SCIREA Journal of Traffic and Transportation
ISSN: 2995-7818
http://www.scirea.org/journal/TTE
March 5, 2024
Volume 6, Issue 1, February 2024
http://dx.doi.org/10.54647/tte580069

Comprehensive Anomaly Detection in Ghana: Addressing Social, Economic, and Security Challenges Through Transfer Learning in Traffic Surveillance

Isaac Opei [1,*], **Isaac Azure** [2], **Anas Musah** [3]

[*1,2,3] Regentropfen College of Applied Sciences, Ghana

Email: [*,1] isaacopei91@gmail.com, [2] azureike@yahoo.com, [3]anas.musah@recas.edu.gh

Abstract

The abstract presents a comprehensive overview of the social and economic challenges faced by Ghanaians, including high inflation, currency depreciation, and a surge in violent crimes such as armed robberies and murders. It highlights the alarming increase in road accidents, emphasizing their significant impact on mortality rates and the country's GDP. The document expresses concern about the poor response time of security agencies to these anomalies, leading to reputational consequences and a potential deterrent for foreign investors. The Inspector General of Police aims to reduce response time, but existing solutions focus on specific types of anomalies. The abstract proposes using transfer learning, specifically the DenseNet121 method, to detect various anomalies in traffic surveillance videos. The study details the training process, dataset augmentation, and model evaluation. While the proposed model shows positive results in identifying anomalies in certain conditions, the abstract suggests the need for improved models to handle long-distance, poor coverage, and hazy environment scenarios in future research.

Keywords: Ghana, Anomalies, Transfer learning, DenseNet121, Security agencies, Response time

1.0 Introduction

The introduction to the proposed research begins by drawing parallels between human learning processes and the field of machine learning. It likens the process of a newborn learning to speak its mother tongue through observation and trial-and-error to the concept of machine learning, which seeks to imitate human pattern recognition using data. The study emphasizes the significance of transfer learning, a machine learning research domain focused on leveraging knowledge gained from solving one problem to address another. It clarifies that transfer learning allows the retention and application of valuable information from previously trained models, addressing the challenge of insufficient training data and promoting knowledge transfer among models.

The existing system section provides a comprehensive overview of previous research efforts aimed at addressing the issue of response time to various anomalies, including robberies and accidents. Noteworthy studies are cited in the investigation into police response time to road crashes in Iran.[1] and findings on the effectiveness of CCTV cameras in deterring crime [2]. The limitations of these existing systems are highlighted, underscoring the need for a more comprehensive and adaptable approach to anomaly detection.

Challenges of the existing system are then discussed, pointing out the limitations of some studies that focused solely on specific types of anomalies. For instance, [3] .'s study concentrated only on knife detection[3], while other models were limited to detecting certain classes of anomalies like robbery or assault. The inability of these systems to provide real-time detection and raise alarms for swift responses is emphasized, highlighting the importance of overcoming these challenges in developing an effective anomaly detection system.

The proposed system section introduces the main thrust of the research. It details the envisioned real-time anomaly detection model designed to overcome the limitations of existing systems. The proposed system is framed as a solution capable of detecting fourteen classes of anomalies, including robbery, violence, arrests, arson, burglaries, fights, shootings, car accidents, shoplifting, theft, and vandalism. The use of transfer learning with the DenseNet121 algorithm is specified as the chosen methodology for the proposed model, emphasizing its potential to offer a holistic and adaptable approach to anomaly detection.

In summary, the introduction navigates through human and machine learning parallels, reviews existing research endeavors, critiques their limitations, and sets the stage for the proposed system. The proposed model is positioned as a comprehensive and real-time anomaly detection solution, aiming to enhance public safety and security.

1.1 General Objective

This research aims to create a real-time anomaly detection system for surveillance in highways and other monitored areas.

1.2 Specific Objectives

i. Develop and implement a DenseNet121 Transfer Learning (TL) model for the real-time detection of anomalies.

ii. Extend the implemented DenseNet121 model to effectively identify various classes of anomalies within the monitored environment.

iii. Evaluate and compare the performance of the proposed anomaly detection model against existing models, assessing its efficiency and effectiveness in real-world scenarios.

2.0 Research Methodology

System analysis has become integral to the development of websites, mobile apps, and various software applications. While machine learning programs fall within the broader category of software, they deviate from the traditional Software Development Lifecycle (SDLC) methodology. In the realm of computer science and engineering, preliminary to system development is the imperative step of conducting system analysis. According to the Merriam-Webster dictionary, system analysis is defined as "the process of studying a procedure or business to identify its goals and purposes and create systems and procedures that will achieve them in a cost-effective manner." Alternatively, it is a problem-solving method that deconstructs a system into its fundamental elements to scrutinize how each component functions independently and collaboratively to accomplish the system's objectives [4].

Conventional software development projects adhere to SDLC, employing well-defined processes to yield high-quality software. Various SDLC models, such as the Integration and Configuration model, Waterfall model, Incremental model, Iterative model, and Spiral model, guide developers through distinct phases to ensure the efficient creation of software. On the other hand, the Machine Learning Life Cycle (MLLC) constitutes a unique cycle that teams must traverse to construct and manage high-quality machine learning models.

2.1 Research Design

The study employed the Design Science Research Methodology, a framework rooted in critical thinking and problem-solving design with the primary goal of enhancing the performance of objects. As outlined by the CDC in 2003, this methodology generates novel guidelines, contributing to the interpretation of real-world problems and the expansion of existing knowledge and theory. Given that the study's objective centered around developing a model for the real-time detection of various anomalies and signaling the relevant agencies to minimize response time, the Design Science Method was deemed the most fitting approach. Through the Design Science Research Methodology, the study sought to establish the validity and consistency of the model, ensuring it provides practical solutions to the identified problem.

This qualitative research project focuses on roadside anomalies and their detection. The formulation of research questions and associated issues significantly influences the design of research methodologies [5] . The central questions of What, How, and Who constitute the subjects of inquiry in this study. These types of inquiries align with the characteristic approach of qualitative research, aiming to provide descriptive insights into the phenomena being investigated. Unlike quantitative research, the research question in this study is oriented towards Why, exploring contrasts between groups or seeking connections between variables within the subjects under investigation [6] . The nature of this research is open-ended, emphasizing a comprehensive understanding of participants' experiences rather than seeking specific and limited measurable or visible data [6] . Consequently, this study is framed as a qualitative investigation.

