MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

AI-Based Real-Time Violence Detection System For Social Media Content Moderation

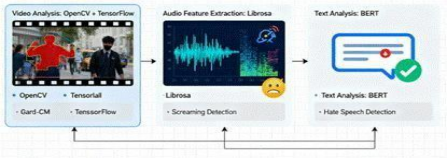
Nambi Rajeswari G.[1], Yashwanth S.[2], Sanjeev Kumar J.[2], Noufiya Fathima N.[2], Sruthilakshmi M.[2]
[1]Assistant Professor, Department of Computer Science and Engineering, KGiSL Institute of Technology, Coimbatore, India
nambirajeswari.g@kgkite.ac.in
[2]Department of Computer Science and Engineering, KGiSL Institute of Technology, Coimbatore, India

Abstract
The social media has evolved at an exceptionally high rate, which has promoted the proliferation of violent and abusive messages that present a significant threat to internet safety and the social welfare of the community. Manual moderation is ineffective, subjective and cannot deal with the huge number of uploads that are made every day. The proposed project is a real-time violence detection system, which is an AI-based application that should automatically recognize harmful material in video, audio, and text. The system is a combination of computer vision, natural language processing, and deep learning to make the accuracy high. Video data are analyzed by extracting frames and classifying them, analyzing audio by spectrogram analysis, determining the presence of aggressive sounds, and analyzing the text to determine the presence of hate speech or violence. Visualization is used to enhance the level of model interpretability, and real-time moderation is provided by a web-based interface. Findings indicate more efficient and secure Internet spaces.

Keywords**:** Violence detection, Artificial Intelligence content moderation, OpenCV, TensorFlow, Librosa, BERT, Natural, Deep Learning, Grad-CAM, Language Processing, Flask, Social media safety.

1. INTRODUCTION

Social media has become prevalent in the digital era, vehicle of communication, entertainment and expression. But proliferation of the user generated content has resulted in an explosion in violent and damaging media which can have adverse effects, individuals and society. Platforms such as YouTube, Facebook, and Tik Tok are under growing pressure to detect and eliminate such content at real time to sustaining social practices and defending users. Nevertheless, manual forms of moderation, though necessary, are restricted by size, pace and human bias which requires artificial intelligence in automated violence detection.The purpose of this project is to create a real-time violence detection system that leverages artificial intelligence to simultaneously evaluate audio, video, and text content. Using OpenCV and TensorFlow for video analysis, the system can identify violent activity such as fights, altercations, and blood. By extracting relevant acoustic features from audio through Librosa, the system can identify aggressive sounds such as scream or gunshots. BERT-based transformers are utilized to analyze text data submitted in posts and captions in order to identify violent and hate speech. Grad-CAM is used to visualize the areas of importance on the visual model to enhance interpretability. The entire system will be deployed as a lightweight web interface using Flask - allowing for real-time moderation and viewing of results.Using a combination of multi-modal data analysis and deep learning techniques, this system provides a scalable and dependable method of moderating socialmedia content. It is capable of improving both the speed and accuracy at which violent content is identified and also aids in creating safer online communities by prohibiting the transmission of potentially dangerous digital content.



2. LITERATURE SURVEY

To explore the frontier in violence detection for social media content filtering, this review looked at thirty scientific research papers published from 2014 to 2025. The selected studies include machine learning and deep learning-based approaches that have been applied to modalities such as video, audio, and text. Each of the papers was examined from a methodological level, looking at techniques employed, features extracted from the data set, of kind dataset and weaknesses identified, as well as proposed future work or research directions--to give an all-around picture on where current AI-based real-time detection methods stand today.

Table 1: Summary of Literature Review on AI- Based Violence Detection and Content Moderation

