Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.1533

ORIGINAL RESEARCH ARTICLE

A deep learning approach for forensic handwriting analysis

Adeyinka Oluwabusayo Abiodun[1,2,*], Olalekan M. Awoniran[3], Kingsley Eghonghon Ukhurebor[4,*], Ozichi Nweke Emuoyibofarhe[5], Adetoye Adeyemo[5], Idemudia Edetalehn Oaihimire[6]

[1] *Department of Computer Science, National Open University of Nigeria (NOUN), Abuja 900001, Nigeria*

[2] *Africa Centre of Excellence on Technology Enhanced Learning (ACETEL), National Open University of Nigeria (NOUN), Abuja 900001, Nigeria*

[3] *Bristol Medical School, University of Bristol, Bristol BS8 1NU, United Kingdom*

[4] *Department of Physics, Edo State University Uzairue, P.M.B. 04, Auchi* 312001, *Edo State, Nigeria*

[5] *Department of Computer Science, Bowen University, Iwo 232101, Osun State, Nigeria*

[6] *Faculty of Law, Edo State University Uzairue, P.M.B. 04, Auchi 312001, Edo State, Nigeria*

*** Corresponding authors:** Adeyinka Oluwabusayo Abiodun, aabiodun@noun.edu.ng; Kingsley Eghonghon Ukhurebor, ukeghonghon@gmail.com, ukhurebor.kingsley@edouniversity.edu.ng

ABSTRACT

Bayesian inference, which stems from Bayes' theorem, has been the major means to identify and recognize forensic biometric (BMT) traits over the years. Parameter consideration for this theorem differs from one examiner to another based on their level of expertise and subjectivity. Issues have been raised concerning this way of identifying and recognizing BMT traits in the forensic environment; therefore, there is a need to apply deep learning models to the recognition and identification of these BMT traits. Hence, in this research, various deep learning algorithms were adopted for the classification of handwriting. The handwriting was divided into different classes. The convolutional neural network (CNN) employed for this research was trained from scratch and also off-the-shelf", Support Vector Machine (SVM), Deep Neural Network (DNN), and Extreme Gradient Boosting (XGBoost) algorithms were also employed. Each of these algorithms performed well in various classes of these handwritings and gave varying performances in predicting the classes handwritings, with CNN having a 0.82 F-measure score and 96% accuracy leading, SVM having 79% accuracy, XGBoost having 73% accuracy, and DNN having 77% accuracy. However, CNN recorded the best result among the employed algorithms. Implicatively, CNN accurately predicted the class's handwriting. The results obtained from this study will further assist in figuring out the factors that explain examiners' determinations of sufficiency for individualization.

Keywords: CNN; SVM; DNN; Bayesian inference; forensic handwriting analysis

ARTICLE INFO

Received: 10 January 2024
Accepted: 25 January 2024
Available online: 11 July 2024

1. Introduction

Handwriting in the Pattern Recognition Community (PRC) is quite different from that of the forensic community. Identification, classification, and recognition in both communities differ. PRC can use image descriptors built for object recognition in natural images, like those of the SIFT[1], Adaptive SIFT/SURF[2], as well as the histogram of directed gradients (HOG)[3], for image and object recognition, detection, and segmentation. For quantifying the strength of evidence, the forensic community solely relies on the probability ratio from Bayesian inference. One of several reasons seems to be that the forensic community's testimony is often used in courts of law to make decisions.

Forensic science can be described as the application of scientific methods and processes to the investigation of criminal offenses. The methods used for evidence collection, interpretation, and presentation, as well as the underlying scientific basis for forensic examinations of some types of evidence, were among the concerns posed in a 2009 National Research Council (NRC) report[4]. Statistics is a crucial topic for the forensic community, although statistical inference is among the standard elements of statistical analysis.

In applying Bayesian inference to classify and recognize forensic biometric (BMT) features in individuals, considerable progress has been made over the past decades. Bayesian inference is based on the Bayes theorem[5,6], and it uses the likelihood ratio to identify handwriting[7–11]. These have been used in fingerprints [12,13] and other hard and soft BMT traits. The likelihood ratio makes use of posterior odd and prior odd for analysis. In contrast to frequentist inference, Bayesian inference interprets probabilities as subjective degrees of belief in order to state and evaluate beliefs. It can be seen as the probability of the degree of belief where probability statements are parameters θ of random variables with no guaranteed frequencies and are given a prior distribution $\pi(\theta)$ representing subjective beliefs about θ computing the posterior distribution for θ from data using Bayes theorem. It provides the tools to update their beliefs based on the evidence of new data.

There are many aspects to Bayesian inference. Subjective Bayesians are more into the probability interpretation of degrees of personal belief. Objective Bayesians reply on known data (prior odd) with the hope of leading into an objective posterior. Contrary to frequentist Bayesians, who use Bayesian methods only when the corresponding posterior has better frequency behaviour, empirical Bayesians estimate the prior distribution from the data. As a matter of fact, the difference between Bayesian and frequentist inference can be a little unclear, leading to a lot of drift in statistics, machine learning, and science. The differences between frequentist and Bayesian statistical reasoning have long been debated[14]. Throughout the 20th century, frequentists dominated statistical practice. Frequentist techniques are used in many popular machine learning algorithms, such as logistic and linear regression, to carry out statistical inference. Although before the 20th century Bayesians dominated statistical practice, numerous algorithms from the Bayesian schools, such as Expectation-Maximization, Bayesian Neural Networks, and Markov Chain Monte Carlo, have become more and more prevalent in machine learning. For more on their differences and relationships in the context of machine learning, see Causevic[14].

