Advanced Weapon Detection and Classification Using Fine-Tuned Transfer Learning Models

MSc Research Project
Programme Name

Brandon Craig D’souza
Student ID: x23100125

School of Computing
National College of Ireland

Supervisor: Prof Abdul Qayum

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Brandon Craig D'souza
Student ID:	X23100125
Programme:	Msc Data Analytics **Year:** 2023-24
Module:	Msc Research Project
Supervisor:	Prof. Abdul Qayum
Submission Due Date:	16/09/2024
Project Title:	Advanced Weapon Detection and Classification Using Fine-Tuned Transfer Learning Models
Word Count:	6433 **Page Count** 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Brandon Craig D'souza

Date: 16/09/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Advanced Weapon Detection and Classification Using Fine-Tuned Transfer Learning Models

Brandon Craig D'souza
x23100125

Abstract

Safety and security are big problems in today's modern world. With the increasing number of criminal activities, automatic control systems have become very important for security purposes. One of the most serious activity is use of illegal weapons. Many systems today have a number of disadvantages as they tend to employ a lot of manpower in the monitoring and controlling of processes and also in managing of knowledge in this field. This project will present a weapon detection system by using Deep Learning (DL) models and a Flask-based web application. The dataset has been sourced from images.cv which includes labeled weapon images categorized into four classes: knife, pistol, rifle, and sword. This study uses different models which includes Convolutional Neural Networks (CNN), Xception, EfficientNet-B2 and EfficientNet-B2 with Attention Baseline. The performance of these models have been evaluated based on accuracy and F1-score. Among them EfficientNet-B2 with Attention Baseline has been achieved the highest accuracy of 95% and a macro average F1-score of 0.95 which has an increased performance and accuracy. A web app has been developed using Flask for a user interface, which allows users to upload weapon images and obtain real-time predictions. The application interface gives users an easy image upload and displays the classification results which includes the weapon type detected. This investigation will thus help the law enforcement and security personnel to easily identify and classify weapons quickly so as to increase safety of people and decrease the extent to which human intervention is required.

Keywords: Weapon Detection, Deep Learning, CNN, Xception, EfficientNet-B2.

1 Introduction

1.1 Background

Intelligent Surveillance System is a technological software which is designed to detect and classify the weapons from images with high efficiency (Ramya et al. , 2023). This system uses the EfficientNet-D0 network to categorize the images of weapons as it offers balance of speed and accuracy. The model of EfficientNet-D0 is well suited as it has high accuracy in the detection of images while at the same time the model size is relatively small compared to others. They are then used to predict on whether it is a fire arm or a knife among others. The system has been trained on large datasets having images of different types of weapons which allows it to learn and recognize any differences in appearance. This can be used in several fields like airports, at the border control and at other crowded places where it can increase security by identifying weapons that might be concealed in form of other objects.

1.2 Aim of the study

The purpose of this study is to enhance the intelligent weapon detection framework to employ DL approaches for classification between weapon pictures and to integrate this into a simple web interface. The quantitative assessment of the study will also involve comparing the performance of the above mentioned DL models: CNN, Xception, EfficientNet-B2 and EfficientNet-B2 with Attention Baseline to determine the best performing model for weapon classification. The study is achieved with a labeled dataset which uses knife, pistol, rifle, and sword images thereby ensuring accurate and reliable weapon detection. Another objective of the project will include developing a web application using Flask where users can upload weapon images and videos and the DL models could be demonstrated in practice by providing real-time predictions of the same. The intended benefit of the proposed method is to enhance the current security and surveillance systems by providing an accurate model of weapon detection and play a part in enhancing safety measures even if it is a basic form of threat exposure.

1.3 Research Questions & Objectives

1.3.1 Research Questions

The two research questions for this research are as follows:

1. Which deep learning model—CNN, Xception, EfficientNet-B2, or EfficientNet-B2 with Attention Baseline—achieves the highest accuracy and robustness in weapon classification?
2. What are the specific strengths and weaknesses of EfficientNet-B2 with Attention Baseline as compared to other models in weapon classification?

1.3.2 Research Objectives

There are several research objectives are as follows:

1. For weapon classification, it includes to design, train, as well as evaluate deep learning models namely CNN, Xception, EfficientNet-B2, and EfficientNet-B2 with Attention Baseline.
2. To evaluate and compare the performances using the given evaluation parameters for accuracy and F1-score to select the weapon detection model with the highest accuracy and reliability.
3. To create a convenient and functional web application based on the Flask tool for classifying weapons, based on the images or videos uploaded by a user.
4. To apply advanced techniques on the model to improve its detection ability and reliability to distinguish various categories of weapons.

2 Related Work

2.1 ML models in weapon image classification detection

In the domain of weapon image classification and detection, different machine learning techniques have been used to increase accuracy and performance. So there is a study which is given by (Kannur, 2022) who has proposed a novel method using hierarchical support vector machines (HSVM) for weapon identification based on images of repeated stab wound patterns from homicidal cases. The approach includes segmenting the region-of-interest in the images using a transition region-based segmentation algorithm which is been followed by extracting texture, shape and size features from the segmented wounds. The methodology consists of three main stages: generating non-overlapping segments from the wound patterns, extracting features from these segments and identifying the wound patterns along with their corresponding weapons. This process uses HSVM as a classifier for multiple classes by giving a good and accurate alternative to traditional forensic pathology methods. One of the main challenges which is been faced in this study is the accurate segmentation and feature extraction from complex wound patterns which is very important for reliable weapon identification. Also ensuring the robustness of the classifier across different wound types and weapon classes gives another important challenge. Support Vector Machines (SVMs), Decision Trees and other techniques have been used for evaluation which shows a good identification accuracy of 96.71%. But the limitations are that the model was trained only on 250 images. Another study given by (Gupta et al., 2024) who introduced an advanced weapon recognition system using a dual-framework approach includes SVM-P for robust detection and classification. The study has some challenges in balancing the performance across different weapon types by dealing with class imbalances and ensuring the model's robustness. The model has achieved good results like for Class 1 weapons an accuracy having 89.86%, a recall having 97.01% and an F1 score having 93.30% with a good total accuracy of 99%. Class 2 weapons showed a precision score having 91.61%, a recall score having 95.78% and an F1 score having 93.65% with a final accuracy level of 98%. Even though the model was tested for the adaptability to different environmental conditions the study highlights challenges in maintaining its accuracy in different lighting conditions and obstructions. Whileproposing an advanced automated model for weapon detection in closed-circuit television (CCTV) videos (Kalla and Suma, 2022) suggested to increase public safety due to rising criminal activities. The main challenge lies in having texture features while optimizing them to reduce system complexity and training time without compromising any accuracy. Results have showed that the HIPSO-SVM model achieves good accuracy rates of 95.34% and 98.60% on the YouTube and Gun movies databases. The paper given by (Gelana and Yadav, 2019) proposed an automated system for detecting "Active Shooters" carrying non-concealed firearms and alerting CCTV operators to any type of threats both visually and audibly. The main challenge in this study is ensuring the system's accuracy and reliability in real-time detection across different environments and changing video quality. Also the system must process and analyze huge amounts of video data without any computational resources. Despite these challenges the proposed method has achieved a high classification accuracy of 97.78%. The system's automated detection capability reduces the reliability on human operators which is thereby improving surveillance performance and response times during the incidents. Another study given by (Rahman et al., 2020) who proposed a ML based approach to increase accuracy and performance in identifying shooting errors and predicting shooter performance. The proposed methodology comprises two main modules. The first module is using different algorithms to determine that a random forest classifier that best recognizes error patterns by achieving an average accuracy of 96.8%. The second module uses an AdaBoost classifier to predict shooters' scores and performance with an accuracy of 69%. The main challenge is the lack of real-time data which has some previous DL solutions impractical for real-world things. The study given by (Rajesh et al., 2020) mainly focused on the important role of timely and accurate diagnosis of leaf diseases in agriculture to prevent any losses in productivity and the quality of agricultural products. The proposed system is using a decision tree algorithm to identify and classify leaf diseases by increasing detection accuracy and reducing the time required for diagnosis compared to existing methods. The main challenge solved by this system is the need for an accurate, fast and reliable technique to detect early signs of diseases on plant leaves.