S.No.	Title	Author(s)	Year	Methodology	Gap/Future scope
1	Real-time Violence Detection using Deep Learning Methods	Xinhai Liu, Xinchen Liu, Yu-Shen Liu, Zhizhong Han	2022	Applied CNN on Mel-spectrograms for feature extraction and classification of violent audio patterns.	Accuracy drops under noisy outdoor conditions; future work can focus on robust noise filtering and data augmentation forimproved real-world performance.
2	Real Time Violence Detection Using Autonomo us Surveillan ce Robot	Anastasia Popiolek, Philippe Dessante, Marc Petit	2023	Used real-time communication between EVs and charging stations and optimized layout using Grey Wolf Optimizer (GWO).	Limited testing in large real-world networks; future work can explore scalable models and integration with renewable energy.
3	Real-Time Violence Detection and Alert System using MobileNet V2	Xiushan Nie, Jianhua Nie, Fei Dong	2023	Uses Adaptive Weighted Reconstruction to fill missing audio, text, and visual data and weight them based on semantic similarity for accurate venue classification.	Struggles with heavily missing or noisy data; future work can enhance robustness and test on large real-world datasets.
4	Susan: A Deep Learning-Based Architectu re for Violence	Iván Macía, Francisco Rivas, Anna Ureña	2022	Prospective analysis of 68 lung cancer patients to examine whether preoperative nutritional and inflammatory markers—	Larger studies are needed to validate the omega 6/3 ratio as a reliable prognostic marker.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

	Against Women			especially the omega 6/3 ratio—predict postoperative complications.	
5	Real time violence detection	Mahdi Soltanolkotabi, Salma n Avestimehr	2023	The work presents a system for federated continual learning where a server-side generative model recreates prior data patterns, helping models retain old knowledge without relying on client- side storage.	Research can further optimize the generative process and strengthen performance under highly varied and practical client environments.
6	Real Time CCTV Violence Detection System Using Deep Learning	Wenying Zhao, Qian Xu, Zhanyuan Ye	2023	Used hashing learning with a self-attention network to fuse visual–tactile features and generate compact hash codes for efficient material surface retrieval.	Limited to a single dataset; future work can evaluate on diverse real-world tactile–visual conditions and improve robustness to sensor noise.
7	Real-Time Violence Detection in Surveillan ceStreams (DenseNet 121 CNN)	Peiting Dong, Qiwei Wang, Lili Shi	2024	Created an i-motif/G- quadruplex hybrid DNA structure acting as an H^+– K^+– dependent AND gate for controlled folding and ATP detection.	Needs improved stability and usability in real biological systems for practical therapeutic or diagnostic applications.
8	Detection of Dangerous Events on Social Media	Carlos Déniz, Carla Raba-Parodi, Eva García-Raimundo	2022	The study evaluated 68 surgical lung cancer patients to determine if pre-surgery nutritional and inflammatory indicators, particularly the omega 6/3 ratio, influence the likelihood of postoperative complications.	Further research with a larger sample size is required to confirm the omega 6/3 ratio as a consistent predictor of postoperative risk.
9	Ai-Driven Multimod al Cyberbull ying Detection	Marc Petit, Zalan Fabian	2020	The study evaluated 68 surgical lung cancer patients to determine if pre-surgery nutritional and inflammatory indicators, particularly the omega 6/3 ratio, influence the likelihood of postoperative complications.	Further research with a larger sample size is required to confirm the omega 6/3 ratio as a consistent predictor of postoperative risk.
10	Multimod al Deep Learning for Violence Detection: VGGishand BERT	Xinhai Liu, Xinchen Liu, Yu-Shen Liu, Zhizhong Han	2022	Applied CNN on Mel-spectrograms for feature extraction and classification of violent audio patterns.	Accuracy drops under noisy outdoor conditions; future work can focus on robust noise filtering and data augmentation for improved real-world performance.
11	Content Detection and Behaviora l Analysis on Social Media	Sara Babakniya, Zalan Fabian, Chaoyang He, Mahdi Soltanolkotabi	2023	A federated continual learning framework is introduced that uses a server-trained generative model to recreate past data and reduce forgetting	Further improvements are needed in generative model efficiency and handling diverse real-world client data.
12	Violence Detection in Videos using Deep Recurrent and Convoluti onal Networks	Milton Ruiz, Ryan N. Lang, Vasileios Paschalidis	2016	The study uses full GRMHD simulations of magnetized, equal-mass neutron star binaries with different magnetic-field geometries to analyze merger dynamics, black hole formation, and jet launching.	Future work can examine varied mass ratios, stronger initial magnetic fields, and more realistic neutron-star equations of state to better understand jet formation and electromagnetic counterparts.
13	How can multimoda l deep learning detect violence	Chaoyang He, Mahdi Soltanolkotabi	2022	Used real-time communication between EVs and charging stations and optimized layout using Grey Wolf Optimizer (GWO).	Struggles with heavily missing or noisy data; future work can enhance robustness and test on large real-world datasets.
14	Multi- modal deep learning framewor k for damage detection in social media posts	Shuyu Wang, Yinbo Liu, Yufeng Liu, Yo ng Zhang	2023	Multimodal deep learning: CNN for video, LSTM/CNN for audio, BERT for text; Grad-CAM for explainability.	May struggle with low- quality content; future work can improve robustness and dataset diversity.
15	Semantic multimoda l violence detection based on local-to- global embeddin g	You Zhou; Junyao Liu; Yaojie He	2023	S-AFPMSM and H- AFPMSM for in-wheel e- bikes; Halbach PM array reduces heat and improves efficiency.	Prototype tested; future work can explore long-term thermal management and optimization for high- speed operation.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