Deep learning techniques may be able to help clear up this confusion. In the forensic community, deep learning models can be used to classify and recognize BMT. There are a lot of techniques for using these models effectively for image classification; for instance, Convolutional Neural Network (CNN) features can also be used as complementary information sources to existing hand-crafted image features by utilizing "off-the-shelf CNN" features (without retraining the CNN) or training the "CNN from scratch"[14–17].

2. Related works

Forensic science refers to the body of scientific expertise and technical techniques used to answer questions about criminal, civil, and administrative law[18]. According to Gialamas[19], criminalistics is among the most important branches of forensic science, which is a career and scientific discipline that deals with the detection, identification, individualization, and assessment of physical evidence using natural science to solve jurisprudence issues.

A recent study brought attention to essential forensic science research, especially in the field of identification. The study concentrated on monographs that deal with either detection or identification, with specific attention on finger-mark detection, which includes formation mechanisms, secretion residue composition and properties, optical methods, detection techniques, and sequential processing[20–22]. BMT authentication is commonly used in computer science as a form of identification and to specify accesses, according to Wang et al.[23]. It is often used to identify individuals in groups that are under observation.

Physiological versus behavioural traits are commonly used to categorize BMT identifiers. BMT's contribution to security in today's digital world cannot be overstated, as it has gained popularity due to its multiple advantages. According to Mishra[24], forensic graphology is the study of handwriting, and it can be helpful in identifying handwriting in ransom notes in kidnapping cases, blackmailing letters, pen poisoning letters, and other similar situations. Fisher et al.[25] used automated handwriting analysis to classify personality traits, especially those that contribute to violent actions. Pugnaloni and Federiconi[26] worked on handwriting, believed that data such as the dynamically captured direction, stroke, distance, scale, pressure, and shape of an individual's signature allow handwriting to be a reliable indicator of an individual's identity.

Also, because the ear is visible, ear images are easy to take, and the ear structure may not change drastically over the years, Ear Recognition may be a good solution among several techniques[27]. Researchers are interested in ear identification because of its distinct physiological appearance and long-term structure[28–30].

Finger-marks are directly important in criminal investigations for individualization purposes, according to de Ronde et al.[13], and they also play a significant role in forensic science. This is supported by the fact that each individual has a distinct pattern of friction ridge skin on their hands, which can be used to identify them. A connection between the donor and a crime scene can be identified by identifying the source of the finger-mark. There has been a lot of research into enhancing the friction ridge pattern for individualization purposes by visualizing latent fingerprints[21,31]. Also, several researchers have focused on fingerprints, including Ulery et al.[32]; Ulery et al.[33]; Ulery et al.[34]; Ulery et al.[35]; Haraksim and Meuwly,[36]; Haraksim et al.,[37]; Liu et al.[38]; among other BMT security methods, fingerprint-based BMT authentication systems have been among the most widely used, common, and efficient authentication techniques for both identity recognition and verification. This BMT authentication method was established based on the natural evidence that each person has unique fingerprints on their hands that differentiate them from others. The interpretation of fingerprint data, according to Liu et al.[38], is dependent on the judgments of fingerprint examiners. However, Guo et al.[39] reported that forensic science and digital authentication both depend on fingerprint biometrics. They are predicated, nonetheless, on the untested premise that no two fingerprints—not even those from the same person's various fingers—are the same. This makes them unusable in situations where the fingerprints that are being displayed come from fingers that are not listed on the file[39]. In contrast to the commonly held belief, Guo et al.[39] recently demonstrated that there is a 99.99% confidence level in the similarity of fingerprints from distinct fingers belonging to the same individual. It was discovered that these commonalities exist across all pairs of fingers within the same person, even when filtering for spurious features like sensor modality. They extract fingerprint representation vectors using deep twin neural networks. Additionally, they discover evidence that ridge orientation—particularly in the vicinity of the fingerprint centre—explains a significant portion of this similarity, while details included in conventional approaches are essentially nonpredictive. According to our research, this link can sometimes lead to an almost two-fold boost in forensic inquiry efficiency.

Face BMT technologies are commonly utilized in our everyday endeavours, but no entirely automated face recognition (FR) system is currently approved by the judicial system, according to Arbab-Zavar et al.[40], which led to the implementation of manual and computer-aided forensic FR and describes the variations between automatic FR systems (BMT) and forensics, as well as summarizing the present advancement towards addressing the difficulties existing in FR. Other authors that have worked on face BMT are Wei et al.[41]; Reid et al.[42]; Alsaadi[43].

Ear BMT is also examined by Arbab-Zavar et al.[40] as a genuinely valuable BMT feature, and there has been a lot of research development. The current situation of formal recognition of ears as a forensic device is addressed, and a collection of morphological traits is provided, as well as an overview of their discriminatory forces. These attributes are critical in determining whether enough information is available for identification in the event of missing features. The language associated with these characteristics can also make it easier to communicate ear comparison reports to courts, which is a significant step in making such evidence useful for prosecution.

