National College of Ireland

Project Submission Sheet

Student Name:	Anif Zarus Andrew Thomas, Khamalesh Ramesh and Prashanth Kumar
Student ID:	x23307609, x23325216 and x23298201
Programme:	Master's in Data Analytics **Year:** 2024 - 2025
Module:	Deep Learning and Generative AI (MSCDAD_A)
Lecturer:	Mohit Garg
Submission Due Date:	7 Apr 2025
Project Title:	Detecting Suspicious Behavior with Deep Learning
Word Count:	3225

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 references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Anif Zarus Andrew Thomas, Khamalesh Ramesh and Prashanth Kumar

Date: 7 Apr 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

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

AI Acknowledgement Supplement

[Insert Module Name]

[Insert Title of your assignment]

Your Name/Student Number	Course	Date

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
[Insert Description of use]	
[Insert Sample prompt]	[Insert Sample response]

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

[Place evidence here]

Additional Evidence:

[Place evidence here]

Detecting Suspicious Behavior with Deep Learning

Anif Zarus Andrew Thomas
Master of Science in Data Analytics_A
x23307609@student.ncirl.ie

Khamalesh Ramesh
Master of Science in Data Analytics_A
x23325216@student.ncirl.ie

Prashanth Kumar
Master of Science in Data Analytics_A
x23298201@student.ncirl.ie

This study emphasis how Deep learning models and computer vision can be used in the field of image classification. The dataset that was used for this study was the UCF crime dataset, which contains images classified based on the crime. In the study, we focus on three crimes, that is shoplifting, vandalism and burglary. We aim to build models that is able to identify and capture images based on criminal activity.

We had used a Custom Convolutional Neural Network (CNN), ResNet50, and a CNN combined with Bidirectional Long Short-Term Memory (BiLSTM) networks to train the dataset and the following models achieved an accuracy of 97%, 94% and 99% in the same order. These findings highlight the effectiveness of deep learning in enhancing image classification tasks, particularly in the context of real-world applications such as crime detection and prevention.

Furthermore, we conducted an evaluation of model performance using metrics such as precision, recall, and F1-score, which helped us understand each model's strengths and weaknesses and showcasing how the models are beneficial for image classification tasks.

Keywords— Deep Learning, Image Classification, Computer Vision, UCF Crime Dataset, Convolutional Neural Networks (CNN), Bidirectional LSTM (BiLSTM), ResNet50

I. INTRODUCTION

Deep Learning is a subset of Machine Learning but unlike the latter, Deep Learning can train itself from large amounts of unstructured data. It processes data in a way that is similar to how the human brain processes data. Deep Learning consists of different models that are suited to handle different types of data and solve their unique objectives. It has been widely used in industries where there are a lot of unstructured data like images, video, 3D, audio and text. Every data that is collected is split into training and testing where the training data is fed into the model and testing data is used to perform predictions.

Computer vision is a subset of artificial intelligence that is used to extract information from raw images and videos. It is mainly used to recognize objects and detect oddities. The Convolutional Neural Network model is widely used for image analysis. This project explores the use of Deep Learning in the computer vision domain, by using the following models Custom CNN, ResNet50, CNN + BiLSTM to process and analyze image data. The dataset consists of images of different categories such as shoplifting, burglary, and vandalism. The data would be explored to analyze odd behaviors from people that could potentially lead to shoplifting, burglary, and vandalism.

The objectives of this project are as follows:

- To Identify suspicious behavior that could potentially lead to shoplifting, vandalism or burglary
- Exploring Deep Learning models such as CNN and Pre-trained CNN models to predict the outcome of the training data
- Evaluating the model performance

This paper includes the following sections:

- Related work: This section focuses on discussing literature reviews of papers that use similar data and models that are relevant to the project.
- Data methodology: This section focuses on the methodology by which the project is executed.
- Conclusion and future work: This section focuses on the overall summary of the project, the key findings and suggestions.
- References: This section includes citing all the resources utilized in this project.

