Social Media Image Classification and Effect of Local Features on Classification Performance

Vedpriya Dongre[1*] and Pragya Shukla[2]

[1,2] Institute of Engineering & Technology, Devi Ahilya Vishwavidyalaya, Indore, India
*[1]vdongre@ietdavv.edu.in
[2]pshukla.ietdavv.edu.in

Abstract. User-generated contents drive social media platforms. Users share heterogeneous data types, i.e., text, image, and video. Among them, a significant amount of data has been posted in image format. Therefore, analyzing and classifying social media-based images have many essential applications. In this paper, we are addressing the issue of social media-based image classification. Social media images have different characteristics than the normal images used in other applications. To simulate this issue, in this paper two scenarios of the experiment have been demonstrated. First, the experiment utilizes three variants of deep learning models to traditionally classify image data. Additionally, the second experiment includes image classification based on textual and visual features. This model is developed using a deep learning concept namely multi-model feature fusion. The implemented models are experimented with a publicly available social media MEME image dataset, which is obtained from Kaggle. Based on the results, we found that only image-visual feature-based classification provides 43% accuracy. On the other hand, when we utilize both textual and visual features of images using the multi-model fusion. Then the classification accuracy has been improved up to 2%. Thus, shortly for social media image classification the multi-model feature fusion techniques are recommended to use.

Keywords: Social Media Data, Social Media Images, Feature Selection, Local Features, Deep Features, Image Classification.

1 Introduction

Social media is one of the popular online platforms. It is also a huge source of fresh information. The data is generated and consumed by the social media platform users [1]. Therefore, social media information can be used to develop social welfare and lifesaving applications [2]. In recent years, social media data has proven its use in various applications like disaster management [3], disease outbreaks [4], and hate speech detection [5]. These applications utilize social media data streams to process using machine learning and relevant techniques. Additionally, the recovered insights of the ML algorithms are utilized in applications [6]. The social media data may be in text, images, and video format. Additionally, the data processing techniques are highly dependent on the formats of data [7]. However, a significant amount of study has been conducted on social media data analysis, but most of them are focused on text-based data processing. Moreover, limited work has been available, which are focused on social media image data analysis.

In this paper, social media image data analysis and classification techniques are explored. First, an understanding is provided of "how the traditional image classification task is different from social media image analysis task". For this purpose, two experiments are involved: (1) First experiment is motivated to apply traditional ML techniques on social media images to perform sentiment-based image classification. (2) Secondly, the image visual content and image text features are combined and then used with the ML algorithm to perform classification. In addition, the performance of both experiments has been compared. In the first experiment, the aim is to demonstrate "how the traditional image classification techniques are working on social media image classification problems". On the other hand, the second experiment includes a deep learning approach namely the multi-model fusion technique. This technique self-extracts and combines text features and visual features obtained from

S. Bhalerao et al. (eds.), *Proceedings of the International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2025),* Advances in Intelligent Systems Research 192,
https://doi.org/10.2991/978-94-6463-716-8_81

social media images. The aim is to validate the potential of a deep multi-model feature fusion approach in social media image classification. By using both experiments, we are trying to demonstrate that social media image classification is different from normal image classification problems therefore this type of analysis needs some special treatment.

To discuss the efforts of this paper, in this section an overview of the proposed work is given. The next section includes a discussion of recently developed noteworthy applications, which are consuming social media images. By using these applications, we are trying to identify the application area, type of dataset used, ML approach, and results. Further, the proposed experimental models have been discussed. Further, the performance analysis of both experiments has been discussed using different performance matrices. Finally, the conclusion of the work has been discussed and the plan is described.