In August and September 2022, interviews were conducted with concerned citizens. Initial contact with respondents was established through phone calls or in-person meetings, following the specified data collection methods. Appointments were confirmed based on the availability of interview subjects. To ensure an environment conducive to accurate information capture, the researcher requested meeting places devoid of distractions, and these locations were arranged on the outskirts of Navrongo station [6] . The interviews were conducted in English, each lasting for a minimum of one hour. Listening and documentation techniques were employed to capture the essence of the interviews. Ethical considerations related to research involving human participants will be further discussed in subsequent sections of this paper.

Each interview was conducted face-to-face by the researcher using an interview schedule and an unstructured interview format. The interview questions were open-ended, including veiled inquiries designed to extract additional information about the subject under study [7] . The interviewing techniques employed with this sample group are elaborated upon in the subsequent sections. The methodology adopted for this thesis involved the use of unstructured or non-standardized interviews. An unstructured interview is a method of gathering interview data with minimal supervision over the interaction between interviewers and interviewees [8]. Unstructured interviews offer greater flexibility and permit more interaction between interviewers and respondents compared to structured interviews, which employ predefined questions and probes in a predetermined order [8] . However, specific questions were occasionally introduced during the interview, depending on the topics raised throughout the sessions.

The specific questions aimed to explore the planning procedures utilized and the backgrounds of the respondents. Questions related to the attitudes and knowledge of the respondents in the planning process were carefully crafted to facilitate open discussions and the sharing of personal experiences. As suggested by [8] , framing questions in this manner encourages participants to express their perspectives on the issue, share views on the occurrence, and communicate their emotions about it. Interviewees can influence both the study's direction and the content of the interview [9]. To provide interviewees or respondents with the optimal opportunity to speak openly, the dynamic between the interviewer and respondents was considered during the interview session. [9] suggests that treating the interviewee as an active agent contributes to a positive interaction between the interviewer and interviewee. This interaction is described as a "sense-making activity" by [9] . The "sense-making activity" occurs when interviewees actively participate and engage in interpreting the interviewer's questions and responses. This engagement can lead to a shift in the direction of the conversation during the interview, deepening and broadening the interviewer's understanding of the relevant issues [9].

2.2 Population and Sampling Technique

The study's population, encompassing all potential elements, subjects, or observations within a specific area that are of interest to the researcher, is outlined by [10]. The population, further clarified as the complete collection of observations from which a sample is selected, is discussed by [11]. We intended to survey a population of 100 contact persons to capture the perspectives and concerns of the general public regarding the subject under investigation. However, this sample size was constrained by resource limitations, as defined by [12].

This study aims to develop a model for real-time detection of various anomaly classes and promptly alert the relevant agencies to minimize response time. The sample selection employed a purposive sampling technique, as described by [13]. Purposive sampling involves the deliberate selection of participants based on specific characteristics and qualities they possess. The criteria for sample selection may vary, such as seeking a particular narrative to explore, having a shared experience with an event, or contributing to theory development [14]. The common thread is that each participant is chosen because they possess a specific quality relevant to the investigator's interest. Ultimately, this approach allows the researcher to grasp the fundamental concept of the learning situation and understand the implications of the experiment's outcomes.

A subset was chosen from the overall population to serve as an accurate reflection of the entire group. The sample size, which refers to the number of elements selected for statistical analysis in a study, is crucial for drawing meaningful conclusions. A sample should typically represent 10% to 30% of the population. In this study, the sample comprised 20 individuals out of a total population of 100 [15].

2.3 Data Collection Technique and Pre-processing

The data collection method employed is contingent on the nature of the project under development. While datasets can be sourced from various outlets such as files, databases, and sensors, utilizing them directly for analysis may pose challenges, including substantial instances of missing data, exceedingly high values, disorganized text data, or noise [16]. In this research endeavor, the primary data source relied upon by the researcher is predominantly secondary data obtained from security organizations. The reliance on secondary data is necessitated by resource and infrastructure constraints for capturing real-time data on highways. Consequently, data preparation becomes imperative to address these challenges.

To address compatibility issues and enhance model generalization, several modifications were applied to the photos in multiple stages prior to model training. Preparing an input image for compression using Keras involved initially rearranging or switching the color channels (RGB) to ensure compatibility. Subsequently, the provided image was scaled to a predetermined size, such as 224 x 224, without considering the aspect ratio. The final step in image preprocessing, mean subtraction, included calculating the mean for each pixel value and subtracting it from the pixel values. This step aimed to establish a baseline for comparison with other factors like weights and biases. Once an image underwent preprocessing, it could be fed into a DenseNet121 model. The same set of procedures was replicated during the testing phase of the study, where images were incorporated into the model testing and looped throughout the testing process.

In this stage, the data scientist is tasked with researching and determining the appropriate algorithm or model to utilize, taking into account the specific problem at hand and the nature of the collected data. For instance, if the target variable is categorical, it constitutes a classification problem. Consequently, the output may be categorized into distinct classes, such as Class A, Class B, or another designated category.

In the training phase, a model is initially divided into three segments: "Training data," "Validation data," and "Testing data." The "training dataset" is employed to train the classifier, the "validation set" aids in refining the classifier's parameters, and the "test dataset" is utilized to assess the classifier's performance. It is important to note that during the training of the classifier, only the training and/or validation set is accessible, and the test data set cannot be employed in the classifier's training process. The test set is reserved exclusively for evaluating the classifier's performance after training completion.

2.4 DenseNet121 Transfer Learning (TL) Model

In this research paper, we have utilized the DenseNet121 transfer learning (TL) model, a densely connected convolutional network model. The DenseNet121 model addresses the challenges associated with traditional feed-forward convolutional networks (CNNs), where each convolutional layer, except the initial input layer, takes the output of the preceding layer and generates an output feature map, passing it to the subsequent layer [17] . As the layer count in a CNN increases, the "vanishing gradient" problem arises, leading to the potential loss of information as the path from input to output layers lengthens. This limitation hampers the model's effective learning capacity.