According to Kavitha and Priyatha[44], as the need for images rises in our daily lives, the motivation to develop forged images expands at the same time. The digital picture has rapidly replaced the previous analog photograph since the advent of digital technology. They said that human faces are retrieved from illuminated maps and that the modern information age has advanced to the point where technologies are being replaced by state-of-the-art digital counterparts[45–50].

Bansal and Kaushal[51], concluded that FR is the most difficult field of pattern recognition, and they also noted that SVM algorithms failed in many instances in the detection of forged pictures. Lund and Iyer[52] affirmed that experts from many forensic laboratories summarize their findings in the form of a likelihood ratio.

Human BMT characteristics such as face, finger, iris scanning, speech, signature, and other features, according to Alsaadi[43], provide a reliable level of protection for both personal and public usage.

According to Ulery[35], latent print evaluators use their expertise to evaluate if the information present in a comparison of two fingerprints (or palmprints) is appropriate to conclude that the prints were from the same source. When fingerprint identification is provided in court, the examiner's decision is presented rather than an objective metric. Ulery et al.[35] therefore planned a study to identify the variables that account for examiners' judgments of what is sufficient for individualization. The analysis showed notable variations in the annotations made by the examiners. We are unable to determine if this is because examiners differ in how they view and analyse the data, or just in how they record those discrepancies. Throughout the test, there may be disagreements on interpretation at several points: When identifying the boundaries of the region of interest to be utilized, an examiner examining an unclear print must assess whether there is adequate continuity; when evaluating a ridge inside an unclear region, an examination must determine whether characteristics are present; and an examiner must make a decision during comparative analysis to determine whether possibly similar features fall within a range that allows for acceptable variances in appearance. Every one of these choices might lead to variations in interpretations and, consequently, in annotations. Furthermore, there were several instances where examiners failed to identify any correspondences between the prints, leading them to render ambiguous conclusions on mated pairings. Much of the observed variety in annotations may have resulted from differences in interpretation as well as unclear criteria in the latent print discipline for when and how to identify features [6,21,22,53]. The efficacy of research like this is restricted by the absence of widely recognized and comprehensive guidelines for specifying and documenting the foundation for findings. Examiners are becoming more often required by courts to provide evidence supporting their judgments (during finding out what is admissible and a conviction)[35]. Hence, the aim of this study was to figure out what factors explain examiners' determinations of sufficiency for individualization. Consequently, in this research, various deep learning algorithms were adopted for the classification of handwriting. The handwriting was divided into different classes. The CNN employed for this research was trained from scratch and also off-the-shelf. Support Vector Machine (SVM), Deep Neural Network (DNN), and Extreme Gradient Boosting (XGBoost) algorithms were also employed so as to determine these algorithms performances in various classes of these handwritings as well as to give varying performances in predicting the classes handwritings.

3. Methodology

3.1. Image gathering

Our images were obtained from 10 different individuals. The initial images are pages of handwritten words obtained from the individuals. Each page was then cropped out line-by-line to give the final images to be classified. Each image is a line of cursive writing that is to be classified into the 10 categories of the individuals[54,55].

3.2. Image preprocessing

The sizes of the different handwritings, as well as the line-by-line cropping, resulted in the images with varying sizes. The highest width and length of the image is about 512 × 64 pixels [54,55]. Hence, other image with white spaces were padded to have equal sizes of images. This was our first stage of preprocessing. Thereafter, we converted all the images to grayscale and rescaled them to 256 × 32 pixels for easier computation[54,55].

Image augmentations were carried out on the images to obtain various versions of the handwritten images. It also creates variations of the desired pattern to avoid overfitting[56]. The augmentations performed on the handwritten images are resizing the handwriting inside the image while maintaining the size of the image, horizontal flips, vertical flips, blurring, histogram equalizations, 90-degree clockwise and counterclockwise rotations, 180-degree rotations, horizontal flips, and vertical flips[54,55].

The final image data size after the augmentations was 18,820. **Figure 1** shows one of the image files.

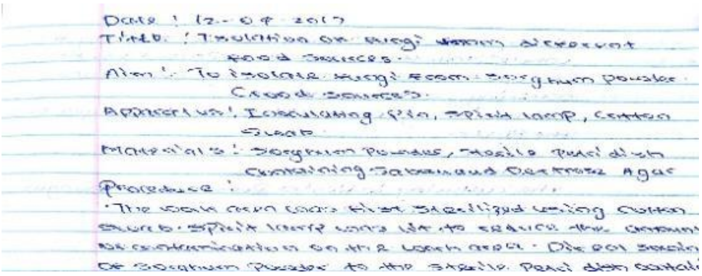

Figure 1. A sample image data.

Machine and deep learning techniques were used in building our model. For machine learning, SVMs and extreme gradient boosting algorithms were used for training the handwriting classification model. For deep learning, DNN and CNNs were used in training the model. However, before using the machine learning methods, the HOG feature extraction algorithm was used to extract features from the images.

3.3. Histogram of Gradients (HOG)

A HOG is an image feature extraction technique used for object detection. The method localizes portions of an image whose features are to be extracted and counts occurrences of gradient orientations. HOG can identify edges as well as provide the direction of the edges[57]. This is done by computing the orientation histograms of the edge intensities. To compute this, firstly, an appropriate filter mask (Laplacian, Sobel, Prewitt filters, etc.) is used to extract the edge gradients and orientations.

