Scientific Research Publishing

Open Journal of Applied Sciences, 2024, 14, 3490-3499
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Comparative Study of Four Classification Techniques for the Detection of Threats in Baggage from X-Ray Images

Boka Trinité Konan[1*], Hyacinthe Kouassi Konan[2], Jules Allani[1], Olivier Asseu[1,2]

[1]INPHB, EDP-STI, Yamoussoukro, Côte d'Ivoire
[2]LASTIC, ESATIC, Abidjan, Côte d'Ivoire
Email: *triniteboca03@gmail.com, oasseu@yahoo.fr, hyacinthekonan2000@yahoo.com, julallani@yahoo.fr

How to cite this paper: Konan, B.T., Konan, H.K., Allani, J. and Asseu, O. (2024) Comparative Study of Four Classification Techniques for the Detection of Threats in Baggage from X-Ray Images. *Open Journal of Applied Sciences*, **14**, 3490-3499.
https://doi.org/10.4236/ojapps.2024.1412228

Received: November 5, 2024
Accepted: December 7, 2024
Published: December 10, 2024

Abstract

Baggage screening is crucial for airport security. This paper examines various algorithms for firearm detection in X-ray images of baggage. The focus is on identifying steel barrel bores, which are essential for detonation. For this, the study uses a set of 22,000 X-ray scanned images. After preprocessing with filtering techniques to improve image quality, deep learning methods, such as Convolutional Neural Networks (CNNs), are applied for classification. The results are also compared with Autoencoder and Random Forest algorithms. The results are validated on a second dataset, highlighting the advantages of the adopted approach. Baggage screening is a very important part of the risk assessment and security screening process at airports. Automating the detection of dangerous objects from passenger baggage X-ray scanners can speed up and increase the efficiency of the entire security procedure.

Keywords

Deep Learning, Baggage Control, Convolutional Neural Networks, Image Filtering, Object Detection Algorithms, X-Ray Images, Autoencoder, Random Forests

1. Introduction

Identifying dangerous objects in passenger baggage is essential to ensure aviation security. Currently, this process relies on X-ray imaging and human inspection, which is laborious and prone to errors, with an accuracy of 80% - 90% [1]. Challenges include the low number of dangerous objects present, the diversity of items, and the need for rapid decisions, especially during peak periods. To improve the

efficiency and accuracy of detection, automation via advanced image processing and object detection techniques is recommended.

Research on threat detection in baggage security can be grouped according to three imaging modalities:

- single-view X-rays [2],
- multi-view X-rays [3] [4], and
- computed tomography (CT) [5].

Classification performance improves with the number of views used, from 89% true positive rate (TPR) and 18% false positive rate (FPR) in single-view imaging [2] to 97.2% TPR and 1.5% FPR in full-view CT imaging [5]. However, it is generally recommended that these methods allow a more complex classification of radiographic images than that of visible spectrum images, and that methods commonly used for natural images, such as SIFT or HoG, are not always effective for radiographs [6].

However, identification performance can be improved by exploiting features of X-ray images:

- by augmenting multiple views;
- by using a colored material image or
- by employing simple density (gradient) histogram descriptors [7]-[9].

The authors of [10] address the challenges of learning features on images of varying sizes and out-of-plane rotations. Their work aims to create an automation framework for detecting firearms from X-ray scanners, focusing on the classification of steel barrel bores to assess the probability of weapons presence in X-ray images, using different classification methods.

Two datasets of images from dual-view X-ray scanners are used to evaluate the performance of the classifiers:

- the first dataset contains images of hand-held travel luggage, while
- the second dataset includes courier parcel scans.

To handle the variable image size, two views are combined into a single sample. Although the out-of-plane rotation problem is not directly addressed, data augmentation techniques and a dataset containing threat objects in different poses are used. Two deep learning techniques are studied, convolutional neural networks (CNNs) and stacked autoencoders, as well as two popular classification models, namely feedforward neural networks and random forests. Their implementation results are compared and critically discussed (**Figure 1**).