2.2 DL models in weapon image classification detection

These models address many issues like the variety of datasets, real-time processing and detection in complex environments. Some of the methods such as sliding window, region proposal, and object detection frameworks have been applied in order to enhance the accuracy of DL in weapon detection and all of them have demonstrated the capability of DL in weapon detection. (Ahmad et al. ,2021) developed a new IADC system using CNNs to enhance security by automating the monitoring process. The system has been developed to detect people who are holding weapons through analysis of videos from the installed CCTV cameras in offices, houses and public places. The main problem that is solved by the system is the problem that a single operator has when he tries to monitor several CCTV feeds. The proposed IADC system is using CNNs for analyzing visual imagery to detect firearms. The results have revealed that the proposed system yields better results than the other models such as VGGNet, OverFeat-1, OverFeat-2, OverFeat-3 in terms of detection accuracy and dependability. One other study which is provided by (Kaya et al. , 2021) who develops a new model for identifying and categorizing 7 sorts of weapons using deep learning technique with the help of VGGNet. This model has been developed to detect assault rifles, bazookas, grenades, hunting rifles, knives, pistols, and revolvers. The performance of the proposed model is then compared with the existing models like VGG-16, ResNet-101 and ResNet-50 for classification. The proposed model has achieved the success accuracy of 98.40% and that of the VGG-16 model has been accomplished to 89.75%, and in case of ResNet-50 model, it has been achieved 93.70% and the ResNet-101 model's has achieved 83.33%. Another study by (Bhatti et al. , 2021) who used two primary approaches: sliding window/classification and region proposal/object detection. Certain algorithms like VGG16, Inception-ResnetV2, Inception-V3, Faster-RCNN Inception-ResnetV2 (FRIRv2), SSDMobileNetV1 and all of them were experimented. YOLOv4 was the best by attaining an F1-score of 91 % and the mean average precision of 91.73%. (Arif et al., 2022) have proposed an approach which is EfficientNet algorithm. The studies have indicated that the EfficientNet algorithm has obtained the accuracy of 98.12%. In the paper titled: 'Audio data for violence detection in environments where video surveillance is not feasible: A review of the state-of-the-art' by (Duraes et al. , 2023), the authors discuss the application of audio data for violence detection where using video surveillance is impossible. By converting the audio into Mel spectrogram images the study has been able to demonstrate that EfficientNet-based models especially EfficientNetB1 has a good accuracy of 95.06% of the time in identifying violent audio instances. Likewise, (Velayudhan et al. , 2022) have been conducted the study using deep low-rank features and few-shot learning for X-ray imagery analysis. It also enhances the mean intersect over union and the mean average precision by 19.85% and 8.33% over old methods. Subsequently, in (Kalvankar et al. , 2020), EfficientNetB5 is used for the classification of Galaxy Morphology where a good accuracy of 93% is obtained. 7% and F1 score of 0.8857 in classifying the images of galaxies from the Galaxy Zoo 2 competition. Finally, (Hnoohom et al. ,2021) addresses and manages the problem of weapon detection in surveillance systems using ARMAS and IMFDB datasets with the help of TensorFlow's Object Detection API. The Faster R-CNN Inception V2 model stands out by giving a mean Average Precision of 0.540 and AP scores of 0.793 and 0.627 at IoU thresholds of 0.5 and 0.75.

Table 1: Comparison Table: ML and DL Models in Weapon Image Classification Detection

Study	Proposed Approach	Methodology	Results	Strengths	Weaknesses/Limitations
(Kannur, 2022)	Hierarchical SVM	Segmentation, feature extraction, HSVM classifier	96.71% accuracy	Better than traditional methods; accurate identification	Complex segmentation; robustness across weapon types
(Gupta et al., 2024)	Dual-framework SVM-P	Feature extraction, performance profiling on precision, recall, F1-score	Class 1: 99%; Class 2: 98%; overall metrics ~93%	High accuracy for various classes	Balancing performance across different weapon types
(Kalla and Suma, 2022)	HIPSO-SVM for CCTV weapon detection	Data collection, feature extraction using AlexNet, ResNet 18, SIFT, HIPSO, SVM	95.34% (YouTube) and 98.60% (Gun movies)	Effective feature optimization; high accuracy	Complexity in feature optimization; training time
(Gelana and Yadav, 2019)	Automated "Active Shooter" detection	Feature extraction, real-time alerting	97.78% classification accuracy	Improves surveillance performance and response times	Real-time detection accuracy; processing large data
(Rahman et al., 2020)	ML-based error pattern recognition and performance prediction	Random forest for error patterns; AdaBoost for performance prediction	Error patterns: 96.8%; Performance prediction: 69%	High accuracy in error pattern recognition	Lack of real-time data; performance prediction challenges
(Rajesh et al., 2020)	Decision tree for leaf disease detection	Decision tree algorithm for leaf disease identification	Enhanced detection accuracy; faster diagnosis	Accurate, fast, and reliable for agricultural productivity	Requires large and well-annotated datasets
(Ahmad et al., 2021)	CNN	CNNs for analyzing visual imagery from CCTV	Better than VGGNet, OverFeat models; high accuracy	Effective for automated monitoring	Implementation complexity; diverse training data needed
(Kaya et al., 2021)	VGGNet-based model	Detection and classification of 7 weapon types	98.40% accuracy; higher than VGG-16, ResNet models	High accuracy; effective for diverse weapon types	Dependence on large, well-annotated datasets

(Bhatti et al., 2021)	YOLOv4, Faster R-CNN, SSDMobileNet V1	Sliding window/classification and region proposal/object detection	YOLOv4: F1-score of 91%; mAP of 91.73%	High precision and recall; efficient real-time processing	Handling occlusions and angle variations
(Arif et al., 2022)	EfficientNet	Weapon detection with EfficientNet	98.12% accuracy	High accuracy; improved performance	Requires extensive training; dataset limitations
(Duraes et al., 2023)	EfficientNetB1, MobileNetV2	Audio data converted to mel spectrogram images	EfficientNetB1: 95.06% accuracy	Effective for audio-based violence detection	Less effective with MobileNetV2; data representation needed
(Kalvankar et al., 2020)	EfficientNetB5	Galaxy Morphology Classification	93.7% accuracy; F1 score of 0.8857	Effective for galaxy classification	Limited to galaxy morphology; generalization issues
(Hnoohom et al., 2021)	Faster R-CNN Inception V2	Object Detection API with TensorFlow	mAP: 0.540; AP scores: 0.793 (IoU 0.5), 0.627 (IoU 0.75)	Good precision for weapon detection	Challenges with IoU thresholds; dataset limitations