Afterward, a histogram grid is created using the gradients and orientations previously obtained. Each histogram divides the respective gradients into a small spatial area called a cell using its length. The amplitude of the gradients of the first order of each cell is calculated in both the horizontal and vertical directions. This amplitude is divided into nine bins with tri-linear interpolation. Each cell-block histogram concatenates 36-D vector characteristics. These characteristics are normalized using "L2-Hys" so that the influence of the local variation caused by illumination and contrast in the foreground will be reduced[58].

3.4. Support Vector Machine (SVM)

SVMs are a supervised learning technique used for classification and regression analysis. SVM can perform binary and multiclass classifications. This is done by constructing a set of hyperplanes in three-dimensional space. These hyperplanes are then used for classification and regression outlier detection. Class separations are derived from the distances of the hyperplanes to the nearest training data point of any class[54,55].

SVM uses mathematical functions referred to as kernels[59]. These kernel functions turn the data input space into a higher-dimensional space, and with this, not all the data points will be explicitly mapped. For our research, the radial basis function (RBF) was employed. Using RBF, the SVM kernel function for a binary classification can be written as follows:

$$2\ k(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|) \tag{1}$$

where $k(x_i, x_j)$ is the kernel function and $\| x_i - x_j \|^2$ is the squared Euclidean distance between two data points. γ is calculated using a free parameter[60]. Extending this to a multiclass problem involves breaking it down to multiple binary classification cases called the one-vs-one.

3.5. Extreme Gradient Boosting (XGBoost)

XGBoost is an ensemble machine learning algorithm based on the decision-tree method and gradient boosting framework. Decision tree methods are quite accurate at predicting structured data. The XGBoost method created a better classifier from less accurate classifiers. The method iteratively stacks models on top of each other, which results in stepwise error correction until a more accurate model is achieved[61].

In using the algorithm, the "objective was to "multi: softmax", "min_child_weight" to "0.2", "max_delta_step" to "0.1" and "reg_alpha" to "0.5".

3.6. Deep Neural Network (DNN)

A DNN is a machine learning technique that is based on a collection of connected units or nodes, which are referred to as artificial neurons. The artificial neurons model the natural neurons in a biological brain. The connections of the artificial neurons, just like the synapses in a biological brain, relay messages through signals to other neurons in the network[11].

In using DNN for the image classification, we created three layers of artificial neurons: the input layer, the hidden layer, and the output layer. All layers are dense, which means that they are tightly connected. The rectified linear activation function (ReLU) was applied in the first two layers, while the softmax activation function was applied in the last layer. Also, a "flatten" layer was added after the second layer. **Table 1** shows the model summary of the DNN model.

Table 1. Model summary of the DNN.

Layer (type)	Output shape	Parameters
dense_3 (Dense)	(None, 32, 256, 12)	24
dense_4 (Dense)	(None, 32, 256, 8)	104
flatten_1 (Flatten)	(None, 65,536)	0
dense_5 (Dense)	(None, 10)	655,370
Total params: 655,498		
Trainable params: 655,498		
Non-trainable params: 0		

3.7. Convolutional Neural Network (CNN)

A CNN is a deep learning algorithm that takes images as input, assigns learnable weights and biases to aspects of the image, and uses these weights to identify or classify the image[62]. Although CNNs could be slower in computation than other classification methods, feature engineering is unnecessary when using this method as it can learn the characteristics needed for image identification and classification. CNNs easily reduce images into forms that are easier to compute without losing the distinguishing features of the particular image[63]. In using CNN for our image classification, three convolutional layers were created. We set the first convolutional layer to have a filter of 16, a kernel size of 5 by 5, and "same" padding. Also, the ReLU activation function, maximum pooling of size 2 by 2, and strides of 2 were added. The parameters for the second convolution layer were quite similar to the first, but the filter size was set to 32, and the same values for activation and maximum pooling were added as with the previous layer. The final convolution layer was also quite similar, but the filter size was set to 64. The activation layer was added, and the ReLU activation function was equally used.

Finally, flatten, dense, and softmax activation layers were added. **Table 2** shows the model summary of the CNN model.

Table 2. Model summary of the CNN.

Layer (type)	Output shape	Parameters
conv2d_3 (Conv2D)	(None, 32, 256, 16)	416
activation_4 (Activation)	(None, 32, 256, 16)	0
max_pooling2d_2 (MaxPooling2)	(None, 16, 128, 16)	0
conv2d_4 (Conv2D)	(None, 16, 128, 32)	12,832
activation_5 (Activation)	(None, 16, 128, 32)	0
max_pooling2d_3 (MaxPooling2)	(None, 8, 64, 32)	0
conv2d_5 (Conv2D)	(None, 8, 64, 64)	51,264
activation_6 (Activation)	(None, 8, 64, 64)	0
flatten_3 (Flatten)	(None, 32,768)	0
dense_7 (Dense)	(None, 10)	327,690
activation_7 (Activation)	(None, 10)	0
Total params: 392,202		
Trainable params: 392,202		
Non-trainable params: 0		

4. Results

The final image data size after the various image augmentations were performed is 18,820. After performing our feature extraction using HOG, our image data has 8192 columns. Therefore, our image data has a matrix of shape 18,820 by 8192. We used principal components analysis to reduce the size of this 18,820 by 500 to reduce the computational cost incurred in our experimentation while also maintaining the image features.