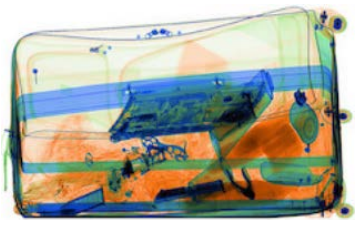

Figure 1. Dual-view raw X-ray of image containing weapon.

The remainder of the paper is organized as follows. Section 2 describes the datasets used in the empirical experiment and illustrates the proposed framework; Section 3 presents the details of the experiments performed and the results, while the conclusion and future work are presented in Section 4.

2. Threat Identification Framework

The framework that is proposed for automated weapons detection includes three modules: preprocessing, data augmentation and threat detection (see Figure 2).

The preprocessing stage consists of four steps:

- green channel extraction,
- grayscale smoothing,
- black-and-white thresholding, and
- data augmentation.

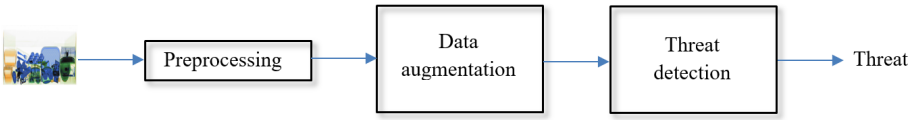

Figure 2. Framework that is proposed for automated threat detection.

The original dataset contains over 22,000 images, of which about 6000 show threat elements (firearms or components).

The images come from a dual-view X-ray machine, which can scan objects from two different angles. X-ray scanners create layered images, allowing operators or algorithms to detect suspicious objects through information such as shape, density, and material characteristics. These images include metadata about their classification: benign or threat (e.g., Figure 1 and Figure 3). And weapon components (such as the barrel or the complete weapon). A sample of 3546 threat images with barrels and 1872 benign images was selected.

A threat element in a security context is an element considered dangerous (e.g., a gun barrel, a spare part of a rifle, etc.). For classification, a security expert or operator could be tasked with classifying images based on their content. This includes visually identifying firearm components (such as a barrel or the complete weapon itself). These objects are categorized as threatening or non-threatening. During preprocessing, each image is processed individually and the two views are combined. The raw images are resized to 128 × 128 pixels and converted to grayscale, extracting the green channel to maximize contrast in dense materials. This allows for better filtering and makes it easier to recognize guns.

A smoothing algorithm is applied to the grayscale image to reduce noise while preserving edges. Among the different algorithms tested, a 3 × 3 kernel Gaussian blur gave the best results. Next, a thresholding technique is used to isolate dense materials, with a threshold equivalent to 2 mm of steel, which allows for the preservation of metallic objects such as gun barrels. This step eliminates a large part of benign backgrounds, such as organic materials.

Figure 3. Image from the dual-view raw X-ray scan.

The image is normalized so that the densest material appears black, and areas below the threshold become white. Thus, images without significant dense material can be classified as benign. This is particularly effective for baggage, as a significant proportion of samples can be filtered out, unlike parcels where there is a greater diversity of metallic objects. Finally, to improve the robustness of classification under class imbalance, it is often useful to introduce noise and realistic variations into the training data [11].

There are several ways to achieve this:

- object volume scaling: scaling the object volume by a factor;
- object flips/shifts: objects can be flipped/shifted in the x or y direction to increase the variation in appearance.

For each image in the training set, multiple instances are generated by combining various augmentation procedures, which are then used during training. The two views of each sample are stacked vertically to create a final image.

Four machine learning methods are compared: two from deep learning (convolutional neural networks, CNNs, and stacked autoencoders) and two classical techniques (neural networks and random forests).