3 Research Methodology

3.1 Methodology

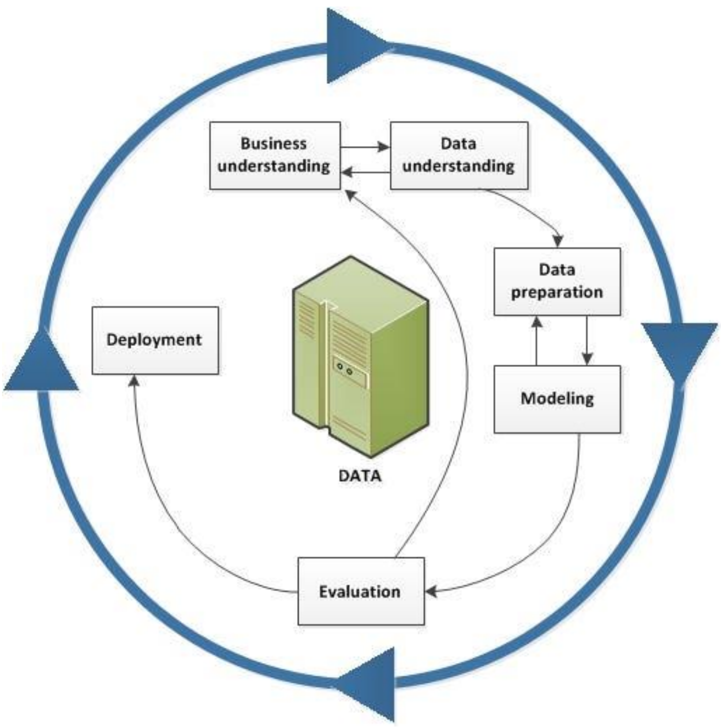

Figure 1: CRISP-DM architecture (IBM (2021))

CRISP-DM (Cross-Industry Standard Process for Data Mining) as seen in Figure 1, provides a structured framework for developing the Weapon Image Classification System by guiding the entire project with a systematic approach. It ensures with understanding, preparation, and modelling of data leading to good evaluation and deployment of the classification model. There are six phases which are as follows:

1. **Business Understanding Phase:** In the Business Understanding phase the main aim is to design an image classification model that will be able to identify and categorize knives, pistols, rifles, swords and other associated weapons (Sivakumar et al., 2024). This system will help enhance security and surveillance applications, as well as the enhancement of fire arm safety training simulation and identification of right measures on violence.

2. **Data Understanding Phase:** The Data Understanding phase includes examining the dataset that is having labelled images of weapons. This phase begins with data collection by ensuring that the dataset is good and representative of different weapon types. Describing the data includes understanding the number of images per class—2500 images each for knives and pistols, 1928 for rifles and 590 for swords.

3. **Data Preparation Phase:** The Data Preparation phase starts with loading the dataset of weapon images with OpenCV. The next step includes pre-processing the images which includes labelling the data to assign class labels (knife, pistol, rifle, sword), resizing images to a consistent dimension and converting categorical labels into numerical values using LabelBinarizer. Addressing class imbalance is very important in this phase so SMOTE (Synthetic Minority Over-sampling Technique) is been used to balance the dataset by ensuring that the models are not biased towards any class Chawla et al., 2002).

4. **Modelling Phase:** The Modelling phase involves choosing and applying various DL architectures that are effective in categorizing images of weapons. The main purpose of this phase is to develop models capable of learning from the prepared sample and making high-quality type of predictions on new, unknown data. It is employing popular models like CNNs, Xception, EfficientNet-B2 and Inception with an attention mechanism.

5. **Evaluation Phase:** The Evaluation phase is to determine the performance of the trained models to be aware of how accurate and reliable the models are in the classification of weapon images. This phase mainly involves assessment of the models to check whether they are capable of delivering good results on unseen data and to check for any areas of improvements. Among the used evaluation models some of them are Classification Report, Confusion Matrix, Sensitivity and Specificity.

6. **Deployment Phase:** The Deployment phase will be to integrate the trained and evaluated models into a GUI to apply the models in practical scenarios in security and surveillance systems. Thus for this project, a Flask based web application will be developed to enable users to test the classification system on image inputs.

3.2 Project workflow

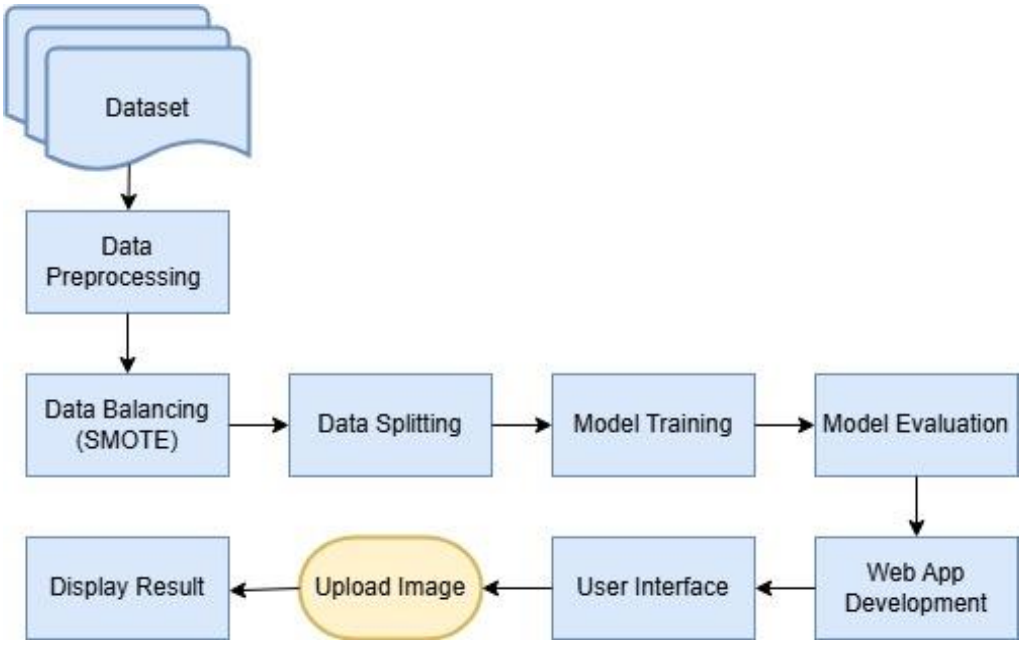

Figure 2: Proposed Workflow

Figure 2 shows the projects overall workflow. All these steps helped to achieve the end results for the research. These steps are explained below with some of them explained in other sections of the report.

3.3 Data Visualization

Figure 3 shows the Exploratory Data Analysis (EDA) of the weapon image dataset which shows the distribution of images across different classes such as knife, sword, rifle, and pistol.

Figure 3: EDA and Data Visualization

3.4 Libraries Imported

In the project for weapon image classification, there are many libraries which have been used for different tasks. For image processing and resizing OpenCV (cv2) is been used, on the other hand NumPy and Pandas are used for the data handling and analysis. Another useful library in scikit-learn is the data pre-processing toolbox which contains LabelBinarizer for conversion of categorical data, train_test_split for splitting of data and evaluation measures such as classification_report and confusion_matrix. Here, imbalanced-learn (imblearn) is utilized for handling the class imbalance problem with the help of SMOTE. TensorFlow/Keras is used in building and training models with keras for defining and managing deep learning models which has layers and call-backs. Specifically, the models used are EfficientNet and Xception from efficientnet.keras and tensorflow.keras.applications have been utilized for their pre-trained architectures and transfer learning mechanisms. The two used for data visualization are Matplotlib and Seaborn. Pickle has been used in the process of model serialization and deserialization.