After feeding the data into SVM, a maximum precision (positive prediction value) of 0.90 is recorded at the 6th grade and a minimum of 0.64 at the 10th grade. It also recorded a maximum recall (true positive rate) of 0.91 at the 3rd grade and a minimum of 0.67 at the 9th grade. The maximum harmonic mean of precision and recall (F-measure) was 0.86 in the 8th grade, and the minimum was 0.74 in the 9th and 10th grades. The overall accuracy of SVM is 0.79. These are shown in **Table 3**.

Table 3. Performance analysis of SVM.

Class	Precision	Recall	F-Measure	Support
1	0.81	0.69	0.75	127
2	0.87	0.76	0.81	216
3	0.77	0.91	0.83	210
4	0.78	0.74	0.76	189
5	0.79	0.89	0.84	187
6	0.90	0.69	0.78	151

Table 3. (*Continued*).

Class	Precision	Recall	F-Measure	Support
7	0.84	0.72	0.78	224
8	0.82	0.90	0.86	203
9	0.83	0.67	0.74	182
10	0.64	0.89	0.74	193
Accuracy	-	-	0.79	1882

Furthermore, after feeding the data into XGBoost, a maximum precision of 0.83 is recorded at the 6th grade and a minimum of 0.61 at the 1st, 9th, and 10th grades. It also recorded a maximum recall of 0.83 at the 10th grade and a minimum of 0.46 at the 1st grade. The maximum F-measure is 0.73 at the 3rd grade level, and the minimum is 0.53 at the 1st grade level. The overall accuracy of XGBoost is 0.67. These are shown in **Table 4**.

Table 4. Performance analysis of XGBoost.

Class	Precision	Recall	F-Measure	Support
1	0.61	0.46	0.53	127
2	0.74	0.66	0.70	216
3	0.68	0.78	0.73	210
4	0.69	0.72	0.71	189
5	0.70	0.74	0.72	187
6	0.83	0.53	0.65	151
7	0.62	0.58	0.60	224
8	0.71	0.71	0.71	203
9	0.61	0.60	0.60	182
10	0.61	0.83	0.70	193
Accuracy	-	-	0.67	1882

Again, as discussed earlier, the data was fed into our DNN model. The results obtained show that a maximum precision of 0.79 is recorded at the 2nd class and a minimum of 0.66 at the 3rd class. It also recorded a maximum recall of 0.83 at the 10th grade and a minimum of 0.60 at the 1st grade. The maximum F-measure is 0.77 in the 9th and 10th grades, and the minimum is 0.65 in the 1st grade. The overall accuracy of DNN is 0.71. These are shown in **Table 5**.

Table 5. Performance analysis of DNN.

Class	Precision	Recall	F-Measure	Support
1	0.72	0.60	0.65	127
2	0.79	0.66	0.72	216
3	0.66	0.76	0.70	210
4	0.68	0.70	0.69	189
5	0.73	0.75	0.74	187
6	0.71	0.64	0.67	151
7	0.72	0.64	0.68	224
8	0.67	0.72	0.70	203
9	0.76	0.77	0.77	182
10	0.72	0.83	0.77	193
Accuracy	-	-	0.71	1882

Finally, our data was fed into the CNN model. The results obtained show that a maximum precision of 0.90 is recorded at the 10th grade level and a minimum of 0.83 at the 7th grade level. It also recorded a maximum recall of 0.97 in the 5th grade and a minimum of 0.76 in the 1st grade. The maximum F-measure is 0.96 at the 5th grade level, and the minimum is 0.82 at the 1st grade level. The overall accuracy of CNN is 0.90. These are shown in **Table 6**.

Table 6. Performance analysis of CNN.

Class	Precision	Recall	F-Measure	Support
1	0.87	0.76	0.82	127
2	0.90	0.91	0.90	216
3	0.94	0.96	0.95	210
4	0.95	0.91	0.93	189
5	0.94	0.97	0.96	187
6	0.96	0.91	0.94	151
7	0.83	0.83	0.83	224
8	0.85	0.94	0.89	203
9	0.91	0.91	0.91	182
10	0.90	0.89	0.89	193
Accuracy	-	-	0.90	1882

Figure 2 shows a plot of the various accuracies obtained from applying the various algorithms to the image data.

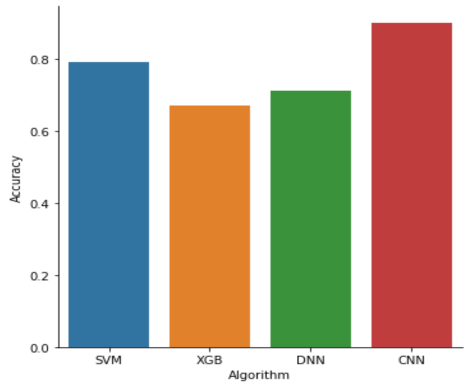

Figure 2. Plot of the various accuracies obtained from applying the various algorithms to the image data.