CNNs, considered the best for image recognition, consist of an input layer (pixel matrix), an output layer (class label), and several hidden layers. These include convolution, activation, and pooling functions, with the final layers being fully connected with a softmax output function. e.g.: from a variety of problems related to image recognition and object detection [12], unmanned helicopter control [13], X-ray cargo inspection [7], and many others.

The convolutional layer applies a 2D filter on the image to learn representations, while pooling reduces the size of the inputs by aggregating the information. After several layers, the output is flattened and passed to a classical layer for the final classification.

Stacked autoencoders, or autoassociative neural networks, are an unsupervised machine learning technique that allows learning features at different levels of abstraction. They consist of two parts:

- The encoder, which reduces the input to a lower dimensional space.
- The decoder, which reconstructs the original input from this representation.

The new representation can then serve as input to another autoencoder, hence the term "stacked". Each layer is first trained independently to learn a hierarchical representation, and then the entire network is fine-tuned by backpropagation in

a supervised manner to distinguish classes. In this study, sparse autoencoders are used, which incorporate regularization to learn a sparser representation of the input.

3. Experimentation and Results

After image preprocessing and filtering, 1850 samples were obtained for baggage classification and 1760 for parcels.

- Baggage dataset: 670 benign images and 1180 containing threats.
- Parcel dataset: 580 benign images and 1190 threatening images.

Each set was split into 80% for training and 20% for testing, with baggage and parcel scans trained and tested separately due to their different operational contexts.

3.1. Presentation of the Techniques Used

The techniques presented below are often used in the field of computer vision and machine learning to extract relevant information from complex data.

A. CNN (Convolutional Neural Network): This is a type of neural network mainly used for image processing. CNNs are designed to recognize local patterns using convolution layers, often followed by pooling layers that reduce dimensionality.

B. Autoencoder: This is a type of neural network used for unsupervised learning. It is composed of two parts: an encoder that compresses the input into a lower-dimensional representation, and a decoder that reconstructs the input from this representation. Autoencoders are often used for dimension reduction, image denoising, or generating new data.

C. OBIFs + NN (OBIFs with Neural Networks): "OBIFs" refers to "Oriented BIFs" (Basic Image Features), which are image features extracted for recognition and analysis. Using "OBIFs" with Neural Networks (NN) implies that these features are used as inputs for a neural network to improve the performance of tasks like classification or object detection.

D. OBIFs + RF (OBIFs with Random Forest): In this case, "OBIFs" are also used, but here with a classifier based on Random Forests. Random Forests are a set of many decision trees and are used for classification and regression tasks. The idea is to use the OBIFs as input features to train the Random Forest model.

3.2. Experimentation

In this experiment, a three-layer stacked autoencoder with 200, 100, 50 neurons respectively, followed by a softmax output function to predict the probability of the classes was used.

For the CNN, a topology with three convolutional layers (with 128, 64 and 32 neurons) followed by a fully connected neural network and a softmax output function was used.

- 3 convolutional layers with 3 × 3 filters, activated by ReLU.

- Batch Normalization after each convolution to stabilize the training.
- MaxPooling after each convolution layer to reduce the size of the feature maps.
- Flatten to convert 2D data to 1D vector.
- Dense layers for final classification, with 128 neurons in the hidden layer and 2 neurons for the output (one for each class).
- 50% dropout after the dense layer to prevent overfitting.
- Softmax activation function in the output layer for multi-class classification.
- Adam optimizer and sparse_categorical_crossentropy loss for classification.

When training our model, the learning curves (Loss Curve and Accuracy Curve) were not optimal. To improve the performance of the CNN on the dataset, several modifications were made, including:

- Data augmentation: This helps to avoid overfitting by generating variations of the training images.
- Normalization: Ensuring that the data is well normalized.
- Adding regularization layers: Like dropout, to reduce overfitting.
- Using Batch Normalization: This can help stabilize and speed up training.

See **Figure 4** for CNN performance.

