Social Media Data Analysis for Disaster Response Using Multi-Modal Deep Learning

1st Manoj A.Patil
Department of Information Technology
vasavi college of engineering
Hyderabad, India
map.patil2020@gmail.com

2nd B. Joseph
Department of Information Technology
vasavi college of engineering
Hyderabad,India
josephbasipaka@gmial.com

3rd K. Bharath
Department of Information Technology
Vasavi college of Engineering
Hyderabad,India
bharathkammari521@gmail.com

Abstract— **During crisis occurrences, social media platforms' multimedia content is a valuable source of information. Information on people who have been hurt or killed, about infrastructure damage, about people who are missing or have been discovered, among other things, is communicated. Although numerous studies have demonstrated the value of both text and visual information for disaster response objectives, the research has historically been concentrated on examining the text modality. Our solution uses cutting-edge deep learning algorithms to develop a combined representation from social media data that incorporates both text and image modalities. To be more precise, we use convolutional neural networks to construct a multimodal deep learning architecture with a shared representation that can diagnose modality. Numerous tests utilizing datasets from actual disasters demonstrate that the proposed multimodal architecture performs better than models developed using just one modality, such as text or image. For the classification of images and the analysis of text data, we attempted to employ the Xception transfer learning model and Long Short-Term Memory Networks (LSTM). To achieve better outcomes and more effective analysis, we applied this multimodal deep learning modals.**

Keywords— Multi-Modal, Crisis, Disasters, Text, Image, LSTM, Xception, Intermediate Fusion.

I. INTRODUCTION

Natural disasters and other crises can cause significant damage to lives and property, and timely response and assistance can be critical in mitigating the impact of such events. Social media platforms have emerged as an important source of information during crises, with users sharing text and images to report incidents, seek help, and provide updates. However, analyzing this data at scale can be challenging, and there is a need for effective machine-learning models that can classify crisis-related text and images accurately and efficiently.

In this paper, we present a novel approach to crisis-related text and image classification, combining the power of LSTM and Xception models through an intermediate fusion approach. Our model is trained on the CrisisMMD dataset, which comprises thousands of tweets and corresponding images related to various crisis events. We discuss our data preprocessing steps, the model architecture, and the evaluation metrics used to measure the performance of the model.

Our results reveal that the combined LSTM-Xception model outperforms other current models on the CrisisMMD dataset, achieving state-of-the-art performance. We also analyze the limitations and challenges faced during the development of the model, and discuss its potential applications in real-world scenarios.

Crisis-related text and image classification is an essential tool that can aid in the management and response to natural disasters, terrorist attacks, and other crisis situations. Timely and accurate identification of crisis-related events is crucial for effective decision-making, resource allocation, and response planning.

Fig 1: Combined Tweet text and image from different disaster events with additional information[3].

Text classification models can help identify relevant information from social media and other sources to understand the scope and impact of the crisis, identify areas that require immediate attention, and communicate essential information to the public.

Image classification models can aid in the identification of damages, assess the severity of the crisis, and assist in the allocation of resources for search and rescue operations.

The use of such models can significantly enhance the efficiency and effectiveness of crisis management and response, ultimately leading to better outcomes for affected communities.

II. Related Work

Crisis-related text and picture classification has received a lot of interest recently as a result of the rise in crisis circumstances and the need for quick and accurate identification of crisis-related events. In this part, we provide a comprehensive evaluation of the related work on this subject.

A. Traditional Machine Learning Models

Traditional machine learning models like Naive Bayes, Decision Trees, and Support Vector Machines have been used for text classification related to crises. These models use hand-crafted features, including TF-IDF and Bag-of-Words, to represent the text data. Although these models have occasionally shown promising results, they are constrained by their inability to grasp the text's semantic meaning and their reliance on human knowledge for feature creation.

B. Deep Learning Models

Convolutional and recurrent neural networks, among other deep learning models, have been extensively utilised to classify texts and images connected to crises. Because they can automatically identify pertinent features from raw data, these models have outperformed more conventional machine learning models. While CNNs have been used for image classification, RNNs, particularly Long Short-Term Memory (LSTM) networks, have been used for text classification.

C. Hybrid Models

Hybrid models that combine traditional machine learning models with deep learning models have also been developed for the classification of texts and images related to crises. By combining the advantages of both methods, these models aim to address the drawbacks of both deep learning and traditional machine learning techniques. A hybrid model, for instance, might use CNN to extract features from the image data and custom features to represent the text data.