2 Literature Review

This section highlights significant contributions of machine learning (ML) to social media image classification across various applications. C.N.S. Silva et al. [8] introduced an image classification model for fish size estimation using a dataset from MyCatch and FishSizeProject, achieving 50% accuracy and revealing opportunities for improvement in ecological studies. E.J. Hoffmann et al. [9] developed a filtering pipeline for geo-tagged images, achieving an F1 score of 0.51 on the Flickr dataset, with improvements from human-validated labels. Similarly, Z. Zou et al. [10] combined image and text data for disaster image classification, boosting accuracy from 84.4% to 87.6% using multimodal techniques on the CrisisMMD dataset. In marketing and engagement, A.M. Goh et al. [11] used CNNs with transfer learning to classify food images, aiding restaurants in promotion and distribution. Y. Li et al. [12] and M. Philp et al. [13] analyzed social media engagement, showing how image quality, embedded text, and specific content features drive user interaction, particularly in industries like restaurants and transportation. D.C. Gkikas et al. [14] explored branded Facebook images, linking text readability and hashtags to engagement. G. Dhanesh et al. [15] emphasized narrative components such as "image narrative" and "point of view" as critical for interaction on Facebook and Instagram. In disaster management, M. Jing et al. [18, 20] developed frameworks for flood scene recognition and image retrieval using ontologies and Bag-of-Words models. Z. Niu et al. [19] addressed weakly annotated tags with semi-supervised relational topic models, improving classification on multimedia datasets. K. Ahmad et al. [22] combined social media and satellite data for real-time flood response, assessing road passability with CNNs and transfer learning.
These studies demonstrate the versatility of ML in enhancing image analysis for real-world challenges, from disaster response to user engagement, underscoring the growing importance of integrating visual and textual data for impactful insights.

Table 1. Recent Social Media Image Classification Task Using ML

Ref.	Application	Dataset	Algorithms	Results
C. NS-Silva et al. [8]	ML based fish species and size identification	MyCatch and FishSizeProject	20 separate model training	Accuracy of ~50%
E. J. Hoffmann et al. [9]	filtering pipeline to high-quality,ground -level imagery	Flickr with more than 28 million images	Fine-tuned architectures	F1 scores of up to 0.51

Z. Zou et al .[10]	identify useful disaster images	CrisisMMD	integrates image and text data by multimodal fusion	84.4% to 87.6%
A. M. Goh et al.[11]	food-image classifications	food-image	CNN	-
Y. Li et al.[12]	examine the influence of image content on engagement	Twitter and Instagram	-	Show: how to improve social media engagement.
M. Philp et al.[13]	Visual characteristics of product and social media engagement.	restaurants' Instagram posts	Google Vision AI	-
D. C Gkikas et al.[14]	Whether text features in image posts associate with engagement and brand awareness.	135 image posts with description from a Fashion retail Facebook	i) readability indices, ii) text length, and iii) number of hashtags.	Long and many hashtags tends to achieve higher engagement
G. Dhanesh et al.[15]	Which characteristics of visual content are associated with higher levels of public engagement?	visuals posted on Instagram and Facebook	visual social semiotic framework	Narrative and point of view enhance engagement, composite feature rally engagement.
M. F. Hasmawati et al. [16]	Extract context and content features from Instagram, and influence on user engagement.	university Instagram posts	LASSO regression	Context features and content features, and image receive more likes and comments.
M. Kilian et al.[17]	analyze how Twitter content influences engagement	N=216,828 Tweets collected from HR Twitter account	Employ computational linguistics	Job-focused content is less engaging than content focusing on relations.
M. Jing et al.[18]	flood scene recognition system	flood related images from US Federal Emergency Management Agency Facebook pages	"Squiral" (square spiral) Image Processing	-
Z. Niu et al.[19]	recognizing images with weakly annotated tags	social multimedia datasets	semi-supervised relational topic model (ssRTM)	
M. Jing et al.[20]	Social media data is incorporated with image analysis to enhance image retrieval	Flood images from US Federal Emergency Management	image features are shows by a histogram of visual words using the	Improve performance of image recognition after incorporating the text features.

		Agency Facebook pages	Bag-of-Words (BoW)	
A. Capatina et al.[21]	How users of AI software perceive capabilities to differentiate solution		integer valued regression	Software classified in category:image,Audience, and sentiment analysis
K. Ahmad et al. [22]	floods classification and floods aftermath detection	MediaEval 2018	pre-trained multiple deep models on ImageNet and Places	Significant improvement in the performance.

Summary and Research Gap: The review of literature highlights the extensive use of social media images with machine learning (ML) across diverse domains such as disaster management, marketing, and transportation. Two primary approaches are identified: one focusing on visual features and the other integrating visual and textual features through multimodal feature fusion. While the latter significantly enhances classification accuracy by 2-5%, it introduces challenges in effectively representing and combining these features, impacting model performance. Despite advancements, the optimal integration of multimodal data remains a critical research gap. Addressing this, the proposed research will explore deep learning-based approaches to improve the accuracy and efficiency of multimodal feature fusion, paving the way for enhanced applications of social media data in real-world scenarios.

