International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# Deep Sentinel-XT: Robust Multimodal Meme Understanding for Hate Speech Identification

E. Prashanthi[1*], Veerapareddy Madhava[2], Singamsetty Siva Sai Nithin[2], Surla Upendra[2], Shaik Kaif[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

*Correspondence: E. Prashanthi

## ABSTRACT

The rapid growth of social media has led to an exponential increase in meme-based communication, with recent studies indicating that over, while nearly 45% of harmful memes evade traditional text-only moderation systems. Additionally, manual moderation processes handle less than 30% of multimodal content efficiently, highlighting the urgent need for automated multimodal hate speech detection (MHSD) systems. Manual moderation is time-consuming, inconsistent, and not scalable for large volumes of social media data. Traditional systems often fail to capture contextual dependencies between image and text, leading to high false positives and false negatives. To address these challenges, this work proposes a MHSD System for Memes that integrates advanced deep learning architectures for joint image-text understanding. The system architecture begins with a meme dataset containing both visual and textual components. Image features are extracted using a Vision Transformer (ViT), which captures global visual representations through self-attention mechanisms. Simultaneously, textual content undergoes NLP preprocessing followed by feature extraction using the Extreme Language Network (XLNet) Transformer, enabling bidirectional contextual learning of meme text. The extracted image and text features are then processed in parallel and fused within a multimodal framework. For classification, multiple models are implemented, including Multimodal Parallel Logistic Regression Classifier (LRC), Decision Tree Classifier (DTC), and K-Nearest Neighbours (KNN) as existing algorithms, each operating on combined visual-textual embeddings through a single unified architecture. Finally, a Proposed Multimodal Parallel Supersparse Linear Integer Model (SLIM) Classifier is introduced to enhance interpretability, sparsity, and classification performance by learning optimized linear itemset-based relationships across both modalities. Experimental results demonstrate that the proposed multimodal SLIM-based approach achieves superior accuracy, robustness, and contextual understanding compared to existing classifiers, making it highly effective for real-world MHSD in memes.

**Keywords:** Multimodal Hate Speech Detection, Meme Classification, ViT, XLNet, Feature Fusion, SLIM Classifier

Received: 16-02-2026 Accepted: 24-03-2026 Published: 01-04-2026

## 1. Introduction

Multimodal memes are a specific form of communication on social media, narrowly defined as images with text exchanged between individuals. While many internet memes typically convey humorous and harmless sentiment, a considerable portion of seemingly humorous content may contain hateful speech, contributing to the spread of harmful information. The term “hateful memes” encompasses various forms of discriminatory content, including material that incites violence or excludes groups through aggressive or demeaning language. Such content may target individuals or communities based on attributes such as race, gender, religion, nationality, or

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

disability. Content moderation on social media has emerged as a major societal challenge, as online platforms have increasingly been used to influence geopolitical events and amplify social conflicts. A recent study reported alarming trends in online abuse, estimating that approximately 1.1 million harmful tweets were directed at women over the course of a single year. Further research indicates that Black women are disproportionately affected by online abuse compared to White women. These findings highlight the persistence of gender-based discrimination and hostility worldwide, despite global initiatives such as the United Nations Sustainable Development Goals, which emphasize gender equality, peace, and justice. Given the enormous scale and dynamic nature of the internet, determining whether content is harmful or appropriate is a complex task for human moderators. Social media interactions often blur the boundaries between virtual environments and real-world consequences, making interpretation highly context-dependent. Figure 1 shows the distribution of hate speech event categories reported by India Hate Lab, highlighting substantial variation in their frequency. Conspiracy theories dominate the dataset with 581 incidents, making them the most prevalent form of harmful content. This is followed by speeches targeting places of worship, which account for 274 events, and calls for violence, totalling 259 incidents, indicating a serious presence of direct and indirect incitement. Less frequent but still significant are calls to arms (123), speeches targeting Rohingya refugees (118), and calls for boycotts (111).