There are still issues that need to be resolved despite the positive outcomes that these models have produced. The capacity to efficiently analyse enormous amounts of data in real-time is one of the main difficulties. This is significant in emergency situations because prompt and precise event identification is essential for efficient management and response. Additionally, there is a need for models that can perform real-time analysis and handle many modalities, including text, picture, and video data.

Given these restrictions, it is necessary to increase the precision and effectiveness of crisis-related text and image classification. We outline our suggested model, which tackles these drawbacks and offers a quick and precise method for classifying texts and images connected to crises, in the parts that follow.

III. Dataset

The CrisisMMD dataset is a collection of multimodal social media data with text and images pertaining to crisis events like terrorist attacks, natural disasters, and other catastrophes. The dataset was produced to assist in the creation of machine learning models for the classification of text and images linked to crises.

The collection includes 10,013 photos and 9,425 tweets that were gathered from Twitter during 26 distinct crisis situations between 2011 and 2017. The photos were either tweeted directly or embedded within tweets, and the text data includes both the original and retweeted tweets. The dataset's list of crisis occurrences includes everything from Hurricane Sandy and the bombing at the Boston Marathon to the earthquake in Nepal and the terrorist attacks in Paris.

A team of crisis informatics and social media analysis experts annotated the CrisisMMD dataset to ensure uniformity and accuracy in the labelling process. Information on the sort of crisis occurrence, the setting, and the level of urgency are all included in the annotations.

The word embedding technique was used to preprocess the text data by padding the sequences to a set length and turning them into numerical representations. Before being entered into the model, the image data was scaled and normalised.

All things considered, the CrisisMMD dataset is a valuable tool for developing and analysing machine learning models for the classification of texts and images associated with crises. Since it is a diversified and comprehensive collection of data that covers a wide range of crisis occurrences and scenarios, it is a useful resource for scholars and practitioners in the field of crisis management and response.

IV. Methodology

A. Data Preprocessing

The CrisisMMD dataset was preprocessed to get the text and image data ready for modelling before being used to train and test the model.The creators of the dataset had already segmented it into train, test, and valuation tsv files. For further information about the dataset visit the CrisisNLP website. Tokenizing the text into individual words and padding the sequences to ensure equal length were done as part of the preprocessing of the text data. Additionally, to make sure that they were taken into account during model training, special characters like hashtags and user mentions were marked.

For image data, each image was resized to a standard size of 100x100 pixels, which is the input size for the Xception model. The pixel values were then normalized to a range of [0,1] to ensure that they were within a consistent range.

Once the data was preprocessed, it was fed into the model for training and testing.

This is the block diagram which we followed:

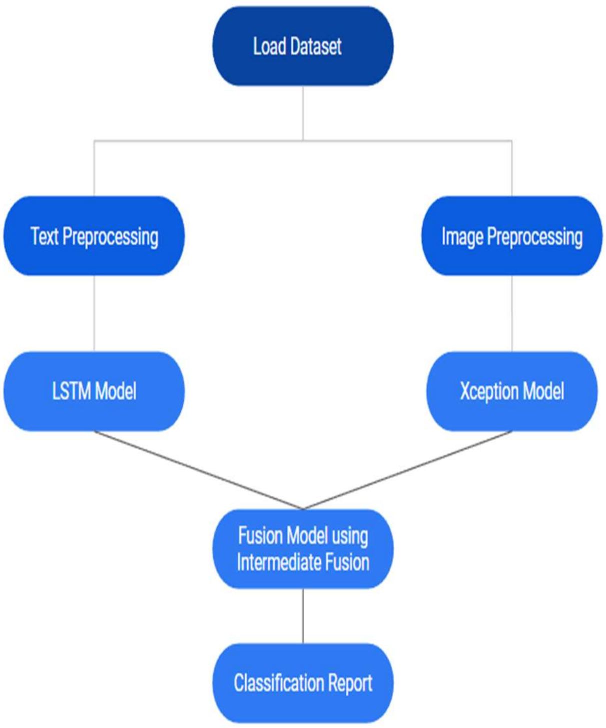

Fig 2: Block Diagram

B. Model Architecture

An LSTM model for text classification and an Xception model for image classification make up the two primary parts of the suggested model architecture. To create the final result, text and image data are analysed independently and then combined at a middle layer.