3 Dataset & Methodologies

The problem of social media image classification is addressed because social media carries fresh information and sometimes the information is available in real-time [23]. However, in social media text posts are popular but image and video posts are taking more attention [24]. In addition, images with text posts on social media are a common way to get higher user engagement [25]. However, in traditional image classification techniques, the images only are used for classification, but social media images contain additional information in the form of the text inside images. Due to this, the single feature image feature or text feature is not able to classify the images accurately. Therefore, social media-based image classification needs both types of feature fusion (visual and textual) [26]. To simulate this problem, the proposed work involves two sets of experiments.

1. **Traditional social media image classification technique:** The aim is to demonstrate "how the visual features of an image can be used for performing the social media image classification."
2. **Multi-model fusion of Visual and textual features for social media image classification:** The deep learning technique is used to extract textual and visual features from image and then combine to prepare a common feature representation. The common features are then used to classify the social media images.

The results of both experiments are compared to demonstrate: "How only image classification techniques are not effective for classifying social media images?". This comparison is also used to simulate "the influence of combining textual and visual features" for accurate social media image classification. The details of both experiments are discussed as:

A.Dataset Used

A publicly accessible dataset of political meme images from social media was used in the studies. The Kaggle repository is the source of that dataset [27]. The dataset is divided into two sections: the

Instagram meme images are included in the first section, and the extracted language that goes with the social media photographs is included in the second section. The dataset contains 6992 Instagram photos in total. The published dataset description states that text extraction from social media photos has been accomplished using the Optical Character Recognition (OCR) technique [28]. A corrected version of the OCR extracted text is also provided in this dataset because the extracted text may have a considerable amount of noise. The image text and images with the relevant sentiment class labels can be used for performing the classification task. There are five total classes available. These labels are: 'very positive', 'positive', 'neutral', 'negative', and 'very negative'.

1. B. Only Visual Features of Images

This section evaluates the impact of visual features on social media image classification using a convolutional neural network (CNN). The implemented three-channel CNN processes normalized color images to reduce computational costs by scaling pixel values between 0 and 1. Developed using Python, the model consists of 10 layers. The input layer is a 2D convolutional layer with 360 neurons and ReLU activation, accepting 60x60 three-channel images, followed by a max-pooling layer. A second 2D convolutional layer with 180 neurons and ReLU activation is followed by another max-pooling layer and a dropout layer, which reduces the feature shape by 50%. A flattened layer converts features into a linear structure for further learning through three dense layers (128, 64, and 32 neurons), all using ReLU activation. The output layer includes 5 neurons with a softmax function for classification. This configuration highlights the CNN's capacity to extract and learn visual features for image classification effectively.

Table 2. Three channel CNN model

Layers	Configuration
Model type	Sequential ()
Layer 1	Conv2D, Neurons 360, activation Relu, input shape = (60, 60, 3)
Layer 2	MaxPooling2D
Layer 3	Conv2D, Neurons 180, activation = Relu
Layer 4	MaxPooling2D
Layer 5	Dropout, dropout (%) 50%
Layer 6	Flatten
Layer 7	Dense, neurons 128, activation = Relu
Layer 8	Dense, neurons 64, activation = Relu
Layer 9	Dense, neurons 32, activation = Relu
Layer 10	Dense, neurons 5, activation = Softmax

Then, another CNN model was put into practice. The visual features that were extracted individually using conventional methods are classified using that model. The Sobel operator is used to extract features from images. To extract edge fetableatures, the Sobel operator is employed. In conventional picture classification methods, edge features are commonly employed. Greyscale images are used in the Sobel edge detection technique. Therefore, we must convert the three-channel CNN model that was previously in use to a single-channel CNN model. Table 3 specifies the single-channel CNN model's necessary setup. The input layer's size and channel count are the only variations between the two models.

