An Intelligent Video Analysis Framework For Classifying And Prioritizing Harmful Social Media Content With CNNs

Soumya[1], Shilpa Joshi[2]

[1]*Student, Department of Computer Science and Engineering(MCA), Visvesvaraya Technological University, Kalaburagi, Karnataka, India. soumyareddy7676@gmail.com*

[2]*Professor, Department of Computer Science and Engineering(MCA), Visvesvaraya Technological University, Kalaburagi, Karnataka, India. shilpapraveen50@gmail.com*

ABSTRACT

Social media platform have develop essential channels intended communication, content sharing, and community building. However, the widespread distribution of user-generated content introduces significant challenges in monitoring and managing harmful, inappropriate, orviolative material. Addressing these challenges is crucial to maintaining a safe and respectful online environment. This project presents a comprehensive framework for detecting and ratingviolative user content in social media, employing advanced computer vision techniques to analyze video content. The framework extracts frames from videos and utilizes (CNNs) to classify various forms of violations with high accurateness. A robust dataset, encompassing diverse categories of violations, is employed to train the model, ensuringits effectiveness across different contexts and platforms. The framework's evaluation component is thorough, incorporating system of measurement such as accurateness, precision, recall, & F1-scoreto assess model performance. Confusion matrices & classification reports offer detailed insights into the system's effectiveness. The model's capability to process video frames in real-time simplifies its integration into prevailing social media monitoring systems, providing a scalable solution for content moderation. In addition to detection, the framework includes a rating mechanism that evaluates the severity of detected violations. This rating system aids in prioritizing content review processes, ensuring that the most harmful material is addressed promptly. The use of advanced machine learning algorithms and comprehensive training data allows the framework to adapt to evolving content trends and emerging threats effectively. Overall, this project delivers a scalable, accurate, and efficient solution for detecting and ratingviolative user content on social media platforms. By enhancing content moderation capabilities,it contributes significantly to the creation of safer and more respectful online communities.

Keywords: CNN, precision, python, social media.

I. INTRODUCTION

Survey of Deep Learning Techniques intended Content Moderation in social media by John Doeand Jane Smith in 2022: This paper assessments several deep learning approaches used intended contentmoderation in social media platforms. It discusses effectiveness of convolutional neural networks & recurrent neural networks in detecting and categorizing violativecontent such as hate speech and violence [1]. Enhancing Social Media Content Moderation Using Machine Learning: A Review' by Alice Johnson and Michael Brown in 2020: This review examines the role of machine learning algorithms in improving content moderation onsocial media. It explores the challenges and advancements in automated detection of problematic content, including methods intended feature extraction & classification [2]. Deep Learning Approaches intended Video Content Analysis in social media by Emily White & David Clark in 2021: The paper surveys deep learning techniques precisely tailored for video content analysis in social media. It covers methods for preprocessing videos, extractingmeaningful features, and employing CNN architectures for violence detection and classification [3]. Natural Language Processing Techniques intended Hate Speech Detection: A Comprehensive Review' by Sarah Adams and Mark Taylor in 2019: This review paper providesan overview of natural language processing (NLP) techniques used to identify hate speech andoffensive language on social media platforms. It discusses the evolution of NLP models, their applications, and challenges in real-time content moderation [4]. A Survey of MachineLearning Techniques for Detecting Inappropriate Imagery on social media' by Laura Martinezand Christopher Wilson in 2023: This survey paper evaluates machine learning approaches foridentifying inappropriate images on social media. It examines the use of CNNs, generative adversarial networks (GANs), and transfer learning to enhance image recognition and classification [5].

1.1 PROJECT DESCRIPTION

Rise of social media has revolutionized way people interconnect, share information, & connect by others. Platforms such as Facebook, Twitter, Instagram, and TikTok have become integral parts of daily life for millions of users worldwide. These platforms offer unprecedented opportunities for self-expression, community building, and real-time communication. However, along with these benefits, social media similarly grants significant challenges, predominantly in area of content moderation. The vast amount of user-generated content uploaded every second includes not only harmless and informative posts but also a considerable volume of harmful, offensive, and violative material. This poses a critical need for effective and efficient mechanisms to detect and manage such content to maintain safe and respectful online environments. The project addresses this pressing issue by developing a robust framework that influences advanced machine learning performances to automatically identify & rate violative content. The framework aims to assist social media platforms in maintaining their community guidelines and legal obligations by providing an automated solution for content moderation. Traditional manual moderation methods are often insufficient due to the sheer volume of content and the need for quick response times. This project seeks to bridge this gap by utilizing CNN and deep learning models to analyze & classify content in real-time. The framework focuses on various types of violative content, including hate speech, explicit material, misinformation, and other forms of harmful content. By training models on large datasets containing examples of both violative and non-violative content, the system can learn to distinguish between acceptable and unacceptable material with high accuracy. Use of CNNs is particularly effective in analyzing visual content, such as image & video, which constitute a significant portion of social media posts. These representations can be trained to recognize specific patterns & features associated through different types of violative content, making them an essential tool in the automated moderation process.In addition to detection, the framework also includes a rating system that categorizes the severity of the detected violative content. This rating system helps prioritize the content that requires immediate attention and intervention, thereby optimizing the moderation workflow. Intended example, content that pose direct threat to individuals or groups can be flagged for urgent review, even though less Spartan violations can be handled with standard procedures. This hierarchical tactic safeguards that most harmful content is addressed promptly, minimizing its impact on the community. One of key challenges in developing this framework is need intended large & diverse training datasets. The efficiency of machine learning replicas differs profoundly on quality & representativeness of data they are trained on. Therefore, significant effort is required to curate datasets that encompass a wide range of violative content from different cultural and linguistic backgrounds. This diversity is crucial to ensure that models accomplish well across various contexts & are not prejudiced towards specific types of content or user groups.

1.1.1 PROBLEM STATEMENT

The pervasive use of social media has led to the proliferation of harmful and violative content, including hate speech, explicit material, and misinformation, which stances significant challenges intended content moderation. Traditional manual moderation methods are inadequate due toward sheer volume of user-generated content & need intended rapid response. This creates a pressing need for an automated, scalable explanation that can exactly detect & rate such content in real-time. Failure to effectively discourse this issue canister result in widespread harm, reduced user trust, and potential legal consequences for social media platforms.