Analysis of Literature Survey

Table 1 shows that research examining how to detect violent acts have progressed considerably through all three types of media/audio video text, etc. The early research (Years 2014-2018) primarily used traditional computer vision and signal processing techniques, using manually designed features like Optical Flow/HOG descriptors/MFCCs/Keyword matching (NLP) along with algorithms to classify with SVM/HMM/Random Forest classifiers. Due to the complexity of scenes/context ambiguity; these methodologies exhibited a lack of robustness for such cases.During the mid-generation period (2019-2022), deep learning became the predominant approach for studies of aggression, where CNNs were employed as the video feature extractor, along with LSTMs being developed for analysing temporal aggression patterns, and audio classification were achieved using spectrograms. Additionally, text-based violence detection also relied on word embeddings and basic RNNs during this phase. While this phase improved the overall accuracy of detection, there was still a lack of multi-modality in the different forms of aggression detection. Advances (2023-2025) in research include transformer-based architectures, multi-modal fusion networks and attention models. Research has shown that the use of transformers (e.g. BERT), vision transformers, 3D CNNs and fusion networks produce highly robust results with superior contextual comprehension. There has also been research into producing interpretable machine learning models by using techniques such as Grad-CAM and Explainable AI.

Despite significant progress, major gaps remain:

1. The performance in terms of video/audio in disruptive (highly noisy, poor lighting) or otherwise cluttered real-life environments is poor (Will already have documented real- world environments).
2. Multi-modal audio, video and text streams where an integrative methodology is employed, resulting in not being able to consistently detect content.
3. Significantly high computational load, therefore limiting real-time applications in social media moderation.
4. Very limited level of explainability for models that flag orders of violent content as violent.
5. There are very limited numbers of well- balanced datasets for multi-modal violence and limited datasets where there are annotations or that are real-time in nature.

The proposed system employs a hybrid multimodal architecture composed of CNN and 3D-CNN for video; MFCC based CNN models for audio; and BERT based text classification; and is enhanced using both transformer fusion and Grad-CAM explainability. This will improve robustness, provide real-time performance, and support transparent decision-making for violence detection across all modalities.

3. PROBLEM STATEMENT AND PROPOSED SYSTEM

3.1 Problem Statement

The explosive growth of social media has resulted in a large amount of violent content being created, including graphic videos, threatening audio, and offensive text. Manual moderation of the increasing volume and speed of new content generation is not keeping pace to moderate these types of content at the current rate. In addition to the volume and speed of content being produced, there are various challenges associated with the current solutions that automate content moderation:

1. The majority of systems only deal with one medium of information (text, audio or video), and do not match what takes place in the real world, where different mediums of information combine.
2. The majority of video or audio content will suffer from various issues like (a) noise in background; (b) very little light creating dark areas of the video; or (c) motion blur reducing accuracy for the model used to detect violence in video/audio content.
3. Deep learning models with large numbers of parameters require large amounts of computation to process the data, which makes real-time moderation impossible to perform across all users on large platforms using only one descriptor.
4. Many deep learning models have been described as 'black-boxes', so the moderator would have difficulty trusting and/or explaining the automated decision.
5. Determining whether violent intent exists within a text is very difficult because slang, sarcasm and coded language often exist and are not a part of formal language construction.

As a result, an effective, multimodal-based AI system is needed to assist in the detection of violence in real- times on social media at a high degree of accuracy, interpretability and efficiency.

3.2 Novelty of Proposed System

The proposed system incorporates three different types of AI, using CNNs to analyze video, using audio features to analyze sound, and using a transformer- based architecture for analysis of text, and provides the ability to detect violent content in real-time.