Table 3. Single channel CNN model

Layers	Configuration
Model type	Sequential ()
Layer 1	Conv2D, Neurons 360, activation Relu, input shape = (60, 60, 1)
Layer 2	MaxPooling2D
Layer 3	Conv2D, Neurons 180, activation = Relu
Layer 4	MaxPooling2D
Layer 5	Dropout, dropout (%) 50%
Layer 6	Flatten
Layer 7	Dense, neurons 128, activation = Relu
Layer 8	Dense, neurons 64, activation = Relu
Layer 9	Dense, neurons 32, activation = Relu
Layer 10	Dense, neurons 5, activation = Softmax

The learning process is then accelerated by configuring an additional CNN model. As a result, the configuration of the layer has undergone certain modifications. Additionally, some layers from the single-channel and three-channel CNN models that were previously in use are eliminated. Table 4 illustrates how the improved CNN model is configured. Because there are fewer layers and neurons in this model than in the two earlier models, it operates more quickly. There are only eight layers in this single-channel CNN.

In order to illustrate the conventional image classification methods, three CNN configurations have been developed. The following is the reason behind the development of these models:

1. The first model uses color channels to illustrate image feature learning.

2. The edge feature learning is being demonstrated by the second model.

3. The CNN model can be advanced using the third model without sacrificing accuracy.

These three deep learning combinations have been the subject of studies, and the outcomes of those tests have been documented. The findings section discusses the obtained experimental results.

Table 4

. Single Channel CNN

Layers	Configuration
Model type	Sequential()
Layer 1	Conv2D, Neurons 32, activation Relu, input shape = (60, 60, 1)
Layer 2	MaxPooling2D
Layer 3	Conv2D, Neurons 64, activation = Relu
Layer 4	MaxPooling2D

Layer 5	Flatten
Layer 6	Dense, neurons 32, activation = Relu
Layer 7	Dense, neurons 32, activation = Sigmoid
Layer 8	Dense, neurons 5, activation = Softmax

4 Proposed Work

For precise social media picture classification, this section has considered the textual and visual characteristics of the photographs. To produce a common feature in this situation, both kinds of features must be extracted and combined. Consequently, a deep learning model has been set up to carry out learning tasks, feature merging, and feature extraction. The multi-model feature fusion technique is the name given to this kind of learning. Along with extracting text features from social media photographs, this model requires us to extract deep learning features. Additionally, the categorization is carried out by combining the extracted deep learning characteristics. Figure 1 illustrates the classification method utilizing both visual and textual social media image elements.

The provided diagram indicates that the model has two input sources. Information about the image is accepted from the first input source, and the text related to the image is accepted from the second. Both dataset components are employed in this model. However, a few local feature selection methods were used before to the dataset's utilization. Initially, text pre-processing processes were applied with the text data. The goal of pre-processing is to make the text less noisy. Consequently, special characters and stop words are eliminated first. After then, the extraction of text features was completed.

The term frequency and inverse document frequency (TF-IDF) have been employed for this purpose. This feature selection method has taken into account a maximum of 2000 features (2000 keywords).

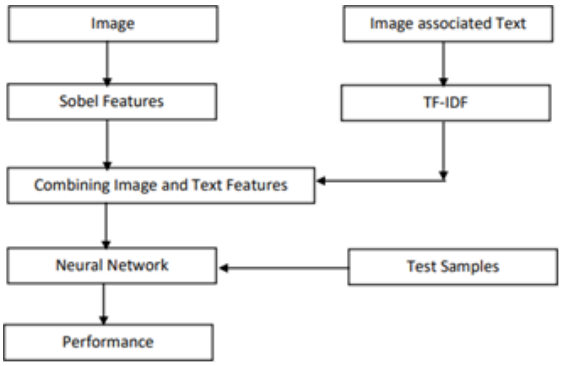


Fig 1. Proposed text and image classification model

The image data was then taken into account. The dataset's image section includes color pictures that include text. As a result, the image's edge features aid in representing its content. In this case, the image's edge feature has been retrieved. Images are first grayscaled, and then edge characteristics based on the Sobel operator are computed. A common vector is produced by combining the two retrieved features—textual and visual. Furthermore, a fully connected neural network layer receives