3.5 Data Pre-processing

Some of the key steps of data pre-processing that are followed for weapon image classification system are as follows: Data Labelling is been done using a LabelBinarizer which transforms the categorical class labels (for example ‘knife’, ‘pistol’, ‘rifle’, ‘sword’) into numerical form which are understood by the ML models. This process is very important for converting class labels into a form that are understandable by the DL models (Chaki et al. (2021)). The LabelBinarizer is then stored to disk to maintain its usage in the future for model prediction through Pickle. Image resizing is done using OpenCV, which is used to make the dimensions of all the input images to be of the same size since this is very crucial when it comes to training the model. This resizing is important because it cuts down the computational load and also allow the model to learn from the images of good size in a proper way (Hangün and Eyecioğlu, 2017). Also images are changed to an array format using img_to_array from Keras which enables them to be input to a neural network.

3.6 Data Splitting (Training and Testing the Model)

To avoid the over-fitting of the weapon image classification model the data is split into training, validation and test data in the ratio 8:1:1. First of all the dataset is divided into training and testing with the help of train_test_split function with 10% of test data as the final model evaluation is carried out using this set. This leads to the formation of training set X_train, and a testing set X_test, their respective labels being y_train and y_test. Also the training set is then divided into training and validation subsets then using train_test_split with a test size of 10% to form the validation set.

3.7 Data Balancing with SMOTE

Before applying SMOTE the class distribution is quite imbalanced as seen in Figure 4, especially for the sword class which contains only 500 images out of the total. This was done using count plot in which the class distribution was represented with the sword class colored red to signify its under-representation. To solve this problem, the dataset applies the balancing procedure with SMOTE. First, the feature set X was transformed from the format of 2D matrix into 2D vector where each image is represented as a vector. SMOTE was then used to create synthetic samples in the sword class in the hope of having equal number of samples between the classes (Chawla et al. , 2002). The class distribution was again checked by creating count plot where it was good to see that the count of sword class has been increased from 500 to 2500 which has balanced the dataset in a good manner as seen in Figure 5.

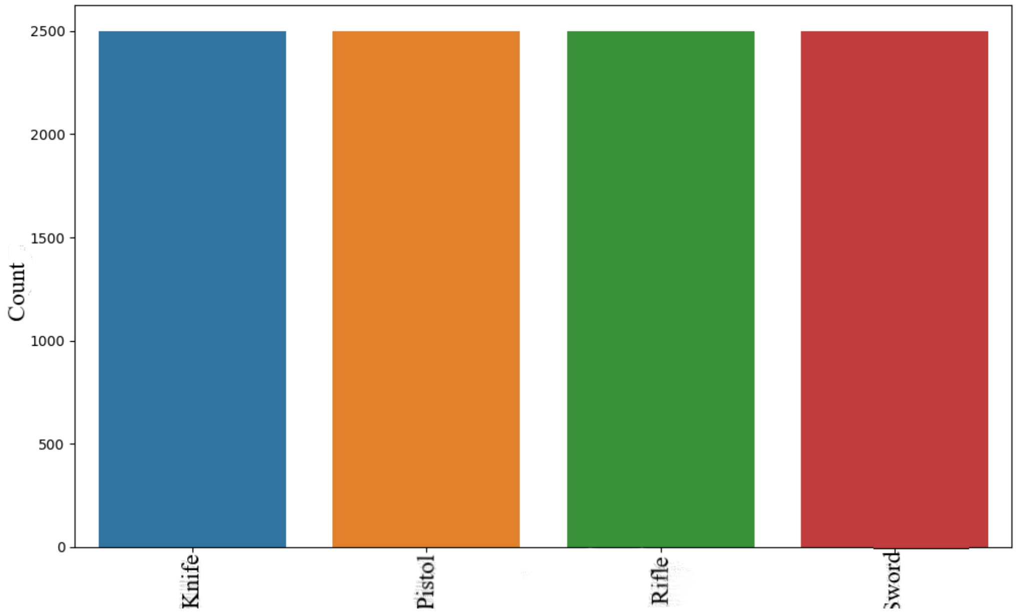

Figure 5: Data is Balanced using SMOTE

3.8 Dataset Description

The dataset which is used for the weapon image classification project is sourced from images.cv which is mainly designed for advanced object detection and image classification tasks within security and surveillance. This dataset includes labeled images of various weapons by giving a robust resource for increasing security systems, firearm safety training simulations and developing good measures for violence prevention. Each image in the dataset is labeled to ensure high accuracy in model performance. The dataset features four distinct class labels: 'knife,' 'pistol,' 'rifle,' and 'sword.' The original dataset is having a total of 3446 images for knives, 3710 for pistols, 1928 for rifles, and 590 for swords. To create a balanced and manageable training set a maximum of 2500 images per class is selected which results in 2500 images each for knives and pistols, 1928 images for rifles and the full 590 images for swords.

3.9 ModTraining

The first step of model training involved the preparations of input data, which involved resizing of images, normalization of pixel intensity, and data augmentation to apply additional transformations like random rotations, zooming and flipping to name but a few. In order to overcome the imbalance in classes, the Synthetic Minority Over-sampling Technique (SMOTE) was used where the training set was divided into 80% training, 10% validation and 10% test set. Transfer learning was used through using pre-trained models like Xception, Efficient Net-B2, and Efficient Net- B2 with the Attention Baseline, this means the models had to rely on the features that were learned from large datasets. Each of these models was then tweaked with layers added in order to optimize it to the weapon classification task. The training, which adopted the Adam optimizer on the initial learning rate of 0.001 therefore categorical cross-entropy was selected as the loss function. Networks were trained for 10 epochs and early stopping technique was used to avoid over training. The use of Batch normalization and dropout layers was to help with the generalization of the data. The models were trained while keeping track of the accuracy and F1-score performance indicators.

EfficientNet-B2 with Attention Baseline had maximum accuracy that was 0.95% and F1-score of 0.95. Among all these models, this one provided the best performance and was adopted in Flask-based web application for real-time detection and classification of weapons.

4 Design Specification

The image classification system is of a quantitative type. It includes pixel values and other quantitative measures. The evaluation of the model is also done with the help of quantitative measures like accuracy, precision F1-score and recall which are numerical measures. But there can be certain qualitative elements in the project. For example, for classification in the models, there can be some qualitative analysis such as which features of the image the model focuses on. Misclassification also involves quantitative analysis in order to determine the weaknesses or bias of the model.

This section also indicates the procedures, design and structure which have been employed in the development of the weapon image classification system and the total structure of the system. The focus of the system is on several DL models to employ pre-trained architectures to improve feature extraction and classification performance. These comprise of CNNs, Xception, EfficientNet-B2, and EfficientNet-B2 with Attention. These architectures are chosen for their good performance in image classification tasks with EfficientNet and Xception by providing SOTA accuracy having advanced network designs and pre-trained weights. EfficientNet-B2 has gained a reputation by applying the compound scaling method in order to adjust the network depth, width and resolution. The attention mechanisms in the EfficientNet-B2 model are beneficial for the network which primarily targets some features only by enhancing the classification accuracy.

The Xception model which incorporated depth-wise separable convolutions provided higher performance as it is able to identify various patterns in images. SMOTE has been used to solve the problem of class imbalance where new samples are created for the minority classes to improve training. The dataset is then divided into training, validation and test datasets in the proportion of 80:10:10 for using the model evaluation and tuning. The system architecture outlined has many important elements like data pre-processing that involves resizing and normalization, model training using different architectures and evaluation of the model using metrics.

5 Implementation

CNN, EfficientNet and Xception models were employed in this project since they are known to be efficient on image classification tasks. EfficientNet-B2 is famous for having efficient scaling that means that a depth, width and resolution of a model is in balance and due to these parameters of one of the most suitable models for this research.