II. RELATED WORKS

In the research paper [1], "Video anomaly detection system using deep convolutional and recurrent models" (2023) by M. Qasim and E. Verdu, the UCF crime dataset was used. The dataset contained 14 crime categories and one normal category. ResNet-18/34/50 was used for spatial extraction, and Simple Recurrent Units were used for the temporal modeler. 75% of the data was used for training purposes with ResNet-50 and Simple Recurrent Units. It achieved an accuracy of 91.24%, and AUC achieved an accuracy of 91.64%. While the CNN models performed better than other models, there were still some issues with wrong classification.

In [2], "Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach" (2024) by R. Kolaib and J. Waleed, MobileNet-V2 and EfficientNet-B7 were used on the UCF Crime dataset. MobileNet-V2 achieved 98% accuracy, whereas EfficientNet-B7 achieved 97% accuracy. This research emphasizes the use of annotated data to improve the accuracy of the model and develop new annotated data.

In [3], "Anomalous Human Action Recognition with Deep Learning Technique" (2024) by M. Pallewar, V. Pawar, and A. Gaikwad, this study focused on using hybrid CNN-LSTM deep learning on the UCF Crime dataset. Six types of crimes can be classified using this model (abuse, arson, assault, burglary, fight, and robbery). CNN was used for feature extraction, and sequence classification was done using LSTM. 80% of the data was used for training, 10% was used for testing, and 10% for validation. The model achieved an accuracy of 97.8%. This research pinpoints how the computational cost increases with the advancements of the models.

In [4], "Real World Anomalous Scene Detection and Classification using Multilayer Deep Neural Networks" (2021) by A. Jan and G. M. Khan, this study uses a hybrid CNN-LSTM model for training on data. The dataset contained various real-world hazards like accidents, fire, theft, and violence. Frame extraction and feature engineering were done prior to implementing the model. CNN was used for spatial feature extraction, and LSTM was used for temporal sequence learning. 70% of the data was used for training, 15% was used for validation, and 15% was used for testing. The model achieved an accuracy of 94.6%. The research lacked more real-time training data.

In [5], "Detection of Anomalous Events Based on Deep Learning - BiLSTM" (2022) by Z. K. Abbas and A. Al-Ani, the dataset contains surveillance videos with labelled anomalous events, which are pre-processed using frame extraction and data augmentation techniques. Before feeding into the BiLSTM for sequential classification, features were extracted through convolutional layers. This model achieved 94.2% accuracy, outperforming traditional LSTM architecture. Computational efficiency needs to be improved, especially when deploying in real-time surveillance systems.

In [6], "An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos" (2021) by W. Ullah et al., spatial features were extracted from video sequences using MobileNetV2 and processed through residual LSTM. UCF Crime, UMN, and Avenue were the datasets used to evaluate the model. The research also involved creating a manual model. The benchmark achieved 78.43%, 98.20%, and 98.0%, outperforming prior methods by 1.77%, 0.76%, and 8.62%, respectively. It emphasizes exploring 3D CNNs and graph neural networks for broader anomaly detection.

In [7], "Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach" (2024) by R. Kolaib and J. Waleed, a crime detection system was developed using surveillance footage, by using CNNs and EfficientNet-B7. It achieved an accuracy of 99.48% and an F1 score of 99.44% on UCF Crime and DCSASS datasets. Although it outperforms existing methods, the lack of high-quality video data in training the model could affect performance in real-world scenarios.

In [8], "Attention-based bidirectional-long short-term memory for abnormal human activity detection" (2023) by M. Kumar et al., they combined InceptionResNet-V3 for spatial feature extraction, Bidirectional LSTM (Bi-LSTM) for temporal modelling, and an attention mechanism to focus on critical spatiotemporal features. UCF11, UCF50, and sub-UCF Crime datasets were used and achieved accuracies of 98.9%, 96.04%, and 61.04%, respectively, outperforming existing methods like CNN+LSTM and I3D Siamese. Challenges faced include handling long, uncut surveillance videos with complex backgrounds.