DenseNets offer a solution to this issue by modifying the conventional CNN architecture and enhancing connectivity between layers. Unlike standard CNNs, where each layer connects only to the preceding and subsequent layers, DenseNets establish direct connections between every layer, minimizing the number of paths. The utilization of the DenseNet121 model in this research ensures clear and sharp image output. Compared to regular CNNs or ResNet models, DenseNet121 requires fewer parameters, facilitates feature reuse, generates more compact models, achieves state-of-the-art performance, and outperforms competitors across various datasets.

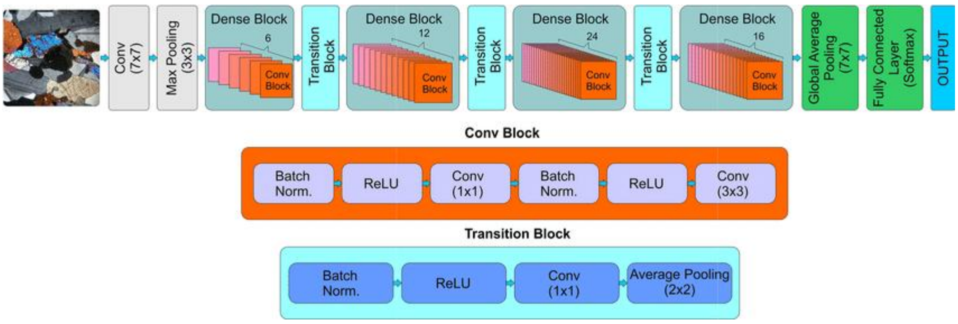

Figure 3.2: Layered Architecture of DenseNet121 model

Source: [17]

In the case of DenseNet121, each layer consists of two parameters: weights and biases. The total number of parameters utilised is the sum of all biases and weights. DenseNet121 ranks among the top categories of neural networks for image classification, processing, and recognition, as evidenced by a literature survey. The initial layer of the model is a convolutional layer with 64 filters, each sized 7 by 7, and a stride of 2. This layer serves to increase the number of channels while simultaneously reducing the spatial dimensions of the input image.

Subsequent to the application of the input image, convolutional layers are constructed. Each layer employs a set of filters on the input image to extract its features. The output of each convolutional layer consists of a series of feature maps, which are then provided as input to the subsequent layer. A convolutional layer can be mathematically represented as

$$y = f(W * x + b)$$

where x represents the input feature map, W is the collection of learnable weights, b is the bias factor, and f denotes an activation function such as ReLU.

To further reduce the spatial dimensions of the feature maps, a max pooling layer with a stride of two (2) and a pool size of 3 x 3 is applied after the initial convolutional layer. The DenseNet121 architecture is composed of four dense blocks, each comprising multiple layers that are densely interconnected. The quantity of feature maps generated by each layer within a dense block is dictated by the growth rate, a fixed number of channels for every dense block in DenseNet121. In the case of DenseNet121, the growth rate is set at 32.

The mathematical representation for a dense block is expressed by the formula:

$H = ([x_0, x_1, x_2, \dots, x_n - 1])$

where H is a composite function representing a sequence of convolutional layers with non-linear activation functions, and $[x_0, x_1, x_2, \dots, x_n-1]$ denotes the input feature maps concatenated from all preceding layers. After each convolutional layer, the feature maps undergo batch normalization and ReLU activation.

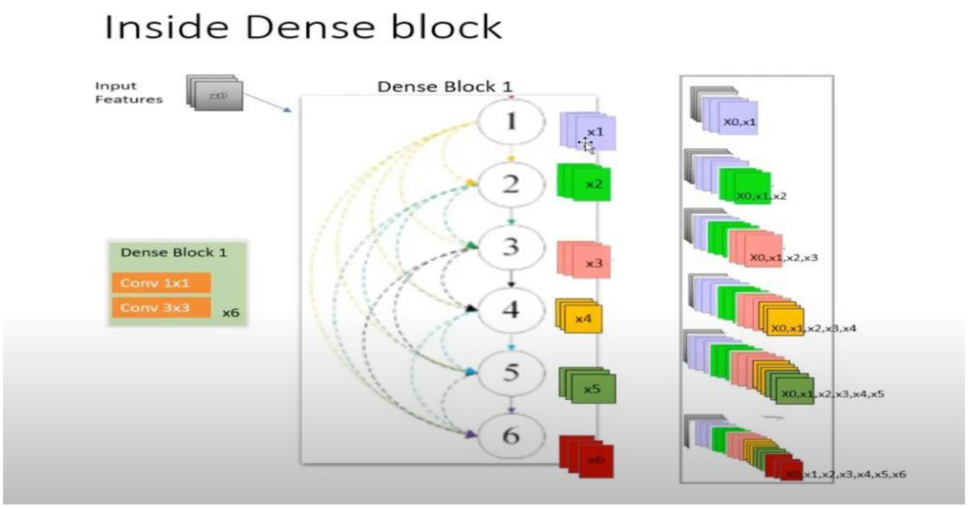

Figure 3.3: The Architecture of Dense Block

Source: *https://www.youtube.com/watch?v=hCg9bolMeJM&t=1053s*

Transition layers are employed to decrease the spatial dimensions of the feature maps and control the number of channels. These layers consist of a 2 x 2 average pooling layer, a batch normalization layer, and a 1 x 1 convolutional layer. The average pooling layer is utilized to reduce the spatial dimensions, while the 1 x 1 convolutional layer is employed to decrease the number of channels. The mathematical model for a transition layer can be expressed using the equation

$$y = f(W * x + b)$$

where x is the input feature map, W is the set of learnable weights, b is the bias term, f is an activation function such as ReLU, and pool is a pooling operation like average pooling, with y' = pool(y) representing the output.

Global Average Pooling (GAP) stands as a widely employed pooling operation in Convolutional Neural Networks (CNNs), particularly for image classification tasks [18]–[20]. This technique entails calculating the average of the output feature maps from each convolutional layer across the entire spatial extent, resulting in a global average value for each feature map [18]–[20]. The outcome of the global average pooling process is a feature vector that serves as the input for a fully connected layer, facilitating classification or subsequent tasks [18]–[20].