5. Discussion

The various algorithms adopted for the classification of the handwritings gave varying performances in predicting the classes of the handwritings. For example, SVM did quite well in predicting the 5th and 8th grades; XGBoost had its best result in the 3rd grade, although the scores are relatively low; DNN had its best results in the 8th and 9th grades; and CNN had its best result in the 3rd grade. However, if the actual scores are taken into consideration, the least-performing class in CNN (0.82 F-measure score in the 1st class) is higher than the best-performing class in both XGBoost (0.73 F-measure score in the 3rd class) and DNN (0.77 F-measure score in the 9th and 10th classes). SVM seems to be in between; however, it recorded the worst results in the classes, with relatively good results in both CNN and DNN. The variations in the results recorded across the various algorithms employed can be attributed to the peculiarities of each handwriting class and the principles on which the specific algorithms are built. In other words, the mathematical principles of the various algorithms favour some of the handwriting classes over others.

However, by considering the overall accuracy of the various algorithms, it can be concluded that CNN recorded the best result among the employed algorithms (0.9). Implicatively, CNN will accurately predict the classes each handwriting falls into 90% of the time.

6. Conclusion

The technology and document examiners of today centre on forensic examination. This has increased the rate at which document examination is required in the military, financial institutions, organizations, and courts of law. Over the years, forensic examiners have carried out their examinations using Statistical Bayesian inference, creating a huge problem in estimating various proposals for probability ratios, rendering it very complicated and difficult to locate the writer of a report and to produce a probability ratio free of nuisance parameters. This research established a template for forensic handwriting recognition using a deep learning approach, and the outcomes of these results were quite impressive.

It is, however, recommended that future research in this field should focus on incorporating past domain information into the deep learning approach for forensic handwriting analysis, as well as a comparison table that will provide the various performance measurement matrices to justify the analysis with an extended analysis with more state-of-the-art methods.

Author contributions

Conceptualization, AOA; methodology, AOA; software, AOA; validation, AOA, OMA, KEU, ONE, AA and IEO; formal analysis, AOA; investigation, AOA; resources, AOA and KEU; data curation, AOA; writing—original draft preparation, AOA and KEU; writing—review and editing, AOA, OMA, KEU, ONE, AA and IEO; visualization, XX; supervision, AOA and KEU; project administration, AOA and KEU; funding acquisition, AOA, OMA, KEU, ONE, AA and IEO. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors appreciate the authors and publishers, whose articles were used as guides for this study. Also, the authors express gratitude to their respective institutions and the Africa Centre of Excellence on Technology Enhanced Learning (ACETEL), National Open University of Nigeria, Abuja, for supporting this study.

Conflict of interest

The authors declare no conflict of interest.

References

1. Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 2004; 60(2): 91-110. doi: 10.1023/b: visi.0000029664.99615.94.
2. Adeyemo AB, Abiodun AO. Adaptive SIFT/SURF algorithm for off-line signature recognition. Journal of Egyptian Computer Science. 2015; 39(1): 50-56.
3. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005). pp. 886-893.
4. Peterson J, Sommers I, Baskin D, Johnson D. The Role and Impact of Forensic Evidence in the Criminal Justice Process Document No.: 231977. Available online: https://www.ojp.gov/pdffiles1/nij/grants/231977.pdf (accessed on 10 October 2023).
5. Martire KA, Kemp RI, Watkins I, et al. The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength, and the weak evidence effect. Law and Human Behavior. 2013; 37(3): 197-207. doi: 10.1037/lhb0000027.
6. Thompson WC, Newman EJ. Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. Law and Human Behavior. 2015, 39(4): 332-349. doi: 10.1037/lhb0000134.