In [9], "Deep BiLSTM Attention Model for Spatial and Temporal Anomaly Detection in Video Surveillance" (2025) by S. Nath et al., the Composite Recurrent Bi-Attention (CRBA) model combines DenseNet201 for spatial feature extraction, BiLSTM for temporal dependencies, and a multi-attention mechanism to focus on critical spatiotemporal regions. The UCF Crime dataset and a custom Road Anomaly Dataset (RAD) were used and achieved 86.2% and 92.2% accuracy, respectively. Challenges faced include computational complexity.

In [10], "Anomaly Detection in Video Surveillance using SlowFast Resnet-50" (2022) by M. Joshi and J. Chaudhari, a deep learning technique based on ResNet detects and categorizes surveillance accidents. Three types of ResNet networks like ResNet50, ResNet101, and ResNet152 were used to extract high-level spatial features from video frames, and recurrent networks were used for temporal analysis. An accuracy of 98.34% was achieved on UCF Crime and HWID12 datasets. Real-time processing requires attention as researchers planned to enhance computational speed to deploy systems for live surveillance systems.

In [11], "Residual Network (ResNet) Based Deep Learning Method for Detection and Classification of Accidents in Surveillance Scenes" (2023) by R. V. et al., three versions of the ResNet network were used (ResNet50, ResNet101, and ResNet152) to extract high-level spatial features from video frames and used a recurrent network for temporal analysis. An accuracy of 98.34% was achieved on UCF Crime and HWID12 datasets using ResNet152. Computation was an issue when trying to deploy in real-time systems.

In [12], "Understanding of Convolutional Neural Network (CNN): A Review" by P. Purwono et al., a deep learning method known as Convolutional Neural Networks (CNNs) is examined with respect to structure and functionality in this research paper. It explains CNN layers, including convolution, pooling, fully connected, and activation components, and their functionality for feature extraction. The paper discusses popular CNN architecture designs along with their application scope for LeNet, AlexNet, and VGGNet. CNNs are widely used for image processing and object detection, as well as natural language processing (NLP) and speech recognition applications because of their high accuracy and efficiency rates.

III. Data Methodology

We are using a CRISP-DM methodology since the approach it provides is suitable for our project. The CRISP-DM consists of 6 phases, and they are as follows.

A.Tools and Library

1) NumPy: This Library is used to handle numerical data and multi-dimensional array operations.

a) array(): Returns arrays.

b) reshape(): Reshape arrays to any given dimensions.

c) inspace(): Return evenly spaced values over an interval.

d) random.rand(): Returns random arrays.

2) Pandas: More suitable for handling and analyzing structured data, structured datasets.

a) readd_csv(): Imports CSVs to a dataframe.

b) DataFrame(): Handles creating and manipulating tables of data.

c) groupby(): Tranforms or aggregtes grouped data.

3) Matplotlib: Library to create plots and visualizations.

a) plot(): Draws line plot.

b) timshow(): Displays pictures.

c) bar(): Plots bar charts.

d) scatter(): Plots scatter plots.

4) Scikit-learn: A User Friendly library used for machine learning that provides data preprocessing tools such us modeling, evaluation, and metrics .

a) train_test_split(): Splits data into a training set and a test set.

b) StandardScaler(): Normalizes numerical features.

c) classification_report(): Prints precision, recall, F1 score, and accuracy for classifiers.

d) troc_auc_score(): Computes the Area Under the Curve (AUC) for Receiver Operating Characteristic (ROC).

e) GridSearchCV(): Optimizes hyperparameters with cross-validation.

5) OpenCV (cv2): Specialized in image processing and computer vision operations.

a) imread(): Reads images.

b) resize(): Resizes image sizes.

c) cvtColor(): Transforms images from one color space to another.

