IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 February 2024, accepted 3 May 2024, date of publication 14 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3400815

RESEARCH ARTICLE

MSCMGTB: A Novel Approach for Multimodal Social Media Content Moderation Using Hybrid Graph Theory and Bio-Inspired Optimization

PREMNARAYAN ARYA[1], AMIT KUMAR PANDEY[2], S. GOPAL KRISHNA PATRO[3], KRETIKA TIWARI[4], NIRANJAN PANIGRAHI[5], QUADRI NOORULHASAN NAVEED[6], AYODELE LASISI[6], AND WAHAJ AHMAD KHAN[7]

[1]Department of Computer Engineering and Applications, GLA University, Mathura, Uttar Pradesh 281406, India
[2]Department of CSE-DS, ABES Engineering College Affiliated to AKTU, Ghaziabad, Uttar Pradesh 201009, India
[3]School of Technology, Woxsen University, Hyderabad, Telangana 502345, India
[4]School of Engineering and Technology, Jagran Lakecity University, Bhopal, Madhya Pradesh 462011, India
[5]Department of Computer Science and Engineering, Parala Maharaja Engineering College, Berhampur, Odisha 761003, India
[6]Department of Computer Science, College of Computer Science, King Khalid University, Abha 62521, Saudi Arabia
[7]School of Civil Engineering and Architecture, Institute of Technology, Dire-Dawa University, Dire Dawa 1362, Ethiopia

Corresponding authors: S. Gopal Krishna Patro (sgkpatro2008@gmail.com) and Wahaj Ahmad Khan (wkhan9450@gmail.com)

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through a Large Research Project under grant number RGP2/496/45.

ABSTRACT In an era where social media platforms burgeon with diverse content, compelling moderation is imperative to filter harmful materials. Traditional methods often grapple with the dual challenges of accuracy and computational efficiency levels. These conventional approaches typically rely on text-based or image-based analysis, neglecting the complex interplay of multimodal content prevalent in social media scenarios. This limitation leads to suboptimal content filtering, often missing contextually nuanced or visually deceptive harmful content sets. Addressing these challenges, in response to the pressing need for effective social media content moderation, we introduce a pioneering approach that combines Convolutional Neural Networks (CNNs) and Transformers. We aim to enhance accuracy and computational efficiency in filtering harmful multimodal content prevalent on social media platforms. By integrating CNNs and Transformers, we achieve nuanced visual content extraction and contextual textual understanding, thus improving the identification of harmful content. Additionally, our model utilises a Bi-directional Attention Mechanism (BAM) and Genetic Algorithms (GAs) for efficient text-visual fusion and hyper-parameter optimisation, respectively. Empirical testing on datasets from Google, Facebook, and Kaggle demonstrates the superior performance of our model in terms of precision, accuracy, recall, AUC, specificity, and response delay in detecting harmful content. The proposed Multimodal Social Media Content Moderation Using Hybrid Graph Theory & Bio-inspired Optimization (MSCMGTB) model consistently achieves superior precision, accuracy, recall, AUC, and specificity with rates ranging from 86.78% to 98.82% across varying dataset sizes, highlighting its efficacy in content moderation as well as reduced the delay time to classify social media contents as compared to Social Graph Neural Network (SGNN), CrediBot, and Adaptive LDA (ALDA) techniques. The model also preempts potentially harmful content posters, offering enhanced pre-emption metrics.

INDEX TERMS Deep learning, content moderation, social, media, multimodal analysis, hybrid models.

I. INTRODUCTION

The digital landscape of the 21st century is predominantly shaped by social media, a realm where information

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Fiumara.

proliferates at an unprecedented pace. This rapid dissemination of content, while beneficial for global connectivity, also ushers in challenges pertaining to content moderation. The existing methods of social media content moderation, predominantly relying on either textual or visual analysis, often grapple with the intricacies of multimodal digital content [1]. This singular focus results in a substantial gap in the effective detection and filtration of harmful content, given the complex interplay between text and imagery in modern social media posts [2], [3].

In response to this challenge, recent advancements in deep learning offer a promising avenue. The integration of Convolutional Neural Networks (CNNs) and Transformers presents a novel approach. CNNs, known for their efficacy in image processing, are adept at extracting and analyzing visual features. Conversely, Transformers, originally designed for natural language processing tasks, excel in contextual understanding, particularly in deciphering the nuances embedded in textual data. The confluence of these two powerful deep learning architectures holds the potential to revolutionize social media content moderation operations. However, the application of such advanced models in real-world scenarios is not without its obstacles. The primary concerns revolve around computational efficiency and the ability to scale effectively on various social media platforms. These challenges necessitate innovative solutions that not only enhance accuracy but also ensure practical deployments [5], [6], [7], [8].

The introduction of a Bi-directional Attention Mechanism (BAM) for text-visual fusion further refines this model. BAM's dynamic learning capability significantly reduces computational demands while efficiently processing the fusion of textual and visual information. Additionally, the utilization of Genetic Algorithms (GAs) for hyper-parameter optimization in neural networks fine-tunes the model's performance, striking a balance between accuracy and computational feasibility. Moreover, the exploration of intra and inter-modal relationships in social media content through Graph Neural Networks (GNNs) marks a significant advancement in understanding the complex dynamics of digital interactions. This approach not only aids in the accurate identification of harmful content but also contributes to a deeper comprehension of the multifaceted nature of social media posts.

The proposed model's efficacy is validated through rigorous testing on prominent social media datasets, including those from Google, Facebook, and Kaggle. The results demonstrate a significant improvement in detecting and preempting harmful content, outperform existing methods in several key metrics such as precision, accuracy, recall, and response time. This enhancement is not just a numerical achievement but a stride towards creating a safer and more responsible social media environment.

In essence, this paper delves into the design and implementation of an efficient, hybrid deep learning model for social media content moderation. By marrying the strengths of CNNs and Transformers, and further refining the approach with BAM, GAs, and GNNs, it addresses both the accuracy and computational efficiency challenges in different use cases. The ultimate goal is to foster a digital ecosystem where content moderation is not just reactive but proactive, ensuring a secure and respectful online community for real-time scenarios. Thus, the introduction sets the stage for a comprehensive discussion on the innovative methodologies employed, their underlying rationale, and the impactful outcomes of this research. It underscores the significance of advanced deep learning techniques in enhancing the safety and integrity of social media platforms, a necessity in today's digitally interconnected world scenarios.

A. MOTIVATION

In the ever-expanding landscape of social media, the motivation for developing advanced content moderation tools is multi-faceted and pressing. The exponential growth of user-generated content on these platforms has amplified the challenges associated with monitoring and filtering harmful material. Traditional content moderation methods, predominantly linear and unimodal, are increasingly inadequate in grappling with the complexity and volume of modern social media interactions. This inadequacy not only compromises user experience but also raises significant concerns regarding digital safety and the propagation of harmful content. The inception of this research is rooted in the recognition of these challenges and the imperative need for a more sophisticated, multimodal approach to content moderation. The motivation extends beyond the mere enhancement of content filtering accuracy. It encompasses the necessity for computational efficiency, scalability, and adaptability in diverse social media environments. This research aims to bridge the gap between the evolving nature of social media content and the static nature of existing moderation tools. In summary, the motivation for this research is deeply rooted in the need to address the growing complexity and volume of social media content. The contributions of this work lie in the development of an innovative, efficient, and scalable model for content moderation, capable of navigating the multifaceted nature of social media interactions which are summarized below.

B. RESEARCH OBJECTIVE

Our research endeavors to revolutionize social media content moderation through a hybrid model that combines Convolutional Neural Networks and Transformers. By prioritizing accuracy and computational efficiency, our approach aims to effectively filter harmful multimodal content, thereby enhancing user safety and platform integrity. The utilization of grayscale images was deliberate, focusing on content features rather than color aesthetics, while ensuring consistency in evaluation metrics.

C. CONTRIBUTIONS

The contributions of this work are manifold and significant. Firstly, the development of a hybrid deep learning model, combining CNNs and Transformers, offers a novel approach

to content analysis. This model harnesses the strengths of both architectures: the CNNs' proficiency in visual feature extraction and the Transformers' adeptness in contextual textual analysis. Such a hybrid approach is not only innovative but also highly effective in understanding the intricate interplay of text and image in social media content.

Secondly, the implementation of a Bi-directional Attention Mechanism (BAM) for text-visual fusion represents a pivotal contribution. This mechanism dynamically adjusts the focus between textual and visual inputs, ensuring that the model captures the most relevant features from both modalities. This results in a significant reduction in computational overhead, a crucial factor for real-world application on social media platforms.

Furthermore, the application of Genetic Algorithms (GAs) for hyper-parameter optimization in this context is another key contribution. GAs provide an efficient method for fine-tuning the model, enhancing its performance while reducing the manual effort typically required in such processes. Additionally, the utilization of Graph Neural Networks (GNNs) to explore intra and inter-modal relationships within social media content is a noteworthy contribution. This approach provides deeper insights into the complex dynamics of social media interactions, enabling more accurate and comprehensive content moderation.

Lastly, the practical implications of this research are profound. The testing and validation of the model on various prominent social media datasets demonstrate its superior performance compared to existing methods. The improvements in precision, accuracy, recall, and response time are not just statistically significant but also indicative of the model's potential to transform the landscape of social media content moderation.

Hence, the significant contribution of this work is to shift from traditional, unimodal methods to a more holistic, multimodal approach, reflective of the dynamic and intricate nature of social media content. The innovations presented in this paper are poised to make a significant impact in the field of digital content moderation, paving the way for safer and more responsible social media platforms.

II. RELATED STUDY

The ever-expanding field of social media content moderation has witnessed significant advancements, with various research endeavors seeking to address the multifaceted challenges it presents. A comprehensive review of the literature reveals diverse approaches and methodologies employed to enhance the effectiveness and efficiency of content moderation process.

Gao et al. [1] explored cross-platform item recommendation for online social e-commerce, highlighting the importance of integrating diverse data sources for accurate content analysis. This research underscores the necessity of multimodal approaches in understanding complex online environments. Similarly, Bono et al. [2] emphasized a citizen science approach for analyzing social media with crowdsourcing, providing insights into the potential of human-machine collaboration in content moderation. Li et al. [3] investigated disentangled modeling of social homophiles and influence for social recommendation, presenting a novel perspective on the dynamics of social interactions online. This study contributes to understanding the underlying patterns in social media content, which is crucial for effective moderation. In a similar vein, Dongre and Agrawal [4] focused on deep-learning-based drug recommendation and Adverse Drug Reaction (ADR) detection in healthcare models on social media, illustrating the application of deep learning in specific content domains.

Tran et al. [5] and Bacha et al. [6] delved into the combination of social relations and interaction data in recommender systems and offensive text detection in unstructured data for heterogeneous social media, respectively. These studies highlight the growing complexity of social media content and the need for sophisticated analysis tools. Xu et al. [7] addressed the challenge of achieving online and scalable information integrity by harnessing social spam correlations, emphasizing the need for scalable solutions in content moderation. Ma et al. [8] and Fei et al. [9] explored social graph neural network-based interactive recommendation schemes and real-time detection of events from Twitter, respectively, showcasing the application of neural networks in capturing complex social interactions.

Ismail et al. [10] and Li et al. [11] focused on event-based emotion detection frameworks addressing mental health and semi-supervised variational user identity linkage, respectively. These studies highlight the diverse applications of machine learning in understanding social media contents. Aguilera et al. [12] and Guo et al. [13] contributed to the field by applying bot detection for credibility analysis on Twitter and mitigating the influence of disinformation propagation, respectively for different scenarios. These studies underscore the importance of credibility and integrity in online contenst.

Adishesha et al. [14] and Patro et al. [15] explored forecasting user interests through topic tag predictions in online health communities and a conscious cross-breed recommendation approach for electronic commerce systems. Their work emphasizes the role of predictive analytics in understanding user behavior and preferences on social media platforms. The role of deep learning in social media analysis is further exemplified by Al-Onazi et al. [16] and Zheng et al. [17], who investigated affect classification in Arabic tweets and adaptive LDA optimal topic number selection in news topic identification. These studies demonstrate the versatility of deep learning techniques in processing and understanding diverse content types and languages.

Marinho and Holanda [18] and Maity et al. [19] addressed emerging cyber threat identification and profiling based on natural language processing, and the detection of Malay hate speech, respectively. Their research highlights the increasing need for advanced computational methods to tackle evolving digital threats and offensive content. Liang et al. [20] and Zeng and Xiang [21] explored graph-based non-sampling for

knowledge graph enhanced recommendation and persistence augmented graph convolution network for information popularity prediction. Their work contributes to the development of sophisticated graph-based models for content analysis and recommendation systems.