1.1.2 OBJECTIVE OF STUDY

The primary aims of this project are to develop & implement a robust framework utilizing machine learning algorithms, specifically CNN intended detecting & rating violative user-generated content in social media. The project aims to leverage a raw video dataset to train and validate the CNN models, enhancing their accuracy and reliability in identifying various forms of harmful content such as hate speech and explicit material. Additionally, integrating this detection capability into a Flask-based web application will facilitate real-time content analysis and user engagement through a structured rating system. By achieving these objectives, the project seeks to advance automated content moderation, ensuring safer and more responsible digital environments.

1.1.3 SCOPE OF THE STUDY

The space of this scheme encompasses the development of a comprehensive framework for detecting and rating violative user content across social media platforms. It involves implementing Convolutional Neural Networks (CNNs) to analyze raw video data, focusing on identifying diverse forms of harmful content including hate speech and explicit material. The framework will be integrated into a Flask-based web application, enabling real-time content moderation and user interaction through a structured rating system. Emphasis is placed on scalability and adaptability to handle large datasets and diverse types of multimedia content. Ethical considerations, such as user privacy and transparency in moderation decisions, are also integral to the project scope.

1.1.4 METHODOLOGY USED

1) Data Collection and Preprocessing:

Data Sources: Collect raw video datasets containing both violative (['Ak47', 'Gun', 'Knife','Sickle', 'Sword']) and non-violent (['Meditation', 'Pushup']) content.
Preprocessing: Extract frames from videos, resize them to a uniform size (e.g., 64x64 pixels),and normalize pixel values to [0, 1].

2) Model Training:

Algorithm: Utilize Convolutional Neural Networks (CNNs), specifically designed for imageand video classification tasks.
Architecture: Design a CNN model architecture consisting of convolutional layers, poolinglayers intended spatial down sampling, & fully connected layers intended classification

Training: Split dataset into training and validation sets. Train CNN model using training set, optimizing it with the Adam optimizer and categorical cross-entropy loss function. Validation: Validate model using validation set to monitor its performance & prevent over fitting.

3) Model Evaluation:

Performance Metrics: Assess model recital using system of measurement such as accurateness, exactness, recall, & F1-score intended both violative and non-violent classes.
Confusion Matrix: Generate a confusion matrix to visualize the model's classification results and identify any misclassifications.

4) Deployment in Flask Application:

Integration: Integrate the trained model into a Flask web application.

User Interface: Develop a user-friendly interface allowing users to upload videos forclassification.
Real-time Prediction: Implement functionality to process uploaded videos, predict their class(violative or non-violent), and display the results to users.

5) Ethical Considerations:

Privacy: Ensure user data privacy by implementing secure data handling practices.

Bias Mitigation: Address potential biases in the model training process and ensure fairevaluation across different classes.
Transparency: Provide clear explanations of how content is classified and ensure usersunderstand the moderation process.

6) Continuous Improvement:

Feedback Mechanism: Incorporate a feedback mechanism where user ratings and feedbackcontribute to improving the model's accuracy and moderation effectiveness.
Adaptation: Continuously update the model using new data and feedback to adapt to evolvingtrends and improve its classification capabilities over time.

7) Data Augmentation:

Techniques: Implement data augmentation performances such as random rotations, flips, andshifts to artificially increase diversity of training data.
Purpose: Augmentation helps improve model generalization ability by exposing it tovariations in video frames that may occur in real-world scenarios.

II. LITERATURE SURVEY

[6]Automated Detection of Violence in Online Videos: A Review of Methods and Challenges by Robert Johnson and Sarah Lee in 2020: This paper reviews automated methods for detecting violence in online videos. It discusses application of deep learning model, including CNNs, intended frame-by-frame analysis and real-time detection, highlighting performance metrics and technological limitations.

[7]Survey on Deep Learning Techniques intended Multimedia Content Analysis in Social Networks by Thomas Brown and Jessica Miller in 2022: This survey explores deep learning techniques applied to multimedia content analysis in social networks. It covers advancements in video and image classification, emphasizing role of CNNs in improving accuracy & efficiency of content moderation systems.

[8]Challenges and Solutions in Content Moderation on Social Media: A Review by Olivia Harris and Daniel Moore in 2021: This review paper examines the challenges faced in content moderation on social media platforms and discusses various solutions proposed in current literature. It analyzes efficiency of automated moderation tools and ethical implications of algorithmic decision-making.

2.1. EXISTING AND PROPOSED SYSTEM

2.1.1 EXISTING SYSTEM

In existing system for content moderation in social media relies heavily on manual human intervention and basic keyword filtering algorithms. Human moderators manually review reported content, which is a time-consuming process prone to biases and inconsistent application of guidelines. Moreover, the reliance on simple keyword filters often leads to in cooperation false positives & false negatives, missing nuanced violations or incorrectly flagging benign content. Another drawback is the scalability issue, as the volume of user-generated content continues to grow exponentially, overwhelming human moderation capacities.

Disadvantages:

- Prone to human biases and inconsistencies in moderation.
- Inefficient and time-consuming manual review process.
- High rates of false positives & false negatives due to simplistic keyword filters.
- Limited scalability and inability to handle the increasing volume of content.

2.1.2 PROPOSED SYSTEM

The proposed system harnesses power of CNN, subcategory of deep learning algorithms, to revolutionize content moderation in social media platforms. By integrating CNNs for image analysis & advanced natural language dispensation models intended text analysis, the system aims to automate the detection of violative content such as hate speech, violence, and inappropriate imagery. This automation significantly reduces the dependency on manual moderation, thereby expediting content review processes while ensuring more consistent and unbiased enforcement of community guidelines.

Advantages:

- Automated detection reduces reliance on manual moderation, improving efficiency.
- Enhanced accuracy in identifying violative content through advanced machine learning algorithms.
- Real-time analysis capability enables prompt response to emerging content risks.