Mathematically, consider a convolutional layer with m feature maps of size H x W, where each feature map is denoted as a matrix

$$F = [f - ij]$$

with i and j representing the row and column indices, respectively. The convolutional layer's output is a 3D tensor of size m x H x W, represented by

$$A = [aijk]$$

The global average pooling operation is applied to this output tensor A, following the formula:

$$gk = \frac{1}{H * W} * sum(sum(aijk * delta - k))$$

where k = i, and delta-k is the Kronecker delta function, equaling 1 otherwise. Essentially, this formula computes the average of all values in the k-th feature map of A, yielding a scalar value g-k for each feature map.

The result of the global average pooling process is a feature vector G = [g-1, g-2,..., g-m] of size m. This feature vector can be employed as input for a fully connected layer, enabling classification or other downstream activities.

The conclusive output of the DenseNet121 model is derived by subjecting the output of the last dense block through a fully connected layer activated by softmax and a global average pooling process. Mathematically, this is expressed as

$$y = softmax(W * pool(x) + b)$$

where x represents the input feature map, pool denotes the global average pooling operation, W signifies the collection of learnable weights, b stands for the bias term, and softmax is an activation function responsible for generating a probability distribution across the output classes.

2.5 Dataset for Testing

To assemble a dataset encompassing various highway anomalies, an exhaustive review of commonly utilized secondary data sources was conducted. These sources included datasets related to Road Accidents, Burglary, Fighting, Shoplifting, Robbery, Stealing, Shooting, Arson, Arrest, Assault, and Abuse. The dataset comprises 1,377,896 surveillance videos featuring a diverse range of documented anomalies, along with an additional 1,013,000 videos depicting normal scenarios. Table 3.1 provides a comparative analysis of datasets based on parameters such as the number of images, training and testing images, dataset length, and types of anomalies. Anomalies covered in the dataset include Road Accidents, Abuse, Arson, Assault, Burglary, Explosions, Fighting, Shoplifting, and Vandalism.

Table 3.1: System specification

Dataset	No. Of images	Training Images	Testing Images	Anomaly Type
Road Accidents DS	26,163	23.5k	2663	Road Accident
Abuse DS	19,297	19.1k	297	Abuse
Assault DS	13,057	10.4k	2657	Assault
Arrest DS	29,765	26.4k	3365	Arrest
Arson DS	27,193	24.4k	2793	Arson
Explosions DS	25,310	18.8k	6510	Explosions
Vandalism DS	14,711	13.6k	1111	Vandalism
Burglary DS	47,157	39.5k	7657	Burglary
Fighting DS	25,931	24.7k	21231	Fighting
Shoplifting DS	32,423	24.8k	7623	Shoplifting
Shooting DS	14,770	7140	7630	Shooting
Stealing DS	46,784	44.8k	1984	Stealing
Robbery DS	42,335	41.5k	835	Robbery

Normal DS	1,013,000	948k	65.0k	Normal

Data analysis encompasses a series of procedures conducted on the data acquired during the study, aiming to condense, scrutinize, and structure it to achieve the research objectives. The dataset utilized in this study originated from the UCF-crime dataset. Rigorous verification procedures were employed to ensure the accuracy and completeness of the collected information. The presentation of data involved the use of tables, percentages, and graphical representations.

3.0 Results and Discussion

In pursuit of developing a road anomaly detection model, the solution was conceived, strategized, and implemented using Python ML libraries. The system architecture and its progression involved various enhancements. The model was crafted using open-source programming.

This model was implemented using a virtual computational environment known as Kaggle. The specifications of this environment are detailed in Table 4.1 below.

Table 4.1: System specification

GPU T4 x 2 -1	14.8GB
GPU T4 x 2 -2	14.8 GB
CPU	32 CORE
RAM-1	13GB
DISK	73.1GB

Python, known for its robustness, flexibility, and user-friendly nature, is highly recommended for applications in the domains of AI, ML, DL, and NLP. Essential deep learning libraries within the Python ecosystem include Keras, matplotlib, sklearn, imutils, NumPy, argparse, pickle, cv2, and OS. These libraries play a crucial role in evaluating and constructing deep learning models, conducting general scientific computing, visualizing data, and managing data organization. Due to these advantages, a classifier was developed in Python to detect anomalies in input videos and photos.

The Excel file used for labeling the data was populated with the names of each image and video from the data folder. Labels were assigned to specify the type of anomaly for each image in the dataset.

3.1 Implementation of the DL Approach

A controlled procedure known as Densely Connected Network was employed to train the classifier in the implementation of the Deep Learning (DL) approach. The choice of DenseNet121 was driven by its notable performance in image and video categorization, as well as object detection, as indicated in the literature review. The implementation was carried out using Python (Khan et al., 2022).

For a distinctive approach to the process, the top layers of the DenseNet121 algorithm were set to false, and custom layers were added on top of it. This novel implementation included the following custom layers, building upon the DenseNet121 algorithm [21]:

- Average Pooling Layer.
- Dense Layer.
- Drop Out.
- Dense Layer.
- Drop Out.
- Dense Layer.

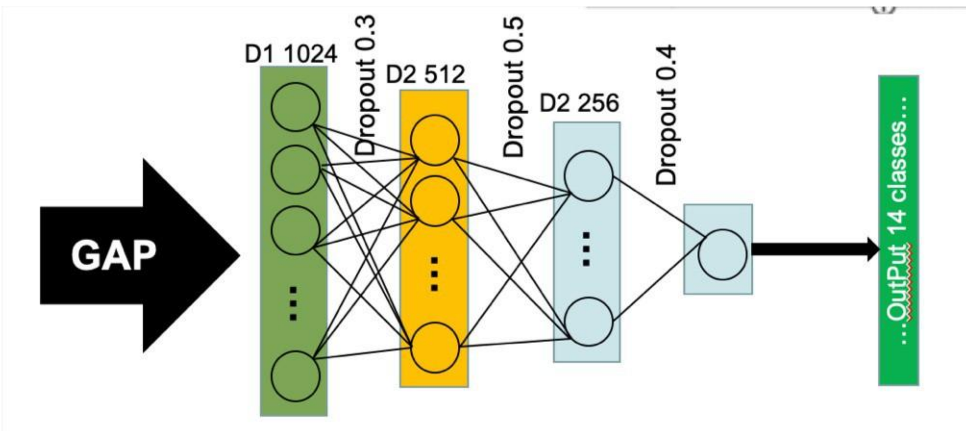

Figure 4.1: The architecture of the custom layers defined on top of the densenet121

Data augmentation techniques were employed to enhance the model's generality. Two data augmentation (DA) objects were instantiated—one for training and another for testing. The training object underwent random rotations, shearing, shifting, and zooming on the data to augment it. It's worth noting that data augmentation techniques were not applied to the test data; instead, only the mean subtraction approach was utilized.