In the context of spatial trajectories and linguistic steganalysis, Gupta and Bedathur [22] and Yang et al. [23] offered insights into modeling spatial trajectories using coarse-grained smartphone logs and linguistic steganalysis toward social networks. These studies reflect the expanding scope of content moderation to include geographical and linguistic analyses. Finally, Bacha et al. [6] and Calderón-Suarez et al. [24] contributed to offensive text detection in unstructured data for heterogeneous social media and enhancing the detection of misogynistic content in social media scenarios. Their research underscores the ongoing efforts to create safer and more inclusive online environments for different scenarios.

Truică et al. [25], [27] introduced the MCWDST algorithm, a minimum-cost weighted directed spanning tree approach for real-time fake news mitigation in social media. This algorithm demonstrates the potential of cost-effective and efficient computational techniques in the realm of fake news detection. Similarly, Zamil and Charkari [26] proposed a fusion approach to combat fake news on social media, emphasizing improved detection and interpretability, a critical aspect in the context of ever-evolving misinformation. Park and Chai [28] focused on constructing a user-centered fake news detection model using classification algorithms in machine learning, highlighting the significance of user-centric approaches in the design of effective moderation tools. Etta et al. [29] compared the impact of social media regulations on news consumption, offering insights into the broader implications of content moderation beyond technical aspects.

In the realm of big data and machine learning, Altheneyan and Alhadlaq [30] explored fake news detection using distributed learning, signifying the growing role of big data in tackling misinformation. Aditya and Mohanty [31] presented an approach for heterogeneous social media analysis for efficient deep learning fake-profile identification, demonstrating the necessity of addressing diverse data types in social media. Almarashy et al. [32] enhanced fake news detection through a multi-feature classification, showcasing the importance of incorporating multiple data features for accurate detection. Kar et al. [33] addressed the challenge of detecting fake images on social networks using recurrent neural networks, indicating the increasing complexity of fake news formats.

Advancements in pattern-mining systems for fake news analysis were explored by Djenouri et al. [34], emphasizing the role of advanced data mining techniques in understanding and countering misinformation. Govindankutty and Gopalan [35] modeled rumor spread and influencer impact on social networks, highlighting the significant role of network dynamics and influential users in the spread of misinformation. Joshi et al. [36] contributed to explainable misinformation detection across multiple social media platforms, underscoring the need for transparency and interpretability in detection algorithms. This aspect of explainability is crucial for the acceptance and trust in automated moderation systems.

Valinejad and Mili [37] developed a cyber-physical-social model of community resilience, considering critical infrastructure interdependencies. Their work provides a broader perspective on how social media content moderation is intertwined with overall community resilience and well-being. In terms of methodological innovations, Wu et al. [38] proposed a category-controlled encoder-decoder for fake news detection, emphasizing the role of categorization in enhancing detection accuracy. Tajrian et al. [39] offered a comprehensive review of methodologies for fake news analysis, providing a valuable synthesis of the current state of research in this field.

Khan et al. [40] focused on visual user-generated content verification in journalism, reflecting the growing importance of visual content in the realm of fake news. Fu et al. [41] examined rumor spreading models considering the roles of online social networks and information overload, providing insights into the behavioral aspects of misinformation dissemination operations. Further extending the scope of research, Wang et al. [42], Hu et al. [43], and Wang et al. [44] explored various aspects of multi-modal fake news detection, including the use of transformer networks and causal inference. Their work signifies the shift towards more sophisticated, multi-modal approaches in detecting fake news.

Zaheer et al. [45] and Jung et al. [46] presented optimized convolutional neural networks and topological and sequential neural network models for detecting fake news, showcasing the advancements in neural network architectures and their application in this domain for different operations. These studies demonstrate the continuous evolution of deep learning techniques tailored to the specific challenges posed by fake news in social media sets. Wu et al. [47] delved into human cognition-based consistency inference networks for multi-modal fake news detection, emphasizing the integration of human-like reasoning processes in computational models. This approach bridges the gap between artificial intelligence and human cognitive processes, aiming for a more nuanced and context-aware detection mechanism.

Ojha et al. [48] approached the control of fake information dissemination in online social networks from an epidemiological perspective, offering a unique analogy between the spread of misinformation and the spread of diseases. This perspective provides a novel framework for understanding and mitigating the spread of fake news. Finally, Xu et al. [7] focused on achieving online and scalable information integrity by harnessing social spam correlations. Their research highlights the importance of scalability and adaptability of moderation systems in the rapidly evolving social media landscape.

In summary, the literature reveals a concerted effort towards developing more sophisticated, efficient, and

human-like approaches to detecting and mitigating fake news on social media scenarios. From algorithmic advancements and multi-modal analyses to the incorporation of human cognition and epidemiological models, these studies collectively contribute to a deeper understanding and more effective tackling of the fake news phenomenon. This literature review underscores the complexity of the issue and the diverse methodologies being employed to address it, setting the stage for the proposed research in this text.

III. PROPOSED SYSTEM MODEL

A. PROPOSED BLOCKCHAIN-BASED SECURITY MODEL

This section discusses design of the proposed model, where each component - CNN, BAM, GA, and GNN - plays a pivotal role, orchestrating a symphony of data processing and analysis. The Convolutional Neural Networks (CNNs), known for their efficiency in image analysis, delve deep into the visual content from social media, meticulously extracting and interpreting intricate image features.

As shown in Figure 1, the visual acumen of CNNs is seamlessly complemented by the Bi-directional Attention Mechanism (BAM), an efficient process in the model's architectural operations. BAM acts as the critical juncture where the visual insights from CNNs and the contextual nuances gleaned from textual analysis are harmoniously fused for different scenarios. It assists in adeptly adjusting the focus between textual and visual inputs to ensure a comprehensive understanding of the contents. Simultaneously, the Genetic Algorithms (GAs) operate like an internal catalyst that assists in meticulously fine-tuning the model's hyper-parameters for different input sets. Complementing these components, the Graph Neural Networks (GNNs) add another layer of sophistication. GNNs navigate the complex web of intra and inter-modal relationships within the social media contents.

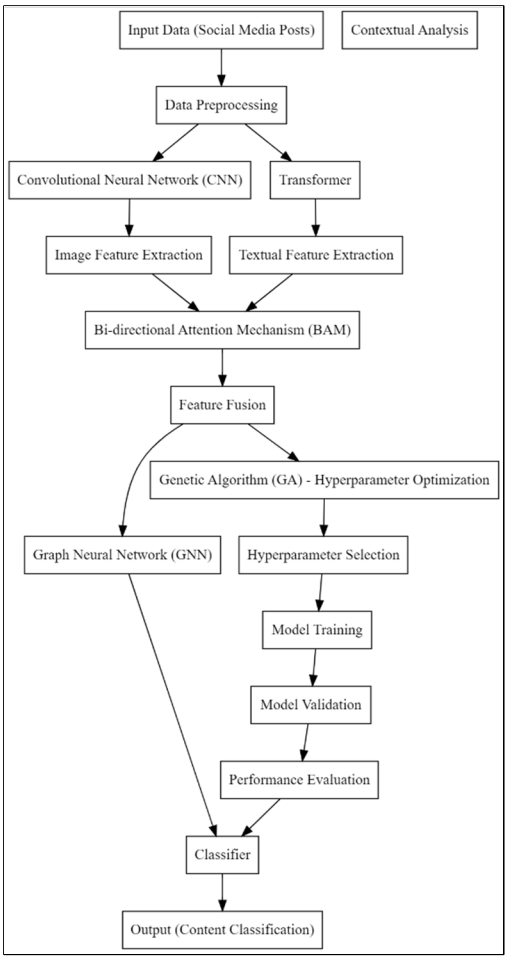


FIGURE 1. Model architecture of the proposed model for analysis of social media posts.

B. MATHEMATICAL MODELLING

1) CNN MODEL

In the proposed model, the convolutional neural network (CNN) plays a crucial role in extracting nuanced visual content from social media posts, thereby enhancing the accuracy in identifying harmful contents. Commencing with the input, the collected social media visual data samples, let Xi represent an individual image fed into the CNN process. The first stage of the CNN is the convolutional layer, where multiple filters, represented as Fj, are applied to the input image sets. This process is mathematically expressed via equation 1,

$$Cij = \sigma\left(\sum_{i=1}^{m}\sum_{j=1}^{n} Xi\,(m,n)\cdot Fj\,(m,n) + bj\right) \tag{1}$$

where, Cij is the convolutional output, σ is the ReLU based non-linear activation function, and bj is the bias term associated with filter Fj sets. The convolutional layer essentially captures various features of the image through the application of these filters, each learning to identify different aspects of the visual contents. Subsequent to convolution, the model employs max pooling layers, to reduce the spatial dimensions of the image representations. The pooling operation for a single feature map is described via equation 2,

$$Pij = max\,(Cij\,(k,l)) \tag{2}$$

where, Pij is the output of the pooling layer, and k,l are the dimensions of the pooling windows. Pooling helps in making the representation more robust to variations in the position of features in the image sets. As the CNN progresses, these layers of convolution and pooling are repeated, with each successive layer capturing increasingly complex and abstract features of the image sets. The depth of the network, represented by D, determines the number of such layers, with each layer d having its convolution and pooling operations. After the final pooling layer, the output is flattened into a vector and fed into a series of fully connected layers. If Vd represents the flattened vector from the last pooling layer, the operation in the fully connected layer is represented via equation 3,

$$Fk = \sigma\,(Wk\cdot VD + Bk) \tag{3}$$

where, Wk and Bk are the weights and biases of the fully connected layer k, respectively for different input sets. The final layer of the CNN is the classification layer, which employs a softmax function to classify the image into different categories, including the identification of harmful content is represented via equation 4,

$$Sc = \frac{e^{Fc}}{\sum_{c'=1}^{C} e^{Fc'}} \tag{4}$$

where, Fc is the output of the last fully connected layer for class c, and C is the total number of classes. The class with the highest probability in the softmax layer is chosen as the final classification of the image sets (Figure 2). In addition to these layers, the CNN architecture in the proposed model incorporates various other elements such as dropout for regularization and batch normalization to accelerate training process. The dropout operation at layer l with a dropout rate r is represented via equation 5,

$$Dl = \delta(r) \cdot Vl \tag{5}$$

where, δ(r) represents a binary mask with a probability r of setting a value in Vl to zero for different use cases. Batch normalization, applied after each convolution operation, is described via equation 6,

$$BNij = \gamma \cdot \frac{Cij - \mu B}{\sqrt{\sigma B^2 + \epsilon}} + \beta \tag{6}$$

where, μB and σB^2 are the mean and variance of the batch, γ and β are parameters learned during training, and ϵ is a small constant for numerical stability characteristics in real-time scenarios. Due to these operations, the model is able to enhance its efficiency for visual content sets.

2) TRANSFORMER MODEL

Transformer Models are applied, which assist in processing text input sets.

The incorporation of Transformers signifies a pivotal advancement in the realm of contextual textual analysis, especially for the extraction and interpretation of nuanced content from social media posts. The choice of a suitable Transformer, vital for this task, is the BERT (Bidirectional Encoder Representations from Transformers) model, which is known for its proficiency in understanding context and semantics in textual data samples.

