



# Human Trafficking Detection System Using EfficientNet Model

Emmanuel Waribo Otiti, Chidiebere Ugwu, Linda Uchenna Oghenekaro, Augustine Ugbari

Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria

Email: fammava@gmail.com, chidiebere.ugwu@uniport.edu.ng, linda.oghenekaro@uniport.edu.ng,

augustine.ugbari@uniport.edu.ng

**How to cite this paper:** Otiti, E.W., Ugwu, C., Oghenekaro, L.U. and Ugbari, A. (2023) Human Trafficking Detection System Using EfficientNet Model. *Open Access Library Journal*, 10: e10716.  
<https://doi.org/10.4236/oalib.1110716>

**Received:** September 8, 2023

**Accepted:** November 25, 2023

**Published:** November 28, 2023

Copyright © 2023 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Human trafficking is the world's most prevalent and growing crime. Law enforcement agencies have been faced with a lot of challenges worldwide in this area. Over the years different works of literature have tried to sort out methods and techniques for detecting and controlling human trafficking by making use of different types of data available. In this research, an EfficientNet model was developed to help in the detection of human trafficking cases, by classifying an image into a class of chains. This helps to narrow the range of possibilities for an investigator. The object-oriented system analysis and design methodology were adopted and python was used in implementing the system. Data images and CSV files containing 97556 samples were obtained from the Kaggle repository and pre-processed by making use of the TensorFlow ImageDataGenerator function for normalization. After performing pre-processing, the top layers of the pre-trained model are replaced with the own data in fine-tuning. The data was used to train the model with a batch size of thirty-two (32), and the model was trained for 25 epochs. The model was evaluated against the accuracy, recall, and precision metrics, each of which recorded a score of 0.802. The metrics demonstrated that the model is robust in classifying hotel chains in detecting human trafficking activities.

## Subject Areas

Artificial Intelligence, Computer Vision, Image Processing, Machine Learning

## Keywords

EfficientNet, Human Trafficking, Machine Learning, Hotel Images, Detection System, Normalization, Mobile application

## 1. Introduction

Human trafficking is the world's most prevalent and growing crime. Law enforcement agencies have been faced with a lot of challenges worldwide in this area. Victims are individuals who seek for a greener pasture and fall easily for human traffickers' plots. Thus, currently, the third-largest organized crime worldwide is human trafficking [1]. The term modern slavery is seen as an ensemble of human trafficking, forced labour, domestic servitude, and sexual exploitation [2]. An alarmingly high rate of modern slavery was estimated by the International Labour Organization (ILO) to have reached at least 1.3 million people by 2016 [3]. The victims are either recruited, transported, transferred or harboured by force, fraud or deceived. The traffickers exploit their victims for profit. The exploitation could be in the form of sex, child abuse or forced labour. Human trafficking usually starts in origin countries or locations where extreme poverty, war, high crime rates are prevalent (usually Southeast Asia, sub-Saharan Africa, and Eastern Europe). Also, the proliferation of social media has made a lot of people vulnerable to human traffickers [4]. And this has resulted in human traffickers gaining more ground over law enforcement agencies. In Zimbabwe, social media has been used to promote human trafficking due to the absence of laws that will deal with such crimes committed online [5]. Human rights violations which are mainly as a result of human trafficking can be monitored through social media, which offers a vast base of crowdsourced information. The only challenge is quantifying, and interpreting the streams of unstructured data in a broader context which is still subject to investigation [6]. Currently, dynamic features of individuals or certain groups can be derived from various web sources. These sources are either websites, search engines, online job vacancy ads, mobile applications, e-commerce platforms, surveillance cameras, sensors, or social media platforms. These characteristics are data that could be obtained and used for detects of the activities of human traffickers using machine learning and data mining. Human trafficking is a multifaceted issue, it lacks sufficient data, adequate knowledge of the traffickers and the knowledge of system complexities. Therefore, it is an issue that requires knowledge of the key players and their relationships in a complex world, across-the-board database. Data mining and machine learning provide analytical tools and techniques for managing huge volumes of data. Natural language processing (NLP) and computer vision, which is a subset of machine learning has been used to detect and report advertisements suspected of trafficking [7]. The data that are used are mostly textual and image data. There are different types of images that are related to human trafficking, which could be pictures of victims taken in various locations including hotel rooms, passports of victims and even tattoos pictures on the body of victims. Given the best results can be achieved by combining features from various data sources [8]. However, pinpointing the precise location of a hotel room is a difficult and important task in the fight against human trafficking, since most victims are photographed in hotel rooms.