7. Abiodun AO, Adeyemo AB. The Reproducibility and Repeatability of Modified Likelihood Ratio for Forensics Handwriting Examination. International Journal of Computer and Information Engineering. 2021; 15(5): 322-328.
8. Abiodun AO, Adeyemo AB. A Neural Network Approach to Writer's Model for Full Likelihood Ratio in Handwriting Analysis. American Journal of Neural Networks and Applications. 2020; 6(2): 29. doi: 10.11648/j.ajnna.20200602.13
9. Abiodun AO, Adeyemo AB. An Exhaustive Mapping Model for Modified Likelihood Ratio for Handwriting Recognition in Forensic Science. Journal of Forensic Science & Criminology. 2019; 7(3): 301.
10. Chen XH, Champod C, Yang X, et al. Assessment of signature handwriting evidence via score-based likelihood ratio based on comparative measurement of relevant dynamic features. Forensic Science International. 2018; 282: 101-110. doi: 10.1016/j.forsciint.2017.11.022.
11. Abiodun OI, Jantan A, Omolara AE, et al. State-of-the-art in artificial neural network applications: A survey. Heliyon. 2018; 4(11): e00938. doi: 10.1016/j.heliyon.2018.e00938.
12. Adamu LH, Taura MG. Application of likelihood ratio and posterior probability density in sex estimation from level two fingerprint features among Hausa ethnic group. Egyptian Journal of Forensic Sciences. 2017; 7(1). doi: 10.1186/s41935-017-0026-6.
13. de Ronde A, Kokshoorn B, de Poot CJ, et al. The evaluation of finger-marks given activity level propositions. Forensic Science International. 2019; 302: 109904. doi: 10.1016/j.forsciint.2019.109904.
14. Causevic S. Frequentist vs. Bayesian Approaches in Machine Learning: A Comparison of Linear Regression and Bayesian Linear Regression. Available online: https://towardsdatascience.com/frequentist-vs-bayesian-approaches-in-machine-learning-86ece21e820e (accessed on 23 October 2024).
15. Pan Y, Huang W, Lin Z, et al. Brain tumor grading based on Neural Networks and Convolutional Neural Networks. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Published online August 2015. doi: 10.1109/embc.2015.7318458.
16. Shen W, Zhou M, Yang F, et al. Multi-scale Convolutional Neural Networks for Lung Nodule Classification. Information Processing in Medical Imaging. Published online 2015: 588-599. doi: 10.1007/978-3-319-19992-4_46.
17. Wolterink JM, Leiner T, Viergever MA, et al. Automatic Coronary Calcium Scoring in Cardiac CT Angiography Using Convolutional Neural Networks. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. Published online 2015: 589-596. doi: 10.1007/978-3-319-24553-9_72.
18. Tistarelli M, Grosso E, Meuwly D. Biometrics in Forensic Science: Challenges, Lessons and New Technologies. Lecture Notes in Computer Science. Published online 2014: 153-164. doi: 10.1007/978-3-319-13386-7_12.
19. Gialamas DM. Criminalistics. Encyclopedia of Forensic Sciences. Published online 2000: 471-477. doi: 10.1006/rwfs.2000.0464.
20. Kasper SP. Latent Print Processing Guide. Elsevier; 2016.
21. Bleay SM, Croxton RS, Puit M. Fingerprint Development Techniques. Published online March 25, 2018. doi: 10.1002/9781119187400.
22. Daluz HM. Fingerprint Analysis Laboratory Workbook, 2nd ed. Taylor and Francis Group; 2018.
23. Wang Z, Yang J, Zhu Y. Review of Ear Biometrics. Archives of Computational Methods in Engineering. 2019, 28(1): 149-180. doi: 10.1007/s11831-019-09376-2.
24. Mishra A. Forensic Graphology: Assessment of Personality. Foresic Research & Criminology International Journal. 2017; 4(1). doi: 10.15406/frcij.2017.04.00097.
25. Fisher J, Maredia A, Nixon A, et al. Identifying Personality Traits, and Especially Traits Resulting in Violent Behavior through Automatic Handwriting Analysis. Pace University; 2012.
26. Pugnaloni M, Federiconi R. Forensic Handwriting Analysis: A Research by Means of Digital Biometrical Signature.
27. Štruc V, Žganec-Gros J, Vesnicer B, et al. Beyond parametric score normalisation in biometric verification systems. IET Biometrics. 2014; 3(2): 62-74. doi: 10.1049/iet-bmt.2013.0076.
28. Anwar AS, Ghany KKA, ElMahdy H. Human ear recognition using SIFT features. 2015 Third World Conference on Complex Systems (WCCS). Published online November 2015. doi: 10.1109/icocs.2015.7483254.
29. Pflug A. Busch C. Ear biometrics: a survey of detection, feature extraction and recognition methods. IET Biometrics. 2012; 1(2): 114-129.
Zhang L, Li L, Li H, et al. 3D Ear Identification Using Block-Wise Statistics-Based Features and LC-KSVD. IEEE Transactions on Multimedia. 2016, 18(8): 1531-1541. doi: 10.1109/tmm.2016.2566578.
30. Carneiro G, Nascimento J, Bradley AP. Unregistered Multiview Mammogram Analysis with Pre-trained Deep Learning Models. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Published online 2015: 652-660. doi: 10.1007/978-3-319-24574-4_78.
31. Ulery BT, Hicklin RA, Buscaglia J, et al. Accuracy and reliability of forensic latent fingerprint decisions. Proceedings of the National Academy of Sciences. 2011, 108(19): 7733-7738. doi: 10.1073/pnas.1018707108.
32. Ulery BT, Hicklin RA, Buscaglia J, et al. Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. Hsiao CK, ed. PLoS ONE. 2012, 7(3): e32800. doi: 10.1371/journal.pone.0032800.
33. Ulery BT, Hicklin RA, Kiebuzinski GI, et al. Understanding the sufficiency of information for latent fingerprint value determinations. Forensic Science International. 2013, 230(1-3): 99-106. doi: 10.1016/j.forsciint.2013.01.012.