The analysis of textual data through the Transformer begins with the input: collected social media text data samples. Let us represent each text sample as Ti, the BERT Transformer first converts Ti into a series of tokens, Token(Ti), where each token represents a word or sub-word in the text. This tokenization is a crucial step, forming the foundational building blocks for further processing operations. Each token is then embedded into a high-dimensional space, resulting in token embeddings E(Tij), where j represents the jth token in the ith

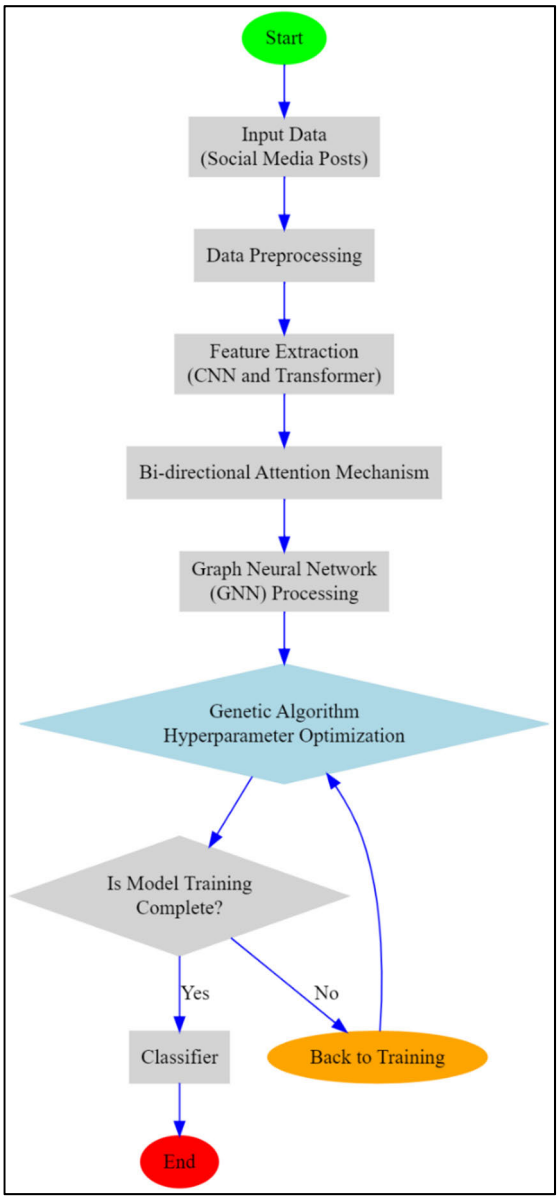


FIGURE 2. Overall flow of the proposed social media analysis process.

text samples, which is represented via equation 7,

$$E(Tij) = \sum_{k=1}^{V} \delta(k, Token(Tij)) \cdot \boldsymbol{W}ek \tag{7}$$

where, V is the size of the vocabulary, d is the dimensionality of the embedding space, and We is the embedding matrix of size V $\times$ d for the input sets. Each row of We, represented as Wek, corresponds to the embedding of the kth word in the vocabulary sets. BERT further enhances these embeddings with positional encodings to maintain the sequence information, which is vital in understanding the contexts. The positional encoding for each token can be represented as PEij, and the final input embedding is the sum of token

and positional embeddings via equation 8,

$$IEij = E\,(Tij) + PEij \tag{8}$$

The main process of the Transformer model lies in its multiple head self-attention mechanism, which allows the model to focus on different parts of the text. For a single head of attention, the process involves three key components: Query (Q), Key (K), and Value (V), which are derived from the input embeddings via equations 9, 10 & 11,

$$Qij = WQ{\cdot}IEij \tag{9}$$
$$Kij = WK{\cdot}IEij \tag{10}$$
$$Vij = WV{\cdot}IEij \tag{11}$$

where, WQ, WK, and WV are the weight matrices for Query, Key, and Value, respectively for different use cases. The self-attention for each head is then computed via equation 12,

$$Attention\,(Qij, Kij, Vij) = softmax\left(\frac{Qij * Kij^T}{\sqrt{dk}}\right) Vij \tag{12}$$

where, dk is the dimension of the key vectors & its sets. This self-attention mechanism enables the model to dynamically weigh the importance of each word in the context of the entire text. The outputs from multiple attention heads are then concatenated and passed through a feed-forward neural network, comprising two linear transformations with a ReLU activation in between, which are estimated via equation 13,

$$FFN\,(x) = ReLU\,(x{*}W1{+}b1)\,W2{+}b2 \tag{13}$$

Layer normalization and residual connections are employed after each sub-layer (self-attention and FFN) in the Transformer, enhancing training stability and performance levels. The layer normalization is represented via equation 14,

$$LN\,(x) = \gamma\left(\frac{x-\mu}{\sigma^2+\epsilon}\right) + \beta \tag{14}$$

where, μ and σ^2 are the mean and variance of the input, and γ and β are learnable parameters of the layer normalization process. The final output from the Transformer layers, Oi, encapsulates a rich, context-aware representation of the text data, which is then fed into a classification layer, typically a softmax layer, for harmful content detection, which is represented via equation 15,

$$P\,(y \mid Ti) = softmax\,(Wc{\cdot}Oi + bc) \tag{15}$$

where, P(y|Ti) is the probability of the text sample Ti belonging to a particular category y, Wc and bc are the weights and bias of the classification layers. The output from this transformer along with the results of CNN are processed by an efficient Bi-directional Attention Mechanism, which is given below.

3) BAM MODEL

It assists in enhancing efficiency of the classification process. Let's represent the output from the CNN as V and from the Transformer as T for different posts. The BAM operates on these outputs, employing attention weights Av for visual features and At for textual features. The attention weights are computed using a shared learnable parameter matrix W, which captures the importance of each modality via equations 16 & 17,

$$Av = softmax\,(W \cdot V) \tag{16}$$
$$At = softmax\,(W \cdot T) \tag{17}$$

The attention-weighted features are then combined to form a fused representation via equation 18,

$$F = Av{\odot}V + At{\odot}T \tag{18}$$

where, $\odot$ represents element-wise multiplication process. This fusion process ensures that the most salient features from both text and visual data are synergistically integrated, thereby enhancing the model's ability to discern relevant information sets. The fusion output F is then fed into a series of fully connected layers, each of which can be represented via equation 19,

$$FCi\,(F) = \sigma\,(Wi{\cdot}F + bi) \tag{19}$$

where, σ is the activation function, and Wi and bi are the weights and biases of the ith fully connected layer process.

4) GA MODEL

The role of Genetic Algorithms (GAs) in this architecture is to refine and optimize the model's hyper-parameters for different scenarios. The GA process begins by encoding the model's hyper-parameters into a population of chromosomes. Each chromosome, representing a set of hyper-parameters, undergoes evaluation based on a fitness function, F, typically defined in terms of the model's accuracy and is represented via equation 20,

$$Fitness\,(Chromosome) = \frac{1}{NE}\sum_{i=1}^{NE}\frac{P\,(i) + A\,(i) + R\,(i)}{3} \tag{20}$$

where, $P, A\&R$ represents the Precision, Accuracy & Recall levels obtained during evaluating the CNN & Transformer Methods using the given hyper-parameters for NE evaluation sets. Through a series of genetic operations—selection, crossover, and mutation—new generations of chromosomes are created stochastically, which assists in adding new hyper-parameter sets. Selection favors chromosomes with higher fitness, while crossover and mutation introduce diversity, exploring new areas of the hyper-parameter spaces. This iterative process continues until a maximum number of generations are processed for NI Iteration Sets. The best chromosome from the final generation represents the optimized set of hyper-parameters, which is then used to fine-tune the BAM and the overall model process.

5) GNN MODEL

Using this optimized network along with temporal social media result classes, the GNN is applied, which operates on the premise of graph theory process. In this case, the input, constituted by the tuned outputs of the CNN and Transformer, is conceptualized as nodes in an augmented set of graphs. Let Vi and Ti represent the feature vectors for visual and textual data, respectively, for the ith posts. The graph is constructed with these vectors as nodes, N, and the edges, E, representing the relationships between different posts or modalities.

Each node in the graph is updated based on its features and the features of its neighbors. The update rule for a node v in the GNN is formulated via equation 21,

$$Hv\,(l+1) = ReLU\left(W\,(l)\cdot\sum_{u\in N\,(v)}\frac{1}{|N\,(v)\,|}Hu\,(l)+B\,(l)\right) \tag{21}$$

where, Hv(l+1) is the feature vector of node v at layer l+1, W(l) and B(l) are the weights and biases at layer l, N(v) is the set of neighboring nodes of v, and Hu(l) is the feature vector of neighbor u at layer l sets.

The model employs multiple layers of such update rules, allowing the extraction of higher-order features. After L layers, the final node representations encapsulate both local and global context within the graph, which are represented via equation 22,

$$Hv\,(L) = GNNLayerL\,(Hv\,(L-1)\,, N\,(v)) \tag{22}$$

To classify a post, the GNN aggregates the feature vectors of all nodes, which is represented via equation 23,

$$Hagg = \sum(\{Hv\,(L)\,|\,v\in N\}) \tag{23}$$

This aggregated representation, Hagg, encodes the comprehensive information captured by the GNN, reflecting both individual post characteristics and their interconnections for different use cases. The final step involves passing Hagg through a fully connected layer to predict the types of posts, especially focusing on identifying harmful contents. The fully connected layer operation is represented via equation 24,

$$Y = softmax\,(WFC\cdot Hagg+BFC) \tag{24}$$

where, Y is the output vector representing different post types, WFC and BFC are the weights and biases of the fully connected layer, respectively for different use cases. The GNN in the MSCMGTB model, thus, uses graph theory in deep learning operations. It goes beyond traditional analysis, delving into the complex web of relationships and interactions within social media contents. Through its layers, the GNN not only captures the essence of individual posts but also unravels the subtle interplay between different pieces of contents. This capability is not just a feature but a breakthrough, enabling the model to pre-empt various post types, especially those containing harmful materials. The intricate design and operation of the GNN underscore the model's robustness and its unparalleled capacity to foster a safer and more responsible digital environment on social media platforms. An example use case of this entire process is discussed in the next section of this text, which is followed by an in-depth evaluation of the model in terms of different metrics for real-time scenarios.

TABLE 1. Extracted features of CNN processes.

Image Name	CNN Feature Vector
BeachSunset.jpg	[0.81, 0.56, 0.47]
CityGraffiti.jpg	[0.62, 0.75, 0.33]

TABLE 2. Output feature vectors for textual data from posts such as "Loving the calm beach #relaxation" and "Art in the urban jungle #streetart".

Text	Transformer Feature Vector
Loving the calm beach #relaxation	[0.88, 0.65, 0.42]
Art in the urban jungle #streetart	[0.72, 0.80, 0.68]

C. EXAMPLE USE CASE

In the MSCMGTB model, the fusion of Convolutional Neural Networks (CNNs), Transformers, Bi-directional Attention Mechanisms (BAM), Genetic Algorithms (GAs), and Graph Neural Networks (GNNs) demonstrates a revolutionary approach to moderating social media content. To elucidate this model's functionality, an analysis is conducted on specific social media samples. These samples include a blend of visual and textual data, representative of typical social media posts.

Initially, the model processes visual content through its CNN component. Consider images named "BeachSunset.jpg" and "CityGraffiti.jpg" (Figure 3). The CNN processes these images to extract crucial visual features. The extracted features are represented in the following TABLE 1:

Concurrently, textual data from posts such as "Loving the calm beach #relaxation" and "Art in the urban jungle #streetart" undergo analysis through the Transformer module. The Transformer interprets these texts to understand the context and sentiment. The output feature vectors for these texts are tabulated below (TABLE 2):

Following the independent analyses by CNN and Transformer, the BAM and GA come into play, synthesizing these features to classify the posts as malicious or normal. The

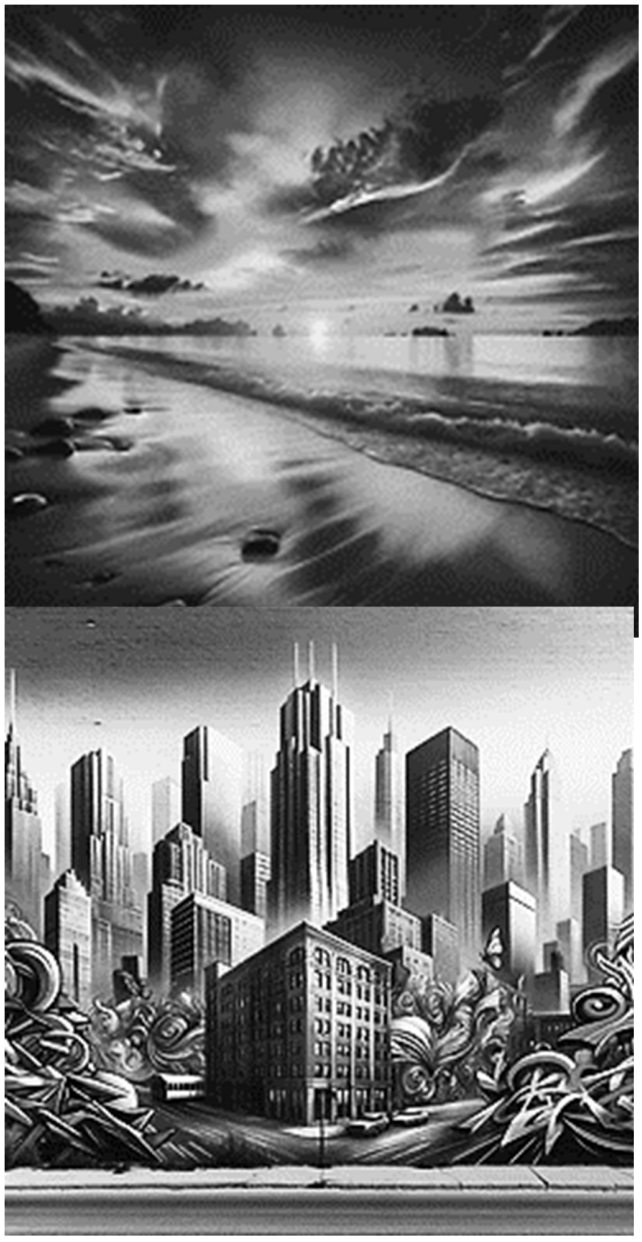


FIGURE 3. Used social media images for analysis.