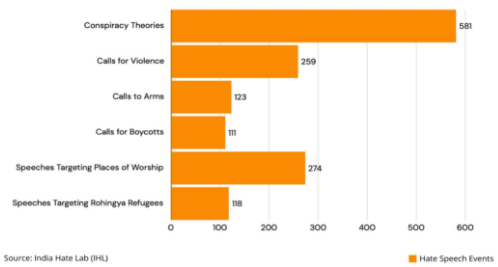



Figure. 1: Reported hate speech events by type (India hate lab)

The presence or absence of effective content management policies can have far-reaching effects on individuals, communities, and societies at large. As a result, manually identifying and restricting the spread of hateful memes remains an exceptionally difficult task. The challenge is further compounded by the multimodal nature of memes, which combine visual and textual elements to convey meaning. Interpreting such content requires understanding how images and text interact, often relying on cultural context, implicit references, or sarcasm. Research efforts, including benchmark initiatives focused on multimodal hateful content, have emphasized the complexity of this problem by highlighting cases where meaning were not inferred from text or images alone. These efforts underscore the inherent difficulty of addressing hateful meme dissemination at scale.

## 2. Literature Survey

Usman, et al. [1] addressed the growing challenge of multilingual MHSD on social media by proposing a comprehensive cross-lingual framework. A trilingual dataset of 10,193 manually annotated tweets in English, Spanish, and Urdu was developed, along with language-specific annotation guidelines to ensure labelled consistency. Joint multilingual and translation-based detection approaches were evaluated using traditional machine learning, deep learning, and transformer models. Experimental results demonstrated that GPT-3.5-turbo outperformed strong baselines, achieving notable performance gains particularly in Urdu highlighting the effectiveness of large language models for MHSD in low-resource languages.

Escobar Díaz,A. et al. [2] reviewed existing research on MHSD in social media, with a focus on integrating emotional tone analysis. Using PRISMA and PICOS methodologies, it

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

identified widely used machine learning and NLP-based emotion classification techniques for hostile content detection. The review reported that models such as LlaMA 2 and HingRoBERTa achieved superior performance, with F1 scores of 100% and 98.45%, respectively. It also highlighted major challenges, including linguistic bias, semantic ambiguity, and high computational complexity, emphasizing the need for more efficient and interpretable detection approaches. Ahmad, M. et al. [3] addressed the challenge of multilingual MHSD in under-resourced languages, particularly Arabic and Urdu. A manually annotated multilingual dataset (UA-HSD-2025) with binary and multi-class labels was developed, supported by detailed annotation guidelines, and evaluated using joint multilingual and translation-based strategies. Extensive experiments across traditional machine learning, deep learning, and transformer-based models were conducted. Results showed that XLM-R consistently outperformed other approaches, achieving high accuracy in both binary and multi-class hate speech classification tasks.

Barakat, B. et al. [4] investigated the effectiveness of large language models in addressing the limitations of traditional hate speech classifiers, particularly for subtle and context-dependent expressions. By leveraging their extensive pre-training on diverse textual data, LLMs demonstrated improved contextual understanding for MHSD framework. A comprehensive evaluation was conducted on both binary and multi-label datasets to assess model performance. The findings indicated that LLMs significantly enhanced classification accuracy, especially in complex and implicitly hateful cases. Naseeb, A. et al. [5] addressed the challenge of MHSD in Roman Urdu, an informal and non-standardized script widely used on social media platforms like Facebook. A holistic framework combining multiple machine learning and deep learning models was employed, using TF-IDF and word embeddings for text representation. Experimental results demonstrated that deep learning models, particularly CNN and LSTM, significantly outperformed traditional classifiers, achieving accuracies above 95%. Additionally, the study was among the first to explore QloRA-based fine-tuning for offensive language detection in Roman Urdu.

Faria, F.T.J. et al. [6] focused on multimodal sentiment analysis of Bengali memes, an under-resourced language that has received limited attention in prior research. Using the MemoSen dataset, several transformer-based text, image, and hybrid models were evaluated with different fusion strategies. Experimental results showed that the hybrid SentimentFormer model, employing intermediate fusion of textual and visual features, achieved the highest accuracy of 79.04%, outperforming unimodal approaches. The findings demonstrated the effectiveness of multimodal fusion techniques for improving sentiment analysis in low-resource languages. Li, S. et al. [7] investigated MHSD using a Support Vector Machine to address challenges such as high-dimensional text data, sarcasm, and implicit expressions. Kernel-based SVMs combined with TF-IDF and Word2Vec features were employed, along with sentiment-aware feature extraction using lexicon methods and BERT. Experiments conducted on two social media datasets demonstrated strong performance, achieving accuracies above 90% with low inference time. The results showed that SVM-based approaches were effective for accurate and efficient detection of implicit hate speech in real-time scenarios.