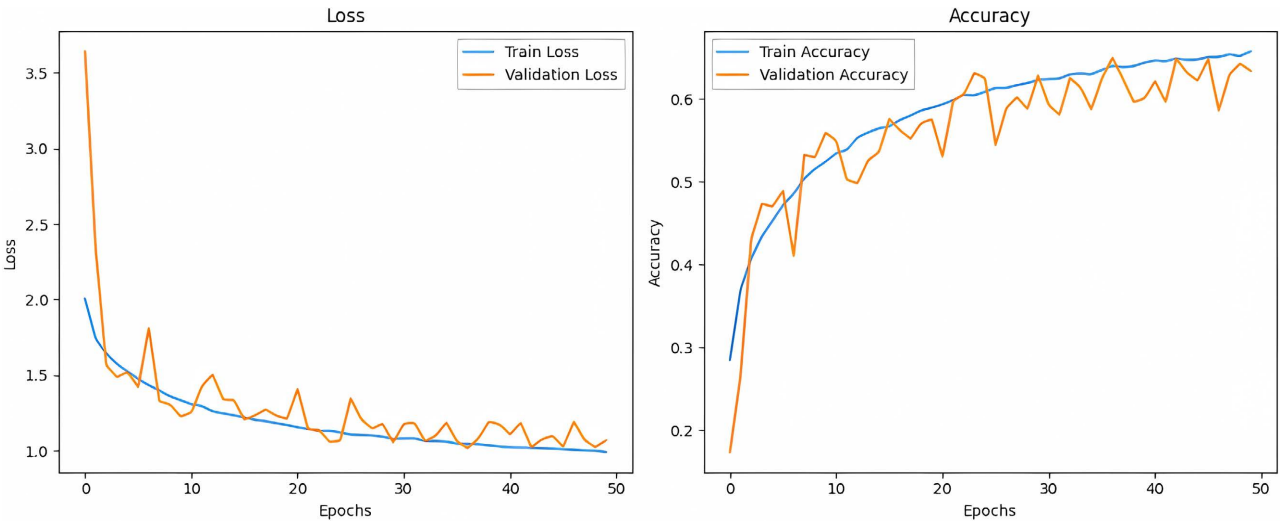

Figure 4. Loss and accuracy.

The Random Forest (RF) was trained with 200 trees, while the shallow neural network (NN) had an n-n-2 topology, where n is the input size. Since the RF and NN cannot be directly trained on raw pixels, a feature extraction step was required.

Oriented basic image feature histograms (BIFs) were used as a texture descriptor (as suggested in [6]), an effective method in computer vision.

BIFs classify each pixel in an image into one of seven categories based on local symmetries, such as: flat, slopes, spots, lines, and saddle-shaped. Oriented BIFs add a quantized orientation dimension to encode compact representations. The OBIF feature vector is then used as input to the RF and shallow NN classifiers.

To evaluate classification performance, three measures were used:

- Area under the ROC curve (AUC),

- False positive rate at 90% of true positives (FPR@90%TPR), and
- F1-score.

AUC is a common metric to evaluate classification performance, while FPR@90%TPR indicates the number of false positives expected when 90% of threats are correctly identified. This 90% threshold is recommended by [6] for X-ray image classification.

F1-score is also used for imbalanced datasets, considering precision (correctly identified threats vs. all identified threats) and recall (correctly identified threats vs. all threats present). A table shows the results for the baggage dataset for these metrics.

Results are reported for the four classification techniques and the three preprocessing steps:

- raw data,
- grayscale smoothing, and
- black and white thresholding.

As shown in **Table 1**, the CNN outperformed the other methods with AUC ranging from 94% to 97%, depending on the preprocessing step.

The second best method was the Shallow NN with AUC values ranging from 86% to 95%, while the worst performance was obtained by the RF with AUC ranging from 67% to 83%.