This section illustrates the distribution of data for the training, testing, and validation of images categorized under each anomaly class. The data distribution is visually presented through bar charts for enhanced organization and clarity. Figures 4.1, 4.2, and 4.3 depict the data distribution for training, testing, and validation, respectively.

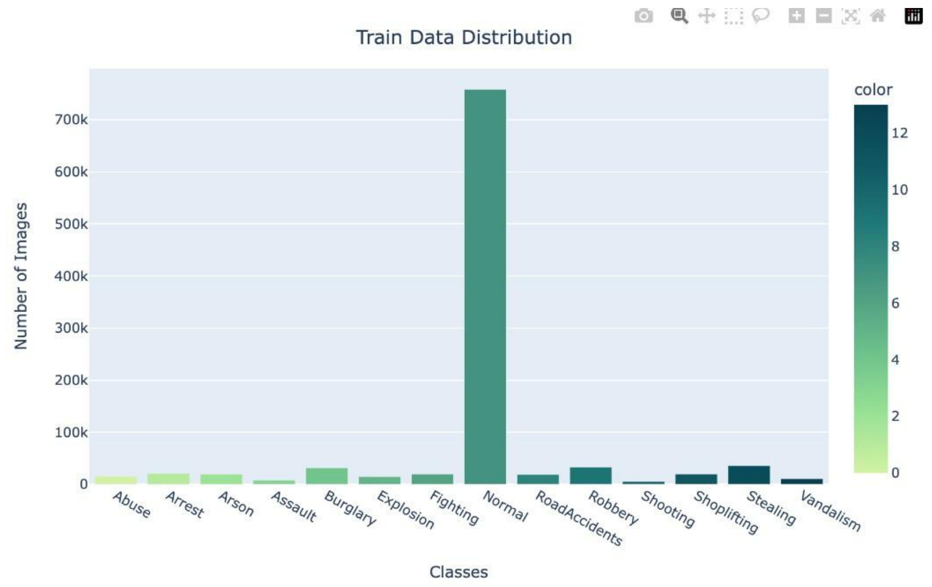

Figure 4.2: Distribution of data for training.

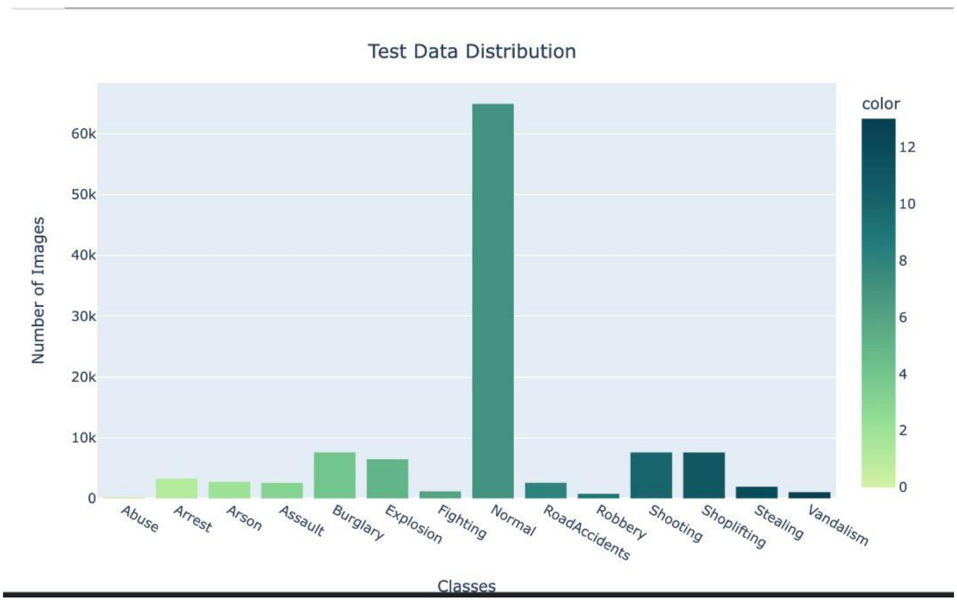

Figure 4.3: Distribution of data for Test

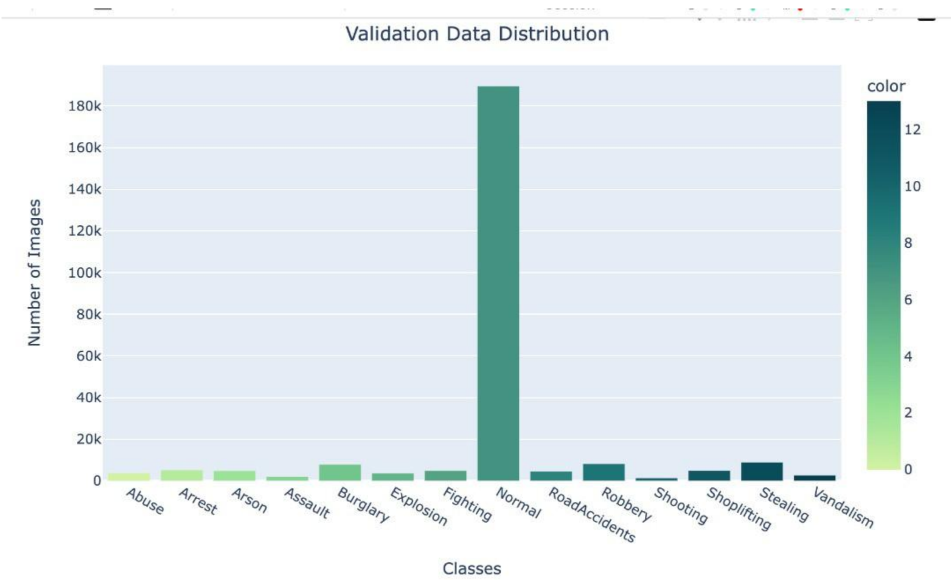

Figure 4.4: Distribution of Validation Data

After designing the architecture of the DenseNet121 model on Kaggle, we constructed our model using an optimizer, specifically stochastic gradient descent (SGD). To compute the training loss for our classes, we employed "binary cross-entropy" [21] . The training data, consisting of trainX and trainY, was submitted to train the model we created. Utilizing 32 batch sizes, 10 epochs, and trainY parameters, our network learned the data through mean subtraction and data augmentation, executed via the "fit generator" function in our model [21]. Finally, we saved our model, as depicted in Fig 4.2.