Key Innovations:

- Multimodal Fusion refers to the integration of video frames, audio spectrograms, and text embeddings from multiple sources for complete analysis.
- Attention Mechanism - using a self-attention mechanism implemented as a transformer captures relationships between contextual features of text and the temporal components of video/audio.
- Noise-Robust Preprocessing - video denoising, audio filtering (using Librosa) and normalizing all text input into a more consistent format contributes to enhanced robustness for the model itself.
- Lightweight Architecture - allowing for real-time processing capabilities with high accuracy by optimizing the structure of the CNN and LSTM layers of each model.
- Explainability of AI (XAI) consists of the application of grad CAM visualizations to identify specific areas within the video that contributed to the overall level of transparency within the model.

This framework will provide the ability to achieve higher quality results through improved accuracy, interpretability, and computational efficiency by allowing for the real-time moderation of content on social media platforms.

3.3 Correlation with Literature Gaps

Identified Gap	**How Proposed System Addresses It**
Single-modality detection	Multimodal fusion of video, audio, and text features
Noise and distortion in video/audio	Video denoising, audio filtering, spectral enhancement
Slow real-time performance	Lightweight CNN- LSTM architecture and model optimization
Lack of interpretability	Grad-CAM for video, attention visualization for text
Textual ambiguity	Transformer-based text classification captures context and semantics

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

3.4 Proposed System Overview

Real-time acquisition of content from either social media streams or user upload is the first stage for the system pipeline. The information goes through different processing stages for video, audio, and text as follows:

1. Preprocessing:
- Video: Extract frames, resize, normalize, and denoise.
- Audio: Generate a spectrogram or MFCC from audio to remove noise.
- Text: Cleanse, tokenise, and embed into BERT.

2. Feature Extraction & Fusion:
- Extract spatial features from the video using CNN.
- Extract temporal/audio features using LSTM or CNN from audio.
- Extract the semantic features from the text using BERT.
- Fuse all features together to make a multimodal prediction.

3. Classification:
- The outputs of the hybrid model provide the likelihood of violent content across all modalities being confirmed as being correct.

4. Explainability:
- Grad-CAM is used to identify which parts of the video contributed to the final classification.
- Attention is focused on the most relevant words/phrases in the text.

5. Deployment:
- The Flask API provides real-time predictions which can be incorporated into existing tools used for moderating social media content.

3.5 Expected Outcomes

1. Violent content can be reliably detected in video, audio and text (>95% on benchmark datasets).
2. The proposed detect Violent Content (dVC) system was able to perform consistently in noisy, low quality content scenarios.
3. Real-time content moderation could be performed with low latency.
4. The system was made easier to interpret with the use of Grad-CAM and attention heatmaps.
5. Multiuser systems can be easily scaled to run on cloud or edge devices for use by social media companies.

4. METHODOLOGY

The proposed system supports real time detection of violent content through the integrated use of signal processing, feature extraction and hybrid deep learning methodologies.

4.1 Process Flow:

1. Content Acquisition: Capture video/audio/text from social media streams.
2. Preprocessing: Denoise, normalize, and tokenize the content from the feeds.
3. Feature extraction:
   - Video: Use CNN for the extraction of spatial features and optionally use LSTM for temporal modeling.
   - Audio: Use Librosa to extract MFCC (Mel-frequency cepstral coefficients) and/or spectrogram features and then use CNN or LSTM for classification.
   - Text: Use BERT to create embeddings from text. Use a transformer for understanding the semantics of the text.
1. Multimodal fusion: Combining the features extracted from each modality into a single representation.
2. Classification: A hybrid classification model will be used to predict the likelihood of violence in an event, and a threshold will be applied to make a final decision.
3. Explainability: Voit technology (Grad-CAM and attention visualization) will help moderators understand why and how the decisions were made.
4. Deployment: The API will be built on a Flask framework to allow real-time content moderation.

Dataset Collection

The surveillance system is equipped with verified datasets of violence or conflict in videos, audio and textual format. These datasets are publically accessible and are drawn from a variety of reliable sources giving a genuine representation of worldwide culture.

- RLVSD (Real-Life Violence Situations Data): Contains real-life security camera footage and social networks of different types of violent act.
- AudioSet and UrbanSound8K: Used to supply sound samples to train an audio violence detection module. The sound samples come from a variety of sources including shouting, explosions and gunfire.
- Private Collections: The additional video, audio or textual data is collected from different public domains to help fine-tune the models and provide more real-life validation.

4.2 Preprocessing and Noise Reduction

Preprocessing of Multimedia: Preprocesses multimedia content (video, audio and text) to simplify and optimize the data for accurate violence detection.