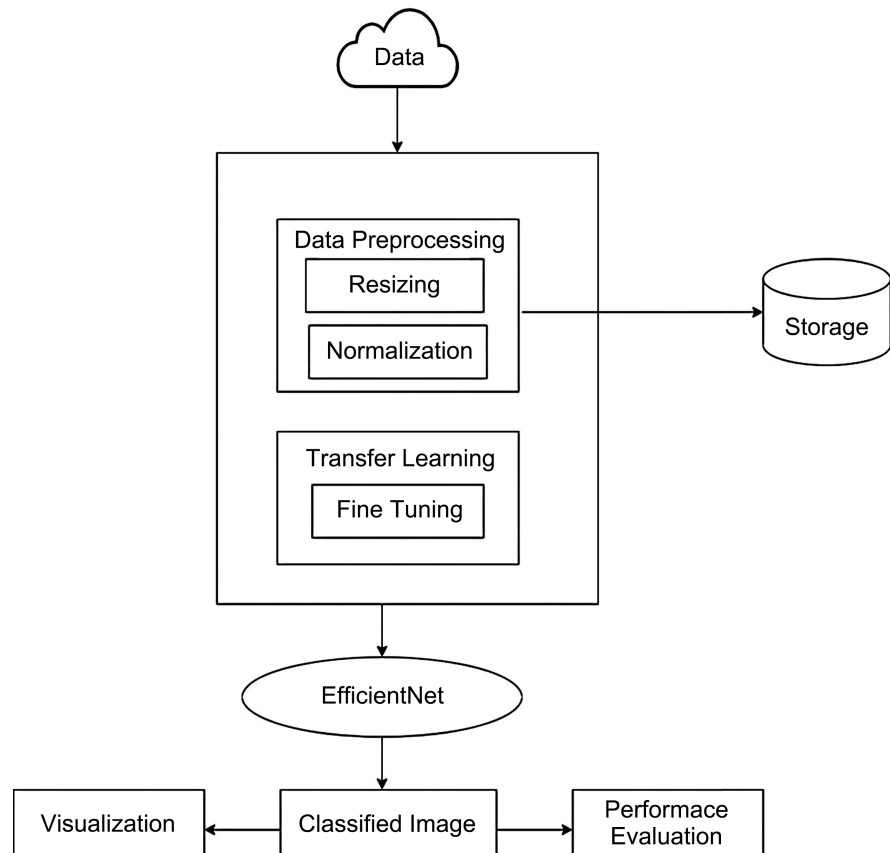
It is evident that 80% of victims are used for sexual exploitation and the remaining 20% are engaged in force labor [9]. However, majority of the works done were carried out for mostly texts, tattooed images and photographed victim images but little or no work has been done for hotel room images. Some of the papers related to the work were studied. In [10], the authors investigated the indication mining problem in the sphere of online sex advertising. FlagIt “Flexible and Adaptive Indicator Generation from Text” was created. It combined the advantages of a lightweight expert system and traditional semi-supervision (heuristic relabeling) with recently announced cutting-edge unsupervised text embedding. This is used to tag millions of words with markers that are significantly connected with human trafficking. In [11], a unique mathematical optimization approach for integrating network topology into content modeling was proposed. The usefulness of their suggested framework in detecting human trafficking-related information was proved by experimental findings on a real-world dataset using Network Structure Information (NSI). In [12], two sets of real-world datasets were utilized in training a developed language model that captures the structural properties of adult service ads, which are; the trafficking-10k and language model datasets. Dataset from the language model contains ads that were posted in the year 2017 on the Backpage.com website. In [13], crawled and extracted expert knowledge from the internet using a text-mining method was used to handle text data emanating from online content, by creating a fundamental architecture and essential modules such as information extraction, data cleansing and deduplication, and an expert recommendation model. In [14], a method for the development of a web scraper was proposed that uses R code program to find files on a website and then pull out and store the filtered data. Modules and algorithm used to navigate a website via links automatically was discussed. In this research, we will be utilizing a pretrained model which was not used in our related works of literature to achieve our aim.