this shared vector in order to conduct training. The configuration shown in Table 5 describes the realized architecture of combining textual and visual information and training the neural network. There are just five layers in this model. The Sobel operator is used to extract the picture edge characteristics, which are then accepted by the first layers of the implemented architecture. The input size of this layer is therefore 60*60. Furthermore, the features recovered using the TF-IDF approach are accepted by the second layer.Both input layers are combined in the third layer of the model. The concatenation process between the first and second levels is carried out by this layer. A dense layer that is fully connected is then traversed by the concatenated features. Learning has been done using this layer. The categorization results are also provided by the final layer. The network's output yields sentiment class labels ranging from 0 to 4. Lastly, all implemented models' performance has been evaluated and contrasted. The models' accuracy and training time performance are shown in the following section.

Table 5. Neural Network Configuration for textual and visual feature classification

Layers	Configuration
Layer 1	Layer Input, shape=(60, 60, 1)
Layer 2	Layer Input, shape=(2000, 1)
Layer 3	Concatenate
Layer 4	Layer Dense, neurons 128, activation = Relu
Layer 5	Layer Dense, neurons 5, activation = Softmax

5 Result Analysis

The findings from both investigations are covered in this section. The experiment's specifics and findings are examined as follows:

A. Classification of visual features

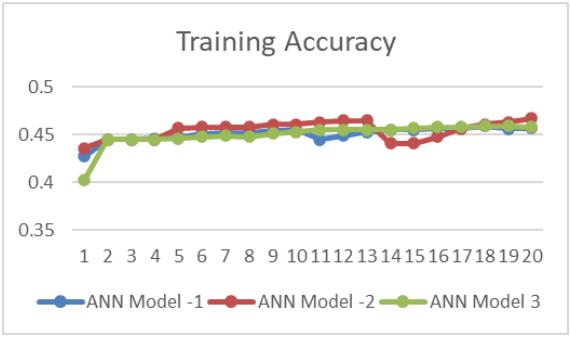
Showing "the operation of traditional image classification techniques" is the aim of this experiment. Thus far, three deep learning models have been implemented. These models are:

1. M1: CNN has three channels. This model uses the color channels to show its training and classification capabilities.

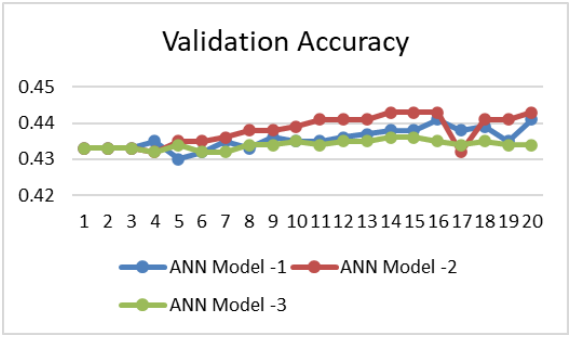
2. M2: CNN has just one channel. This model is used to show how edge features may be trained and classified.

3. M3: CNN has just one channel. The goal is to show how choosing the right neural network settings can increase the algorithm's execution time. Therefore, a model with fewer layers and neurons is set up to minimize training speed and computing cost.

The dataset of meme images is used to train the implemented CNN models [16]. Furthermore, the models have undergone validation. The accuracy of training and validation has been assessed based on experiments that have been undertaken.



(A)



(B)

Fig. 2. shows the performance of the implemented CNN models in terms of (A) Training Accuracy (B) Validation accuracy

The measurement of the categorization models' correctness is called accuracy. Five epochs have been used for the models' training and validation. Additionally, Figure 2 is used to provide the accuracy. The training accuracy of the visual feature categorization methods is shown in Figure 2(A). Furthermore, the classification models' validation accuracy is displayed in Figure 2(B). The Y-axis in both figures displays the classification accuracy as a percentage (%), while the X-axis indicates the number of epochs needed to train and evaluate the CNN models. All three models provide comparable accuracy based on the achieved training accuracy; however, model 2 ultimately offers the highest training accuracy. However, although model M3 yields consistent results, models M1 and M2's validation accuracy varies with epoch. Model M2 is doing better in terms of validation accuracy.

Model M2 is offering in both situations (training and validation). more precise outcomes Additionally, the classification report of the CNN models that have been installed is measured for a more thorough performance analysis. The classifiers' precision, recall, and f-score are included in the classification report. The primary purpose of this report is to evaluate classifier performance. Table 2 presents the models' categorization report in terms of f-score, precision, and recall.

