 2024).

Acknowledgement

The authors acknowledge the technical assistance of the Computing and Information Science Laboratory

Staff of Bamidele Olumilua University of Education, Science, and Technology, Ikere - Ekiti, Ekiti State, Nigeria.

References

- Ahmad, M., Alqahtani, A., Alshammari, N., & Alqhamdi, A. (2022). Deep learning-based banknote authentication and counterfeit detection using convolutional neural networks. *IEEE Access*, 10, 84721-84735. <https://doi.org/10.1109/ACCESS.2022.3198465>
- Albahli, S., Alzahrani, F., Albattah, W., & Alotaibi, R. (2023). Deep learning framework for multi-currency banknote recognition and classification. *Applied Sciences*, 13(7), 4209. <https://doi.org/10.3390/app13074208>
- Baek, J., Lee, B., Han, D., Yun, S., & Lee, H. (2021). Character region awareness for text detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7), 2277-2291. DOI: 10.1109/CVPR.2019.00959
- Baek, J., Lee, B., Han, D., Yun, S., & Lee, H. (2021). Character region awareness for text detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7), 2277-2291. <https://doi.org/10.1109/TPAMI.2020.2973971>
- Chousangsuntorn, C., Tongloy, T., Chuwongin, S., & Boonsang, S. (2021). A deep learning system for recognizing and recovering contaminated slider serial numbers in hard disk manufacturing processes. *Sensors*, 21(8), 6261. <https://doi.org/10.3390/s21186261>
- Holila, H., Pratama, A.R., Lestari, S.A.P., & Indra, J. (2024). Introduction national identification number and name on ID card using OCR (Optical Character Recognition) method. *Jurnal Teknik Information (JUTIF)*, 5(4). <https://doi.org/10.52436/1.jutif.2024.5.4.2242>
- Jocher, G., Chaurasia, A., Qiu, J., & Ultralytics Team. (2024). YOLO by Ultralytics. GitHub Repository. <https://github.com/ultralytics>.
- Jang, U., Suh, K.H., & Lee, E. C. (2020). Low-quality banknote serial number recognition based on deep neural network. *Journal of Information Processing System*, 16(1), 224-237. <https://doi.org/10.3745/JIPS.04.0160>
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, DOI: 10.1109/CVPR.2016.90.
- Koponen, J., Haataja, K., & Toivanen, P. (2022). Recent advancements in machine vision methods for product code recognition: A systematic review. *F1000Research*, 11, 1099. <https://doi.org/10.12688/f1000research.124796.1>
- Li, M., Lv, T., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., Wei, F., & Zhou, M. (2023). TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models. *Proceedings of the AAAI Conference on Artificial Intelligence*,

37(11), 13094-13102. <https://doi.org/10.1609/AAAI.V37i11.26400>

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019. DOI: 10.1109/TNNLS.2021.3084827

Liu, X., Li, Z., Wang, P., & Sun, Y. (2023). Deep document intelligence using Transformer architectures for invoice information extraction. *Knowledge Based Systems*, 275, 110732. <https://doi.org/10.1016/j.knosys.2023.110732>

Monteiro, G., Camelo, L., Aquino, G., Fernandes, R.A., Gomes, R., Printes, A., Torne, I., Silva, H., Oliveira, J., & Figueiredo, c. (2023). A comprehensive framework for industrial stiker information recognition using advanced OCR and object detection techniques. *Applied Sciences*, 13(12), 7320. <https://doi.org/10.3390/app13127320>

Sharma, R., & Gupta, P., & Sun, Y. (2021). Handwritten cheque recognition using deep CNN-

LSTM architecture. *International Journal of Intelligent Systems and Applications*, 13(4), 52-63. <https://doi.org/10.5815/IJISA.2021.04.05>

Shi, B., Bai, X., & Yao, C. (2021). An end-to-end trained neural network for image-based sequence recognition and scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7), 2292-2304. <https://doi.org/10.1109/TPAMI.2020.2965206>

Smith, J. S., Tian, J., Halbe, S., Hsu, Y. C., & Kira, Z. (2023). A closer look at rehearsal-free continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2410-2420).

Subramani, N., Matton, A., Greaves, M., & Lam, A. (2020). A survey of deep learning approaches for ocr and document understanding. *arXiv preprint arXiv:2011.13534*. <https://doi.org/10.48550/arXiv.2011.13534>