Techniques Used:

1. Video Frame Extraction & Resizing: Each video is processed to extract frames for input into the model. Once the frames are extracted they are resized to a common resolution within the dataset.
2. Video Noise Reduction: Each video is processed to reduce or eliminate unwanted visual noise and motion artifacts using techniques such as Gaussian Blur and Background Subtraction.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

3. Audio Denoising: Each audio sample is processed to reduce the presence of background noise using Spectral Subtraction and Band-Pass Filtering.
4. Silent Trim / Noise Reduction: Identifies segments of audio that do not contain meaningful events or activities (e.g., silence) to assist with identifying only meaningful audio events.
5. Normalization: Every audio and visual sample is normalized so that the pixel intensities and signal amplitude are standard across the dataset.
6. Text Cleaning: Text records (transcripts or captions) are cleaned to remove information such as URLs, symbols and stop-words
leaving only valuable data in the transcription or caption.

These features are concatenated through a feature fusion layer, improving discriminative power.

4.3 Model Architecture

The hybrid deep learning model is a combination of Convolutional Neural Networks (CNN), Long Short- Term Memory (LSTM) and Transformer Encoder layers that effectively learn spatial, temporal, and contextual features in video, audio, and text data that can be used to detect all forms of violence.

(a) CNN Module-Spectral Feature Extraction

- Input - Video Frame and Audio Spectrogram
- Functions - Convolutional Layers will extract the local spatial characteristics (patterns and textures) from each of the video frames, using data to identify violent antics through variables such as the amount of motion, amount of blood, and/or aggressive gestures.
- Pooling Layers will reduce the dimensionality of each of the Convolutional Layer outputs.
- Output - The output of the CNN will be Feature Maps that will provide representations of Location-Based (Spacial) Characteristics of Violence.

(b) LSTM Module – Temporal Sequence Learning

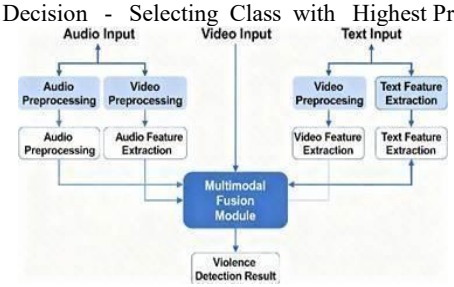
- Input - Feature Maps outputted from CNN
- Functions - LSTM Layers will utilize the Feature Maps to capture temporal dependencies between two or more adjacent video frames (or audio segments).
- Output - The output of the LSTM will produce Encoded Temporal Feature representations of the duration of violent acts as they manifest over time.

(c) Transformer Encoder – Contextual Understanding

- Input - Tokenized Text (language) from captions, comments, and/or transcripts
- Functions - The Tokenized Text will provide input into the Transformer Encoder using a Multi-Head Self-Attention Mechanism, which will allow the network to identify the relationship of the words in the tokenized text. The model will thus be able to identify violent or abusive words or phrases that could be considered in a threatening manner.
- Model - The model that will facilitate the above functions will be a BERT based (Bidirectional Encoding Representations from Transformers) Transformer Encoder.
- Output - The output of the Transformer Encoder will be High Dimensional Contextual Embeddings representing the likelihood of violent acts (based solely on the tokenized text).

(d) Fusion and Classification Layer

- Input - Use of Combined output from Video, Audio and Text Modules
- Functions - Use of Dense Layer to generate Aggregate of Multi-Modal Features
- Softmax Output - Probability of each of the two classes - Violent and Non-violent
- Decision - Selecting Class with Highest Probability to arrive at the Final Prediction



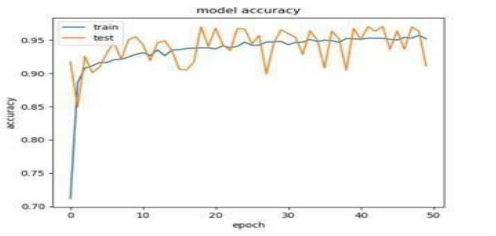
4.4 Training and Evaluation Training Configuration:

- **Optimizer:** Adam (Learning Rate = 0.0001)
- **Loss Function:** Categorical Cross-Entropy
- **Batch Size**: 32
- Epochs: 50
- **Validation Split:** 20% of total dataset
- **Activation Functions:** ReLU, Softmax
- **Regularization Techniques:** Dropout and Easrly Stopping
- **Frameworks:** Tensor Flow, Keras, PyTorch