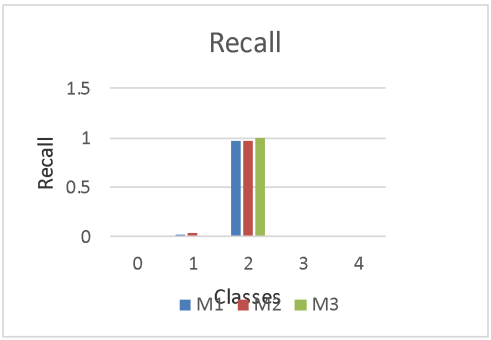
Additionally, bar graphs are used to visually show the data. The accuracy of the implemented CNN models by class is displayed in Figure 3(A). Accuracy and precision are equally important.

Only two of the five classes can be recognized by the models, based on the achieved precision. As a result, the models' categorization outcomes are only trustworthy. Furthermore, the models' recall is also calculated and displayed in Figure 3(B). The Y axis in both figures displays the classification models' precision and recall, while the X axis displays the classes to be recognized. The results show that the models can only identify class 2.

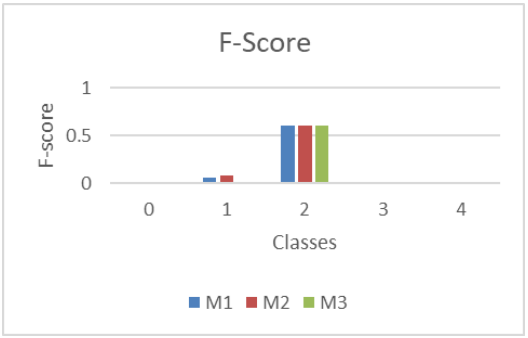
The class-wise f-score of the classifiers is then determined. The f-score is frequently used to calculate the categorization performance. When performance analysis in terms of accuracy and recall is not reliable, the f-score is used to identify which classifier is doing better. The mean precision and recall numbers are essentially provided by the f-score. It can also be used to assess how well a classification is done. Figure 3(C) displays the categorization models' measured f-score. After taking into account each parameter, we can finally say that the models are only able to recognize classes 1 and 2, not the others. The classification accuracy of the multi-model feature fusion technique has been compared with that of the most accurate model, M2. Figure 3(D) shows the relative accuracy of the two methods.

The Y axis in this graphic displays the accuracy as a percentage (%), while the X-axis displays the models that were used for categorization. We may conclude from the results that the feature-level integration of linguistic and visual characteristics of social media photographs can increase the accuracy of classification. We increased the classification accuracy by 2% by employing the straightforward combination. Therefore, in order to achieve a satisfactory classification accuracy, we suggested taking into consideration the multi-model feature fusion technique as soon as possible for the implementation of apps based on social media photos. To achieve extremely precise results, the CNN parameters had to be adjusted throughout deployment.

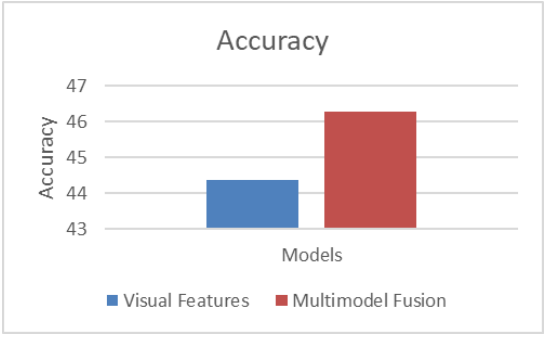
(A)



(B)



(C)



(D)

Fig. 3. performance of the CNN models in terms of (A) Precision and (B) Recall (C) F-score (D) Comparing Accuracy

B. Visual and Textual Features Classification

Comparing the effectiveness of the multi-model feature fusion method, which is based on both textual and visual features, with the classification of simply visual features is the goal of this phase. The classification accuracy of the multi-model feature fusion technique has been compared with that of the most accurate model, M2. Figure 3(D) shows the relative accuracy of the two methods. The Y axis in this graphic displays the accuracy as a percentage (%), while the X-axis displays the models that were used for categorization.