5.1 CNN Model

In the weapon image classification system, the CNN model has been useful in properly identifying and categorizing images of weapons (Kaya et al., 2021). CNNs are mainly used for processing image data by learning multi-layered spatial features with the help of backpropagation. This model applies convolutional layers for instance to detect small features like edges, textures and shapes (Baker et al., 2018) which are very helpful when it comes to distinguishing between the various categories of weapons for instance knives, pistols, rifles and swords. These features are then combined having with pooling layers for reduction of dimensionality and computation while having important information.

5.2 Xception Model

The Xception model is been used in this project of weapon to increase the accuracy and performance of identifying different weapons. Its role is to extract and learn complex features from the weapon images by allowing it to distinguish between classes such as knives, pistols, rifles and swords with high precision.

5.3 Efficient Net B2 Model

EfficientNet-B2 belongs to the EfficientNet series (Tripathy et al. , 2023) which models use the compound scaling technique that scales the network depth, width and resolution simultaneously. This approach enhances feature extraction capacity while at the same time reducing the computational load. In this project the function of EfficientNet-B2 is to identify images of weapons and segregate them into classes, which include knives, pistols, rifles, and swords with the help of the architecture of the model. The weights of the model have been initialized from good training of large datasets which will enable the model to quickly learn the weapon classification task.

5.4 Efficient Net B2 with Attention Baseline Model

In the EfficientNet-B2 with Attention Baseline model, this attention mechanism will improve the model's ability to handle complex and different weapon images by showing important features and delete irrelevant ones. By using pre-trained EfficientNet-B2 weights and combining attention baseline the model achieves higher accuracy and robustness in distinguishing between weapon categories.

5.6 Graphical User Interface (GUI) For Testing

Figure 6 depicts the developed web app for the weapon detection system which has been implemented using DL for the prediction part and Flask for the back end. The interface provides an upload button that enables the users to upload an image or video file. Once it is uploaded the users can click on the "Click to Predict Weapons" button to begin the weapon detection. Next, the image is passed through the trained DL models and the output contains the identification of the detected weapon type, for instance, "knife" accompanied by the image.

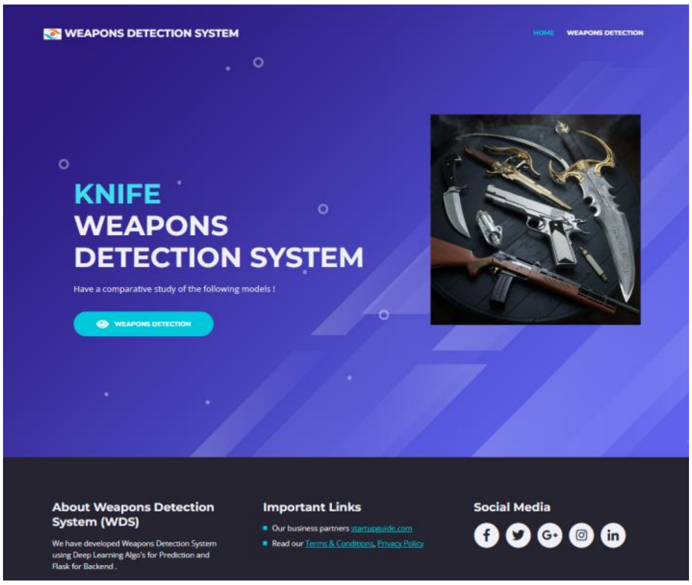

Figure 6: Web App Interface

6 Evaluation

6.1 CNN Model

Figure 7 depicts the training accuracy and loss of the CNN model over the number of epochs. The accuracy plot illustrates the training and validation accuracy of the model giving an indication of the model's progress during the training phase. The loss plot demonstrates training and validation loss by indicating as to how effectively the model is able to reduce mistakes over time. The graphs depict an improvement in the accuracy and a decrease in the loss graph which is an indication the model is training and learning well.

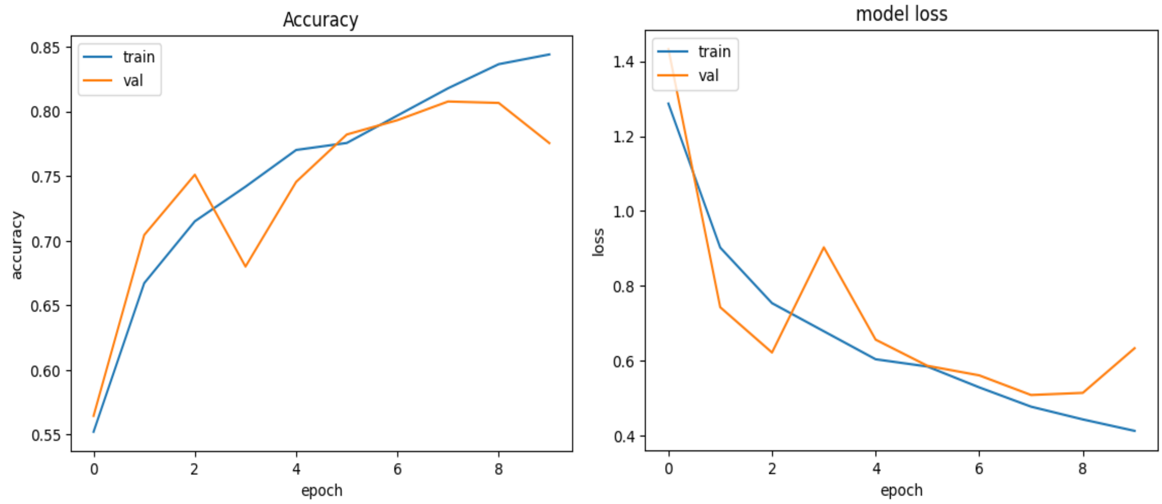

Figure 7: Accuracy and Loss Graph

Figure 8 presents the confusion matrix of the CNN model. The matrix shows the actual classes (rows) versus the predicted classes (columns). For example, the model has been correctly classified 234 images of class 0, misclassifying 18 as class 1, 6 as class 2, and 5 as class 3. Similarly, class 1 had 194 correct predictions with some misclassification in other classes.

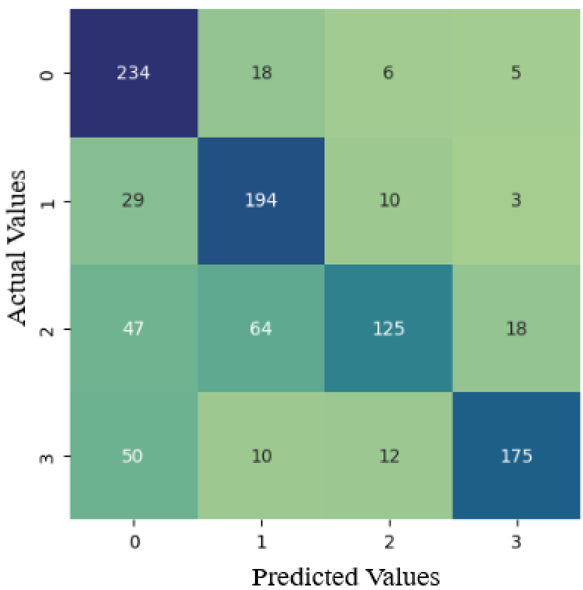

Figure 8: Confusion Matrix

6.2 Xception Model

Figure 9 is showing accuracy and loss graph for the Xception model. The plots suggest that the model is training and learning well as there is a steady increase in in the accuracy and a decrease in the models training loss. However, there can be seen some degree of overfitting in the model.