Table 4.2: Training results table for our trained model

Epochs	Time(s)	Loss	Auc	Val_{loss}	Val_{auc}
1/5	9777	9777	0.9326	0.9326	0.9087
2/5	7414	0.8392	0.9626	1.3474	0.9155
3/5	7334	0.7755	0.9686	1.3885	0.9188
4/5	7306	0.7360	0.9718	1.4135	0.9194
5/5	7170	0.9742	0.9742	1.4297	0.9167

The epoch duration is measured in seconds, representing one complete cycle of the model processing the training data. The loss, reflecting errors made in each epoch, ranged between 0.7 and 1.1 epochs, indicating minimal errors and emphasizing the model's efficiency and effectiveness. The AUC, representing accuracy or performance for each epoch, ranged from 0.9 to 0.99, indicating a high level of accuracy for our model.

3.2 Model Application

The earlier part of this chapter outlines the development of a DenseNet121 model and its training using a set of training images. This section now delves into the procedures for assessing the model's performance through a rolling prediction average applied to the training dataset. The testing dataset comprises images and videos depicting highway anomalies sourced from various outlets. Necessary Python modules and packages such as "deque" and "collections" were imported for implementing the gravitational search method. The process commenced by identifying the optimal images for anomaly detection, and the subsequent loop iterated through images until reaching the peak of efficiency in anomaly detection.

We systematically input the entire testing dataset, consisting of 1,225,140 photos covering all 14 anomaly classes, into our trained model one by one. This process aimed to assess the model's performance on unfamiliar data, specifically the test dataset. Subsequently, we scrutinized the outcomes to gauge the effectiveness of the model. The model was designed to display the input photos on the screen and annotate them with the label corresponding to the detected anomaly.

3.2.1 Experimental Results Scenario One

The model successfully recognized frames as anomalies in each class, accurately categorizing images into classes such as "accident," "Abuse," "Assault," and so forth. However, in some instances, the remaining photos were either misclassified or lacked classification entirely. This discrepancy occurred when the images were clear and of high quality, allowing the model to make accurate classifications. On the other hand, when the camera was positioned too far away, resulting in unclear photos, or when there was a flaw in the model, misclassification occurred, as illustrated in the example below [21].

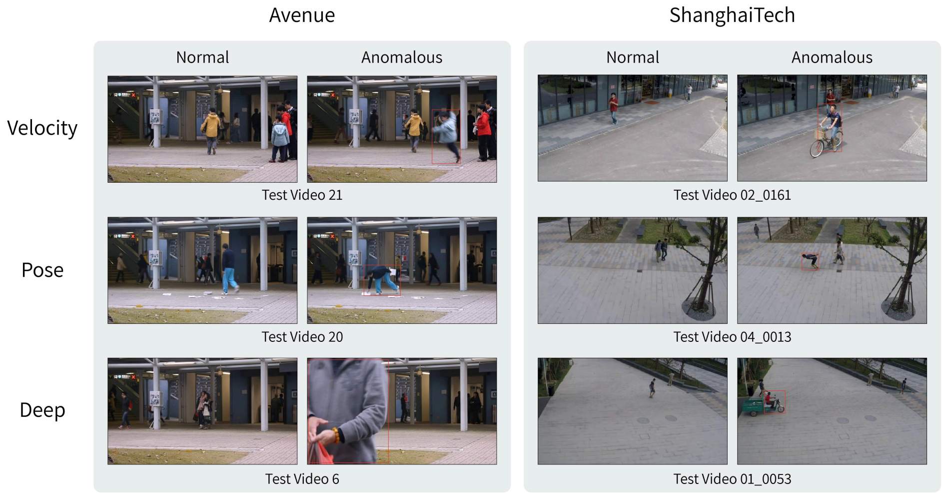

Figure 4.5: Test 1's recording of a result. Although the camera's resolution in this instance was inadequate, the viewing angle was appropriate.

3.2.2 Experimental Results Scenario Two

In this scenario, we provided a set of ordinary photographs captured without any unusual elements. The model correctly classified some of the photos as normal, while others were either misclassified or resulted in errors. A high-resolution camera was employed in this case, positioned at the scene, and the model identified and labeled it as normal, as depicted below.