The findings suggest that feature-level fusion of textual and visual characteristics of social media photographs can increase categorization accuracy. We increased the classification accuracy by 2% by employing the straightforward combination. Therefore, in order to achieve a satisfactory classification accuracy, we suggested taking into consideration the multi-model feature fusion technique as soon as possible for the implementation of apps based on social media photos. To achieve extremely precise results, the CNN parameters had to be adjusted throughout deployment.

Table 6. Class wise Performance of the CNN models for Meme Image classification

Classes	**Precision**			**Recall**			**F-Score**		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.40	0.42	0.0	0.03	0.04	0.0	0.06	0.08	0.0
2	0.43	0.44	0.43	0.97	0.97	1.0	0.60	0.60	0.60
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

6 Conclusion

Social welfare applications leveraging social media data have gained traction due to the vast amount of information available online. However, limitations arise from the diverse data formats used on social platforms. This study focuses on social media image classification techniques, particularly those involving text within images, by reviewing recent developments in this area. The review highlights key applications, machine learning models, datasets, and results. Deep learning approaches, especially convolutional neural networks (CNNs), are identified as the most effective for image classification problems. Using a dataset of meme images from Instagram, experiments compared traditional image classification methods with a multi-model feature fusion approach, which integrates visual and textual elements.

The results revealed that traditional image classification methods, involving three variations of CNNs, achieved a maximum accuracy of 43%, with limited class recognition and poor performance on text-containing images. By contrast, the CNN-based multi-model feature fusion technique improved accuracy by incorporating textual features, achieving up to 46% accuracy. This demonstrates the importance of combining text and image analysis for reliable classification. Future work aims to apply this technique in simulations to enhance the accuracy and applicability of social media image classification.

References

2. J. Hruska, P. Maresova, “Use of Social Media Platforms among Adults in the United States—Behavior on Social Media”, Societies 2020, 10, 27; doi:10.3390/soc10010027
3. G. Lal K V, U. Sait, T. Kumar, R. Bhaumik, S. Shivakumard, K. Bhallad, “Design and development of a smartphone-based application to save lives during accidents and emergencies”, Procedia Computer Science 167 (2020) 2267–2275
4. X. Huang, Z. Li, C. Wang, H. Ning, “Identifying disaster related social media for rapid response:a visual-textual fused CNN architecture”, International Journal of Digital Earth 2020, VOL. 13, NO. 9, 1017–1039, https://doi.org/10.1080/17538947.2019.1633425
5. Q. Jia, Y. Guo, G. Wang, S. J. Barnes, “Big Data Analytics in the Fight against Major Public Health Incidents (Including COVID-19): A Conceptual Framework”, Int. J. Environ. Res. Public Health 2020, 17, 6161; doi:10.3390/ijerph17176161
6. M. A. Bashar, R. Nayak, K. Luong, T. Balasubramaniam, “Progressive domain adaptation for detecting hate speech on social media with small training set and its application to COVID-1 concerned posts”, Social Network Analysis and Mining (2021) 11:69, https://doi.org/10.1007/s13278-021-00780-w
7. A. Gupta, R. Katarya, “Social media based surveillance systems for healthcare using machine learning: A systematic review”, Journal of Biomedical Informatics 108 (2020) 1035
8. S. B. Abkenar, M. H. Kashani, E. Mahdipour, S. M. Jameii, “Big data analytics meets social media: A systematic review of techniques, open issues, and future directions”, Telematics and Informatics 57 (2021) 101517
9. C. NS Silva, J. Dainys, S. Simmons, A. Audzijonyte, “A machine learning based image classification method to estimate fish sizes from images without a specified reference object”, bioRxiv preprint 2022
10. E. J. Hoffmann, K. Abdulahhad, X. X. Zhu, “Using social media images for building function classification”, Cities 133 (2023) 104107
11. Z. Zou, H. Gan, Q. Huang, T. Cai, K. Cao, “Disaster Image Classification by Fusing Multimodal Social Media Data”, ISPRS Int. J. Geo-Inf. 2021, 10, 636.
12. A. M. Goh, X. L. Yann, “Food-Image Classification Using Neural Network Model”, Int. J. of Electronics Engineering and Applications, Vol. 9, No. 3, July-Sep. 2021
13. Y. Li, Y. Xie, “Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement”, American Marketing Association, Journal of Marketing Research 2020, Vol. 57(1) 1-19
14. M. Philp, J. Jacobson, E. Pancer, “Predicting social media engagement with computer vision: An examination of food marketing on Instagram”, Journal of Business Research 149 (2022) 736–747
15. D. C Gkikas, K. Tzafilkou, P. K. Theodoridis, A. Garmpis, M. C Gkikas, “How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook”, International Journal of Information Management Data Insights 2 (2022) 100067
16. G. Dhanesh, G. Duthler, K. Li, “Social media engagement with organization-generated content: Role of visuals in enhancing public engagement with organizations on Facebook and Instagram”, Public Relations Review 48 (2022) 102174
17. M. F. Hasmawati, A. Romadhony, “Content And Context Based Analysis Of University Social Media Engagement”, Jurnal Teknik Informatika dan Sistem Informasi ISSN 2407-4322 Vol. 9, No. 4, Dec 2022, Hal. 3227-3237
18. M. Kilian, T. Andranik, C. Dan, “Don’t be so Emotional: How Social Media Communication Affects Potential Applications Engagement”, Twenty-Eighth European Conference on Information Systems (ECIS2020) – A Virtual AIS Conference.
19. M. Jing, B. W. Scotney, S. A. Coleman, S. Kelly, X. Zhang, K. Ahmad, M. T. McGinnity, A. Schlaf, S. Gr¨under-Fahrer, G. Heyer, “Flood Event Image Recognition via Social Media Image and Text Analysis”, COGNITIVE 2016 : The Eighth International Conference on Advanced Cognitive Technologies and Applications, 2016.
20. Z. Niu, G. Hua, X. Gao, Q. Tian, “Semi-supervised Relational Topic Model for Weakly Annotated Image Recognition in Social Media”, CVPR 2014