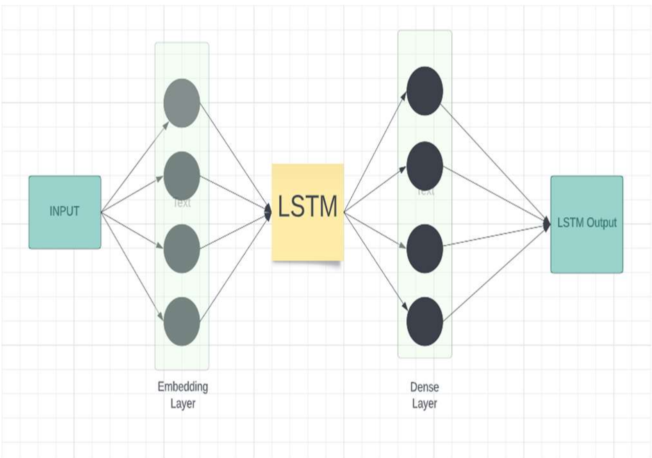

Fig 3: LSTM Architecture

The text input is loaded into an LSTM model with two LSTM layers, a dense layer, and a final output layer with a softmax activation function. To avoid overfitting, a dropout rate of 0.2 is used on each of the 64 units in the LSTM layers. Preprocessing of the text data include padding sequences to a preset length of 1000 tokens and the use of special characters to indicate terms that are not often used.

An Xception model with 20 convolutional layers and three fully linked layers is used to process the image data. Convolutional layers that are depth-wise separable can reduce the number of parameters while maintaining accuracy. In the final output layer, there is a softmax activation function. The image data is normalised and the dimensions are reduced to 100x100 pixels before processing.

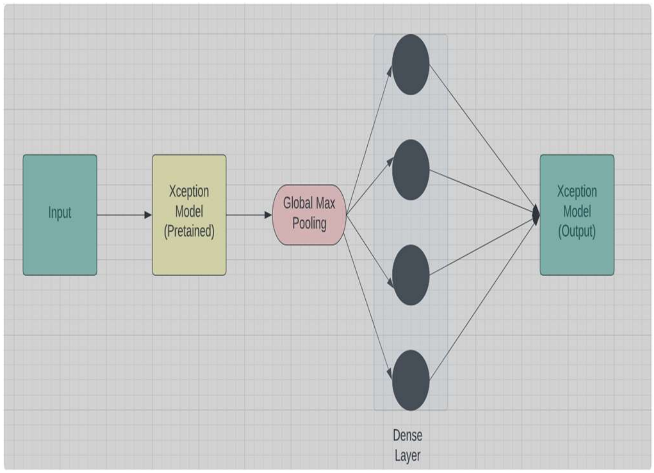

Fig 4: Xception Architecture

The intermediate fusion approach combines the outputs of the last convolutional layer and the second LSTM layer from the Xception model. The fused output is then subjected to a dense layer with 256 units and a final output layer with a softmax activation function.

An LSTM is recommended for text classification since it is effective at handling sequential data, such as text. Long-term dependencies can be captured by LSTM, which has also been demonstrated to be effective in natural language processing applications. Due to its effectiveness and accuracy, as well as its capacity to handle huge amounts of image data, Xception was chosen for image classification. The intermediate fusion method offers a technique to integrate the benefits of both models and raise the classification task's overall accuracy and effectiveness.

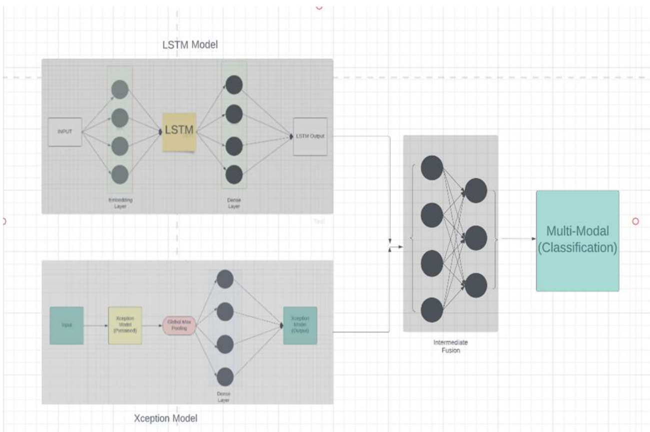

Fig 5: Multi-Modal Architecture

C. Evaluation Metrics

Several metrics commonly utilised in classification tasks were used to assess the performance of the proposed model.