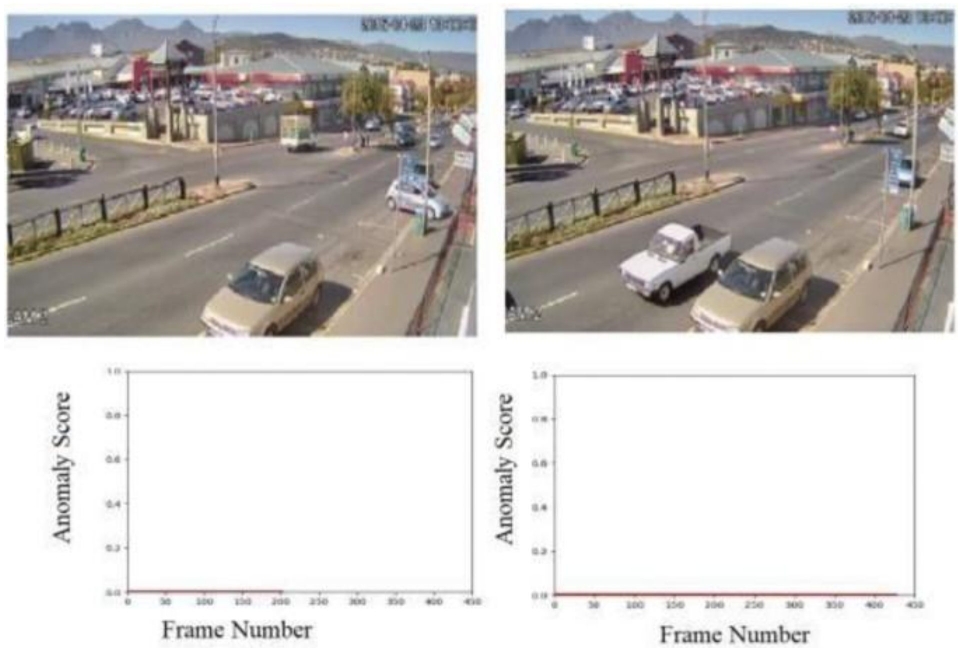

Figure 4.6: Test 2's outcome was recorded. Without labelling the image, it displayed the outcome

This study employs the multiclass ROC curve and AUC metrics, acknowledging the complexity introduced by having multiple classes in the classification problem, each potentially holding varying degrees of importance or priority.

The metrics extend the ROC curve and AUC metric to multiclass scenarios by employing either a micro-averaging or macro-averaging strategy. In micro-averaging, the True Positive Rate (TPR) and False Positive Rate (FPR) are aggregated across all classes, yielding an overall AUC metric. Conversely, in macro-averaging, the AUC metric is calculated individually for each class, followed by averaging these values. The micro-averaging method assigns equal weight to each class, while the macro-averaging method grants more weight to classes with larger sample sizes.

In a multiclass classification problem with K classes, the actual class labels (y) and predictions (y-hat) can be represented as one-hot encoded matrices Y and Y-hat, respectively, both of shape (n-samples, K). The micro-averaged ROC curve and AUC metric can then be computed as follows:

1. Calculate the aggregate confusion matrix C of shape (K, K) as the sum of unique confusion matrices for each class:

$C = (ij) = \sum iy[:,i].(y - hat[:,j] == 1))$

where $y[:,i]y[:,i] and\ y - hat[:,j]y - hat[:,j]$ are the i-th and j-th columns of Y and Y-hat, respectively, and ·· represents element-wise multiplication.

2. Determine the total counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for all classes:

TP = $=\sum iC[i,i]TP = \sum iC[i,i], FP = \sum iC[:j]FP = \sum jC[:,j], TN = C[j,i], and\ FN = \sum i(C[i,:].sum() - C[i,i])FN\sum i(C[i,:].sum() - C[i,i])for\ i,j \in [1,K]i,j \in [1,K]$

3. Calculate the total True Positive Rate (TPR) and False Positive Rate (FPR): TPR=TPTP+FNTPR=TP+FNTP and FPR=FPFP+TNFPR=FP+TNFP .

4. Plot the micro-averaged ROC curve by depicting TPR versus FPR.

5. Compute the micro-averaged AUC metric using the trapezoidal rule or another integration method.

The micro-averaging approach treats all fourteen classes of anomalies equally, yielding an overall TPR and FPR across all classes. The overall accuracy under the curve is reported as 83%.

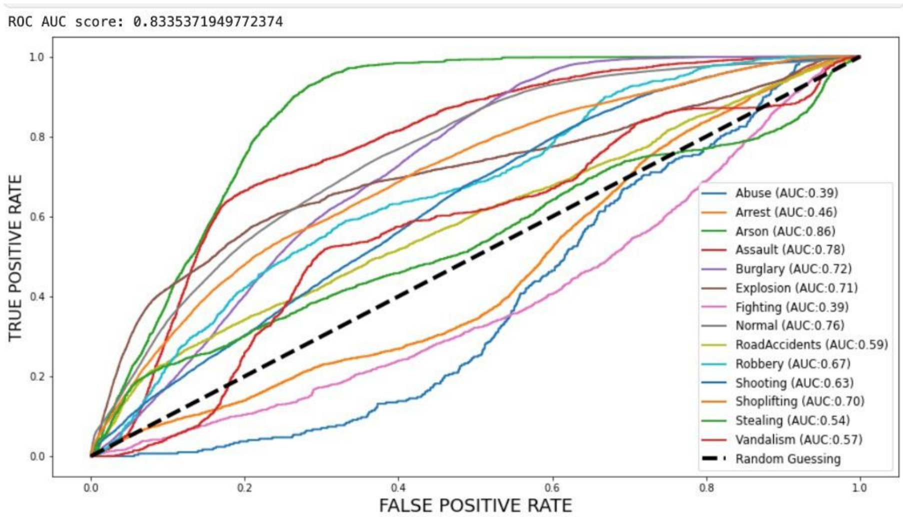

Figure 4.7: Accuracy measurement of test Dataset

4.0 Conclusion

The primary objective of this research was to create a model for the detection of roadside anomalies using an advanced algorithm, specifically the denseNet121, along with the gravitational search algorithm. The denseNet121 model represents an evolution in traditional Convolutional Neural Networks (CNNs), offering heightened efficiency, precision, and improved image processing/detection capabilities. This innovative model is instrumental in addressing roadside anomalies, thereby contributing to a reduction in response time for security agencies attending to incidents.

The study aimed to investigate various ways in which images/videos depicting roadside scenes can be identified as anomalies. Additionally, it sought to explore the advantages of the denseNet121 model over traditional CNNs and illustrate the most suitable and ethical approach to anomaly detection on the roadside.

As per the survey findings, a growing number of security organizations and agencies are adopting the denseNet121 model for roadside anomaly detection, emphasizing its significance in cutting-edge applications. A notable advantage of the denseNet121 model is its ability to overcome the "vanishing gradient" problem commonly encountered in traditional CNNs.

Facial recognition stands out as another prominent feature, leveraging the natural and unalterable nature of faces compared to traditional security measures like PINs or passwords. This enhances the security of sensitive information and documents. The incorporation of the denseNet121 model in roadside anomaly detection plays a crucial role in addressing research gaps that were not adequately covered in previous studies.