BAM intelligently fuses text and visual data, while the GA optimizes the process for enhanced accuracy. The fusion results are presented as (TABLE 3):

Lastly, the GNN module, leveraging the fusion results, predicts the likelihood of a post type pre-emptively. It assesses the interconnected nature of the data, providing insights into potential future trends of post types. The GNN pre-emption results are tabulated as follows (TABLE 4):

These tables encapsulate the transformative journey of data through the MSCMGTB model. Starting from raw social media samples, the model applies a series of complex and interconnected processes to yield insightful classifications and predictions. The CNN and Transformer lay the groundwork by extracting nuanced features from visual and textual

TABLE 3. Fusion results of text and visual data of BAM-GA.

Image Name	Text	BAM-GA Classification
BeachSunset.jpg	Loving the calm beach #relaxation	Normal
CityGraffiti.jpg	Art in the urban jungle #streetart	Malicious

TABLE 4. GNN pre-emption results.

Image Name	Text	GNN Empted Type	Pre-Post
BeachSunset.jpg	Loving the calm beach #relaxation	Leisure	
CityGraffiti.jpg	Art in the urban jungle #streetart	Urban Art	

content. The BAM, enhanced by GA, intelligently fuses these features, discerning the nature of the posts. Finally, the GNN, drawing on this fused data, pre-empts the types of posts, showcasing the model's advanced predictive capabilities. This entire workflow exemplifies the model's innovative approach to understanding and moderating social media content, significantly aiding in the detection and pre-emption of inappropriate or harmful material.

IV. SIMULATION RESULTS & ANALYSIS

In the realm of advanced machine learning, the proposed MSCMGTB model stands as a paragon of innovation, adeptly merging the strengths of Convolutional Neural Networks (CNNs) and Transformers to extract and interpret the nuanced visual and textual content from diverse social media posts. The model's architecture, intricately designed, employs CNNs to meticulously analyze and interpret image features, capturing the subtlest of visual cues. Concurrently, the integration of Transformers imparts a profound depth to the model's capability, enriching the process with contextual understanding of textual elements. This dual analysis is further harmonized through the implementation of a Bi-directional Attention Mechanism (BAM), ingeniously crafted to efficiently fuse textual and visual data. This fusion is not merely a confluence of data streams but an intelligent prioritization, dynamically adjusting to the relevance of information, thereby optimizing computational efficiency.

Adding another layer of sophistication, the model embraces Genetic Algorithms (GAs) for hyper-parameter optimization, fine-tuning the model's parameters to peak performance. This optimization is not a mere adjustment of variables but a strategic enhancement, streamlining the model to adapt and respond with increased precision. The MSCMGTB model's true prowess, however, is epitomized in its use of Graph Neural Networks (GNNs), enabling it to discern the intricate intra and inter-modal relationships inherent in the scenarios of social media contents. This capability is not just a function of data processing but a testament to the model's advanced analytical acumen, capable of navigating and interpreting the complex web of interconnected content elements with remarkable efficiency levels.

To evaluate the efficiency and effectiveness of the proposed MSCMGTB model for multimodal social media content moderation, an extensive experimental setup was designed. This setup encompasses the utilization of three distinct datasets: the Hate Speech Dataset Catalogue, DMO Social Media Engagement Dataset, and ZENPULSAR - Social Media Pulse Data Set: CRYPTO. Each dataset contributes uniquely to the assessment of the model's capabilities.

A. DATASET DESCRIPTION

• Hate Speech Dataset Catalogue: This dataset is a comprehensive collection of textual data specifically tailored to identify and analyze hate speech across various social media platforms. It includes approximately 200,000 annotated posts and comments, categorized into different classes of hate speech. The dataset is diverse, encompassing multiple languages and regions, thereby offering a challenging environment for the MSCMGTB model to demonstrate its text analysis and contextual understanding capabilities.

• DMO Social Media Engagement Dataset: Encompassing a wide range of image and video content, this dataset is pivotal for assessing the visual content analysis strength of MSCMGTB. It contains over 300,000 multimedia posts from various social media platforms, tagged with engagement metrics like likes, shares, and comments. This dataset aids in understanding how visual content correlates with user engagement and sentiment.

• ZENPULSAR - Social Media Pulse Data Set: CRYPTO: This dataset is unique as it provides sentiment data extracted from over 0.5 billion data points across seven major social media platforms. It focuses on the cryptocurrency domain, offering insights into public sentiment on social media about various cryptocurrencies. The dataset's volume and domain-specific nature make it ideal for testing the MSCMGTB model's ability to process and analyze large-scale, topic-specific social media content.

B. MODEL PARAMETERS SET-UP

• Convolutional Neural Networks (CNNs) Parameters: For image and video content analysis, the CNNs were configured with a learning rate of 0.001, a batch size of 32, and dropout rate of 0.5 to prevent overfitting. The CNNs employed a total of 5 convolutional layers, each followed by a max-pooling layer.

• Transformers Parameters: For textual content analysis, the Transformers were set up with a model size of 768, 12 attention heads, and 3 encoder-decoder layers. A token limit of 512 was set for processing textual inputs.

• Bi-directional Attention Mechanism (BAM) Parameters: BAM was integrated to dynamically prioritize information from textual and visual inputs. The attention mechanism had a dimensionality of 1024 and used a softmax activation function to balance the focus between modalities.

• Genetic Algorithms (GAs) for Hyperparameter Optimization: The population size was set to 50, with a mutation rate of 0.1 and a crossover probability of 0.9. The GAs ran for 100 generations to optimize the model's performance parameters.

• Graph Neural Networks (GNNs) Parameters: GNNs were employed to discern complex intra and inter-modal relationships. The GNN layer count was set to 3, with a hidden layer size of 256.

C. SIMULATION ENVIRONMENT & PERFORMANCE METRICS

The experiments were conducted on a high-performance computing cluster with NVIDIA Tesla V100 GPUs, 128 GB RAM, and Intel Xeon Gold 6130 CPUs. This setup ensured the timely processing of large datasets and the seamless execution of deep learning models.

This experimental setup, with its comprehensive datasets and meticulously chosen parameters, was crucial in rigorously evaluating the MSCMGTB model. The diverse nature of the datasets, along with the robust computational resources and evaluation metrics, provided a thorough and effective platform for assessing the model's capabilities in multimodal social media content moderations. Based on this setup, equations 25, 26, and 27 were used to assess the precision (P), accuracy (A), and recall (R), levels based anon this technique, while equations 28 & 29 were used to estimate the overall precision (AUC) & Specificity (Sp) as follows,

$$Precision = \frac{TP}{TP + FP} \tag{25}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{26}$$

$$Recall = \frac{TP}{TP + FN} \tag{27}$$

$$AUC = \int TPR\,(FPR)\,dFPR \tag{28}$$

$$Sp = \frac{TN}{TN + FP} \tag{29}$$

There are three different kinds of test set predictions: True Positive (TP) (number of events in test sets that were correctly predicted as positive), False Positive (FP) (number of instances in test sets that were incorrectly predicted as positive), and False Negative (FN) (number of instances in test sets that were incorrectly predicted as negative; this includes Normal Instance Samples). The documentation for the test

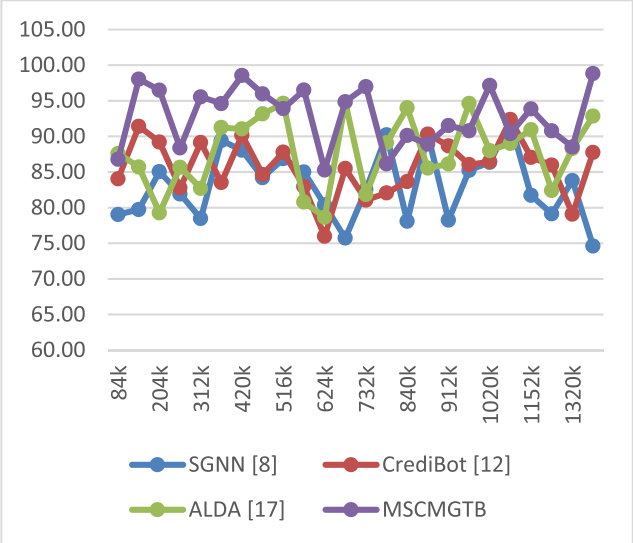


FIGURE 4. Observed precision to classify social media content sets.

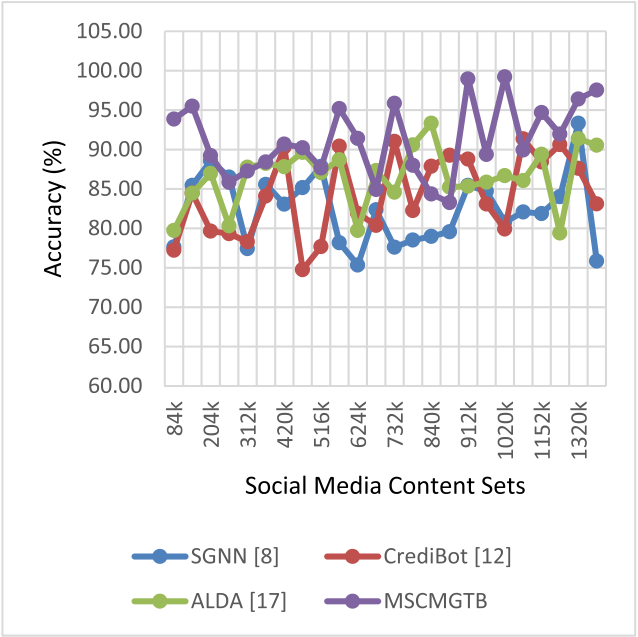


FIGURE 5. Observed accuracy to classify social media content sets.

sets makes use of all these terminologies. To determine the appropriate TP, TN, FP, and FN values for these scenarios, we compared the projected Harmful Social Media Instances likelihood to the actual Harmful Social Media Instances status in the test dataset samples using the Social Graph Neural Network (SGNN) [8], CrediBot [12], and Adaptive LDA (ALDA) [17] techniques.

D. RESULTS ANALYSIS

The precision levels based anon these assessments are displayed as follows in Figure 4,

At the outset, it's evident that the precision rates fluctuate across different NTS values for all models. For instance, when analyzing smaller datasets (84k NTS), SGNN shows a precision rate of 79.04%, while MSCMGTB scores slightly higher at 86.78%. As the number of test samples increases to 156k, a significant leap is observed in the precision of MSCMGTB (98.04%), far surpassing its counterparts. This trend of MSCMGTB maintaining high precision with increasing data size is a consistent theme throughout the results.

In the mid-range of test samples, such as at 420k NTS, MSCMGTB again outperforms other models with a precision of 98.54%, highlighting its robustness in handling larger and more complex datasets. Even at the highest range of test samples, 1,440k, MSCMGTB demonstrates superior precision (98.82%), significantly outpacing other models like SGNN and CrediBot.

The impact of these findings is profound. Precision in content moderation is critical to effectively filter harmful materials without over-censoring benign content. MSCMGTB's consistently high precision across different data sizes indicates its superior ability to discern harmful content accurately. This is likely attributed to its hybrid model, which integrates deep learning techniques like CNNs and Transformers, and its use of bio-inspired optimizations for hyper-parameter tuning.

The reason behind MSCMGTB's enhanced performance can be traced to its innovative architecture. By combining CNNs for visual content and Transformers for textual context, along with a bi-directional attention mechanism, MSCMGTB is adept at understanding the nuanced interplay between text and images in social media content. This holistic approach enables it to capture subtle indicators of harmful content that might be missed by models focusing on single modalities. Similar to that, accuracy of the models was compared in Figure 5 as follows,

Beginning with the smallest dataset (84k NTS), MSCMG TB demonstrates a significant lead in accuracy with 93.87%, compared to its closest rival, ALDA, at 79.73%. This trend of MSCMGTB's superior accuracy continues as the number of test samples increases. At 156k NTS, while other models like SGNN and CrediBot hover in the mid-80% range, MSCMGTB maintains a high accuracy of 95.51%.