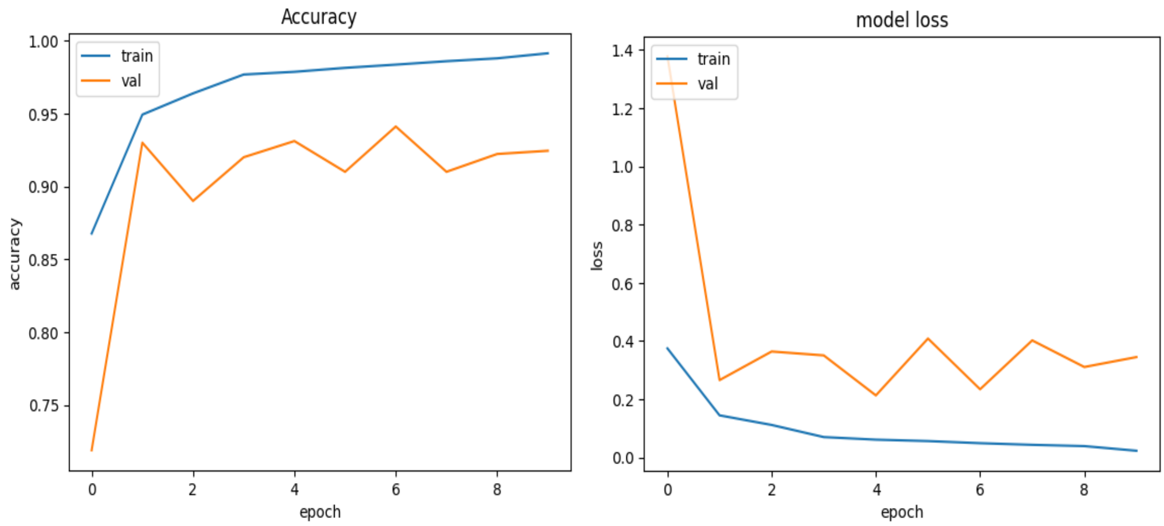

Figure 9: Accuracy and Loss Graph

Figure 10 presents the confusion matrix of the Xception model. The matrix shows the actual classes (rows) versus predicted classes (columns). For example, the model correctly classified 246 images of class 0 with minor misclassification to other classes. Class 1 had 214 correct predictions but 22 which is been misclassified as class 2. Also class 2 showed misclassification with equal numbers (241) classified correctly and incorrectly. Class 3 had 208 correct predictions with some misclassification to other classes.

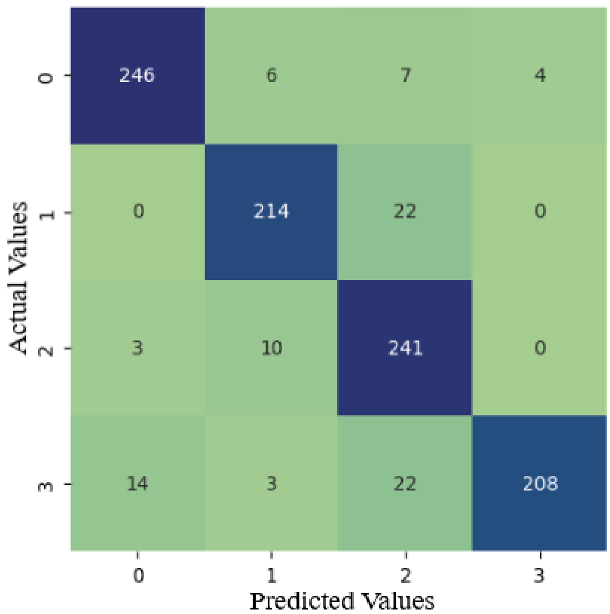

Figure 10: Confusion Matrix

6.3 Efficient Net B2 Model

Figure 12 is showing accuracy and loss graph for the Efficient Net B2 model. The training loss decreases steadily in the graph which shows that the model is successfully minimizing the error on training data. The training accuracy increases steadily which indicates the model is learning in a good manner.

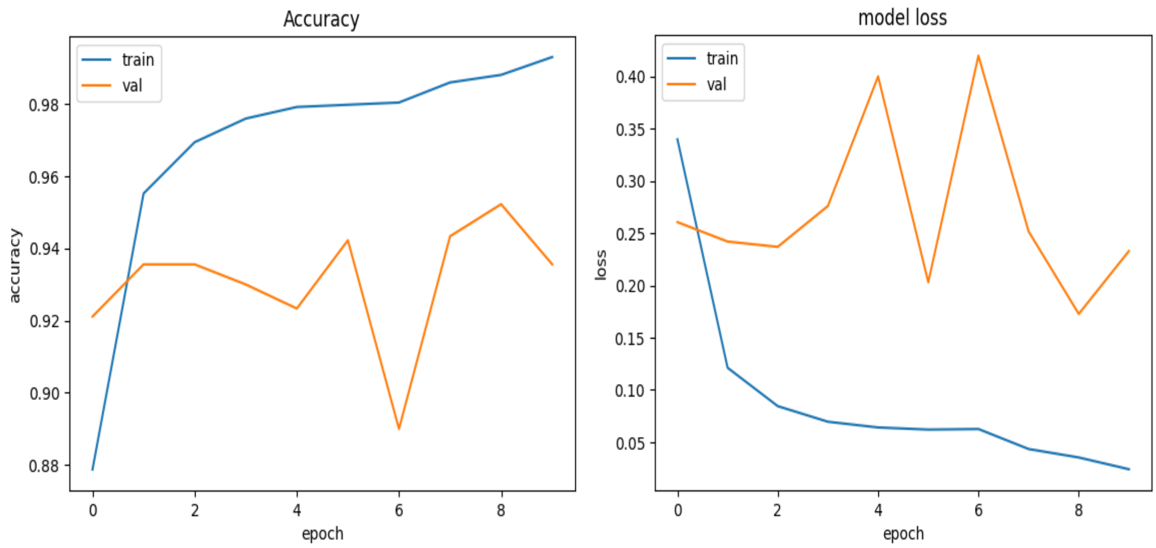

Figure 12: Accuracy and Loss Graph

Figure 13 displays the confusion matrix for the EfficientNet-B2 model. The matrix shows that the model accurately classified 252 images of class 0 with only a few misclassified as other classes. Class 1 had 221 correct predictions and minimal misclassification to class 2 and 3. Class 2 showed 220 correct classifications with some misclassification to class 1 and 3. Class 3 had 243 accurate classifications with few types of errors.

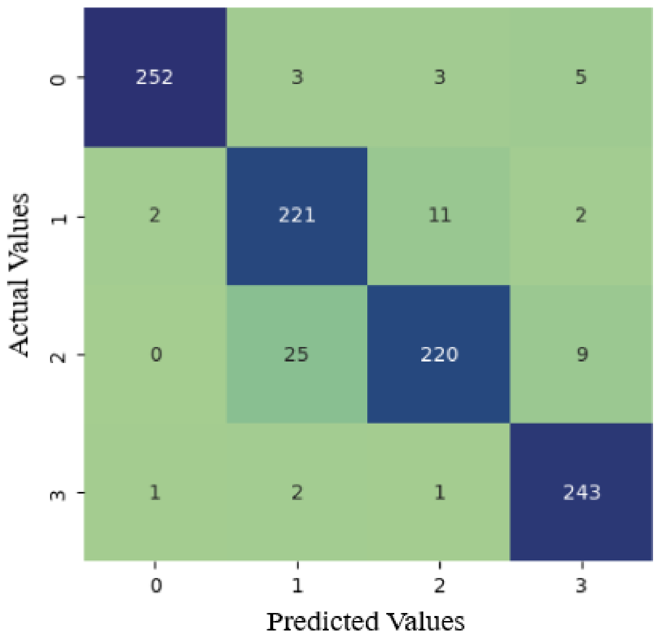

Figure 13: Confusion Matrix

6.4 Efficient Net B2 Model With Attention Baseline

Figure 14 is showing accuracy and loss graph for the Efficient Net B2 with attention baseline model. It can be seen that the training accuracy gradually increases which shows a good sign of learning. The Validation accuracy fluctuates a bit but gets pretty steady after some epochs which also shows that the model is learning pretty well from the training data and is becoming increasingly accurate in making predictions.