A cutting-edge branch of machine learning, known as transfer learning (TL), has emerged as a highly precise approach in recent years. Leveraging deep and robust neural networks, transfer learning has become instrumental in addressing complex problems that conventional statistical and machine learning techniques find challenging or even impossible to solve (Khan et al., 2022). Its efficacy spans various scientific domains, unlocking solutions to previously formidable problems.

Transfer learning proves valuable in identifying abnormalities across diverse datasets, employing algorithms like CNN, DenseNet121, RNN, and LSTM. The DenseNet121, renowned for its effective classification of objects and images, serves as a transfer learning strategy in this study. The model, based on DenseNet121, is applied to detect anomalies in images captured by traffic surveillance systems (VSS) videos.

For efficient neural network training, a substantial amount of labeled data is essential. Recognizing the dataset's limitations, data augmentation techniques such as rotation, shear, zoom, and flip were incorporated during the algorithm's training phase. Validation and assessment of the DenseNet121 model were conducted on a dataset comprising 30 films, with training performed on images. During testing, label switching between photos posed a challenge to accuracy, prompting the resolution of the label flickering issue. The DenseNet121 trained model was optimized for image classification by integrating a rolling prediction average technique.

This study aims to benefit both collision victims and personnel monitoring multiple traffic surveillance displays. The system, through video analysis, can identify anomalies and alert personnel to pertinent actions. Positive outcomes were observed in testing on videos from high-resolution cameras used in VSSs. However, challenges arose in accurately processing films captured from a significant distance or in hazy environments. Addressing these challenges and developing enhanced models for long-distance, low-coverage, and hazy environment films represent promising directions for future research.

5.0 Recommendation

This model exhibits certain limitations, and the preceding chapter offers insightful comments along with recommended enhancements. The following suggestions aim to address these concerns:

1. **Comprehensive Anomaly Detection:** Enhance the model to detect a broader range of anomalies, irrespective of video quality.

2. **Low-Quality Video and Foggy Environment Handling:** Improve the model to intelligently handle situations with low video quality and environments characterized by fog. Consider implementing error alerts for such scenarios.

3. **Integration with Security Alarms:** Install security alarms at checkpoints and establish a seamless connection with the model. This integration should trigger alarms when anomalies are detected, ensuring a swift response.

4. **Exploration of Convolutional Networks:** Encourage further research into convolutional networks, exploring ways to expedite the anomaly detection process and enhance overall efficiency.

Implementing these recommendations is anticipated to refine the model's performance, leading to a more robust and effective anomaly detection system.

Reference

[1] D. Khorasani-Zavareh, H. Khankeh, R. Mohammadi, L. Laflamme, A. Bikmoradi, and B. J. A. Haglund, “Post-crash management of road traffic injury victims in Iran. Stakeholders’ views on current barriers and potential facilitators,” BMC Emerg Med, vol. 9, May 2009, doi: 10.1186/1471-227X-9-8.

[2] J. M. Caplan and L. W. Kennedy, “Police-monitored CCTV cameras in Newark, NJ: A quasi-experimental test of crime deterrence,” J Exp Criminol, vol. 7, no. 3, pp. 255–274, Sep. 2011, doi: 10.1007/s11292-011-9125-9.

[3] A. Glowacz, M. Kmieć, and A. Dziech, “Visual detection of knives in security applications using Active Appearance Models,” Multimed Tools Appl, vol. 74, no. 12, pp. 4253–4267, Jun. 2015, doi: 10.1007/s11042-013-1537-2.

[4] J. , Whitten, L. D. Bentley, and K. C. Dittman, “Systems Analysis and Design Methods 5e,” McGraw-Hill Higher Education, vol. 5, 2000.

[5] W. H. A, “Research design: Qualitative, quantitative, and mixed methods approaches 74(48), 9263–4267.,” 1991.

[6] C. . K. M, “Research designs and variables - sage research methods 85(28),” vol. 40, no. 2, pp. 145–158, 1998.

[7] U. Califonia State, “Quantitative Research - Research Methods Simplified - Research Guides & Library How-To at California State University Sacramento.” Accessed: May 25, 2023. [Online]. Available: https://csus.libguides.com/res-meth/quant-res

[8] M. J, “Research design and dissertation: Unstructured interview methods ,” pp. 28–85, 1995.

[9] F. J, “The relationship between interviewer-respondent rapport and data quality.,” 1993.

[10] A. L. O. A. O. T. M. O. Adeniyi, “Essentials of Business Research ,” Research Gate, 2011.

[11] E. A. O. T. Akinade, “Research Methods: A Pragmatic Approach for Social Sciences, Behavioural Sciences and Education,” 2009.

[12] J. A. Avwokeni, “Research Methods: Process, Evaluation & Critique. Portharcourt: UnicampusTutorial Services,” 2006.

[13] I. Etikan, “Comparison of Convenience Sampling and Purposive Sampling,” American Journal of Theoretical and Applied Statistics, vol. 5, no. 1, p. 1, 2016, doi: 10.11648/j.ajtas.20160501.11.

[14] J. W. Creswell, CRESWELLQualitative-Inquary-and-Research-Design-Creswell. 2013.

[15] A. Mugenda&Mugenda, “ Qualitative and Quantitative Research Methods.Nairobi: Acts Press ,” 2003.

[16] S. J. Niranjan et al., “Bias and stereotyping among research and clinical professionals: Perspectives on minority recruitment for oncology clinical trials,” Cancer, vol. 126, no. 9, pp. 1958–1968, Jan. 2020, doi: 10.1002/cncr.32755.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 2261–2269, Aug. 2016, doi: 10.48550/arxiv.1608.06993.

[18] T.-Y. Lin and S. Maji, “Improved Bilinear Pooling with CNNs,” Jul. 2017, [Online]. Available: http://arxiv.org/abs/1707.06772

[19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization.” [Online]. Available: http://cnnlocalization.csail.mit.edu

[20] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” Nov. 2013, [Online]. Available: http://arxiv.org/abs/1311.2901

[21] S. W. Khan et al., “Anomaly Detection in Traffic Surveillance Videos Using Deep Learning,” Sensors, vol. 22, no. 17, Sep. 2022, doi: 10.3390/s22176563.