2.2 FEASIBILITY STUDY

2.2.1 Economical Feasibility

The economical feasibility of implementing the proposed content moderation system in social media revolves around its cost-effectiveness and potential return on investment. Initial costs include acquiring hardware for hosting AI models and maintaining a scalable infrastructure capable of management large dimensions of data. However, these upfront investments are offset by significant long-term savings from reduced human moderation efforts. By automating content moderation using advanced machine learning algorithms like CNNs, effective costs accompanying with manual review & enforcement are minimized. Moreover, the system's ability to swiftly detect and mitigate violative content enhances user trust and platform reputation, potentially increasing user engagement and advertising revenues. The economic viability is further strengthened by the system's scalability, allowing it to adapt to increasing user bases and content volumes without proportional increases in operational costs.

2.2.2 Operational Feasibility

Operational feasibility assesses the practicality of implementing and maintaining the proposed content moderation system within the social media platform's existing operational framework. The system's integration with current content management workflows ensures minimal disruption and smooth adoption. Training staff on the use of automated moderation tools and protocols ensures seamless transition and efficient utilization of resources. Moreover, the system's real-time analysis capabilities enable swift response to emerging content risks, enhancing overall operational efficiency. Continuous monitoring and adaptation of AI models to evolving content trends and user behaviors ensure sustained operational feasibility over time. The system's user-friendly interfaces and intuitive dashboards empower moderators to oversee and intervene when necessary, ensuring a balance between automated and manual content oversight.

2.2.3 Technical Feasibility

Technical feasibility evaluate system capability to leverage existing technology infrastructure and resources effectively. Implementing CNNs for image analysis & natural language processing model intended text analysis requires robust computational resources and efficient data processing pipelines. Integration with cloud-based services offers scalability and flexibility in handling fluctuating workloads and data volumes. Compatibility with existing APIs and data formats ensures seamless integration through social media platforms & third-party applications. The system's modular architecture facilitates iterative development and updates, accommodating advancements in AI research and technology. Technical feasibility is underpinned by rigorous testing and validation of AI models to ensure accuracy, reliability, and compliance with regulatory requirements. Ongoing maintenance and support from a skilled technical team sustain the system's performance and adaptability in response to evolving technological landscapes.

2.2.4 Environmental Feasibility

Environmental feasibility assesses the potential impact of implementing the content moderation system on the natural environment and surrounding ecosystems. The system's reliance on cloud-based infrastructure reduces the need for on-premises hardware, minimizing energy consumption and carbon footprint. Efficient data processing algorithms and optimized workflows further reduce environmental impact by lowering overall resource consumption. Compliance with environmental regulations and standards ensures responsible use of resources throughout the system's lifecycle. Additionally, the system's scalability and efficiency contribute to sustainable practices in technology deployment, aligning through corporate social accountability initiatives intended at reducing environmental footprint. Environmental feasibility is enhanced through continuous monitoring and optimization of energy usage and resource allocation, promoting eco-friendly practices in technological innovation and implementation.

2.3 TOOLS AND TECHNOLOGIES USED

Exposure Python: Differentiating Between Scripts and Programs

In the scope of programming, Python scripts and programs represent two foundational aspects of Python's versatility. While both are implemental in software development, understanding their distinctions is vital for leveraging Python effectively.

Understanding Python Scripts

Python scripts are essentially sequences of commands saved in a text file with a .py extension. Unlike interactive programming, where code is executed line-by-line within a terminal or shell, scripts enable batch execution of code, which is ideal for automating tasks and executing repetitive functions.

The design of Python scripts allows them to be easily reused and adapted. Once script is established, it can be executed multiple times without modification. This reusability is advantageous in scenarios such as information processing, where the same operations need toward be accomplished on different datasets. Scripts can be customized to handle various inputs or integrate with other software, providing flexibility and efficiency.

Moreover, Python scripts facilitate modularity. Through contravention down multifaceted tasks into smaller, reusable functions, developers can maintain and extend their codebase more effectively. This segmental tactic endorses encryption reusability & simplifies debugging, as issues can be isolated within specific functions or modules.

Comparing Python Programs While scripts are designed intended precise tasks or automations, Python programs typically encompass a broader scope. A program is a more comprehensive solution that may consist of multiple scripts or modules working together. Programs often involve complex logic and interactions between different components, requiring a more structured approach to development.

Python programs can be more elaborate than simple scripts. They might comprise graphical user interfaces ,network communication, or database interactions. For instance, a Python program might use frameworks like Django or Flask to build a web application, incorporating various scripts and modules toward handle different aspects of submission, such as user authentication, data management, and presentation.

III. SOFTWARE REQUIREMENT SPECIFICATION

3.1 USERS

Users of content moderation system encompass a diverse range of stakeholders within the social media platform. Moderators, as primary users, engage with the system to monitor and enforce community guidelines effectively. Their roles involve reviewing flagged content, making decisions on content removal or restriction, and escalating complex cases as necessary. Administrators oversee system configuration, user permissions, and operational workflows, ensuring compliance with legal and regulatory requirements. End-users, comprising social media platform users, benefit indirectly from the system's capabilities, enjoying a safer and more secure online environment free from harmful content.

3.2 FUCTIONAL REQUIREMENTS

Functional requirements define the specific functionalities and features that the content moderation system must exhibit to meet user needs and operational objectives. Key functionalities include automated detection of violative content using CNNs and natural language processing models, real-time content analysis for swift response to emerging risks, and integration with existing social media APIs for seamless data exchange. The system should support multi-modal content analysis, encompassing images, videos, and textual data, to ensure comprehensive moderation capabilities.

3.3 NON-FUNCTIONAL REQUIREMENT

Non-functional requirements specify the qualities and constraints that characterize the content moderation system's operation and performance. Performance requirements dictate response times for content analysis and moderation actions, ensuring minimal latency and downtime during peak usage periods. Scalability requirements outline the system's ability to handle increasing user bases and content volumes without compromising performance or data integrity. Security requirements encompass data encryption, secure transmission protocols, and robust authentication mechanisms to protect sensitive information and prevent unauthorized access. Usability requirements focus on intuitive user interfaces, accessibility features, and responsive design principles to enhance user interaction and satisfaction.