6) Keras: a library which facilitates neural network development and training model which provides numerous features like ANN layers, optimizers etc..

a) Used Layers:

- Conv2D: Extracts spatial features from images through convolution operations.
- MaxPooling2D: Reduces spatial sizes by keeping maximum values in a window.
- Dense: Fully connected layer that processes input for classification or regression.
- Dropout: Randomly deactivates neurons during training to avoid overfitting.
- BatchNormalization: Normalizes layer inputs to stabilize and speed up training.

b) Optimizer Used:

- Adam Optimizer: Combines momentum and adaptive learning rates for efficient training.
- Callbacks:
- EarlyStopping: Stops training if performance does not increase to avoid overfitting.
- LearningRateScheduler: Changes the learning rate dynamically during training.

7) TensorFlow.Keras: Provides advanced deep learning features.

a) Architectures:

- ResNet50: Pre-trained deep convolutional neural network architecture, ideal for image classification tasks.
- Used Layers:
- Flatten: Converts multidimensional data to a 1D vector for dense layers.
- Dense: Processes output features for classification or regression tasks.

8) Bidirectional LSTM: Processes sequences by utilizing forward and backward context, ideal for natural language processing and time-series tasks.

B. Business understanding:

The business understanding focuses on the objectives of our project which are as follows

- To Identify suspicious behaviour that could potentially lead to shoplifting, vandalism or burglary

By achieving the objective, we can

- Reduce financial losses due to shoplifting, burglary and vandalism
- Provide a safe environment for the customers

C. Data understanding:

The data understanding focuses on how the data is being collected, exploring the type of data and verify their quality

Data Collection:

The dataset is obtained from Kaggle. It goes by the name “UCF CRIME DATASET”. This dataset contains real world surveillance footages that is used for crime detection and behaviour analysis.

Exploratory Data Analysis :

The Dataset consists of 4 classes:

- Shoplifting
- Burglary
- Vandalism
- Normal

Class	Images per Class	File format	Pixels
Shoplifting	1738	PNG	64 * 64
Burglary	1099	PNG	64 * 64

Vandalism	1020	PNG	64 * 64
Normal	1061	PNG	64 * 64

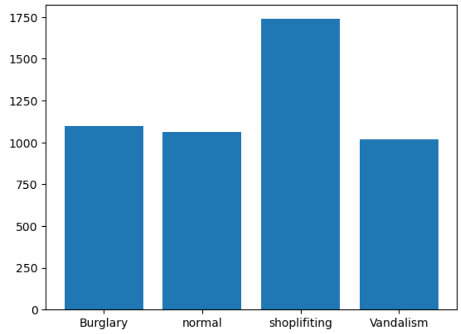

Figure 1- Quantity of classes

The image data was separated into different folders each consisting of the class names. There was little imbalance in the data due to footages being in sequence. All the file formats were stored in PNG and the pixel size of the images were 64 * 64.

D. Data Preparation:

The data preparation phase focuses on implementing pre-processing steps to clean the data and prepare it in such a way that it is suitable for modelling. Fine tuning the pre-processed data will enhance the accuracy of the model.

The images in the dataset that was obtained where already classified and split into test and training. Since there were constraints with training higher numbers of images due to lack of high-power computing, we had to pick limited number of images. The images in its original form were 64 * 64 in resolution. When training the model with 64 pixels, we were able to increase the number of images to train but there was a lack of accuracy from the model since it was not able to accurately predict the correct class. We had to resize the images from 64 pixels to 224pixels so that the model can train better. However, this led to limited training data.

Some of the data quality checks that was done when picking the images for modelling was if the image was corrupted, the quality of the image, if it had brightness and correctly focused the behaviour that we were trying to predict.

The pre-processed data is stored so the pre-processing steps does not need to be followed repeatedly. It is stored as an NumPy array and is directly loaded into the model.

E. Modelling:

To predict the behaviour, we used three different models, and they are as follows:

1.Custom CNN model:

We built a custom CNN model with relu activation function. We used Adam optimiser to adjust learning rate. Maxpooling was used to reduce spatial dimensions without losing crucial information. Dropout rates were adjusted to avoid overfitting

2. ResNet50 model:

Pre-Trained CNN model Resnet50 was used to predict the behaviour. The model was picked due to being resistant to the vanishing gradient problem.