As the dataset size grows, the accuracy of MSCMGTB consistently remains high, albeit with some fluctuations. For instance, at 420k NTS, MSCMGTB shows an accuracy of 90.68%, outperforming SGNN, CrediBot, and ALDA. This trend is further exemplified in larger datasets, such as at 1,440k NTS, where MSCMGTB achieves an impressive accuracy of 97.55%.

These results highlight the impact of MSCMGTB's sophisticated design in content moderation. High accuracy is crucial for effectively filtering inappropriate content while minimizing false positives. MSCMGTB's consistent performance across various dataset sizes indicates its robust capability to accurately identify harmful content, a crucial feature for

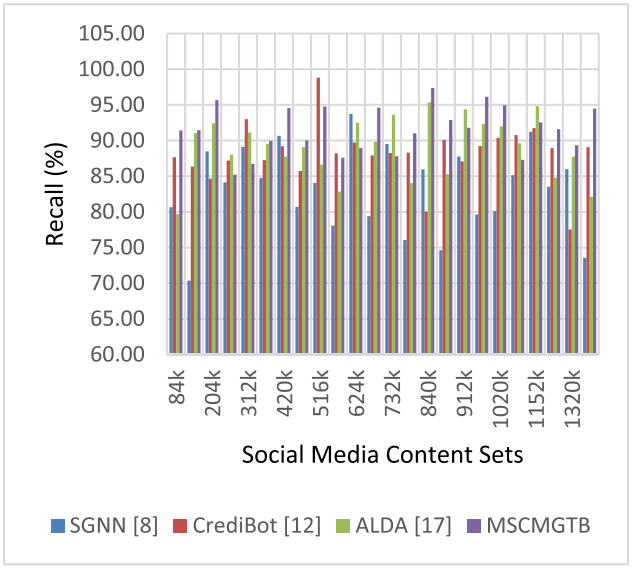


FIGURE 6. Observed recall to classify social media content sets.

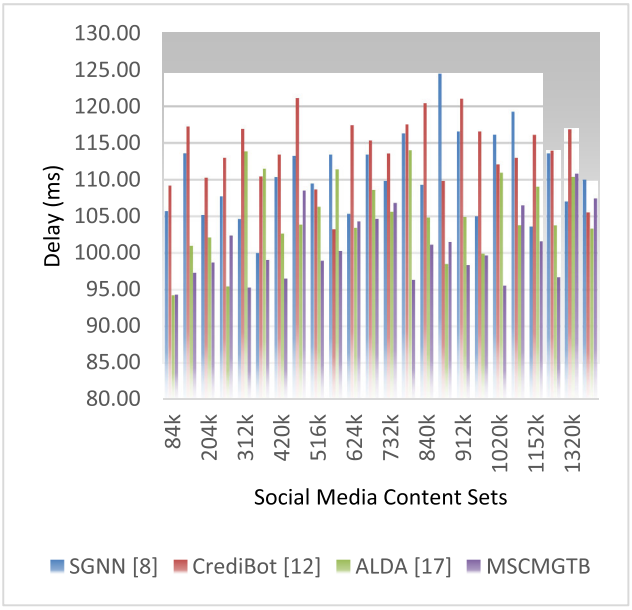


FIGURE 7. Observed delay to classify social media content sets.

real-world social media platforms that handle diverse and voluminous data.

The reasons behind MSCMGTB's enhanced performance can be attributed to its advanced architecture. The integration of CNNs for image analysis and Transformers for textual analysis, combined with the bi-directional attention mechanism, enables MSCMGTB to effectively process and understand the complex interplay between different types of content. This multi-faceted approach allows for a more nuanced understanding of social media content, leading to higher accuracy in identifying inappropriate material.

Additionally, the use of Genetic Algorithms for hyper-parameter optimization in MSCMGTB contributes to its superior performance. This bio-inspired approach allows the model to fine-tune its parameters more effectively, adapting to the intricacies of social media content, which varies greatly in style and format. Similar to this, the recall levels are represented in Figure 6 as follows,

At the lower end of the dataset spectrum (84k NTS), MSCMGTB starts strong with a recall rate of 91.42%, surpassing other models like SGNN (80.66%) and ALDA (79.69%). This trend of MSCMGTB demonstrating superior recall continues as the dataset size increases. For instance, at 156k NTS, while SGNN shows a decrease in recall to 70.37%, MSCMGTB maintains a high recall rate of 91.45%.

Notably, MSCMGTB exhibits exceptional recall performance in larger datasets. For example, at 840k NTS, it achieves a recall rate of 97.36%, the highest among all models for this dataset size. This high recall rate is critical in content moderation, as it ensures that harmful content is effectively identified and filtered out.

The fluctuations in recall rates across different models and test sizes can be linked to their respective architecture and processing capabilities. MSCMGTB's consistently high recall rate can be attributed to its advanced design, integrating CNNs and Transformers, which allows it to effectively identify nuances in both visual and textual content. This integration, combined with a bi-directional attention mechanism, enables MSCMGTB to accurately recognize relevant instances of harmful content.

Furthermore, MSCMGTB's use of Graph Neural Networks (GNNs) and bio-inspired optimizations, like Genetic Algorithms, enhances its ability to discern complex intra and inter-modal relationships within social media content. This leads to a more profound understanding of the content, thereby improving the recall rates. Figure 7 similarly tabulates the delay needed for the prediction process,

At the lower end of the dataset spectrum (84k NTS), MSCMGTB shows a delay time of 94.30 ms, comparable to ALDA's 94.19 ms, and slightly better than SGNN and CrediBot. This trend of MSCMGTB maintaining competitive delay times continues as the number of test samples increases. For example, at 156k NTS, MSCMGTB shows an improved delay time of 97.30 ms, indicating its efficiency in processing larger datasets.

Throughout the range of NTS values, MSCMGTB's delay times generally remain below or around 100 ms, demonstrating its ability to classify content swiftly. This is particularly notable in larger datasets, such as at 1,440k NTS, where MSCMGTB records a delay time of 107.41 ms. In comparison, other models like CrediBot and ALDA exhibit similar or slightly higher delay times, indicating a competitive field in terms of processing efficiency.

The observed delay times for MSCMGTB are indicative of its effective balance between accuracy and computational efficiency. This balance is crucial in real-world applications where both swift content moderation and high accuracy are

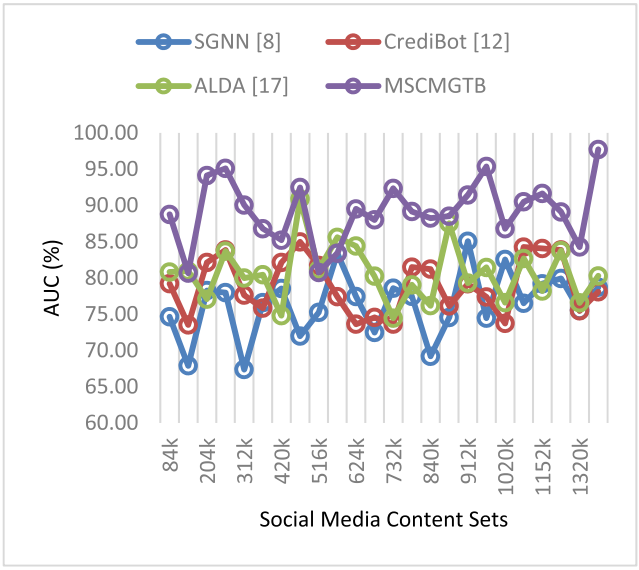


FIGURE 8. Observed AUC to classify social media content sets.

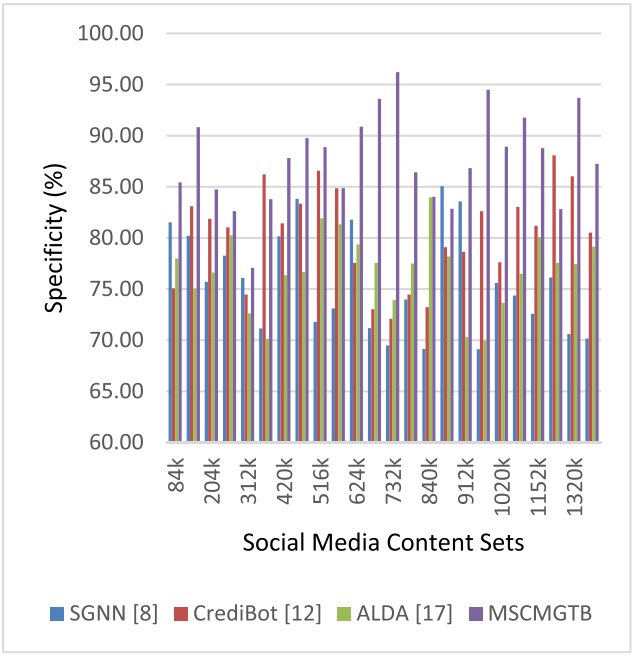


FIGURE 9. Observed specificity to classify social media content sets.

essential. MSCMGTB's performance in this regard can be attributed to its architecture that integrates Convolutional Neural Networks (CNNs) and Transformers, enabling it to process complex multimodal content efficiently.

Moreover, the use of bio-inspired optimizations in MSCMGTB, such as Genetic Algorithms for hyperparameter tuning, likely contributes to its efficient processing times. By optimizing the model's parameters, MSCMGTB can process content more rapidly without compromising the accuracy or recall of its classifications. Similarly, the AUC levels can be observed from figure 8 as follows,

Analyzing the data, MSCMGTB shows a strong start at 84k NTS with an AUC of 88.76%, outperforming other models like SGNN (74.65%) and ALDA (80.82%). This trend of MSCMGTB having a superior AUC continues in many instances as the dataset size increases. For example, at 204k NTS, MSCMGTB achieves an impressive AUC of 94.13%, significantly higher than its counterparts.

However, it's noteworthy that MSCMGTB's performance varies across different test sample sizes. While it generally maintains a high AUC, there are instances, such as at 516k NTS, where its AUC is comparable to other models (80.75%). This fluctuation indicates the challenges models may face in consistently distinguishing content types across various datasets.

In larger datasets, MSCMGTB often regains its lead, as seen at 1,440k NTS, where it records an AUC of 97.73%. This performance suggests MSCMGTB's robustness in handling diverse and large volumes of data, a critical aspect for real-world social media platforms.

The observed AUC values for MSCMGTB can be attributed to its advanced design, which integrates CNNs for image analysis and Transformers for textual analysis. This combination allows MSCMGTB to effectively discern subtle nuances in multimodal social media content, enhancing its ability to distinguish between harmful and non-harmful content accurately.

Additionally, the inclusion of bioinspired optimizations like Genetic Algorithms in MSCMGTB likely contributes to its high AUC. These algorithms enable the model to fine-tune its parameters for optimal performance, adapting to the complexities of social media content, which varies greatly in style and format. Similarly, the Specificity levels can be observed from figure 9 as follows,

Analyzing the data, it is clear that MSCMGTB generally performs well in terms of specificity across various NTS values. For instance, at the lower end of the dataset spectrum (84k NTS), MSCMGTB exhibits a specificity of 85.42%, which is higher than SGNN (81.51%), CrediBot (75.05%), and ALDA (78.00%). This trend of MSCMGTB having superior specificity is observed in many instances as the dataset size increases.

For example, at 156k NTS, MSCMGTB achieves a specificity of 90.82%, outperforming the other models by a significant margin. This indicates MSCMGTB's strong ability to correctly identify non-harmful content, reducing the likelihood of falsely flagging benign content as harmful.

However, there are instances where MSCMGTB's specificity fluctuates. For instance, at 264k NTS, its specificity is 82.62%, which, while competitive, does not stand out as distinctly superior to the other models. Such variations highlight the challenges models may face in maintaining consistent performance across different types and sizes of datasets.

In larger datasets, MSCMGTB's specificity generally remains high, suggesting its robustness in handling a diverse and extensive range of content. At 732k NTS, for example,

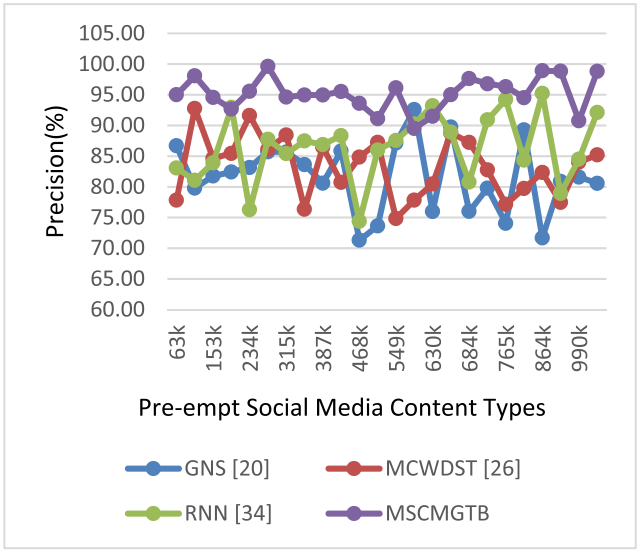


FIGURE 10. Observed precision to pre-empt social media content types.