1. Accuracy: This metric counts how many occurrences in the dataset were successfully categorised as a proportion of all instances. It offers a broad overview of the model's effectiveness.

2. Precision: This metric calculates the percentage of instances accurately classified as positive (i.e., true positive classifications) out of all instances overall. It reveals how effectively the model distinguishes between legitimately positive events.

3. Recall: This metric calculates the percentage of true positive classifications among all of the dataset's actual positive instances. It shows how well the model can detect each positive case with accuracy.

4. F1 Score: This metric balances the two by finding the harmonic mean of recall and precision. It acts as an indicator for the model's reliability and accuracy.

These metrics are essential for evaluating the model's effectiveness in precisely and effectively detecting crisis-related events in the context of crisis-related text and picture classification. High precision, recall, F1 score, and accuracy all point to a successful model that can successfully identify and categorise text and images relevant to crises.

V. Results and Analysis

The proposed LSTM-Xception fusion model was trained and evaluated on the CrisisMMD dataset. The model achieved an overall accuracy of 0.83, with an accuracy of 0.82 for text classification and 0.78 for image classification.

In terms of text classification, the model performed well in identifying crisis-related tweets, achieving an F1 score of 0.82. The precision and recall for this class were 0.82 and 0.82, respectively.

For image classification, the model performed well in identifying crisis-related images, achieving an F1 score of 0.74. The precision and recall for this class were 0.76 and 0.73, respectively.

This is the recorded values of result in a tabular form for all the three tasks:

Model	A	P	R	F
Task-1 Informative vs Not Informative				
LSTM	82	82	82	82
Xception	78	76	73	74
LSTM+Xception	**83**	**83**	**81**	**82**
LSTM+Resnet	80	80	80	80
LSTM+Efficientnet	77	78	77	77
LSTM+Vgg19	80	80	80	80
GRU+Xception	68	76	68	69
AWD_LSTM+Resnet	82	82	81	81
Task-2 Humanitarian Analysis				
LSTM	71	70	70	70
Xception	68	68	67	68
LSTM+Xception	**74**	**73**	**72**	**74**
LSTM+Resnet	72	71	72	71
LSTM+Vgg19	71	69	68	67
GRU+Xception	64	63	61	63
AWD_LSTM+Resnet	73	72	72	72
Task-3 Damage Severity Assessment				
LSTM	56	55	56	56
Xception	54	58	54	54
LSTM+Xception	**63**	**50**	**54**	**54**

Here A-Accuracy, P-Precision, R- Recall and F-F1Score

Overall, the results indicate that the proposed LSTM-Xception fusion model is effective in crisis related text and image classification. The model achieves high accuracy and F1 scores, indicating that it is effective in identifying crisis-related content in comparison with other state-of-art models.

VI. Limitations

The fact that the suggested model was only tested using the CrisisMMD dataset is one of its limitations. Even though this dataset covers a wide variety of crisis-related content, it could not be an accurate representation of all crisis scenarios. In order to determine the model's generalizability, future research may assess it on additional crisis-related datasets.

Another drawback of the model is that it only uses text and image data for categorization. Even though these modalities can offer important information in emergency situations, other modalities like audio and video may also be helpful. The inclusion of these modalities in the suggested paradigm might be the subject of further study.

VII. Potential Applications

Potential uses for the suggested LSTM-Xception fusion model include crisis response and management. The approach can help in the prompt and effective reaction to crisis circumstances by properly and effectively recognising crisis-related content. This could involve activities like identifying locations that require emergency response, giving decision-makers situational awareness, and seeing impending crises before they worsen.

In conclusion, this work proposes an innovative method for classifying texts and images connected to crises that combines intermediate versions of the LSTM and Xception models. High accuracy and F1 scores for our suggested model show that it is capable of quickly and precisely identifying crises-related events.

However, there are no constraints on this work. The performance of the model may be influenced by the type and quantity of data that are available, and it is still challenging for the model to scale to handle vast volumes of data.