The model loss fluctuates a bit which is expected in some models.

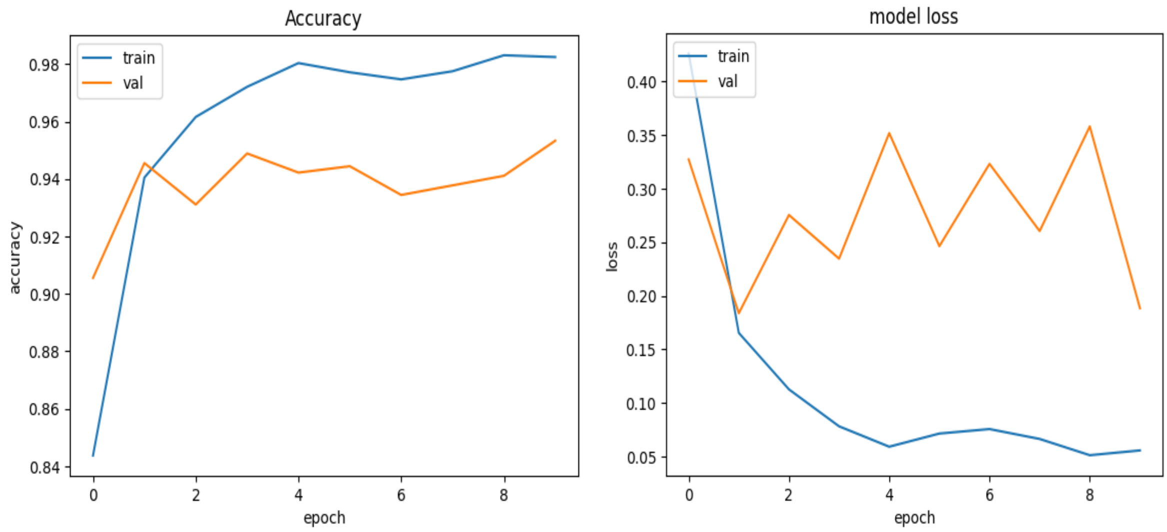

Figure 14: Accuracy and Loss Graph

Figure 15 shows the confusion matrix for the EfficientNet-B2 with Attention Baseline model. The matrix has been showed that the model classified 277 images of class 0 with minimal misclassification in a good way. For class 1 it accurately predicted 229 instances with a few misclassifications to classes 2 and 3. Class 2 had 213 correct classifications with minor errors in classes 1 and 3. Class 3 saw 229 correct predictions and few misclassifications. This matrix showed that the EfficientNet-B2 with Attention Baseline model performs good by showing high accuracy and low misclassification rates across all weapon categories.

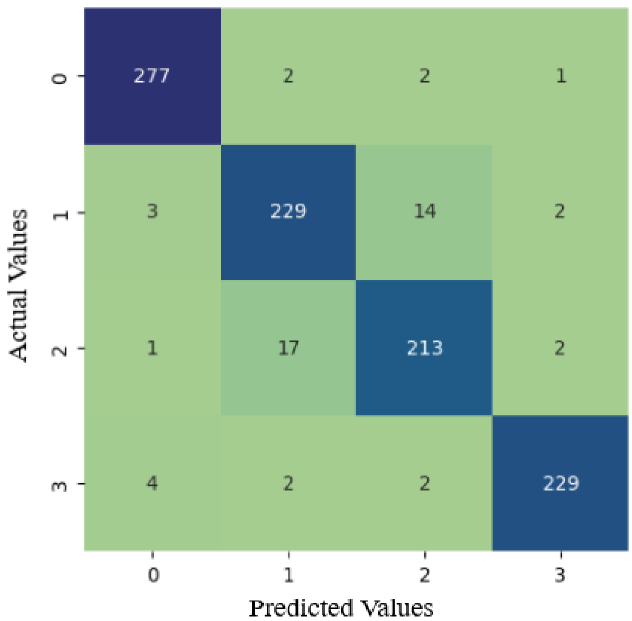

Figure 15: Confusion Matrix

6.5 Classification Performance of Deep Learning Models

The classification performance of the DL models which includes CNN, Xception, EfficientNet-B2, and EfficientNet-B2 with Attention Baseline showed different levels of accuracy and robustness in weapon classification tasks. The CNN model has achieved an accuracy of 77% with a macro average F1-score of 0.72. It was fairly good in its performance in all classes, however, it had particularly better strength in class 3 with an f1-score of 0.78 but relatively poorer in class 2 with an f1-score of 0.61. The Xception model has mostly improved the performance by achieving the accuracy of 91% and the macro average F1-score of 0.91. It performs well in class 2 with an f1-score of 0.88 and class 3 with f1-score of 0.91 which shows that it has perfect feature extraction and classification. The following EfficientNet-B2 model then pushed the accuracy figure a notch higher to 94 percent and the macro average F1-score was 0.94. This has shown very good precision and recall particularly for class 0 (f1-score of 0. 97) and class 3 (f1-score of 0.96) and it has shown how it handles large weapon images. The greatest prediction performance of the proposed model is 95% and the macro average F1-score is 0.95 for EfficientNet-B2 with Attention Baseline when compared to the training accuracy of 87.75% & test set accuracy of 83.49% in the previous research (Ramya et al., 2023). It depicted high values in all classes though the given scores were more focused on classes 0 and 3 with f1-score of 0.98 and 0.97 respectively. CNN is relatively a simple model in terms of its architecture because it has a small number of layers in its design. This affects the ability to capture complex features and patterns. The EfficientNet-B2 with the attention mechanism makes the model to attend or pay more attention to the relevant parts of the image and in addition to this it has a better data augmentation, regularization, and optimization methods which in return results to better or higher accuracy than the traditional CNN architectures (Tan, M. and Le, Q. V., 2019).

Table 2: Comparison of Deep Learning Models

Model	Accuracy
CNN	73%
Xception	91%
EfficientNet-B2	94%
EfficientNet-B2 with Attention Baseline	95%

6.6 Predicted output/results on the created Web Page

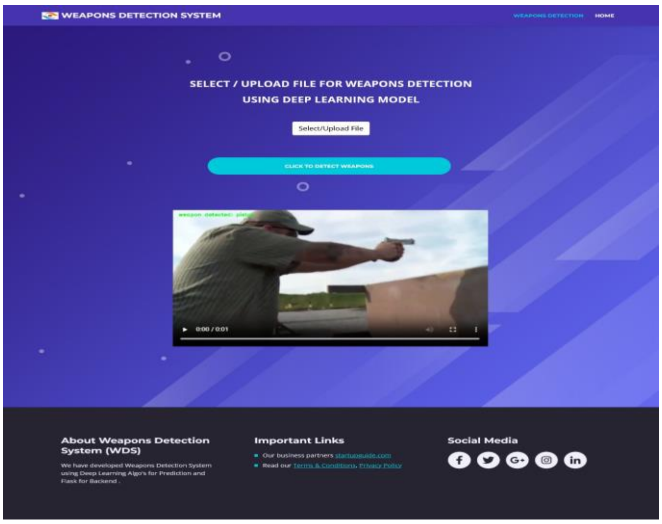

Figure 17: Weapon Detected as Rifle

On input of a video as seen above, it will detect the type of weapon and display it on the webpage as seen in figure 17.