## 2. Materials and Methods

In the proposed system, we addressed the issue of detecting human trafficking activities using hotel images. The architecture design of the system is depicted in **Figure 1** which consists of the dataset explained in section 2.1 that undergoes data pre-processing involving resizing and normalization. A storage unit stores the pre-processed image. Next is the transfer learning phase, EfficientNet model, and the classified image that is visualized and the model evaluation. The system will help to provide a more accurate model for predicting and controlling human trafficking.

### 2.1. Datasets

A mobile application called TrafficCam was used to develop the Hotel-ID dataset [15], and uploaded to Kaggle repository. The dataset is organized in folders, where each folder contains images and csv files. The train\_images folder contains



**Figure 1.** Architecture of the proposed system.

sub folders which are named in numbers which represent the chains of hotel images that are contained in them. The train\_image folder contain a total of 91 chain folders. The images (JPEG) have a max size of 512px. The folder also contain train.csv file which contains four columns namely, image which contains the image name, chain column which contains the chain a hotel image belongs to, hotel\_id which is the id given to the image and timestamp column contains the time the image was taken. The train.csv has a total of 97556 samples (rows). A sample of images and their id is shown in **Figure 2**.

## 2.2. Hotel Image Data Pre-Processing

Each image in the dataset needed to be pre-processed to be viable for the model. Resizing and normalization steps were taken during the preprocessing phase. Tensorflow was used for data pre-processing, by importing ImageDataGenerator class in keras using “from keras.preprocessing.image import ImageDataGenerator”.

1) Resizing: The images were downscaled from 512 to a max size of 224, this is to ensure the most suitable loss and training time tradeoff. The image was resized to 224 so as to ensure that the pretrained model could train on the image properly, as it is the predefined input size for the model. **Figure 3** shows an example of an original room image and the resized image.



Figure 2. Sample dataset of the proposed system.

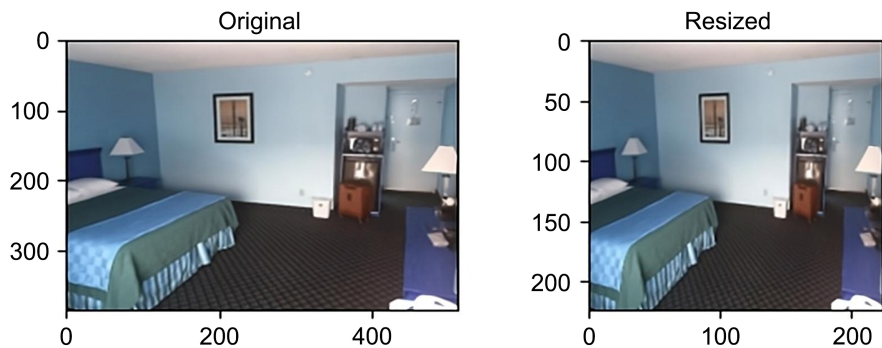


Figure 3. Resized image.

2) Normalization: To make certain all feature values are on the same scale and given equal weight, normalization is crucial. Pixel scaling is also known as normalization where image pixel values are scaled to the range 0(zero)-1(one) from 0-255 preferred for neural network models. Example of normalization done on the resized image can be found in Figure 4, and it was done using datagen = ImageDataGenerator (rescale = 1.0/255.0).