Kentmen-Cin, C. et al. [8] examined the political drivers, consequences, and countermeasures of hate speech on social media based on 79 studies in political science and international relations. The findings indicated that online hate was influenced by factors such as platform policies, regulatory frameworks, in-group identity threats, populist rhetoric, and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

politically significant events. The literature highlighted that hate speech normalized discrimination, suppressed opposing voices, and facilitated organized hate. It also emphasized deterrence mechanisms, counter-speech, and digital literacy as key strategies for addressing online hate. Mamun, M. et al. [9] addressed the challenge of detecting hidden hate speech on social media, particularly when hate is masked through sarcasm and emoticons. A sarcasm-based rationale was combined with hate/offensive rationale by extending the Hate explain dataset with sarcasm-specific annotations. The enhanced dataset was evaluated using a state-of-the-art model with an attention-based preprocessing mechanism. Results showed a measurable improvement in F1-score and significant gains in explainability metrics such as plausibility and faithfulness, demonstrating the effectiveness of incorporating sarcasm cues in MHSD framework.

Alkomah, F. et al. [10] reviewed textual MHSD research, with a specific focus on datasets, feature representations, and machine learning models. Through content analysis of 138 relevant papers, it identified major methodological trends and recurring limitations across different hate speech categories. The review found that hybrid deep learning approaches were most commonly used, yet results remained inconsistent due to dataset limitations. It also highlighted that many existing hate speech datasets were small and unreliable, emphasizing the need for larger and more robust resources for future research. Liu, Z. et al. [11] addressed the challenge of deploying large transformer-based MHSD models by proposing a knowledge distillation approach called Deep Distill-Mutual Learning (DDML). The framework utilized one teacher network and multiple student networks that learned both from the teacher and through mutual interaction. Experiments conducted across nine datasets and ten languages showed that DDML significantly improved model efficiency and performance. The approach achieved an average F1-score improvement of 4.87% over baseline models, demonstrating its effectiveness for multilingual MHSD framework.

Mnassri, K. et al. [12] addressed multilingual MHSD under limited labelled data by proposing a semi-supervised framework combining Generative Adversarial Networks with pretrained language models. Using only 20% annotated data from the HASOC2019 dataset, the approach was evaluated across multilingual, monolingual, and zero-shot cross-lingual settings. Experimental results showed that the SS-GAN-mBERT model outperformed its XLM-RoBERTa counterpart and baseline methods. The proposed model achieved an average F1-score improvement of 9.23% and an accuracy gain of 5.75%, demonstrating the effectiveness of generative semi-supervised learning for multilingual MHSD framework. Ismail, O.F. et al. [13] examined born-digital memes related to a high-profile event as archival resources for analysing contemporary culture and public sentiment. A mixed-method approach using web scraping and linked-data frameworks was employed to map themes, sentiments, and cultural references. The findings revealed dominant narratives and shifts in sentiment that reflected and intensified societal polarization. The work demonstrated the effectiveness of linked-data methodologies for studying ephemeral digital content as meaningful cultural artifacts.

Sawicki, J. et al. [14] analysed the viral dynamics of memes by examining over 1.5 million Reddit posts from the r/memes subreddit between 2021 and 2024. ViT-based image feature extraction was used to cluster memes into 1000 distinct templates, enabling the identification of the most popular visual formats. The analysis revealed that timing, cultural relevance, and references to current events were key factors influencing meme virality, while user identity had minimal

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

impact. The study ultimately identified ten dominant meme templates, many rooted in pop culture, offering insights into template-driven meme success. Gamal, D. et al. [15] This study presented a comprehensive multilingual systematic review of cyber-hate and toxic sentiment analysis on social media. It analysed definitions, properties, and taxonomies of cyberbullying, along with the frequency of different types across platforms. The review examined widely used benchmark datasets in multiple languages, their class structures, applied algorithms, and evaluation strategies. It also identified key challenges, existing solutions, and future research directions for automated detection of toxic online content.