6.1 Discussion

The weapon image classification system has used some advanced DL models like CNNs, Xception, EfficientNet-B2 and EfficientNet-B2 with Attention were each chosen for their good performance in image classification. EfficientNet-B2 with its compound scaling methods and Xception has been used for depth-wise separable convolutions which is good in feature extraction and accuracy. The combination of Attention Mechanism in EfficientNet-B2 further increases classification precision by focusing on the important features of the images. Addressing class imbalance with SMOTE secures balanced training by improving model robustness.

7 Conclusion and Future Work

7.1 Conclusion

This study has successfully developed a complex weapon detection system by using advanced DL models and combining them into a practical web application. The main objective was to evaluate and compare the performance of different DL architectures having CNN, Xception, EfficientNet-B2, and EfficientNet-B2 with Attention Baseline in classifying weapon images. The studies have been indicated that on achieving the accuracy of 95%, the EfficientNet-B2 with Attention Baseline was the best performing model and had good or superior performance in other aspects. This demonstrated that with regards to the discrimination of different weapon types, the model performs very well by making it the most robust type of solution to the problem among all the models tried. Subsequently, there exists a web application that has been created using Flask framework which provides a user-friendly interface for weapon categorization in real time. The real-world use of the DL models is demonstrated through the ability of users to upload images and get the predictions within a short span. This system not only enhances the efficiency of detecting weapons but also enhances the application and installation of such a technology in security and surveillance.

7.2 Future Works and Limitations

Future works for the weapon detection system could include having the knowledge and solution of the problems which have been established in this study. The dataset could be expanded to a much greater quantity of weapon types and the number of images per each class could also be increased to enhance the model's stability and the accuracy of its classifications. Applying the methods for addressing the issues of class imbalance, for instance, enhanced data augmentation or the mixed balancing approaches that may raise the performance of the model as well. To improve the adaptability more the use of transferring the knowledge from the models trained on the large datasets could be considered. It also improves the web application performance and scalability by adding the implementation of better backend processes and enlarging some of the user interface elements, which definitely could make the work of the application smoother for a user. The quality of the dataset was not up to the mark which has class imbalance issues and limited number of images for some weapons such as swords. This imbalance has also influenced the model's performance which is primarily in detecting classes with fewer samples.

References

Ramya, R., Lasya, C., Madhav Sai, N., and Paneerselvam, S., 2023. *An Intelligent Surveillance System for Weapon Detection Based on EfficientDet Algorithm*. Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA), pp.453-459.23

Kannur, A. and Kannur, A., 2022. A Hierarchical Support Vector Machines for Weapons Identification Using Multiple Stabbed Wound Images. *International Journal of Image, Graphics and Signal Processing*, *10*(6), p.84.

Gupta, G., Chattopadhyay, S., Kukreja, V., Aeri, M. and Mehta, S., 2024, March. The Arsenal Algorithm: AI-Driven Weapon Recognition with CNN-SVM Model. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-6). IEEE.

Kalla, K. and Suma, G.J., 2022. Weapon Detection in Surveillance Videos Using Human Inspired Particle Swarm Optimization Algorithm and Support Vector Machine. *International Journal of Intelligent Engineering & Systems*, *15*(4).

Gelana, F. and Yadav, A., 2019. Firearm detection from surveillance cameras using image processing and machine learning techniques. In *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS-2018* (pp. 25-34). Springer Singapore.

Rahman, S., Sharmin, N. and Ahmed, T., Machine Learning-based Approaches in Error Detection and Score Prediction for Small Arm Firing Systems in the Military Domain.

Rajesh, B., Vardhan, M.V.S. and Sujihelen, L., 2020, June. Leaf disease detection and classification by decision tree. In *2020 4th international conference on trends in electronics and informatics (ICOEI)(48184)* (pp. 705-708). IEEE.

Ahmad, G., Alanazi, S., Alruwaili, M., Ahmad, F., Khan, M.A., Abbas, S. and Tabassum, N., 2021. Intelligent ammunition detection and classification system using convolutional neural network. *Computers, Materials and Continua*, *67*(2), pp.2585-2600.

Kaya, V., Tuncer, S. and Baran, A., 2021. Detection and classification of different weapon types using deep learning. *Applied Sciences*, *11*(16), p.7535.

Bhatti, M.T., Khan, M.G., Aslam, M. and Fiaz, M.J., 2021. Weapon detection in real-time cctv videos using deep learning. *Ieee Access*, *9*, pp.34366-34382.

Arif, E., Shahzad, S.K., Iqbal, M.W., Jaffar, M.A., Alshahrani, A.S. and Alghamdi, A., 2022. Automatic Detection of Weapons in Surveillance Cameras Using Efficient-Net. *Computers, Materials & Continua*, *72*(3).

Durães, D., Veloso, B. and Novais, P., 2023. Violence detection in audio: evaluating the effectiveness of deep learning models and data augmentation.

Velayudhan, D., Hassan, T., Ahmed, A.H., Damiani, E. and Werghi, N., 2022, June. Baggage threat recognition using deep low-rank broad learning detector. In *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)* (pp. 966-971). IEEE.

Kalvankar, S., Pandit, H. and Parwate, P., 2020. Galaxy morphology classification using efficientnet architectures. *arXiv preprint arXiv:2008.13611*.

Hnoohom, N., Chotivatunyu, P., Maitrichit, N., Sornlertlamvanich, V., Mekruksavanich, S. and Jitpattanakul, A., 2021, November. Weapon Detection Using Faster R-CNN Inception-V2 for a CCTV Surveillance System. In *2021 25th International Computer Science and Engineering Conference (ICSEC)* (pp. 400-405). IEEE.

Baker, N., Lu, H., Erlikhman, G. and Kellman, P.J., 2018. Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, *14*(12), p.e1006613.

Tripathy, S., Singh, R. and Ray, M., 2023. Automation of brain tumor identification using efficientnet on magnetic resonance images. *Procedia Computer Science*, *218*, pp.1551-1560.

Sivakumar, M., Ruthwik, M.S., Amruth, G. and Bellam, K., 2024, April. An Enhanced Weapon Detection System using Deep Learning. In *2024 2nd International Conference on Networking and Communications (ICNWC)* (pp. 1-8). IEEE.

IBM (2021), 'CRISP-DM Help Overview', https://www.ibm.com/docs/en/spss-modeler/SaaS? topic=dm-crisp-help-overview.

Chaki, N., Pejas, J., Devarakonda, N. and Kovvur, R.M.R., eds., 2021. *Proceedings of International Conference on Computational Intelligence and Data Engineering ICCIDE 2020*. Lecture Notes on Data Engineering and Communications Technologies, 56. Singapore: Springer Nature Singapore. Available at: https://doi.org/10.1007/978-981-15-8767-2

Leelavathy, B., Sri Datta, Y.S.S. & Rachana, Y.S., 2021. Quality Assessment of Orange Fruit Images Using Convolutional Neural Networks. In: N. Chaki, J. Pejas, N. Devarakonda & R.M. Rao Kovvur, eds. *Proceedings of International Conference on Computational Intelligence and Data Engineering*. Singapore: Springer, pp. 343-353. Available at: https://doi.org/10.1007/978-981-15-8767-2_33

Hangün, B. and Eyecioğlu, Ö., 2017. *Performance Comparison Between OpenCV Built in CPU and GPU Functions on Image Processing Operations*. *International Journal of Engineering Science and Application*, 1(2), pp.34-41.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, pp.321-357.

Tan, M. and Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, California, PMLR 97, pp.6105-6114