The ImageDataGenerator.flow\_from\_directory function was then used to load our data in batches (32) into the pretrained model for transfer learning.

3) Storage: The preprocessed image dataset is then stored for further use by researchers that want to make use of the preprocessed dataset to develop more models.



### 2.3. EfficientNet Model

EfficientNet model is a convolutional neural network architecture that makes use of a compound coefficient when scaling. The model scaling approach standardizes the breadth, depth, and resolution of a network using predetermined set of scaling coefficients, alternatively to the variable scaling which is common practice.

Compound scaling works on the premise that a bigger input picture will demand a network with more layers to broaden its receptive field and more channels to pick up finer details present in the larger image. In this study, EfficientNetB0 architecture was fine-tuned, which is explained in our experiment section. The EfficientNetB0 was developed by using a multi-objective neural architecture search that optimizes both floating-point operations and accuracy [16]. In our work, the Imagenet weight was added to the EfficientNetB0 baseline network.

Figure 5 shows the EfficientNetB model architecture.

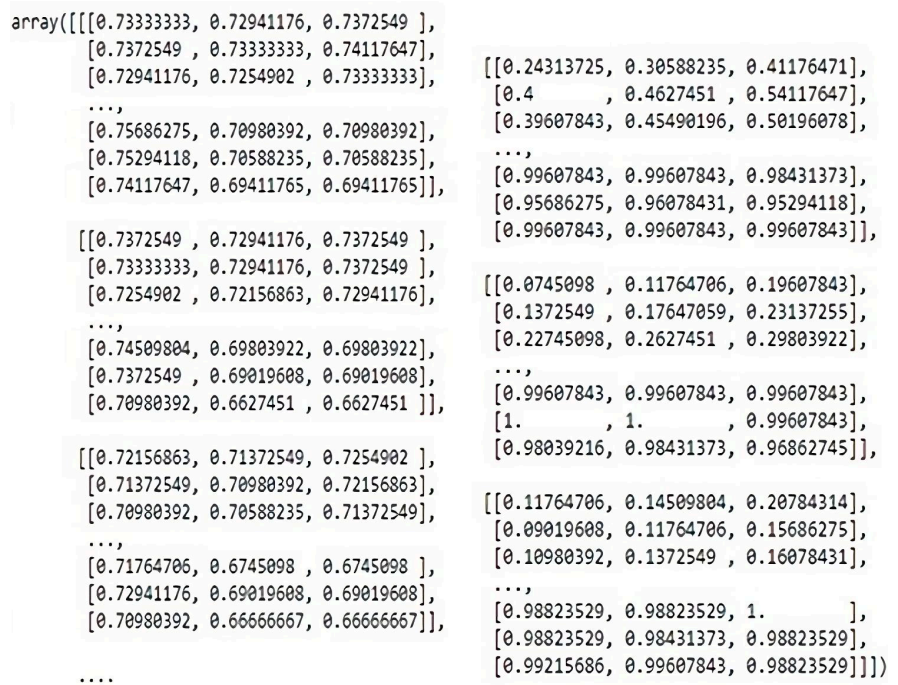


Figure 4. Normalized image output.

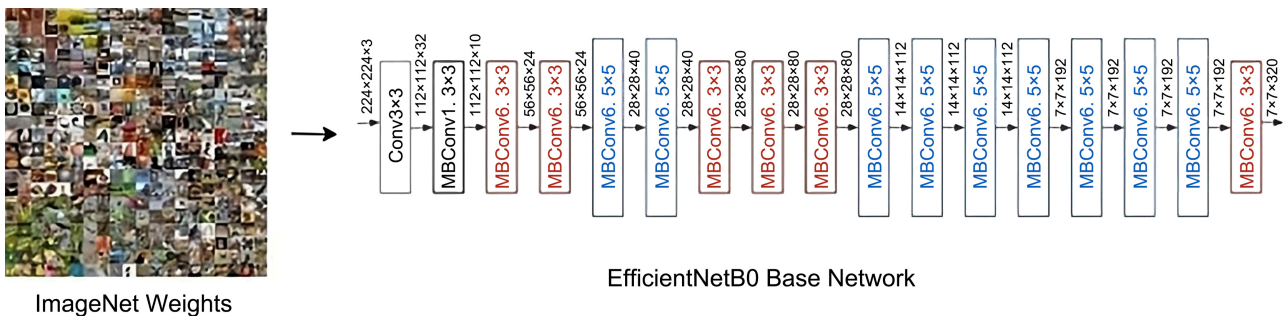


Figure 5. ImageNet weights & EfficientNetB0 model.