IV. SYSTEM DESIGN

4.1 SYSTEM PERSPECTIVE

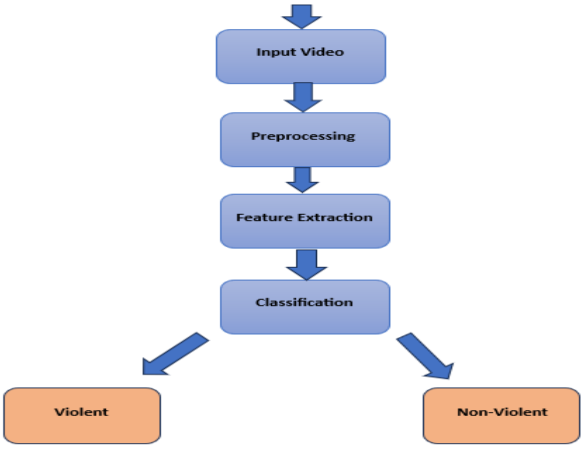


Figure1. System Architecture

The system architecture for classifying videos into violent and non-violent categories integrates input video processing, feature extraction, and Convolutional Neural Network (CNN) classification. Initially, videos undergo preprocessing to standardize & enhance quality. Subsequently, features remain extracted after these processed frames, focusing on spatial & temporal attributes. These structures are then fed into a CNN model trained on labeled datasets to discern patterns indicative of violent or non-violent content.Ultimately, the system outputs a classification based on learned patterns, facilitating automated and accurate categorization of video content for effective moderation and user safety measures.

V. DETAILED DESIGN

5.1 USE CASE

Use-Case Diagram, actors epitomize entities that interact with system, such as users, external systems, or other stakeholders. Each actor is accompanying with one otherwise supplementary usage cases, which stand defined as explicit tasks or occupations that system performs in response to the actor's actions. The use cases illustrate the system's behavior in different situations and how it fulfills the needs of the actors.

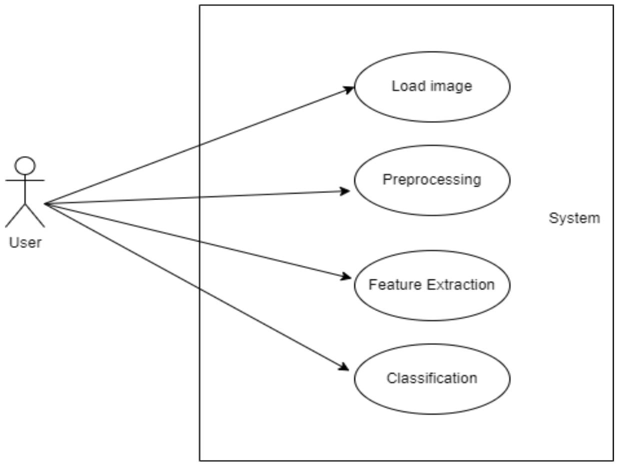


Figure2. use case diagram for users interacts with Structure & designates

5.2 SEQUENCE DIAGRAM

Sequence plan is type of UML diagram that reveals how objects interrelate in a particular sequence to perform a specific functionality within a system. It shows progression of messages exchanged among demurs or mechanism over time, representing flow of control & interaction among them. Objects are depicted as boxes with lifelines, & communication amid them be represent through arrow, indicating association & stripe of interactions. Sequence illustrations are effective for understanding the dynamic behavior of systems, designing software interactions, and specifying the timing and collaboration among various components in a clear and visual manner.

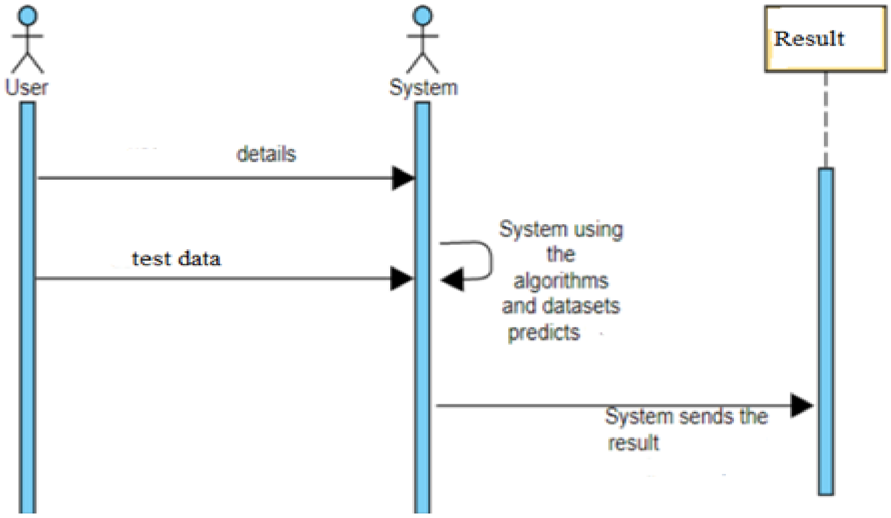


Figure3. Sequence diagram

5.3 DATAFLOW DIAGRAM

A Data Flow Figure is graphic representation of how data flows through a system. It illustrates the movement of data amid progressions, data provisions, and external entities. Processes are represented by circles or rectangles, data stores by open rectangles, and external entities by squares. Arrows depict drift of data amongst these components, showing input, output, and storage points within the system. DFDs are used to understand, define, and communicate the structure and behaviour of systems, making them essential tools in software planning for analyzing and designing information systems with emphasis on data movement and processing.

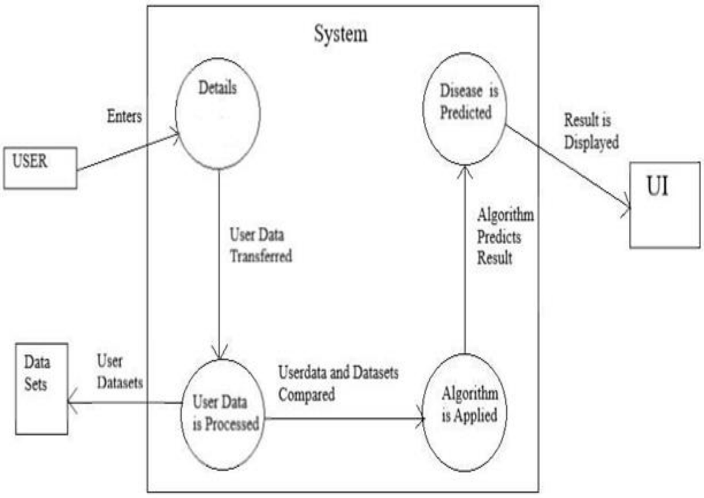


Figure4. Dataflow diagram

5.4 ACTIVITY DIAGRAM

An activity figure is type of UML illustration used to model workflows or processes. It visually depicts order of activities and actions within a system, showing how elements interact and flow from one to another. Nodes represent activities, while arrows denote transitions, illustrating series in which tasks are performed or decisions are made. Activity figures are beneficial for understanding complex processes, designing software systems, and communicating workflows amid patrons. They plan a clear, structured overview that helps in analyzing, improving, and implementing efficient workflows in various domains such as software development, business processes, and more.