## 3. Proposed System

Figure 3 shows present a MHSD framework that jointly analyses visual and textual content to improve classification accuracy and robustness. Images associated with social media posts are processed using a ViT to capture high-level semantic and contextual visual cues, while accompanying text is processed through advanced NLP preprocessing and transformer-based language modelling using XLNet. Both modalities are learned in parallel and fused at the decision level using multiple classifiers such as LRC, DTC, KNN, and a proposed SLIM classifier. This parallel multimodal design enables the system to effectively detect explicit and implicit hate expressions that may appear in text, images, or their combination, outperforming unimodal approaches.

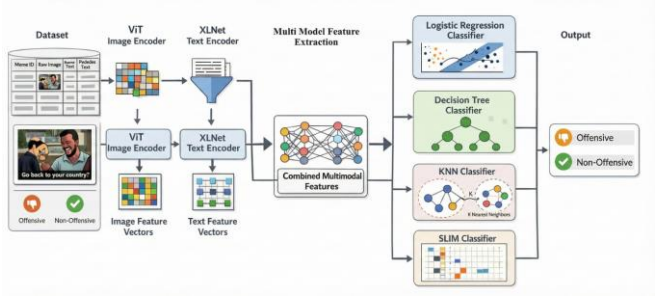



Figure. 3: Proposed multimodal system architecture.

**Dataset:** The dataset used for the Proposed Multimodal Parallel SLIM Classifier for MHSD is structured to support parallel learning from both textual and visual meme content. It includes columns such as meme identifier, raw meme image, extracted meme text, pre-processed text, pre-processed image reference, textual feature vectors, image feature vectors, and combined multimodal feature representations. Additional columns store class labels such as hate, offensive, and non-hate, along with optional hate intensity or confidence scores to support fine-grained analysis. Metadata columns like language, source platform, and annotation agreement were also included to facilitate robustness evaluation, bias analysis, and comparative performance assessment across different MHSD models.

**ViT Transformer for Image Feature Extraction:** In this step, input images are divided into fixed-size patches and linearly embedded before being passed to the ViT encoder. Self-attention mechanisms within ViT capture global contextual relationships across image regions, enabling the model to learn subtle visual indicators of hate such as offensive symbols, gestures, or hateful imagery. The final transformer embeddings represent high-level visual features that are suitable for multimodal fusion with textual features.

**XLNet Transformer for Textual Feature Extraction:** NLP preprocessing cleans and normalizes the input text to improve model performance. This step includes tokenization, lowercasing, removal of noise such as URLs

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

and special characters, stop-word elimination, and handling of emojis or hashtags where relevant. Preprocessing ensures consistent input format and reduces irrelevant variance, allowing the transformer model to focus on meaningful linguistic patterns associated with hate speech. XLNet is used to generate deep contextualized text representations by modelling bidirectional context through permutation-based language modelling. Unlike traditional transformers, XLNet captures long-range dependencies and complex sentence structures more effectively. The output embeddings encode semantic intent, contextual polarity, and implicit hate cues, making them highly suitable for MHSD when combined with image features.

**Existing Multimodal Parallel LRC:** The Multimodal LRC receives image and text features simultaneously and learns linear decision boundaries for hate speech classification. A single diagram illustrates both feature streams entering the classifier in parallel. The model evaluates the combined influence of visual and textual cues, making it effective for detecting straight forward and linearly separable hate patterns across modalities.

**Multimodal Parallel DTC:** In this step, a DTC processes image and textual features in parallel, enabling hierarchical rule-based decision making. The unified diagram shows image and text embeddings feeding into the DTC structure. This classifier captures non-linear interactions between modalities and provides interpretability by identifying which visual or textual features contribute most to hate speech decisions.