3. CNN + BiLSTM:

This is a hybrid model where the CNN is used to deal with spatial features in sequence of data and BiLSTM captures the temporal dependencies

F.Evaluation:

The evaluation was done by comparing different metrics across the three models. These metrics were included since they showcased how well the models performed.

Model Evaluation	CNN	ResNET50	CNN+BiLSTM
Accuracy	0.9787	0.9429	0.9982
Loss	0.2806	0.2483	0.0012
Value Accuracy	0.9980	0.9268	0.8415
Value Loss	0.2210	0.2270	0.5008
Epochs	10	2	3
Precision	0.9987	0.9982	1.0
Recall	0.9998	0.8822	0.9995
F1 score	0.9987	0.9319	0.9998

The Accuracy metric is used as an indicator to showcase the overall performance of the model. The precision indicates the true positive prediction compared to the overall predictions. The F1 metric is the mean of the precision, and the recall showcases the reliability of the model.

The CNN model performed the best in terms of accuracy. The CNN + BiLSTM had the lowest loss which indicates that it is more suitable to the data. ResNet50 wasn’t prone to overfitting based on the recall metric.

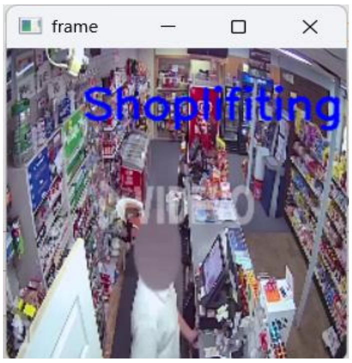

Figure 2

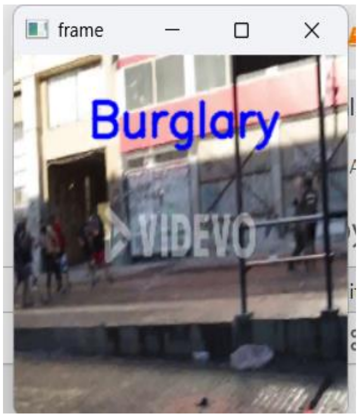

Figure 3

The RoC AUC curve illustrates the relationship between true positives and false positives for the model's outcomes. In our case, the curve approaches a value of 1, indicating that the model demonstrates great performance.

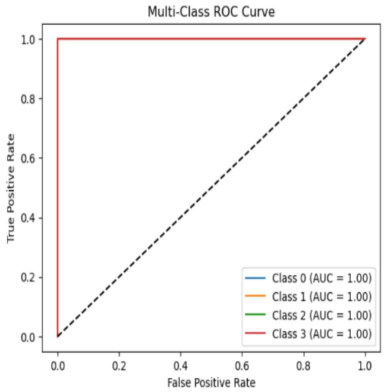

Figure 4

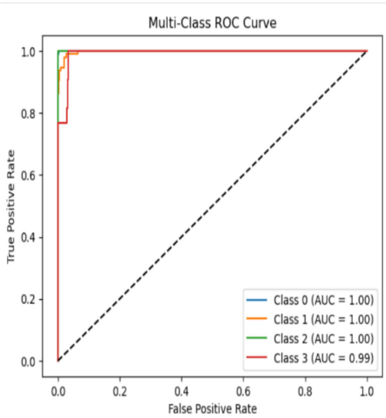

Figure 5

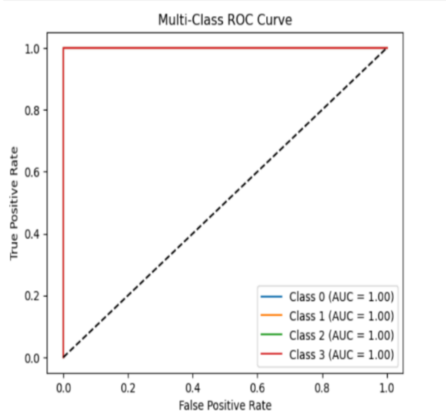

Figure 6

IV. Conclusion

In Conclusion, we were able to explore the various facets of deep learning while integrating the dataset. We were not able to use the full dataset due to computational constraints. Some of the steps we followed were data cleaning, data pre-processing, modelling and evaluation. As explained in the metrics, the CNN + BiLSTM was the better performer amongst the three model but required significant computational resources. The ResNet50 had a good balance between speed and accuracy, making it suitable for variety of scenarios. The suggestions for future works include real time testing using edge computing, using a lot more training dataset to make the model more efficient, tune the hyperparameters to reduce the computational time.