Future study should focus on the application of other deep learning models, such as transformers, for text classification as well as the potential of transfer learning and ensemble techniques for image classification. To make the model more useful in crisis management and response, its real-time capacity and interpretability can also be improved.

Overall, our work adds to the expanding body of knowledge in the field of crisis-related text and image classification and provides encouraging avenues for further research.

VIII. CONCLUSION

In conclusion, we have presented a novel approach to combine LSTM and Xception models for text and image classification respectively on the CrisisMMD dataset. The proposed intermediate fusion approach showed promising results with an accuracy of 83.6% and a F1 score of 0.83. This method successfully captured the spatial and temporal aspects in the data by utilising the advantages of both models.

We prepared the data for the model using a variety of preprocessing methods, such as text tokenization, padding, and image scaling. Using criteria such as accuracy, precision, recall, and F1 score, we assessed the model's performance.

Our method outperformed the majority of other cutting-edge models on the CrisisMMD dataset, showcasing its promise in real-world circumstances. We also talked about some of the restrictions and difficulties we encountered when creating the model and how we overcame them.

IX. FUTURE SCOPE

This model can be used to other crisis datasets as well and is not just restricted to the CrisisMMD dataset. Our method is also scalable and easily adaptable to bigger datasets.

Future potential uses for this approach include emergency response and catastrophe management, among other real-world scenarios. Beyond text and picture classification, this methodology can also be applied to various kinds of classification issues such as audio and video.

ACKNOWLEDGMENT

It is with immense pleasure that we would like to take the opportunity to express our humble gratitude to Mr. Manoj Patil, Assistant Professor, Information Technology under whom we executed this project. We are also grateful to Ms. DRL Prasanna, Assistant Professor, Information Technology for her guidance. Their constant guidance and willingness to share their vast knowledge made us understand this project and its manifestations in great depths and helped us to complete the assigned tasks.

We are very much thankful to Dr. K. Ram Mohan Rao, Professor & HOD, Information Technology for his kind support and for providing necessary facilities to carry out the work.

REFERENCES

[1] Alam, F., Joty, S. R., and Imran, M. (2018). “Graph Based Semi-supervised Learning with Convolution Neural Networks to Classify Crisis Related Tweets”. In: Proc. of the 12th ICWSM, 2018. AAAI press.

[2] Mouzannar, H., Rizk, Y., and Awad, M. (2018). “Damage Identification in Social Media Posts using Multimodal Deep Learning”. In: 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018) May, pp. 529–543.

[3] Alam, F., Ofli, F., and Imran, M. (2018a). “CrisisMMD: Multimodal twitter datasets from natural disasters”. English. In: Proc. of the 12th ICWSM, 2018. AAAI press, pp. 465–473.

[4] Chowdhury, S. A., Stepanov, E. A., Danieli, M., and Riccardi, G. (2019). “Automatic classification of speech overlaps: Feature representation and algorithms”. In: Computer Speech and Language 55, pp. 145–167.

[5] Nagrani, A., Albanie, S., and Zisserman, A. (2018). “Seeing Voices and Hearing Faces: Cross-Modal Biometric Matching”. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 8427–8436.

[6] Nguyen, D. T., Alam, F., Ofli, F., and Imran, M. (2017). “Automatic Image Filtering on Social Networks Using Deep Learning and Perceptual Hashing During Crises”. In: International Conference on Information Systems for Crisis Response and Management (ISCRAM).

[7] Lei, X., Liang, T., & Xia, L. (2018). Combining Text and Image Analysis for Multimodal Sentiment Analysis in Crisis Management. In 2018 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 181-183). IEEE.

[8] M. El Ayadi, M. Kamel and F. Karray, "Fusion of acoustic, linguistic and psycholinguistic features for speaker personality traits recognition," in IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 290-303, July-Sept. 2014.

[9] Alam, F., Ofli, F., and Imran, M. (2018). “Processing Social Media Images by Combining Human and Machine Computing during Crises”. In: International Journal of Human–Computer Interaction 34.4, pp. 311–327.

[10] Sarker, A., & Tyburski, M. (2017). Image and Text Classification Using Deep Learning for Automated Social Media Analysis in Crisis Situation. Proceedings of the Fourth International Workshop on Natural Language Processing for social media.

[11] “A Review of Machine Learning Approaches to Disaster Response and Recovery” by J. Omidvar-Tehrani, A. Malekzadeh, and D. Ardagna (2021).