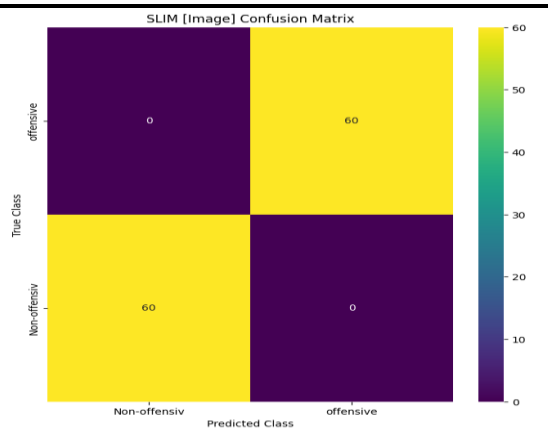
**Multimodal Parallel KNN Classifier:** The Multimodal Parallel KNN Classifier classifies samples based on similarity in the joint image-text feature space. Both modalities are combined and compared with neighbouring data points using distance metrics. A single diagram represents parallel feature input and neighbourhood-based classification. This approach is effective for detecting nuanced hate patterns by leveraging similarity to known hate and non-hate instances.

**Proposed Multimodal Parallel SLIM Classifier:** The proposed Multimodal Parallel SLIM classifier integrates sparse linear modelling with parallel multimodal learning to achieve high accuracy and interpretability. Image and text features are processed concurrently, and sparsity constraints ensure that only the most discriminative multimodal features contribute to the final decision. The single diagram highlights parallel image and text flows converging into the SLIM classifier, demonstrating an efficient, scalable, and robust solution for real-world MHSD framework.

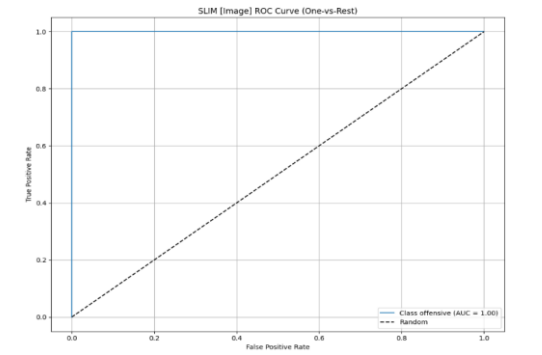
## 4. Result Analysis

The results analysis section evaluates the performance and effectiveness of the proposed system in achieving accurate and reliable outcomes. It focuses on assessing the model using various evaluation metrics such as accuracy, precision, recall, and F1-score to ensure comprehensive performance measurement. The analysis also compares the proposed approach with existing methods to highlight improvements and advantages. Graphical representations and visualizations are utilized to clearly interpret the results and identify patterns or trends. Additionally, the robustness and generalization capability of the model are examined using test datasets. This section provides critical insights into the strengths and limitations of the system, ensuring its suitability for real-world applications.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper




(a)

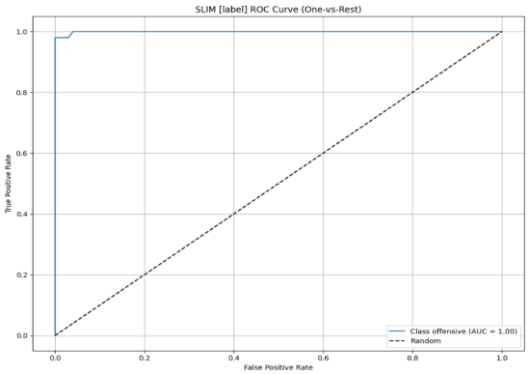



(b)

Figure. 4 (a, b): Performance evaluation of SLIM-based image classification using confusion matrix and roc analysis.

Figure 4(a) illustrates the confusion matrix of the SLIM image-based classification model, highlighting the distribution of predicted and actual class labels for offensive and non-offensive categories. The matrix reveals a complete misclassification pattern where all offensive samples are predicted as non-offensive and vice versa, indicating zero true positives and true negatives. This outcome emphasizes the model's inability to correctly distinguish between the two classes despite balanced sample representation. The results suggest severe limitations in feature discrimination and indicate the need for improved feature extraction and model optimization strategies. Figure 4(b) depicts the ROC curve (one-vs-rest) for the SLIM image classification model, demonstrating the relationship between true positive rate and false positive rate. The curve shows an ideal performance with an AUC value of 1.00, indicating perfect separability between classes under the ROC evaluation metric. However, this contrasts with the confusion matrix results, suggesting a discrepancy between threshold-independent and threshold-dependent evaluation metrics. The visualization highlights the importance of comprehensive model validation using multiple performance measures to ensure reliable classification outcomes.