V. Reference

[1] Qasim, Maryam & Verdu, Elena. (2023). Video anomaly detection system using deep convolutional and recurrent models. Results in Engineering. 18. 101026. 10.1016/j.rineng.2023.101026.

[2] Kolaib, Rasool & Waleed, Jumana. (2024). Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach. Diyala Journal of Engineering Sciences. 98-114. 10.24237/djes.2024.17307.

[3] Pallewar, Megha & Pawar, Vijaya & Gaikwad, Arun. (2024). Anomalous Human Action Recognition with Deep Learning Technique. 904-910. 10.23919/INDIACom61295.2024.10498200.

[4] Jan, Atif & Khan, Gul Muhammad. (2021). Real World Anomalous Scene Detection and Classification using Multilayer Deep Neural Networks. International Journal of Interactive Multimedia and Artificial Intelligence. In Press. 1. 10.9781/ijimai.2021.10.010.

[5] Abbas, Zainab & Al-Ani, Ayad. (2022). DETECTION OF ANOMALOUS EVENTS BASED ON DEEP LEARNING-BILSTM. Iraqi Journal of Information and Communication Technology. 5. 34-42. 10.31987/ijict.5.3.207.

[6] Ullah, W., Ullah, A., Hussain, T., Khan, Z. A., & Baik, S. W. (2021). An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos. *Sensors*, *21*(8), 2811. https://doi.org/10.3390/s21082811

[7] Kolaib, Rasool & Waleed, Jumana. (2024). Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach. Diyala Journal of Engineering Sciences. 98-114. 10.24237/djes.2024.17307.

[8] Kumar, M., Patel, A.K., Biswas, M. *et al.* Attention-based bidirectional-long short-term memory for abnormal human activity detection. *Sci Rep* **13**, 14442 (2023). https://doi.org/10.1038/s41598-023-41231-0

[9] Natha, Sarfaraz & Ahmed, Fareed & Siraj, Mohammad & Lagari, Mehwish & Altamimi, Majid & Ali, Asghar. (2025). Deep BiLSTM Attention Model for Spatial and Temporal Anomaly Detection in Video Surveillance. Sensors. 25. 1-24. 10.3390/s25010251.

[10] Joshi, Mahasweta & Chaudhari, Jitendra. (2022). Anomaly Detection in Video Surveillance using SlowFast Resnet-50. International Journal of Advanced Computer Science and Applications. 13. 10.14569/IJACSA.2022.01310112.

[11] V, Rahul & S, Senthil & Sebastian, Livia & M, Manjusree. (2023). Residual Network (ResNet) Based Deep Learning Method for Detection and Classification of Accidents in Surveillance Scenes. 1-5. 10.1109/RMKMATE59243.2023.10368915.

[12] P. Purwono, A. Ma'arif, W. Rahmaniar, H. Imam, H. Fathurrahman, H. I. Karim, A. Frisky, and Q. M. U. Haq, "Understanding of Convolutional Neural Network (CNN): A Review," International Journal of Robotics and Control Systems, vol. 2, pp. 739-748, 2023, doi: 10.31763/ijrcs.v2i4.888.

[13] Schröer, Christoph & Kruse, Felix & Marx Gómez, Jorge. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science. 181. 526-534. 10.1016/j.procs.2021.01.199.

[14] Shimaoka, Andre & Ferreira, Renato & Goldman, Alfredo. (2024). The Evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. 10.13140/RG.2.2.22493.42721.