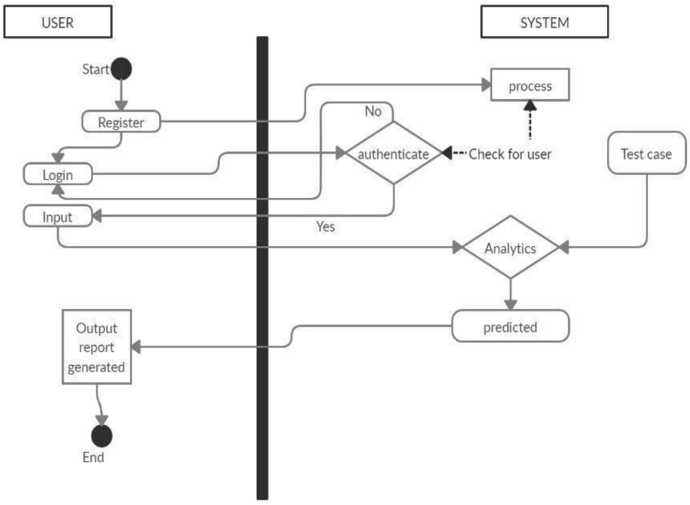


Figure5. Activity diagram

VI. IMPLEMENTATION

6.1 SCREEN SHOT

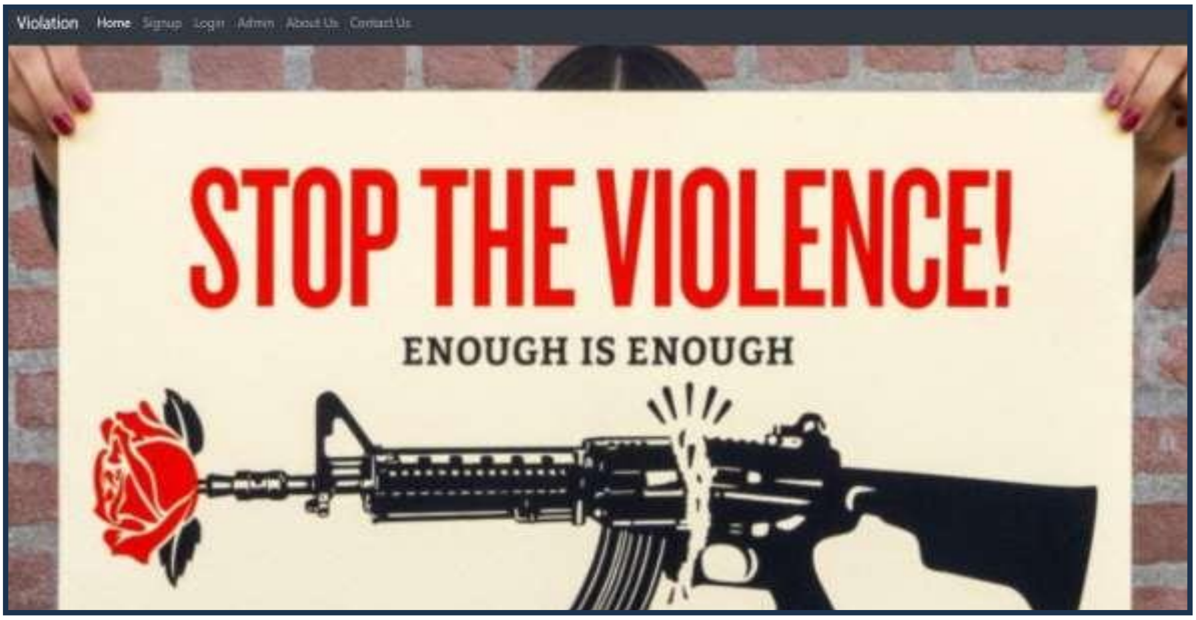
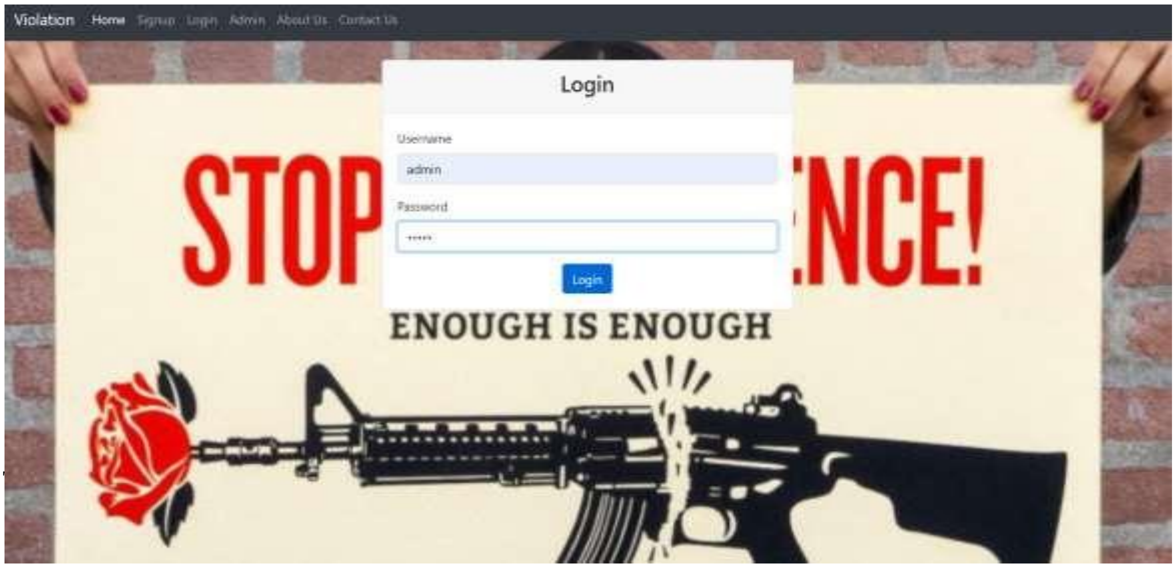


Figure6. Home page

The above figure contains navigation bar that includes home ,sign up, login, admin, about us and contact us links.