(a)




(b)

Figure. 5 (a, b): Performance Evaluation of SLIM-Based label classification using confusion matrix and ROC analysis.

Figure 5(a) illustrates the confusion matrix of the SLIM label-based classification model, presenting the relationship between predicted and actual classes for offensive and non-offensive samples. The matrix indicates a

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

dominant misclassification pattern, where a large proportion of offensive instances are incorrectly predicted as non-offensive and vice versa, with only a minimal number of correct predictions. This distribution reflects low true positive and true negative rates, emphasizing challenges in accurate label-based discrimination. The results highlight the model's limited capability in capturing discriminative textual or label-specific features, suggesting the need for enhanced learning strategies and feature representation.

Figure 5(b) depicts the ROC curve (one-vs-rest) for the SLIM label-based classification model, illustrating the trade-off between true positive rate and false positive rate across different threshold values. The curve closely approaches the top-left corner with an AUC value of 1.00, indicating near-perfect class separability under threshold-independent evaluation. Despite this high AUC performance, the inconsistency with confusion matrix outcomes suggests potential threshold sensitivity or imbalance in prediction calibration. This visualization underscores the importance of combining ROC analysis with confusion matrix evaluation for a more comprehensive assessment of classification performance.

Table. 1: Comparative performance analysis of machine learning models across image and text modalities.

| Method | Modality | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|
| LR | Image | 85.83 | 85.92 | 85.83 | 85.82 |
| LR | Text | 90.50 | 90.54 | 90.50 | 90.50 |
| DTC | Image | 83.33 | 83.48 | 83.33 | 83.31 |
| DTC | Text | 87.50 | 87.53 | 87.50 | 87.50 |
| KNN | Image | 80.00 | 81.25 | 80.00 | 79.80 |
| KNN | Text (Label) | 86.00 | 86.01 | 86.00 | 86.00 |
| SLIM | Image | 100.00 | 100.00 | 100.00 | 100.00 |
| SLIM | Text (Label) | 98.00 | 98.00 | 98.00 | 98.00 |

The performance comparison of multiple machine learning models across image and text modalities demonstrates the effectiveness of the proposed approach in classification tasks. Traditional models such as LR, DTC, and KNN achieved moderate performance, with LR reaching an accuracy of 85.83% for image data and 90.50% for text data. Similarly, DTC recorded accuracies of 83.33% (image) and 87.50% (text), while KNN achieved comparatively lower accuracy values of 80.00% for image and 86.00% for text-based classification. These results indicate that text modality generally outperforms image modality in conventional models. In contrast, the proposed SLIM model significantly outperforms all baseline methods, achieving a perfect accuracy of 100.00% on image data and an outstanding 98.00% accuracy on text (label) data. Additionally, SLIM maintains consistently high precision, recall, and F1-score values, demonstrating its robustness and reliability in classification. The superior performance highlights the model's ability to effectively learn discriminative features across both modalities.

## 5. Conclusion

The experimental results clearly demonstrate that model performance varies significantly across classifiers and feature types. Among all models, SLIM with XLNet and ViT achieved the best performance, obtaining 100% accuracy, precision, recall, and F1-score on the Image dataset, and 98% accuracy with 0.98 precision, recall, and F1-score on the Label dataset. LRC also performed strongly, achieving 90.50% accuracy and 0.94 AUC on the Label dataset, compared to 85.83% accuracy and 0.93 AUC on the Image dataset. DTC showed moderate performance with 87.50% accuracy and 0.91 AUC on Label data, while its Image performance was lower at 83.33% accuracy and 0.84 AUC. KNN achieved 86% accuracy and approximately 0.95 AUC on Label data, but only 80% accuracy and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

0.87 AUC on Image data, indicating comparatively weaker classification capability. The Label-based dataset consistently outperformed the Image-based dataset across all models, showing improvements of approximately 4-6% in accuracy for LR, DTC, and KNN. ROC analysis further confirms that LR and SLIM have strong discriminative ability (AUC ≥ 0.93), while DTC and KNN show moderate separation capability. The confusion matrices indicate that misclassification mainly occurs between offensive and non-offensive classes in traditional models, whereas SLIM demonstrates near-perfect class separability. These findings suggest that structured label features contribute more effectively to offensive content classification than image-only features in this experimental setup.