## 2.4. Transfer Learning Phase

Models have been developed and trained using some of the world largest datasets, which includes MNIST, ImageNet and Wikipedia corpus etc. These trained models were used for specific tasks, using these models as the starting point for tackling another similar problem is known as transfer learning. A pre-trained model can either be used as a starting point for model training or it can be used as feature extraction when developing another model. A pretrained EfficientNetB0 model was used as our starting point, by removing the top layers of the model which is specific to the problem domain it was trained for and replacing these top layers with our dataset and training the model along with the learned weights of the model.

## 3. Experiment

This section entails the implementation of the model's theoretical design in a practical manner. The model is conceived using Object-Oriented methodology and implemented using the Python programming language and Tensorflow library tools.

### 3.1. Data Description

The dataset that was used in this research is made up of a folder with sub-image folders and a CSV file. The image folders are ninety-one in number and contain images of hotel rooms, and the folders are named numerical ranging from 0-91. The numerical names of the folders represent the hotel chain a particular hotel room image belongs to. The CSV file that comes along with the image folders is named train.csv, which contains information about each image, including the time it was taken with a total of 97556 samples (rows). **Table 1** shows the features descriptions.

### 3.2. Developing the Model

The EfficientNetB0 model contains eleven million parameters, however, not all of them were utilized as discussed in this section A hotel chain model is developed

**Table 1.** Features Descriptions.

S/N	Attribute	Description	Dtype
i	Image	The image ID.	Object
ii	Chain	An ID code for the hotel chain. A chain of zero (0) indicates that the hotel is either not part of a chain or the chain is not known. The number of hotels per chain varies widely.	Int64
iii	Hotel_id	The hotel ID.	Int64
iv	Timestamp	When the image was taken. Provided for the training set only.	Object

in this study using a pretrained EfficientNetB0 based on transfer learning and fine-tuning to enhance the prediction accuracy. The proposed model consists of the max-pooling, and classification layers combined together. The feature extractors are made up of  $conv3 \times 3, 128$ ,  $conv3 \times 3, 256$ ,  $3 \times 3, 256$ ,  $conv3 \times 3, 512$ ,  $conv3 \times 3, 1024$ , max-pooling layer of size  $2 \times 2$ , and a Softmax activator. Using an input image size of  $224 \times 224 \times 3$ , the convolution and max-pooling algorithms produced 2D planes called feature maps. And in four different map sizes:  $126 \times 126 \times 32$ ,  $61 \times 61 \times 32$ ,  $28 \times 28 \times 64$ , and  $26 \times 26 \times 128$ .

Fine-tuning was carried out by removing the top layer of the EfficientNetB0 pre-trained model by not including it. For freezing the layer `layer.trainable = False` was used, indicating internal state during training will remain unchanged. The trainable weights of the model will remain unchanged during `fit()` process, and updates to its state will not be executed. The imagenet weight was used as the base during training because it is one of the most diverse datasets that performs well with EfficientNet and have prior knowledge about basic shapes as stated by Tan & Le [17].

### 3.3. Model Training

Training the model entails feeding the pre-trained model with training data of hotel images. This study made use of 5000 image datasets from three hotel chains. The dataset was split into training and validation sets. While 20% was used for the validation set, 80% was used for training. The model executed for twenty-five (25) epochs with thirty-two (32) batch sizes. Total parameter is 6,060,589, trainable parameter being 2,004,618 while non-trainable parameter is 4,055,971.