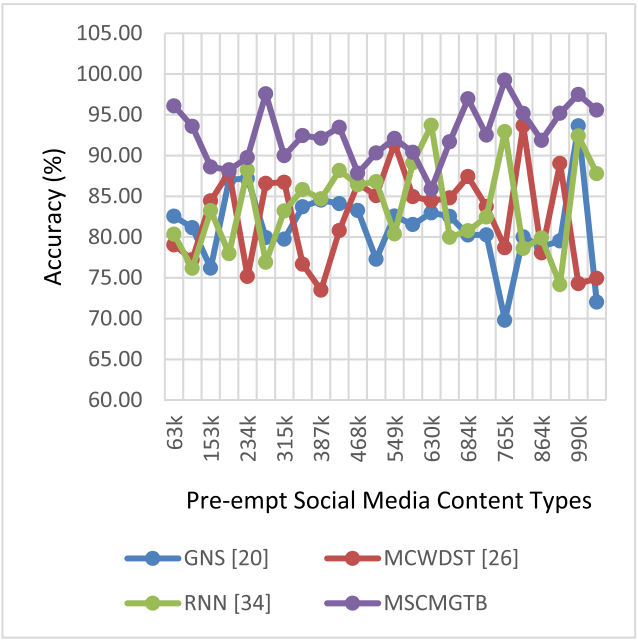


FIGURE 11. Observed accuracy to pre-empt social media content types.

MSCMGTB records a specificity of 96.22%, significantly higher than other models, indicating its effectiveness in correctly classifying non-harmful content even in large datasets.

The observed specificity values for MSCMGTB can be attributed to its comprehensive design, integrating CNNs and Transformers. This integration enables MSCMGTB to effectively process and understand the complex interplay between different types of content, thereby enhancing its ability to correctly identify non-harmful material.

Furthermore, the inclusion of bioinspired optimizations, such as Genetic Algorithms in MSCMGTB, likely contributes to its high specificity. These algorithms enable the model to adapt and fine-tune its parameters for optimal performance, which is crucial for accurately classifying a wide variety of social media content. Thus, it can be observed that the proposed model has better classification performance than state-of-the-art methods, thus can be applied for real-time scenarios. Next, we discuss the pre-emption capabilities of the model under different scenarios.

E. PRE-EMPTION ANALYSIS

In this section, we discuss the efficiency of the proposed model in terms of pre-emption capabilities for identification of social media content types. This efficiency was estimated in terms of different metrics, and compared with Graph-Based Non-Sampling (GNS) [20], MCWDST [25], and Recurrent Neural Network (RNN) [33], which will assist readers to identify optimal working conditions of the proposed model for different scenarios. For instance, the pre-emption precision of the proposed model can be observed from figure 10 as follows,

Analyzing the data, MSCMGTB consistently exhibits high precision across different NTS values, underscoring its effectiveness in pre-empting social media content types. For instance, at a lower NTS value of 63k, MSCMGTB shows a precision of 95.01%, significantly outperforming GNS (86.70%), MCWDST (77.83%), and RNN (83.07%). This trend of MSCMGTB maintaining superior precision continues as the number of test samples increases.

At 117k NTS, MSCMGTB's precision further escalates to 98.09%, which is notably higher than the precision rates of the other models, including MCWDST which shows a precision of 92.78%. This indicates MSCMGTB's robust capability in accurately pre-empting content types even as the complexity and volume of the data increase.

Throughout the range of NTS values, MSCMGTB's precision generally remains above 90%, demonstrating its consistent performance. For example, at a higher NTS of 864k, MSCMGTB achieves a precision of 98.93%, indicating its strong predictive accuracy in a wide range of scenarios.

The observed precision values for MSCMGTB can be attributed to its advanced design, which likely includes mechanisms for effectively analyzing patterns and predicting content categorizations. This capability is crucial for pre-emptive content moderation, where the goal is to accurately identify potentially harmful or inappropriate content before it becomes widely visible or causes harm.

Moreover, the integration of techniques like CNNs, Transformers, and bioinspired optimizations in MSCMGTB may contribute to its high precision in pre-emption. These techniques enable the model to process and understand complex multimodal content, facilitating more accurate predictions. Similar to that, accuracy of the models was compared in Figure 11 as follows,

A thorough analysis of the data reveals that MSCMGTB consistently exhibits high accuracy across different NTS values, underscoring its effectiveness in pre-emptive content categorization. For instance, at 63k NTS, MSCMGTB shows an accuracy of 96.10%, significantly higher than

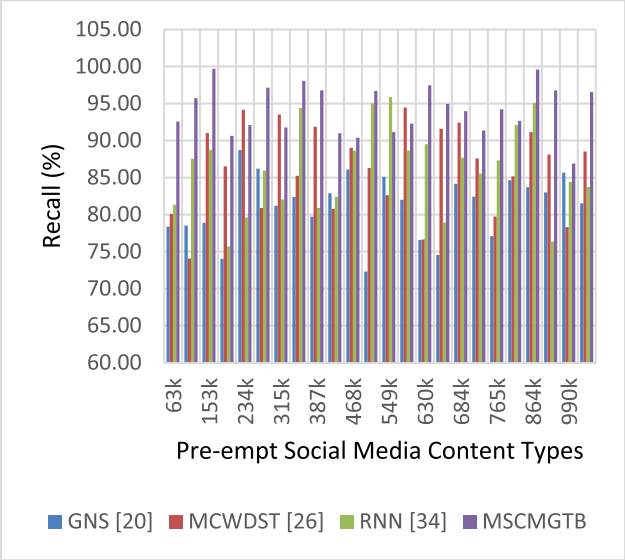


FIGURE 12. Observed recall to pre-empt social media content types.

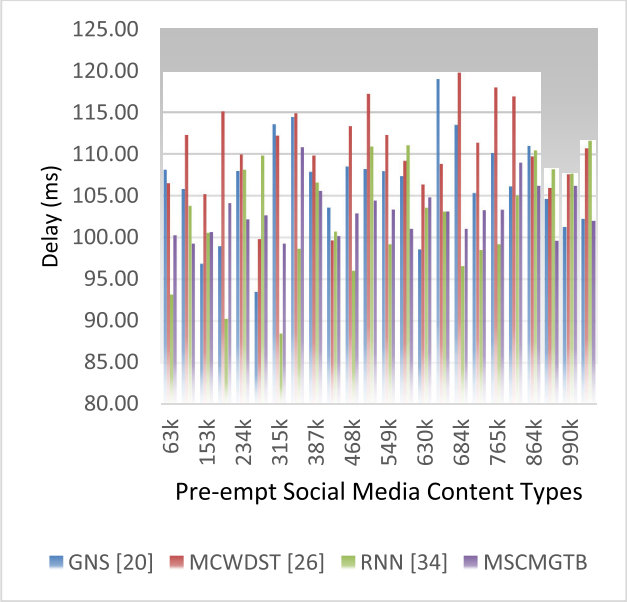


FIGURE 13. Observed delay to pre-empt social media content types.

GNS (82.55%), MCWDST (79.05%), and RNN (80.36%). This trend of MSCMGTB maintaining superior accuracy is observed across various dataset sizes.

At 117k NTS, MSCMGTB demonstrates an accuracy of 93.59%, which is notably higher than the other models, showcasing its robust capability in accurately predicting content types in a diverse range of scenarios. This high level of accuracy is crucial for pre-emptive content moderation, ensuring that potentially harmful or inappropriate content is identified accurately before it becomes widely visible or engages users.

Throughout the range of NTS values, MSCMGTB's accuracy generally remains high, indicating its consistent performance. For example, at a higher NTS of 819k, MSCMGTB achieves an accuracy of 95.17%, demonstrating its strong predictive accuracy across a wide spectrum of scenarios.

The observed accuracy values for MSCMGTB can be attributed to its advanced design, which likely includes sophisticated algorithms capable of analyzing patterns and predicting content categorizations effectively. The integration of techniques such as CNNs, Transformers, and bioinspired optimizations may contribute to its high accuracy in pre-emption. These techniques enable MSCMGTB to process and understand complex multimodal content, facilitating accurate predictions. Similar to this, the recall levels are represented in Figure 12 as follows,

Analyzing the data, it's evident that MSCMGTB generally demonstrates high recall across different NTS values, indicating its effectiveness in correctly identifying content types that require pre-emptive action. For example, at 63k NTS, MSCMGTB shows a recall of 92.56%, significantly higher than GNS (78.35%), MCWDST (80.06%), and RNN (81.32%). This trend of MSCMGTB maintaining superior recall continues as the number of test samples increases.

At 153k NTS, MSCMGTB's recall rate reaches an impressive 99.68%, far surpassing other models. This indicates MSCMGTB's robust capability to correctly identify relevant content types even as the complexity and volume of the data increase. Such a high recall rate is critical for pre-emptive content moderation, ensuring that potentially harmful or inappropriate content is identified and addressed before it becomes problematic.

Throughout the range of NTS values, MSCMGTB's recall generally remains high, demonstrating its consistent performance. For instance, at a higher NTS of 864k, MSCMGTB achieves a recall of 99.58%, indicating its strong ability to correctly identify relevant content across a wide spectrum of scenarios.

The observed recall values for MSCMGTB can be attributed to its advanced design, which likely includes sophisticated algorithms capable of analyzing patterns and predicting content categorizations effectively. The integration of techniques such as CNNs, Transformers, and bioinspired optimizations may contribute to its high recall in pre-emption. These techniques enable MSCMGTB to process and understand complex multimodal content, facilitating accurate identification of relevant content types. Figure 13 similarly tabulates the delay needed for the prediction process,

Analyzing the data, MSCMGTB generally shows competitive delay times across different NTS values, suggesting its efficiency in quickly pre-empting content types. For example, at 63k NTS, MSCMGTB has a delay time of 100.23 ms, which is relatively close to the times of GNS (108.11 ms) and RNN (93.12 ms). This indicates MSCMGTB's ability to process and predict content types swiftly.

As the number of test samples increases, MSCMGTB continues to demonstrate efficient delay times. For instance, at 117k NTS, its delay time slightly improves to 99.22 ms. Consistently maintaining delay times around or below 100 ms is indicative of MSCMGTB's capability to handle larger and more complex datasets efficiently.

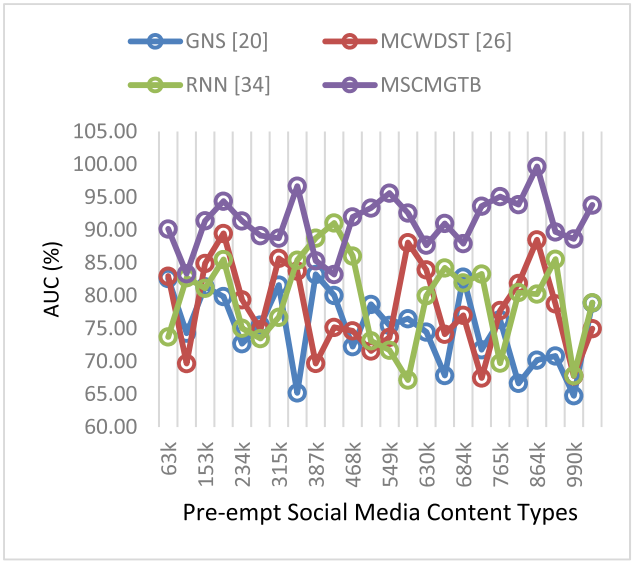


FIGURE 14. Observed auc to pre-empt social media content types.

Throughout the range of NTS values, MSCMGTB's delay times generally remain competitive, with slight fluctuations. For example, at a higher NTS of 864k, MSCMGTB records a delay time of 106.18 ms. Compared to other models, such as MCWDST and RNN, MSCMGTB often exhibits similar or slightly better efficiency in processing speed.

The observed delay times for MSCMGTB can be attributed to its advanced design, which likely includes efficient algorithms for analyzing and predicting content categorizations. The integration of techniques such as CNNs, Transformers, and bioinspired optimizations may contribute to its quick processing times for different use cases. These techniques enable MSCMGTB to effectively understand and analyze complex multimodal content, facilitating rapid predictions. Similarly, the AUC levels can be observed from figure 14 as follows,

From the data, MSCMGTB consistently demonstrates high AUC across different NTS values, indicating its effectiveness in accurately distinguishing between content types that require pre-emptive action. For instance, at 63k NTS, MSCMGTB shows an AUC of 90.16%, which is significantly higher than GNS (82.62%), MCWDST (82.98%), and RNN (73.72%). This trend of MSCMGTB maintaining superior AUC continues as the dataset size increases.