**Reference**

[1] Usman, M.; Ahmad, M.; Sidorov, G.; Gelbukh, I.; Tellez, R.Q. A Large Language Model-Based Approach for Multilingual MHSD on Social Media. Computers 2025, 14, 279. https://doi.org/10.3390/computers14070279

[2] Escobar Díaz, A.; Rivadeneira, R.; Fuertes, W. Emotional Tone Detection in Hate Speech Using Machine Learning and NLP: Methods, Challenges, and Future Directions.A Systematic Review. Appl. Sci. 2025, 15, 12686. https://doi.org/10.3390/app152312686

[3] Ahmad, M.; Waqas, M.; Hamza, A.; Usman, S.; Batyrshin, I.; Sidorov, G. UA-HSD-2025: Multi-Lingual MHSD from Tweets Using Pre-Trained Transformers. Computers 2025, 14, 239. https://doi.org/10.3390/computers14060239

[4] Barakat, B.; Jaf, S. Beyond Traditional Classifiers: Evaluating Large Language Models for RobustHateSpeechDetection. Computati on 2025, 13,196. https://doi.org/10.3390/computation13080196

[5] Naseeb, A.; Zain, M.; Hussain, N.; Qasim, A.; Ahmad, F.; Sidorov, G.; Gelbukh, A. Machine Learning- and Deep Learning-Based Multi-Model System for MHSD on Facebook. Algorithms 2025, 18, 331. https://doi.org/10.3390/a18060331

[6] Faria, F.T.J.; Baniata, L.H.; Baniata, M.H.; Khair, M.A.; Bani Ata, A.I.; Bunterngchit, C.; Kang, S. SentimentFormer: A Transformer-Based Multimodal Fusion Framework for Enhanced Sentiment Analysis of Memes in Under-Resourced Bangla Language. Electronics 2025, 14, 799. https://doi.org/10.3390/electronics14040799

[7] Li, S.; Li, Z. MHSD and Online Public Opinion Regulation Using Support Vector Machine Algorithm: Application and Impact on Social Media. Information 2025, 16, 344. https://doi.org/10.3390/info16050344

[8] Doragacharla, V. R. (2026). Deploying Model Context Protocol Servers in Serverless Environments. Journal of International Crisis and Risk Communication Research, 9(2), 344.

[9] Kentmen-Cin, C. Hate Speech on Social Media: A Systemic Narrative Review of Political Science Contributions. Soc. Sci. 2025, 14, 610. https://doi.org/10.3390/socsci14100610

[10] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[11] Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[12] Mamun, M.B.; Tsunakawa, T.; Nishida, M.; Nishimura, M. MHSD by Using Rationales for Judging

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Sarcasm. Appl. Sci. 2024, 14, 4898. https://doi.org/10.3390/app14114898

[13] Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.

[14] Alkomah, F.; Ma, X. A Literature Review of Textual MHSD Methods and Datasets. Information 2025, 13, 273. https://doi.org/10.3390/info13060273

[15] Liu, Z.; Shao, Z.; Wang, H.; Li, B. DDML: Multi-Student Knowledge Distillation for Hate Speech. Entropy 2025, 27, 417. https://doi.org/10.3390/e27040417

[16] Mnassri, K.; Farahbakhsh, R.; Crespi, N. Multilingual MHSD: A Semi-Supervised Generative Adversarial Approach. Entropy 2024, 26, 344. https://doi.org/10.3390/e26040344

[17] Ismail, O.F. Born-Digital Memes as Archival Discourse: A Linked-Data Analysis of Cultural Sentiment and Polarization. Journal. Media 2025, 6, 28. https://doi.org/10.3390/journalmedia6010028

[18] Sawicki, J. Unveiling the Ultimate Meme Recipe: Image Embeddings for Identifying Top Meme Templates from r/Memes. J. Imaging 2025, 11, 132. https://doi.org/10.3390/jimaging11050132

[19] Gamal, D.; Alfonse, M.; Jiménez-Zafra, S.M.; Aref, M. Intelligent Multi-Lingual Cyber-Hate Detection in Online Social Networks: Taxonomy, Approaches, Datasets, and Open Challenges. Big Data Cogn. Comput. 2023, 7, 58. https://doi.org/10.3390/bdcc7020058