## 4. Results Discussion

When dealing with classification performance, the model is evaluated to identify how well data is classified by a model. The confusion matrix was used for measuring the performance of the model. Confusion matrix basis include precision, recall, and F-measure.

**Accuracy:** This is the percentage of correctly classified observations. The higher the score, the more effective the classification model.

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

**Precision:** This is the percentage of observations that the classifier correctly classified as positive data in the evaluated group (positive).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

**Recall:** This is the percentage of observations labeled positive and classified correctly.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

**F-Measure (F1-Score):** This shows the balance between recall and precision.

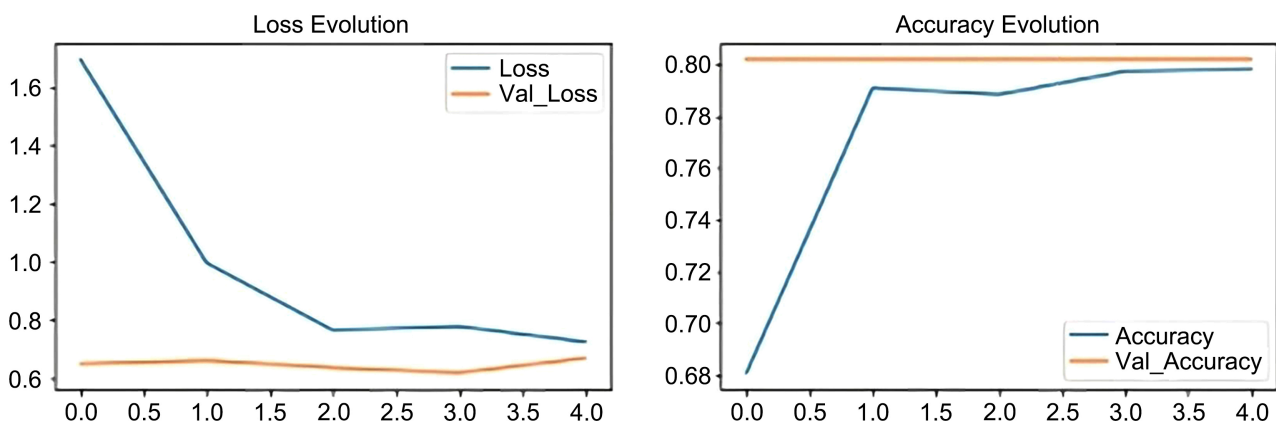


$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The developed model gave an accuracy score of 0.802 which indicates that our model predicted 80.2% of our data as correct when it was tested. Nevertheless, when it comes to an unbalanced dataset like ours, it is advisable to make use of precision and recall to evaluate the model. By using precision which quantifies the amount of positive class predictions truly belonging to the positive class as an evaluation metric, the system has a precision score of 0.802. Furthermore, recall measures the proportion of accurate class predictions made from all positive examples in the training set. Our proposed model had a recall value of 0.802 which clearly shows that all relevant images were retrieved by our system for classification, but it says nothing about how many irrelevant images were also retrieved.

The F1 score, also referred to as the F-measure, is a method used to evaluate recall and precision simultaneously. Due to the fact that outstanding precision can be acquired at the expense of an awful recall, and conversely, terrible precision with an outstanding recall. F1 score enables the expression of both concerns with a singular score. A poor F1 score is 0.0 whereas a perfect F1 score is 1.0. The proposed model F1 score is 0.802 which is close to 1.0. The performance of a classifier is directly proportional to the values of precision and recall. A higher recall indicates that a larger proportion of positive samples are correctly identified, while a higher precision indicates a lower number of false positives. The utilization of micro averaging in aggregating metric scores across all classes is advantageous as it mitigates the risk of generating inaccurate outcomes caused by an imbalanced distribution of the test set.

This result demonstrates that the proposed model has the capability to effectively classify hotel chains. The result in **Figure 6** shows the loss and accuracy. **Figure 7** depicts the precision and recall curves which show the progression of the results identified in the curves. Implying that the training and validation set class distribution is more appropriate for training the proposed model. Furthermore, it validates the effectiveness of transfer learning and fine tuning methods in improving the training process and model efficacy.