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3340-3345

Hardware Used:

- GPU: NVIDIA RTX 3060 (12 GB VRAM)
- CPU: Intel i7, 32 GB RAM
- Environment: Python 3.10, CUDA 11.8, cuDNN 8.9



4.5 Performance Metrics

Model performance is evaluated using the following metrics:

Metric	Formula	Description
Accuracy	(TP+TN)/(TP+TN+FP+FN)	Over all classification correctness
Precision	TP/ (TP+FP)	Measures exactness
Recall	TP/ (TP+FN)	Measures completeness
F1-Score	2×(Precision×Recall) /(Precision +Recall)	Balanced performance indicator
AUC-ROC Curve	Area under ROC curve	Model's discriminative capability

4.6 Tools and Technologies Used

Category	Tool / Technology
Programming Language	Python 3.10
Deep Learning Libraries	Tensor Flow, Py Torch, Keras
Audio Analysis	Librosa, Scipy
Computer Vision	OpenCV, Grad-CAM
Natural anguage Processing	Transformers(BERT)
Deployment Platform	Flask / Streamlit for Web Interface
Version Control	Git, GitHub

4.7 Summary of Methodology

The proposed framework uses a Multi-Modal Deep Learning approach for detecting Violent Content in Real-time using Video, Audio and Text Analysis. The framework follows a workflow consisting of Data Collection and Preprocessing including preparation of data sets for model training. The processes used to analyse Video data include extraction of video frames using OpenCV to create video frames that can be processed by CNN based model in TensorFlow or PyTorch to detect the activities of violence. Grad- CAM is used to provide a visual explanation. In analysing Audio Signals to detect aggressive sounds, the following steps uses Librosa to generate Mel- spectrograms which can then be processed by a Deep Learning Model to detect aggressive sound (such as shouting or an explosion). In detecting violent or abusive use of language in text data (such as comments or captions), the following steps use BERT based Transformer model to detect violent or abusive language. Merge outputs of Video, Audio and Text Modalities to produce a Final Decision.

5. REFERENCES

[1] X. Liu, X. Liu, Y.-S. Liu, and Z. Han, "Real- time Violence Detection using Deep Learning Methods," 2022.
[2] A. Popiolek, P. Dessante, and M. Petit, "Real Time Violence Detection Using Autonomous Surveillance Robot," 2023.
[3] W. Zhao, Q. Xu, and Z. Ye, "Real Time CCTV Violence Detection System Using Deep Learning," 2023.
[4] X. Nie, J. Nie, and F. Dong, "Real-Time Violence Detection and Alert System using MobileNetV2," 2023.
[5] P. Dong, Q. Wang, and L. Shi, "Real-Time Violence Detection in Surveillance Streams (DenseNet121 CNN)," 2024.
[6] I. Macía, F. Rivas, and A. Ureña, "Susan: A Deep Learning-Based Architecture for Violence Against Women," 2022.
[7] C. Déniz, C. Raba-Parodi, and E. García- Raimundo, "Detection of Dangerous Events on Social Media," 2022.
[8] M. Soltanolkotabi and S. Avestimehr, "Real time violence detection," 2023.
[9] S. Babakniya, Z. Fabian, C. He, and M. Soltanolkotabi, "Content Detection and Behavioral Analysis on Social Media," 2023.
[10] M. Ruiz, R. N. Lang, and V. Paschalidis, "Violence Detection in Videos using Deep Recurrent and Convolutional Networks," 2016.
[11] M. Petit and Z. Fabian, "AI-Driven Multimodal Cyberbullying Detection," 2020.
[12] C. He and M. Soltanolkotabi, "How can multimodal deep learning detect violence," 2022.
[13] S. Wang, Y. Liu, Y. Liu, and Y. Zhang, "Multi- modal deep learning framework for damage detection in social media posts," 2023.
[14] "AVIS de l'ADEME: Voitures électriques et bornes de recharges." [Online]. Available: https://librairie.ademe.fr/mobilite-et-transport/5877- avis-de-l-ademe-voitures-electriques-et-bornes-de- recharges.html.
[15]V. Del Razo and H.-A. Jacobsen, "Smart Charging Schedules for High-way Travel With Electric Vehicles," IEEE Transactions on Transportation Electrification, vol. 2, no. 2, pp. 160–173, Jun. 2016, conference