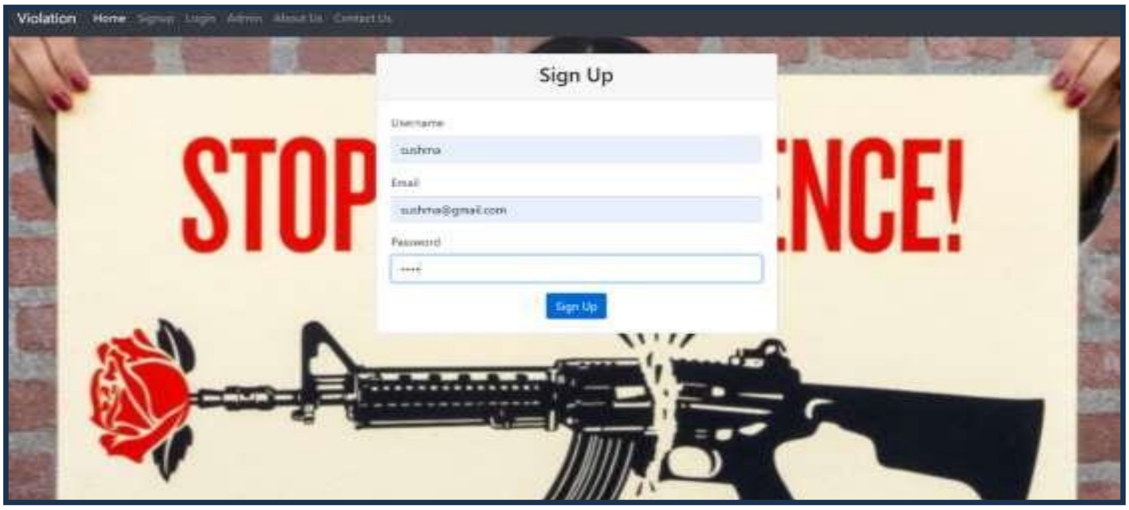


Login

Username: admin

Password: *****

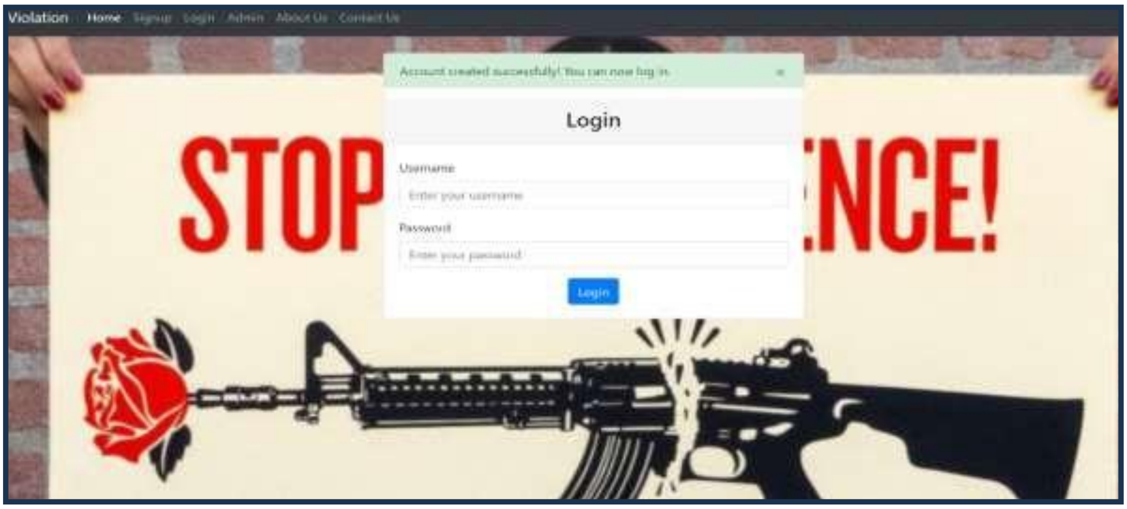
Figure7. User login



Username: sushma
Email: sushma@gmail.com
Password: ****

Figure8. User sign up

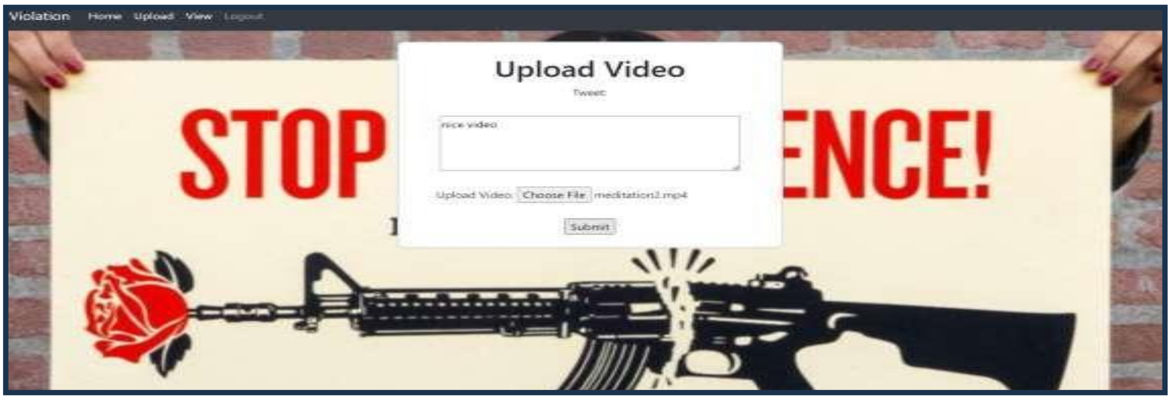
The above figure depicts a signup page which includes fields for username, Email, and password by filling with all the fields and clicking on submit button a new user can signup