**Figure 6.** Accuracy and loss curve.

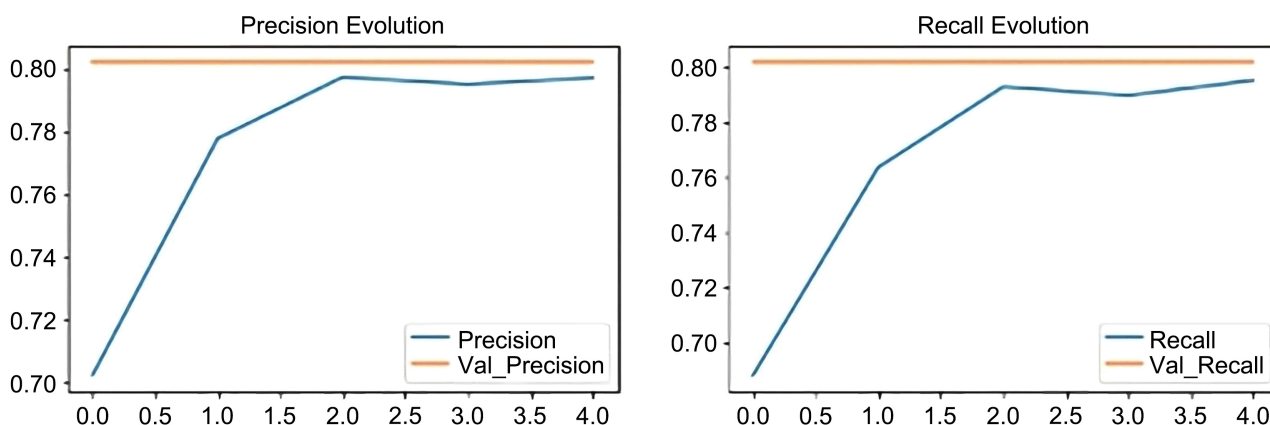


Figure 7. Precision and recall curve.

## 5. Conclusion

This study showed how to classify hotel room image data from a collection of hotel room image dataset. An EfficientNetB0 model based on transfer learning was implemented from scratch, this distinguishes it from other methods that largely depend on data augmentation and data pre-processing. Many of the hotels are independent or part of very small chains, where shared decor isn't a concern. However, the shared standards for their interior decoration for the larger chains mean that many hotels can look quite similar at first glance. Identifying the chain can narrow the range of possibilities. Hence, this study was carried out to develop the model that would be used for classifying selected chains. This study presented a deep learning model for the evaluation, visualization, and classification of hotel room chains based on transfer learning. The motivation of this study was to help law enforcement agencies address some of the challenges in identifying and locating hotel rooms that have been used by traffickers to advertise their victims. The outcome of the obtained accuracy rate of 0.802 shows that our EfficientNetB0 model with transfer learning and finetuning techniques performs well in classifying hotel chains. The model also achieved precision and accuracy of 0.802 respectively. The performance of the implemented model shows that the model could effectively classify hotel chains in practice. Thus, narrowing down the group of possible suspects to a size that can be effectively handled by a human investigator in order to explore every lead.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Amin, S. (2010) A Step towards Modeling and Destabilizing Human Trafficking Networks Using Machine Learning Methods. *AAAI Spring Symposium—Technical Report*, SS-10-01.
- [2] Broad, R. and Turnbull, N. (2019) From Human Trafficking to Modern Slavery: The Development of Anti-Trafficking Policy in the UK. *European Journal on Criminal*