Similar results were obtained when considering the FPR@90%TPR and F1-score metrics. The CNN achieved the best FPR (6%) when trained on the thresholded black-and-white images, while having only 9% FPR when using raw data. On the other hand, while achieving 14% FPR with the last preprocessing step, the NN performance dropped drastically when using the raw and smoothed data, with 50% and 31% FPR, respectively. The same observation can be observed when using the F1-Score: the CNN reaches up to 93%, followed by the stacked autoencoders and the shallow NN with 81% and 79%, respectively. Once again, it is worth noting that the CNN was the only technique able to achieve high classification accuracy in all the preprocessing approaches used, while the other methods required more time spent on the feature engineering and extraction steps.

Table 1. Baggage dataset results for the AUC, FPR@90%TPR and F1-Score metrics.

metric	Technical	Smoothing	Raw data	Black and white threshold
AUC	CNN	94	95	97
	Autoencoder	75	78	90
	oBIFs + NN	86	87	95
	oBIFs + RF	67	72	83
FPR@90%TPR	CNN	9	7	6
	Autoencoder	70	60	26
	oBIFs + NN	50	31	14
	oBIFs + RF	86	66	53

Continued

F1-Score	CNN	91	93	93
	Autoencoder	60	65	81
	oBIFs + NN	64	67	79
	oBIFs + RF	36	41	56

Table 2. Results of the parcel dataset for AUC, FPR@90%TPR and F1-Score measures. The results are presented for the four classification techniques and the three preprocessing steps: raw data, grayscale smoothing and black-white thresholding.

metric	Technical	Smoothing	Raw data	Black and white threshold
AUC	CNN	80	78	85
	Autoencoder	65	66	75
	oBIFs + NN	65	69	84
	oBIFs + RF	63	63	79
FPR@90%TPR	CNN	46	46	37
	Autoencoder	66	69	70
	oBIFs + NN	71	75	40
	oBIFs + RF	91	88	56
F1-Score	CNN	86	83	87
	Autoencoder	40	43	55
	oBIFs + NN	36	32	63
	oBIFs + RF	34	42	58

Table 2 shows the performance metrics on the parcel dataset, illustrating generally lower performance for all techniques. This can be explained by the greater variety of metal objects contained in courier parcels, compared to objects contained in an airport carry-on bag. Again, the CNN outperformed the other methods considered, with an AUC ranging from 78% to 85%, followed by the NN with 65% to 84%, the RF with 63% to 79%, and the stacked autoencoders with 65% to 75%. The AUC obtained on the parcel dataset by the shallow NN, the RF, and the stacked autoencoders is much closer to that obtained on the baggage dataset, where the best-performing method stands out more.

Once again, the CNN achieved the lowest FPR (37%), followed by the shallow NN with 40% FPR, the RF with 56% FPR, and the stacked autoencoders with 70% FPR. Finally, the F1 score measure produced the largest difference in values between the methods, with the CNN reaching up to 87% F1 score, followed by the shallow NN with 63%, the RF with 58%, and the stacked autoencoders with 55%. Moreover, in this case, the CNN was the only technique able to classify threats with high accuracy, using only raw images, where all other techniques performed very poorly (e.g., the AUC on raw data for the CNN was 15 percentage points better than the NN, while maintaining similar performance on the black and white

threshold; 20 percentage points better in the FPR@90% TPR compared to the second best (Autoencoder); and even 46 percentage points better than the Autoencoder for the F1 score).

4. Conclusions

In this study, a deep learning framework was developed to automatically identify steel barrel bores in X-ray images, particularly in the context of airport security checks and parcel inspections. Two deep learning methods (convolutional neural networks, CNNs, and stacked autoencoders) were compared, along with two classical classification techniques (shallow neural networks and random forests) on luggage and parcel datasets.

The performance of the methods was evaluated using three metrics: area under the ROC curve (AUC), false positive rate at 90% true positives (FPR@90%TPR), and F1 score. The results showed that the CNN consistently outperformed the other techniques on all three metrics and both datasets, achieving good accuracy even with raw data, while the other methods required preprocessing steps.