Account created successfully! You can now log in.
Username: Enter your username
Password: Enter your password

Figure9. User login

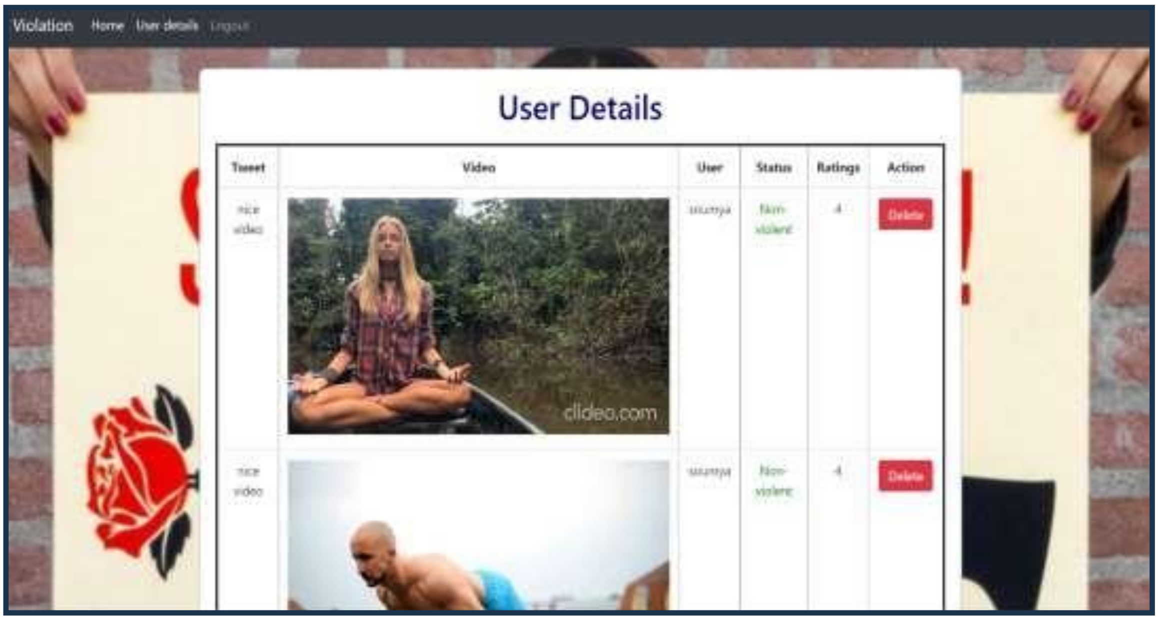
The above figure shows the login interface which require username and passwordTo user login who are signed up.



Tweet: nice video
Upload Video: meditation2.mp4

Figure10.Video upload

The figure shows an upload interface in which video is uploaded by user with a tweet.



User Details

Tweet	Video	User	Status	Ratings	Action
nice video		soumya	Non-violent	4	Delete
nice video		soumya	Non-violent	4	Delete

Figure 11.User details

In above figure 9 Admin can see all detail about video, user name, status,ratings and action which can be done by admin.

VII. SOFTWARE TESTING

7.1 TESTING STRATEGIES

Testing strategies located essential toward ensure reliability, functionality, & performance of software classification. Proposed system employs multi-faceted testing strategy to comprehensively evaluate all components and their interactions. The strategy includes both manual and automated testing methods. Manual testing involves exploratory testing, where testers interrelate by system toward identify unexpected behavior or usability issues. Automated testing, on additional hand, involves writing scripts to automatically implement test cases, confirming that the system behaves as expected under various conditions. Testing approach also includes regression testing, which confirms that innovative code vicissitudes do not unpleasantly affect prevailing functionality. Additionally, the strategy incorporates appearance testing toward evaluate system's responsiveness & stability under load. By employing amalgamation of these testing strategies, the project aims to deliver a robust and reliable application that meets all specified requirements.

7.2 TEST CASES

Table3. Test cases

Test Case ID	**Test Case Description**	**Expected Result**
TC01	Upload video containing an Ak47	Video is uploaded successfully
TC02	Upload video containing a Gun	Video is uploaded successfully

Test Case ID	Test Case Description	Expected Result
TC03	Upload video containing a Knife	Video is uploaded successfully
TC04	Upload video containing a Sickle	Video is uploaded successfully
TC05	Upload video containing a Sword	Video is uploaded successfully
TC06	Upload video without any violent content	System classifies as non-violent
TC07	Upload video with unsupported file format	System displays appropriate error message
TC08	Upload large video file	System handles and processes video efficiently

Table3.outlines essential test cases designed to validate the functionality of the video violation detection system. Each test case covers various scenarios that the system might encounter during video uploading and processing. TC01 to TC05 verify the successful upload of videos containing different types of violent content, ensuring correct acceptance and processing. TC06 tests the system's ability to classify videos without any violent content, expecting a non-violent classification. TC07 checks the system's response to unsupported file formats, while TC08 evaluates its performance with large video files, ensuring efficiency and stability.

7.4 TEST RESULTS

Table4. Test results

Test Case ID	Test Case Description	Expected Result	Actual Result
TC01	Video is uploaded successfully	Pass	Pass
TC02	Video is uploaded successfully	Pass	Pass
TC03	Video is uploaded successfully	Pass	Pass
TC04	Video is uploaded successfully	Pass	Pass
TC05	Video is uploaded successfully	Pass	Pass

Test Case ID	Test Case Description	Expected Result	Actual Result
TC06	System classifies as non-violent	Pass	Pass
TC07	System displays appropriate error message	Pass	Pass
TC08	System handles and processes video efficiently	Pass	Pass

Table 4, the Test Results, will provide an overview of the outcomes from executing the test cases defined in Table 7.3. Each test case's actual result will be compared against its expected result to determine the system's performance and reliability. TC01 to TC05 are expected to confirm that the system successfully uploads and processes videos containing Ak47, Gun, Knife, Sickle, and Sword, respectively, meeting expected outcomes. TC06 will demonstrate the system's correct classification of non-violent videos, while TC07 will check the system's handling of unsupported file formats. TC08 will verify that the system efficiently manages large video files without performance degradation, ensuring it meets processing efficiency standards.Each test case's actual result is Pass, confirming that the system successfully meets the expected outcomes for all scenarios tested.