- Policy and Research*, **25**, 119-133. <https://doi.org/10.1007/s10610-018-9375-4>
- [3] Da Silva Santos, M., Ladeira, M., Van Erven, G.C.G. and Luiz Da Silva, G. (2019) Machine Learning Models to Identify the Risk of Modern Slavery in Brazilian Cities. 2019 18th *IEEE International Conference on Machine Learning and Applications (ICMLA)*, Boca Raton, 16-19 December 2019, 740-746. <https://doi.org/10.1109/ICMLA.2019.00132>
  - [4] Dubrawski, A., Miller, K., Barnes, M., Boecking, B. and Kennedy, E. (2015) Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity. *Journal of Human Trafficking*, **1**, 65-85. <https://doi.org/10.1080/23322705.2015.1015342>
  - [5] Mugari, I. (2020) The Dark Side of Social Media in Zimbabwe: Unpacking the Legal Framework Conundrum. *Cogent Social Sciences*, **6**, Article: 1825058. <https://doi.org/10.1080/23311886.2020.1825058>
  - [6] Garcia, A.A., Britton, M.J., Doshi, D.M., De Choudhury, M. and Le Dantec, C.A. (2021) Data Migrations: Exploring the Use of Social Media Data as Evidence for Human Rights Advocacy. *Proceedings of the ACM on Human-Computer Interaction*, **4**, Article No. 268. <https://doi.org/10.1145/3434177>
  - [7] Tong, E., Zadeh, A., Jones, C. and Morency, L.P. (2017) Combating Human Trafficking with Deep Multimodal Models. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, **1**, 1547-1556. <https://doi.org/10.18653/v1/P17-1142>
  - [8] Vomfell, L., Härdle, W.K. and Lessmann, S. (2018) Improving Crime Count Forecasts Using Twitter and Taxi Data. *Decision Support Systems*, **113**, 73-85. <https://doi.org/10.1016/j.dss.2018.07.003>
  - [9] Barua, S. (2018) Human Trafficking in India. *International Journal of Trend in Scientific Research and Development*, **2**, 2453-2455. <https://doi.org/10.31142/ijtsrd12794>
  - [10] Kejriwal, M., Ding, J., Shao, R., Kumar, A. and Szekely, P. (2017) FlagIt: A System for Minimally Supervised Human Trafficking Indicator Mining. *ArXiv, Nips*, 1-6.
  - [11] Yang, Y., Hu, X., Liu, H., Li, Z. and Yu, P.S. (2018) Understanding and Monitoring Human Trafficking via Social Sensors: A Sociological Approach. <https://doi.org/10.48550/arXiv.1805.10617>
  - [12] Zhu, J., Li, L. and Jones, C. (2019) Identification and Detection of Human Trafficking Using Language Models. 2019 *European Intelligence and Security Informatics Conference (EISIC)*, Oulu, 26-27 November 2019, 24-31. <https://doi.org/10.1109/EISIC49498.2019.9108860>
  - [13] Xie, X., Fu, Y., Jin, H., Zhao, Y. and Cao, W. (2020) A Novel Text Mining Approach for Scholar Information Extraction from Web Content in Chinese. *Future Generation Computer Systems*, **111**, 859-872. <https://doi.org/10.1016/j.future.2019.08.033>
  - [14] Midhu Bala, G. and Chitra, K. (2021) Data Extraction and Scratching Information Using R. *Shanlax International Journal of Arts, Science and Humanities*, **8**, 140-144. <https://doi.org/10.34293/sijash.v8i3.3588>
  - [15] Kamath, R., Rowles, G., Black, S. and Stylianou, A. (2021) Hotel-ID to Combat Human Trafficking Competition Dataset. *Computer Vision and Pattern Recognition*, **1**, 1-5.
  - [16] Khan, Z., Shubham, T. and Arya, R.K. (2022) Skin Cancer Detection Using Computer Vision. In: Mandal, J.K., Hsiung, P.A., Sankar Dhar, R., Eds., *Topical Drifts in Intelligent Computing. ICCTA 2021. Lecture Notes in Networks and Systems*. Springer

ger, Singapore. [https://doi.org/10.1007/978-981-19-0745-6\\_1](https://doi.org/10.1007/978-981-19-0745-6_1)

- [17] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Computer Vision and Pattern Recognition*, 5, 1-11.