Additionally, the CNN achieved higher accuracy than human screening results in the literature, although the datasets were not peer-reviewed. Future work will aim to explore different architectures for the CNN and stacked autoencoders, using larger datasets to further expand on these initial results.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Michel, S., Koller, S.M., de Ruiter, J.C., Moerland, R., Hogervorst, M. and Schwaninger, A. (2007) Computer-Based Training Increases Efficiency in X-Ray Image Interpretation by Aviation Security Screeners. 2007 41*st Annual IEEE International Carnahan Conference on Security Technology*, Ottawa, 8-11 October 2007, 201-206. https://doi.org/10.1109/ccst.2007.4373490

[2] Riffo, V. and Mery, D. (2012) Active X-Ray Testing of Complex Objects. *Insight—Non-Destructive Testing and Condition Monitoring*, **54**, 28-35. https://doi.org/10.1784/insi.2012.54.1.28

[3] Mery, D., Riffo, V., Zscherpel, U., Mondragón, G., Lillo, I., Zuccar, I., *et al.* (2015) Gdxray: The Database of X-Ray Images for Nondestructive Testing. *Journal of Nondestructive Evaluation*, **34**, 1-12. https://doi.org/10.1007/s10921-015-0315-7

[4] Mery, D., Riffo, V., Zuccar, I. and Pieringer, C. (2013) Automated X-Ray Object Recognition Using an Efficient Search Algorithm in Multiple Views. 2013 *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, 23-28 June 2013, 368-374. https://doi.org/10.1109/cvprw.2013.62

[5] Flitton, G., Mouton, A. and Breckon, T.P. (2015) Object Classification in 3D Baggage Security Computed Tomography Imagery Using Visual Codebooks. *Pattern Recognition*, **48**, 2489-2499. https://doi.org/10.1016/j.patcog.2015.02.006

[6] Jaccard, N., Rogers, T.W., Morton, E.J. and Griffin, L.D. (2016) Tackling the X-Ray Cargo Inspection Challenge Using Machine Learning. *SPIE Proceedings*, Baltimore,

12 May 2016, 131-143. https://doi.org/10.1117/12.2222765

[7] Rogers, T.W., Jaccard, N. and Griffin, L.D. (2017) A Deep Learning Framework for the Automated Inspection of Complex Dual-Energy X-Ray Cargo Imagery. *SPIE Proceedings*, Anaheim, 1 May 2017. https://doi.org/10.1117/12.2262662

[8] Li, G. and Yu, Y. (2018) Contrast-Oriented Deep Neural Networks for Salient Object Detection. *IEEE Transactions on Neural Networks and Learning Systems*, **29**, 6038-6051. https://doi.org/10.1109/tnnls.2018.2817540

[9] Shen, Y., Ji, R., Wang, C., Li, X. and Li, X. (2018) Weakly Supervised Object Detection via Object-Specific Pixel Gradient. *IEEE Transactions on Neural Networks and Learning Systems*, **29**, 5960-5970. https://doi.org/10.1109/tnnls.2018.2816021

[10] Bastan, M., Byeon, W. and Breuel, T. (2013) Object Recognition in Multi-View Dual Energy X-Ray Images. *Proceedings of the British Machine Vision Conference* 2013, Bristol, 9-13 September 2013, 1-11.

[11] Zhang, C., Tan, K.C., Li, H. and Hong, G.S. (2019) A Cost-Sensitive Deep Belief Network for Imbalanced Classification. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 109-122. https://doi.org/10.1109/tnnls.2018.2832648

[12] Zhao, Z.-Q., Zheng, P., Xu, S.-T. and Wu, X. (2019) Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 1-21.

[13] Kang, Y., Chen, S., Wang, X. and Cao, Y. (2019) Deep Convolutional Identifier for Dynamic Modeling and Adaptive Control of Unmanned Helicopter. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 524-538. https://doi.org/10.1109/tnnls.2018.2844173