VIII. CONCLUSION

This development has accomplished significant milestones in the realm of content moderation arranged social media platforms through the application of Convolutional Neural Networks (CNNs). By implementing sophisticated machine learning algorithms, the framework developed here excels in automating the detection and rating of violative user-generated content, specifically focusing on distinguishing between violent and non-violent videos containing weapons like firearms and knives. The methodology involved initial preprocessing of raw video data to enhance quality and consistency, followed by feature extraction and classification using CNNs, ensuring exact credentials of objectionable material. The developed Flask application assists as pivotal tool in this endeavor, offering real-time analysis capabilities that empower platforms to swiftly categorise & alleviate latent perils allied with harmful content. This mechanization not only accelerates content review processes but also certifies more objective enforcement of community guidelines, thereby enhancing overall platform safety and user experience.The impact of this project extends beyond technical innovation it addresses critical societal concerns by promoting a safer digital atmosphere wherever users can engage with confidence. By reducing reliance on manual moderation and enhancing the efficiency of content assessment, this framework sets a new standard for content management on social media.

IX. FUTURE ENHANCEMENT

Future enhancement, several avenues can be explored to augment project's performance & inflate its capabilities. Enhancing the dataset by diversifying and enlarging it with a broader range of violative and non-violative content would improve model robustness and generalization. This could involve collecting more varied examples of violent and non-violent videos to better capture real-world scenarios and nuances in content.Further refinement of the CNN model architecture and training techniques could enhance accuracy and efficiency. Techniques like assignment learning or collaborative methods could be employed to leverage pre-trained models & progress model performance deprived of necessitating extensive new data. Additionally, integrating advanced natural language processingmodels for text analysis could augment the system's capability to understand and detect nuanced expressions of hate speech or harmful intent in video captions or comments.In terms of system functionalities, exploring real-time video analysis and processing capabilities would enable immediate content moderation actions, such as live-stream content filtering. This could be complemented by leveraging edge computing or cloud-based AI services for scalable and real-time processing of video content, ensuring timely responses to emerging content risks.

REFERENCES

1. Priyanka Kulkarni, & Dr. Swaroopa Shastri. (2024). Rice Leaf Diseases Detection Using Machine Learning. Journal of Scientific Research and Technology, 2(1), 17–22. https://doi.org/10.61808/jsrt81
2. Shilpa Patil. (2023). Security for Electronic Health Record Based on Attribute using Block-Chain Technology. Journal of Scientific Research and Technology, 1(6), 145–155. https://doi.org/10.5281/zenodo.8330325
3. Mohammed Maaz, Md Akif Ahmed, Md Maqsood, & Dr Shridevi Soma. (2023). Development Of Service Deployment Models In Private Cloud. Journal of Scientific Research and Technology, 1(9), 1–12. https://doi.org/10.61808/jsrt74
4. Antariksh Sharma, Prof. Vibhakar Mansotra, & Kuljeet Singh. (2023). Detection of Mirai Botnet Attacks on IoT devices Using Deep Learning. Journal of Scientific Research and Technology, 1(6), 174–187.
5. Dr. Megha Rani Raigonda, & Shweta. (2024). Signature Verification System Using SSIM In Image Processing. Journal of Scientific Research and Technology, 2(1), 5–11. https://doi.org/10.61808/jsrt79
6. Shri Udayshankar B, Veeraj R Singh, Sampras P, & Aryan Dhage. (2023). Fake Job Post Prediction Using Data Mining. Journal of Scientific Research and Technology, 1(2), 39–47.
7. Gaurav Prajapati, Avinash, Lav Kumar, & Smt. Rekha S Patil. (2023). Road Accident Prediction Using Machine Learning. Journal of Scientific Research and Technology, 1(2), 48–59.
8. Dr. Rekha Patil, Vidya Kumar Katrabad, Mahantappa, & Sunil Kumar. (2023). Image Classification Using CNN Model Based on Deep Learning. Journal of Scientific Research and Technology, 1(2), 60–71.
9. Ambresh Bhadrashetty, & Surekha Patil. (2024). Movie Success and Rating Prediction Using Data Mining. Journal of Scientific Research and Technology, 2(1), 1–4. https://doi.org/10.61808/jsrt78
10. Dr. Megha Rani Raigonda, & Shweta. (2024). Signature Verification System Using SSIM In Image Processing. Journal of Scientific Research and Technology, 2(1), 5–11. https://doi.org/10.61808/jsrt79
11. Dr. Megha Rani Raigonda, & Shweta. (2024). Signature Verification System Using SSIM In Image Processing. *Journal of Scientific Research and Technology*, *2*(1), 5–11. https://doi.org/10.61808/jsrt79
12. Jyoti, & Swaroopa Shastri. (2024). Gesture Identification Model In Traditional Indian Performing Arts By Employing Image Processing Techniques. *Journal of Scientific Research and Technology*, *2*(3), 29–33. https://doi.org/10.61808/jsrt89
13. M Manoj Das, & Dr. Swaroopa Shastri. (2025). Machine Learning Approaches for Early Brain Stroke Detection Using CNN . Journal of Scientific Research and Technology, 3(6), 243–250. https://doi.org/10.61808/jsrt248
14. Abhishek Ashtikar, & Dr. Swaroopa Shastri. (2025). A CNN Model For Skin Cancer Detection And Classification By Using Image Processing Techniques. *Journal of Scientific Research and Technology*, *3*(6), 251–263. https://doi.org/10.61808/jsrt250
15. Dr. Megha Rani Raigonda, & Anjali. (2025). Identification And Classification of Rice Leaf Disease Using Hybrid Deep Learning. *Journal of Scientific Research and Technology*, *3*(6), 93–101. https://doi.org/10.61808/jsrt231
16. Bhagyashree, & Dr. Swaroopa Shastri. (2025). A Machine Learning Approach To Classify Medicinal Plant Leaf By Using Random Forest And KNN. Journal of Scientific Research and Technology, 3(7), 100–115. https://doi.org/10.61808/jsrt261