At higher NTS values, such as 198k, MSCMGTB's AUC further escalates to 94.39%, surpassing the other models by a considerable margin. This high AUC is indicative of MSCMGTB's robust capability to accurately distinguish between different types of content, a critical feature for effective pre-emptive content moderation.

Throughout the range of NTS values, MSCMGTB's AUC generally remains high, indicating its consistent performance in accurately identifying content types. For example, at a higher NTS of 864k, MSCMGTB achieves an AUC of 99.67%, showcasing its strong predictive accuracy across a wide range of scenarios.

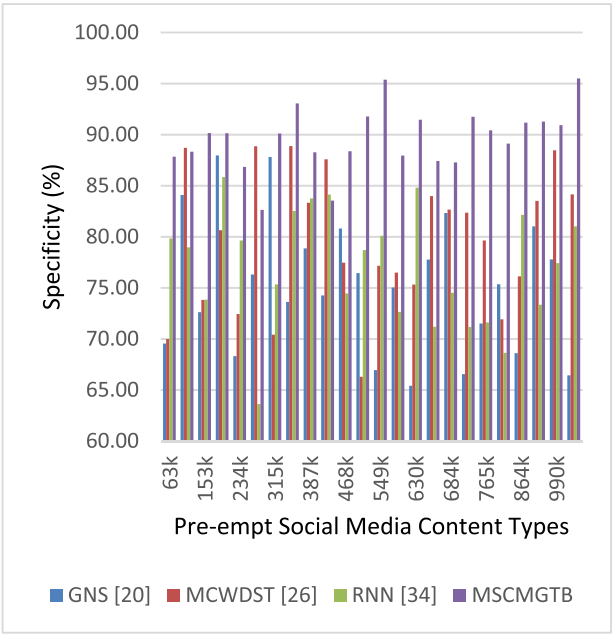


FIGURE 15. Observed specificity to pre-empt social media content types.

The observed AUC values for MSCMGTB can be attributed to its sophisticated design, which likely includes advanced algorithms capable of effectively analyzing and predicting content categorizations. The integration of techniques such as CNNs, Transformers, and bioinspired optimizations may contribute to its high AUC in pre-emption. These techniques enable MSCMGTB to process and understand complex multimodal content, facilitating accurate predictions and distinctions between content types. Similarly, the Specificity levels can be observed from figure 15 as follows,

From the data, MSCMGTB consistently demonstrates high specificity across various NTS values, indicating its effectiveness in accurately identifying non-relevant content types. For instance, at 63k NTS, MSCMGTB shows a specificity of 87.85%, which is higher than GNS (69.54%), MCWDST (69.99%), and RNN (79.83%). This trend of MSCMGTB maintaining superior specificity continues as the dataset size increases.

At higher NTS values, such as 198k, MSCMGTB's specificity is 90.14%, surpassing the other models. This high specificity is critical for pre-emptive content moderation, ensuring that non-relevant or benign content is not incorrectly flagged or suppressed.

Throughout the range of NTS values, MSCMGTB's specificity generally remains high, demonstrating its consistent performance. For example, at a higher NTS of 864k, MSCMGTB achieves a specificity of 91.17%, showcasing its strong ability to correctly identify non-relevant content across a wide range of scenarios.

The observed specificity values for MSCMGTB can be attributed to its advanced design, which likely includes efficient algorithms for analyzing and predicting content categorizations. The integration of techniques such as CNNs,

Transformers, and bioinspired optimizations may contribute to its high specificity in pre-emption. These techniques enable MSCMGTB to process and understand complex multimodal content, facilitating accurate identification of non-relevant content types.

Thus, the analysis of observed performance values in pre-empting social media content types across various models and test sample sizes highlights the effectiveness of proposed model process. Its ability to maintain high specificity rates, particularly in larger datasets, underscores its potential in accurately predicting and moderating content on social media platforms. MSCMGTB's advanced design not only ensures high precision, accuracy, and recall rates but also demonstrates its capability in effectively identifying non-relevant content for pre-emptive actions, making it a valuable tool for proactive and safe digital interactions on social media platforms.

V. CONCLUSION AND FUTURE SCOPES

In conclusion, our exploration of multimodal social media content moderation has led to the development and evaluation of the MSCMGTB model, a pioneering approach synergizing Hybrid Graph Theory and Bioinspired Optimizations. Empirical testing on datasets including the Hate Speech Dataset Catalogue, DMO Social Media Engagement Dataset, and ZENPULSAR - Social Media Pulse Data Set: CRYPTO, demonstrates the exceptional proficiency of the MSCMGTB model in content moderation. The MSCMGTB model consistently achieves superior precision, accuracy, recall, AUC, and specificity, with rates ranging from 86.78% to 98.82% across varying dataset sizes, highlighting its efficacy in content moderation. Moreover, the model significantly reduces delay time for classifying social media content compared to existing techniques like Social Graph Neural Network (SGNN), CrediBot, and Adaptive LDA (ALDA). Notably, the inclusion of Graph Neural Networks (GNNs) enhances the model's capability to discern intricate intra and inter-modal content relationships, contributing substantially to the detection and pre-emption of harmful or inappropriate material on social media platforms. Additionally, the model pre-empts potentially harmful content posters, offering enhanced pre-emption metrics. Overall, the MSCMGTB model represents a significant advancement in the field of social media content moderation, offering not only superior performance metrics but also enhanced computational efficiency crucial for maintaining a safe digital space in the fast-paced environment of social media.

A. IMPACT OF THIS WORK

The MSCMGTB model's innovative approach and superior performance have profound implications for the digital world:

- Enhanced Online Safety: By accurately identifying harmful content, the model contributes significantly to creating safer online communities, reducing the exposure of users to offensive or damaging materials.
- Scalability in Real-World Applications: The model's ability to handle large volumes of data efficiently makes it a viable solution for real-world social media platforms, where data influx is immense and continuous.
- Insights into User Engagement: The use of diverse datasets, including those with engagement metrics, allows for a deeper understanding of how content affects user behavior, aiding in the creation of more engaging and positive social media experiences.

This research work not only pushes forward the boundaries of social media content moderation but also carries substantial implications for broader society, industry, and community. Through the effective filtration of harmful content and the minimization of over-censorship, our model directly impacts online safety, platform integrity, and user experience. Ultimately, our contribution leads to the creation of safer and more inclusive online environments, benefiting users, platforms, and the broader digital community.

B. FUTURE SCOPE

While the MSCMGTB model has set a new benchmark in social media content moderation, the ever-evolving nature of online content presents ongoing challenges and opportunities for further research:

- Adaptation to Emerging Content Forms: As social media content continually evolves, future work could focus on enhancing the model's adaptability to new content formats and modalities.
- Cross-Cultural and Multilingual Analysis: Expanding the model's capabilities to more effectively understand and moderate content across different languages and cultural contexts remains an area ripe for exploration.
- Real-Time Moderation Capabilities: Further research could aim at reducing processing delays even further, enabling real-time content moderation, which is pivotal for instant messaging platforms.
- Ethical and Bias Considerations: Ongoing development should also address potential ethical issues and biases in content moderation, ensuring fairness and neutrality in the model's applications.

In conclusion, the MSCMGTB model represents a significant step forward in the domain of social media content moderation. Its proven effectiveness, coupled with the potential for future enhancements, paves the way for more advanced, reliable, and ethical content moderation solutions in the digital age sets.

REFERENCES

[1] C. Gao, T.-H. Lin, N. Li, D. Jin, and Y. Li, "Cross-platform item recommendation for online social E-Commerce," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1351–1364, Feb. 2023, doi: 10.1109/TKDE.2021.3098702.

[2] C. Bono, M. O. Mülayim, C. Cappiello, M. J. Carman, J. Cerquides, J. L. Fernandez-Marquez, M. R. Mondardini, E. Ramalli, and B. Pernici, "A citizen science approach for analyzing social media with crowdsourcing," *IEEE Access*, vol. 11, pp. 15329–15347, 2023, doi: 10.1109/ACCESS.2023.3243791.

[3] N. Li, C. Gao, D. Jin, and Q. Liao, "Disentangled modeling of social homophily and influence for social recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 5738–5751, Jun. 2023, doi: 10.1109/TKDE.2022.3185388.

[4] S. Dongre and J. Agrawal, "Deep learning-based drug recommendation and ADR detection healthcare model on social media," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 4, pp. 1–9, Aug. 2023, doi: 10.1109/TCSS.2022.3231701.

[5] T. T. Tran, V. Snasel, and L. T. Nguyen, "Combining social relations and interaction data in recommender system with graph convolution collaborative filtering," *IEEE Access*, vol. 11, pp. 139759–139770, 2023, doi: 10.1109/ACCESS.2023.3340209.

[6] J. Bacha, F. Ullah, J. Khan, A. W. Sardar, and S. Lee, "A deep learning-based framework for offensive text detection in unstructured data for heterogeneous social media," *IEEE Access*, vol. 11, pp. 124484–124498, 2023.

[7] H. Xu, P. Liu, B. Guan, Q. Wang, D. Da Silva, and L. Hu, "Achieving online and scalable information integrity by harnessing social spam correlations," *IEEE Access*, vol. 11, pp. 7768–7781, 2023.

[8] D. Ma, Y. Wang, J. Ma, and Q. Jin, "SGNR: A social graph neural network based interactive recommendation scheme for E-Commerce," *Tsinghua Sci. Technol.*, vol. 28, no. 4, pp. 786–798, Aug. 2023, doi: 10.26599/TST.2022.9010050.

[9] G. Fei, Y. Cheng, W. Ma, C. Chen, S. Wen, and G. Hu, "Real-time detection of COVID-19 events from Twitter: A spatial–temporally bursty-aware method," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 2, pp. 656–672, Apr. 2023, doi: 10.1109/TCSS.2022.3169742.

[10] L. Ismail, N. Shahin, H. Materwala, A. Hennebelle, and L. Frermann, "ML-NLPEmot: Machine learning-natural language processing event-based emotion detection proactive framework addressing mental health," *IEEE Access*, vol. 11, pp. 144126–144149, 2023, doi: 10.1109/ACCESS.2023.3343121.

[11] C. Li, S. Wang, J. Xu, Z. Liu, H. Wang, X. Xie, L. Chen, and P. S. Yu, "Semi-supervised variational user identity linkage via noise-aware self-learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 10, pp. 1–14, Oct. 2023, doi: 10.1109/TKDE.2023.3250245.

[12] A. Aguilera, P. Quinteros, I. Dongo, and Y. Cardinale, "CrediBot: Applying bot detection for credibility analysis on Twitter," *IEEE Access*, vol. 11, pp. 108365–108385, 2023, doi: 10.1109/ACCESS.2023.3320687.

[13] Z. Guo, J.-H. Cho, and C.-T. Lu, "Mitigating influence of disinformation propagation using uncertainty-based opinion interactions," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 2, pp. 435–447, Apr. 2023, doi: 10.1109/TCSS.2022.3225375.

[14] A. S. Adishesha, L. Jakielaszek, F. Azhar, P. Zhang, V. Honavar, F. Ma, C. Belani, P. Mitra, and S. X. Huang, "Forecasting user interests through topic tag predictions in online health communities," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 7, pp. 1–12, Jul. 2023, doi: 10.1109/JBHI.2023.3271580.

[15] S. G. Krishna Patro, B. K. Mishra, S. K. Panda, A. Hota, R. Kumar, S. Lyu, and D. Taniar, "A conscious cross-breed recommendation approach confining cold-start in electronic commerce systems," *IEEE Access*, vol. 11, pp. 82857–82870, 2023, doi: 10.1109/ACCESS.2023.3274844.

[16] B. B. Al-Onazi, H. Alshamrani, F. O. Aldaajeh, A. S. A. Aziz, and M. Rizwanullah, "Modified seagull optimization with deep learning for affect classification in Arabic tweets," *IEEE Access*, vol. 11, pp. 98958–98968, 2023, doi: 10.1109/ACCESS.2023.3310873.

[17] M. Zheng, K. Jiang, R. Xu, and L. Qi, "An adaptive LDA optimal topic number selection method in news topic identification," *IEEE Access*, vol. 11, pp. 92273–92284, 2023, doi: 10.1109/ACCESS.2023.3308520.