21. M. Jing, S. Bryan, S. Coleman, T. M. McGinnity, X. Zhang, S. Kelly, K. Ahmad, A. Schlaf, S. Gr¨under-Fahrer, G. Heyer, “Integration of text and image analysis for flood event image recognition”, In Unknown Host Publication IEEE. Download date: 29/04/2023(Accepted/In press)
22. A. Capatina, M. Kachour, J. Lichy, A. Micu, A. E. Micu, F. Codignola, “Matching the future capabilities of an artificial intelligence-based software for social media marketing with potential users’ expectations”, Technological Forecasting & Social Change 151 (2020) 119794
23. K. Ahmad, K. Pogorelov, M. Riegler, O. Ostroukhova, P. Halvorsen, N. Conci, R. Dahyot, “Automatic detection of passable roads after floods in remote sensed and social media data”, arXiv:1901.03298v1 [cs.IR] 10 Jan 2019
24. J. Fromm, K. Eyilmez, M. Baßfeld, T. A. Majchrzak, S. Stieglitz, “Social Media Data in an Augmented Reality System for Situation Awareness Support in Emergency Control Rooms”, Information Systems Frontiers (2023) 25:303–326
25. P. Chakrabarti, E. Malvi, S. Bansal, N. Kumar, “Hashtag recommendation for enhancing the popularity of social media posts”, Social Network Analysis and Mining (2023) 13:21
26. T. Reimer, “Environmental factors to maximize social media engagement: A comprehensive framework”, Journal of Retailing and Consumer Services 75 (2023) 103458
27. I. K. S. Al-Tameemi, M. R. F. Derakhshi, S. Pashazadeh, M. Asadpour, “Interpretable Multimodal Sentiment Classification Using Deep Multi-View Attentive Network of Image and Text Data”, IEEE Access, VOLUME 11, 2023
28. https://www.kaggle.com/datasets/hammadjavaid/6992-labeled-meme-images-dataset
29. A. Fateh, M. Fateh, V. Abolghasemi, “Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection”, Engineering Reports. 2024; 6:e12832, 1 of 26, https://doi.org/10.1002/eng2.12832
30. I. Amerini, C. T. Li, R. Caldelli, “Social Network Identification Through Image Classification With CNN”, IEEE Access, Volume 7, 2019