34. Ulery BT, Hicklin RA, Roberts MA, Buscaglia J. Measuring what latent fingerprint examiners consider sufficient information for individualization determinations. PLoS One. 2014; 9: e110179.
35. Haraksim, R. Meuwly, D. Fingerprint Evidence Evaluation – Robustness to the Lack of Data, in proceedings EAFS 2012, Den Haag.
36. Haraksim R, Meuwly D, Doekhie G, et al. Assignment of evidential value of a fingermark general pattern using a Bayesian Network. BIOSIG; 2013.
37. Liu S, Champod C, Wu J, Luo Y. Study on Accuracy of Judgments by Chinese Fingerprint Examiners. Journal of Forensic Science and Medicine. 2015; 1(1): 33. doi: 10.4103/2349-5014.157908.
38. Guo G, Ray A, Izydorczak M, et al. Unveiling intra-person fingerprint similarity via deep contrastive learning. Science Advances. 2024, 10(2). doi: 10.1126/sciadv.adi0329.
39. Arbab-Zavar B, Wei X, Bustard JD, et al. On Forensic Use of Biometrics. Handbook of Digital Forensics of Multimedia Data and Devices. Published online July 6, 2015: 270-304. doi: 10.1002/9781118705773.ch7.
40. Wei X, Li CT, Hu Y. Face recognition with occlusion using Dynamic Image-to-Class Warping (DICW). 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Published online April 2013. doi: 10.1109/fg.2013.6553747.
41. Reid DA, Nixon MS, Stevenage SV. Soft Biometrics, Human Identification Using Comparative Descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 36(6): 1216-1228. doi: 10.1109/tpami.2013.219.
42. Alsaadi IM. ''Physiological Biometric Authentication Systems, Advantages, Disadvantages and Future Development: A Review'' International Journal of Scientific and Technology Research. 2015, 4(12).
43. Kavitha VP, Priyatha M. A Novel Digital Image Forgery Detection Method Using Svm Classifier. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2017; 3(2): 56-71.
44. Asanga MP, Essiet UU, Ukhurebor KE, et al. Social media and Academic Performance: A Survey Research of Senior Secondary School Students in Uyo, Nigeria. International Journal of Learning, Teaching and Educational Research. 2023; 22(2): 323-337. doi: 10.26803/ijlter.22.2.18.
45. Hussain A, Sipra HFK, Waheed A, et al. Exploring the academic perceptions of climate engineering in developing countries. ATMÓSFERA. Published online November 21, 2023. doi: 10.20937/atm.53264.
46. Hussaini AR, Ibrahim S, Ukhurebor KE, et al. The Influence of Information and Communication Technology in the Teaching and Learning of Physics. International Journal of Learning, Teaching and Educational Research. 2023; 22(6): 98-120. doi: 10.26803/ijlter.22.6.6.
47. Ndunagu JN, Ukhurebor KE, Adesina A. Virtual Laboratories for STEM in Nigerian Higher Education: The National Open University of Nigeria Learners' Perspective. In: Elmoazen R, López-Pernas S, Misiejuk K, et al. (editors). Technology-Enhanced Learning in Laboratories Workshop; 2023.
48. Nneji CC, Urenyere R, Ukhurebor KE, et al. The impacts of COVID-19-induced online lectures on the teaching and learning process: An inquiring study of junior secondary schools in Orlu, Nigeria. Frontiers in Public Health. 2022, 10. doi: 10.3389/fpubh.2022.1054536.
49. Nwankwo W, Ukhurebor KE. Web Forum and social media: A Model for Automatic Removal of Fake Media using Multilayered Neural Networks. International Journal of Scientific & Technology Research. 2020; 9(1): 4371-4377.
50. Bansal D, Kaushal S. A Novel Analysis of Image Forgery Detection Using SVM. International Journal of Engineering and Applied Sciences. 2016; 3(12).
51. Lund SP, Iyer H. Likelihood Ratio as Weight of Forensic Evidence: A Closer Look. Journal of Research of the National Institute of Standards and Technology. 2017; 122. doi: 10.6028/jres.122.027.
52. Kavitha VP, Priyatha M. A Novel Digital Image Forgery Detection Method Using SVM Classifier. IJAREEIE. 2014; 3(2): 56-71.
53. Billah M, Waheed S. Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection. Multimedia Tools and Applications. 2020; 79(33-34): 23633-23643. doi: 10.1007/s11042-020-09151-7.
54. Islam MT, Islam MR, Uddin MP, et al. A Deep Learning-Based Hyperspectral Object Classification Approach via Imbalanced Training Samples Handling. Remote Sensing. 2023; 15(14): 3532. doi: 10.3390/rs15143532.
55. Hayashi T, Gyohten K, Ohki H, et al. A Study of Data Augmentation for Handwritten Character Recognition using Deep Learning. 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). Published online August 2018. doi: 10.1109/icfhr-2018.2018.00102.
56. Zhou W, Gao S, Zhang L, et al. Histogram of Oriented Gradients Feature Extraction From Raw Bayer Pattern Images. IEEE Transactions on Circuits and Systems II: Express Briefs. 2020, 67(5): 946-950. doi: 10.1109/tcsii.2020.2980557.
57. Alfalou A, Jridi M, Kourtli Y, et al. A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF techniques for face recognition. Alam MS, ed. Pattern Recognition and Tracking XXIX. Published online April 30, 2018. doi: 10.1117/12.2309454.
58. Huang S, Nianguang CAI, Penzuti Pacheco P, et al. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics and Proteomics. 2018; 15(1): 41-51.

59. Cao J, Wang M, Li Y, et al. Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment. Bagci U, ed. PLOS ONE. 2019; 14(4): e0215136. doi: 10.1371/journal.pone.0215136.
60. Chen T, Guestrin C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Published online August 13, 2016. doi: 10.1145/2939672.2939785.
61. Indolia S, Goswami AK, Mishra SP, et al. Conceptual Understanding of Convolutional Neural Network—A Deep Learning Approach. Procedia Computer Science. 2018; 132: 679-688. doi: 10.1016/j.procs.2018.05.069.
62. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Available online: http://proceedings.mlr.press/v97/tan19a.html (accessed on 10 October 2023).