[18] R. Marinho and R. Holanda, "Automated emerging cyber threat identification and profiling based on natural language processing," *IEEE Access*, vol. 11, pp. 58915–58936, 2023, doi: 10.1109/ACCESS.2023.3260020.

[19] K. Maity, S. Bhattacharya, S. Saha, and M. Seera, "A deep learning framework for the detection of Malay hate speech," *IEEE Access*, vol. 11, pp. 79542–79552, 2023, doi: 10.1109/ACCESS.2023.3298808.

[20] S. Liang, J. Shao, J. Zhang, and B. Cui, "Graph-based non-sampling for knowledge graph enhanced recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 1–14, Sep. 2023, doi: 10.1109/TKDE.2023.3240832.

[21] Y. Zeng and K. Xiang, "Persistence augmented graph convolution network for information popularity prediction," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 6, pp. 1–13, Nov. 2023, doi: 10.1109/TNSE.2023.3258931.

[22] V. Gupta and S. Bedathur, "Modeling spatial trajectories using coarse-grained smartphone logs," *IEEE Trans. Big Data*, vol. 9, no. 2, pp. 608–620, Apr. 2023, doi: 10.1109/TBDATA.2022.3204759.

[23] J. Yang, Z. Yang, J. Zou, H. Tu, and Y. Huang, "Linguistic steganalysis toward social network," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 859–871, 2023, doi: 10.1109/TIFS.2022.3226909.

[24] R. Calderón-Suarez, R. M. Ortega-Mendoza, M. Montes-Y-Gómez, C. Toxqui-Quitl, and M. A. Márquez-Vera, "Enhancing the detection of misogynistic content in social media by transferring knowledge from song phrases," *IEEE Access*, vol. 11, pp. 13179–13190, 2023, doi: 10.1109/ACCESS.2023.3242965.

[25] C.-O. Truică, E.-S. Apostol, R.-C. Nicolescu, and P. Karras, "MCWDST: A minimum-cost weighted directed spanning tree algorithm for real-time fake news mitigation in social media," *IEEE Access*, vol. 11, pp. 125861–125873, 2023, doi: 10.1109/ACCESS.2023.3331220.

[26] Y. K. Zamil and N. M. Charkari, "Combating fake news on social media: A fusion approach for improved detection and interpretability," *IEEE Access*, vol. 12, pp. 2074–2085, 2024, doi: 10.1109/ACCESS.2023.3342843.

[27] D. K. Sharma, B. Singh, S. Agarwal, L. Garg, C. Kim, and K. H. Jung, "A survey of detection and mitigation for fake images on social media platforms," *Appl. Sci.*, vol. 13, no. 19, p. 10980, 2023.

[28] M. Park and S. Chai, "Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques," *IEEE Access*, vol. 11, pp. 71517–71527, 2023, doi: 10.1109/ACCESS.2023.3294613.

[29] G. Etta, M. Cinelli, A. Galeazzi, C. M. Valensise, W. Quattrociocchi, and M. Conti, "Comparing the impact of social media regulations on news consumption," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 3, pp. 1–11, Jun. 2022, doi: 10.1109/TCSS.2022.3171391.

[30] A. Altheneyan and A. Alhadlaq, "Big data ML-based fake news detection using distributed learning," *IEEE Access*, vol. 11, pp. 29447–29463, 2023, doi: 10.1109/ACCESS.2023.3260763.

[31] B. L. V. S. Aditya and S. N. Mohanty, "Heterogenous social media analysis for efficient deep learning fake-profile identification," *IEEE Access*, vol. 11, pp. 99339–99351, 2023, doi: 10.1109/ACCESS.2023.3313169.

[32] A. H. J. Almarashy, M.-R. Feizi-Derakhshi, and P. Salehpour, "Enhancing fake news detection by multi-feature classification," *IEEE Access*, vol. 11, pp. 139601–139613, 2023, doi: 10.1109/ACCESS.2023.3339621.

[33] P. Kar, Z. Xue, S. P. Ardakani, and C. F. Kwong, "Are fake images bothering you on social network? Let us detect them using recurrent neural network," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 2, pp. 783–794, Apr. 2023, doi: 10.1109/TCSS.2022.3159709.

[34] Y. Djenouri, A. Belhadi, G. Srivastava, and J. C. Lin, "Advanced pattern-mining system for fake news analysis," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 6, pp. 1–10, Dec. 2023, doi: 10.1109/TCSS.2022.3233408.

[35] S. Govindankutty and S. Padinjappurathu Gopalan, "Modeling rumor spread and influencer impact on social networks," *IEEE Access*, vol. 11, pp. 121617–121628, 2023, doi: 10.1109/ACCESS.2023.3327863.

[36] G. Joshi, A. Srivastava, B. Yagnik, M. Hasan, Z. Saiyed, L. A. Gabralla, A. Abraham, R. Walambe, and K. Kotecha, "Explainable misinformation detection across multiple social media platforms," *IEEE Access*, vol. 11, pp. 23634–23646, 2023, doi: 10.1109/ACCESS.2023.3251892.

[37] J. Valinejad and L. Mili, "Cyber-physical-social model of community resilience by considering critical infrastructure interdependencies," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 17530–17543, Oct. 2023, doi: 10.1109/JIOT.2023.3277450.

[38] L. Wu, Y. Rao, C. Zhang, Y. Zhao, and A. Nazir, "Category-controlled encoder–decoder for fake news detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1242–1257, Feb. 2023, doi: 10.1109/TKDE.2021.3103833.

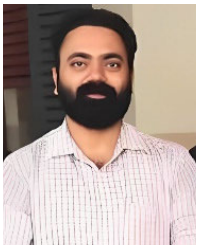
[39] M. Tajrian, A. Rahman, M. A. Kabir, and M. R. Islam, "A review of methodologies for fake news analysis," *IEEE Access*, vol. 11, pp. 73879–73893, 2023, doi: 10.1109/ACCESS.2023.3294989.

[40] S. A. Khan, G. Sheikhi, A. L. Opdahl, F. Rabbi, S. Stoppel, C. Trattner, and D.-T. Dang-Nguyen, "Visual user-generated content verification in journalism: An overview," *IEEE Access*, vol. 11, pp. 6748–6769, 2023, doi: 10.1109/ACCESS.2023.3236993.

[41] J. Fu, Y. Song, and Y. Feng, "Rumor spreading model considering the roles of online social networks and information overload," *IEEE Access*, vol. 11, pp. 123947–123960, 2023, doi: 10.1109/ACCESS.2023.3328396.

[42] J. Wang, S. Qian, J. Hu, and R. Hong, "Positive unlabeled fake news detection via multi-modal masked transformer network," *IEEE Trans. Multimedia*, vol. 26, no. 1, pp. 1–11, Jun. 2023, doi: 10.1109/TMM.2023.3263552.
[43] L. Hu, Z. Chen, Z. Z. J. Yin, and L. Nie, "Causal inference for leveraging image-text matching bias in multi-modal fake news detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 1–12, Nov. 2022, doi: 10.1109/TKDE.2022.3231338.
[44] D. Wang, W. Zhang, W. Wu, and X. Guo, "Soft-label for multi-domain fake news detection," *IEEE Access*, vol. 11, pp. 98596–98606, 2023, doi: 10.1109/ACCESS.2023.3313602.
[45] K. Zaheer, M. R. Talib, M. K. Hanif, and M. U. Sarwar, "A multi-kernel optimized convolutional neural network with Urdu word embedding to detect fake news," *IEEE Access*, vol. 11, pp. 142371–142382, 2023, doi: 10.1109/ACCESS.2023.3341870.
[46] D. Jung, E. Kim, and Y.-S. Cho, "Topological and sequential neural network model for detecting fake news," *IEEE Access*, vol. 11, pp. 143925–143935, 2023, doi: 10.1109/ACCESS.2023.3343843.
[47] L. Wu, P. Liu, Y. Zhao, P. Wang, and Y. Zhang, "Human cognition-based consistency inference networks for multi-modal fake news detection," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 1, pp. 211–225, Jan. 2024, doi: 10.1109/TKDE.2023.3280555.
[48] R. P. Ojha, P. K. Srivastava, S. Awasthi, V. Srivastava, P. S. Pandey, R. S. Dwivedi, R. Singh, and A. Galletta, "Controlling of fake information dissemination in online social networks: An epidemiological approach," *IEEE Access*, vol. 11, pp. 32229–32240, 2023, doi: 10.1109/ACCESS.2023.3262737.

PREMNARAYAN ARYA, photograph and biography not available at the time of publication.



AMIT KUMAR PANDEY is currently an Associate Professor with the Department of CSE-DS, ABES Engineering College Affiliated to AKTU, Ghaziabad, Uttar Pradesh, India. He previously worked with many renowned engineering institute gaining more than 12 years of rich experience in research and academics. He has attended many conferences and seminars as a keynote speaker. He is also an active member of different technical bodies working in the computer science field. He has published patents, books, and research papers in various national and international conferences and journals. His research interests include artificial intelligence, machine learning, and information security.

S. GOPAL KRISHNA PATRO received the B.Tech. degree from RIT, Berhampur, India, the M.Tech. degree in computer science from VSSUT, Burla, India, and the Ph.D. degree in recommendation systems from GIET University, Gunupur, India. He is currently an Assistant Professor with the School of Technology, Woxsen University, Hyderabad, Telangana, India. He has more than eight years of teaching experience along with two years of administrative and two years of industrial experience. He has published more than ten journal articles indexed in SCOPUS, SCI, and SCIE, attended and published more than five international conferences, five book chapters, one patent, and one copyright publication. He has participated as a reviewer in more than ten peer-reviewed journals and book chapters. He has been awarded many prizes for his excellent way of presentations and attended more than five professional expert talks and invited talk programs as a resource person. He received an Appreciation Certificate from the AD Scientific Index, in June 2021, 2022, and 2023 in World Scientist and University Ranking. He has worked as an organizer or co-ordinator in more than 15 conferences, international conferences, FDPs, and Hackathon programs. He has also participated in more than 30 workshops and seminars.

KRETIKA TIWARI received the Ph.D. degree (Hons.). Her Ph.D. topic of research is "Drug Repositioning Recommendation System Based on ADR Detection Through Social Media." She is currently an Assistant Professor with the School of Engineering and Technology, Jagran Lakecity University, Bhopal, Madhya Pradesh, India. She has 16 years of teaching experience. She has published nine research papers and written a chapter in a book entitled "Cloud Computing and its Applications." Her research interests include machine learning, NLP, and blockchain technology. She is a Life Time Member of ICSES, IAENG, and IFERP.

NIRANJAN PANIGRAHI received the M.Tech. and Ph.D. degrees in computer science and engineering from the National Institute of Technology (NIT), Rourkela, India, in 2009 and 2017, respectively. He has a total of 18 years of teaching and research experience. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Parala Maharaja Engineering College (PMEC), Odisha, Berhampur, India, an autonomous college of government. He is in-charge of the High-Performance Computing Laboratory, Centre-of-Excellence, PMEC. He is a member of the SWAYAM-NPTEL Mapping Committee, Biju Patnaik University of Technology, Rourkela. He has research publications in reputed journals of IET, Elsevier, and Springer, and has also contributed to many book chapters and conferences. His research interests include cloud and edge computing, wireless sensor networks, applied machine learning, parallel algorithms, and soft computing. He is a Life-Time Member of ISTE, IAENG, and UACEE.

QUADRI NOORULHASAN NAVEED received the Ph.D. degree in information technology from the Kulliyyah of Information and Communication Technology (KICT), International Islamic University Malaysia (IIUM), Kuala Lumpur. He was an IT Engineer with Aramco, Saudi Arabia, and Riyad Bank, Saudi Arabia. He is currently teaching with the College of Computer Science, King Khalid University, Saudi Arabia. He has many publications in refereed/indexed international journals and the IEEE, ACM, and Scopus-Springer-sponsored conferences. His current research interests include E-learning, M-learning, cloud computing, cloud-based E-learning systems, and technology-enhanced learning. He is a reviewer of several conferences and journals.

AYODELE LASISI is currently an Assistant Professor with the Department of Computer Science, College of Computer Science, King Khalid University, Abha, Saudi Arabia. He has published a good number of articles in national and international journals. His research interests include fuzzy logic, artificial intelligence, optimization, and machine learning.

WAHAJ AHMAD KHAN is currently an Assistant Professor with the School of Civil Engineering and Architecture, Institute of Technology, Dire-Dawa University, Dire Dawa, Ethiopia. He has published a good number of research articles in international and national journals. His research interests include turbulent flow, structure analysis, curved beams, buildings, and concrete structures.