Context-Aware Transfer Learning Approach to Detect Informative Social Media Content for Disaster Management

Saima Saleem*, Monica Mehrotra
Department of Computer Science, Jamia Millia Islamia, New Delhi, India

***Abstract*—In the wake of disasters, timely access to accurate information about on-the-ground situation is crucial for effective disaster response. In this regard, social media (SM) like Twitter have emerged as an invaluable source of real-time user-generated data during such events. However, accurately detecting informative content from large amounts of unstructured user-generated data under such time-sensitive circumstances remains a challenging task. Existing methods predominantly rely on non-contextual language models, which fail to accurately capture the intricate context and linguistic nuances within the disaster-related tweets. While some recent studies have explored context-aware methods, they are based on computationally demanding transformer architectures. To strike a balance between effectiveness and computational efficiency, this study introduces a new context-aware transfer learning approach based on DistilBERT for the accurate detection of disaster related informative content on SM. Our novel approach integrates DistilBERT with a Feed Forward Neural Network (FFNN) and involves multistage finetuning of the model on balanced benchmark real-world disaster datasets. The integration of DistilBERT with an FFNN provides a simple and computationally efficient architecture, while the multistage finetuning facilitates a deeper adaptation of the model to the disaster domain, resulting in improved performance. Our proposed model delivers significant improvements compared to the state-of-the-art (SOTA) methods. This suggests that our model not only addresses the computational challenges but also enhances the contextual understanding, making it a promising advancement for accurate and efficient disaster-related informative content detection on SM platforms.**

Keywords*—Disaster management; twitter; distilBERT; deep learning; multistage finetuning; transfer learning

I. Introduction

According to the “Human cost of disasters 2000-2019” report by the “United Nations Office for Disaster Risk Reduction” (UNDRR) [1], disasters have significantly surged over recent decades. As a result, approximately 1.23 million lives have been lost globally, along with an additional economic loss of 2.97 trillion dollars. To effectively respond and manage such significant ecological disruptions, disaster response organizations increasingly rely on swift and precise human-centric information [1]. In recent years, social media (SM) particularly Twitter, have emerged as invaluable tools for disseminating and obtaining real-time user-generated data during such events [2].

Numerous research works [3-5] have demonstrated that tweets shared on Twitter during disasters often contain informative content, including details about affected people, infrastructure damage, resource needs and availability etc. Such content can be useful for disaster responders in coordinating response efforts, if processed effectively. However, identifying informative content from a substantial volume of irrelevant and noisy tweets during time-critical disaster situations is a challenging task [6]. Moreover, the character limit, uncommon abbreviations, and grammatical mistakes make the detection of informative content even more challenging [7].

Prior research works have explored classical machine learning-based techniques for the automated analysis of SM texts in the context of disasters [8, 9]. However, in recent years, there has been a notable shift towards the application of deep learning (DL) in the analysis of SM text for disaster response [7]. DL has evinced great performance across diverse domains [10-12]. Different DL techniques, including CNN [13, 14], LSTM [15], and Bi-LSTM [16] have been explored for disaster-related tweet classification.

While significant work has been done to analyze SM data related to disasters, the majority of the existing studies rely on traditional non-contextual models within natural language processing (NLP), such as GloVe, Word2Vec (W2V), Bag-of-words (BoW), etc. These models process text sequences in a unidirectional manner, which limits their ability to accurately capture the nuanced context of a tweet sequence. This limitation can lead to inaccurate classification, especially in disaster situations where words like "fire," "flood," and "earthquake" may be used metaphorically. This underscores the need for advanced models capable of understanding the context of disaster-related tweets for the accurate detection of informative content.

In the most recent advancements within NLP, researchers have introduced a spectrum of transfer learning methods and models, notably BERT [17], RoBERTa [18], and DistilBERT [19], among others. These models have demonstrated substantial enhancements across various NLP tasks [20]. They are based on the approach of bidirectional training of transformer-based neural networks to learn the contextual numeric representation of text sequences. This enables them to capture the complete context of a given text sequence, resulting

[1]UNDRR: Human cost of disasters: An overview of the last 20 years 2000-2019. https://www.undrr.org/publication/human-cost-disasters-2000-2019.

in superior performance compared to conventional non-contextual models [21]. However, they often require significant computational resources due to their large size and complexity.

In the context of disaster response, where speed and efficiency are paramount, the DistilBERT transformer emerges as a promising choice among others. It offers significant advantages in terms of computational efficiency due to its smaller size and faster processing. Despite its demonstrated performance and computational efficiency in various real-time applications like fraud detection [22] and network traffic classification [23], the application of DistilBERT to time-critical disaster management applications remain largely unexplored.

This study aims to fill this research gap by harnessing the capabilities of DistilBERT to offer an effective and computationally efficient solution for the specific task of detecting disaster-related informative content on SM. Our novel approach integrates DistilBERT with a Feed Forward Neural Network (FFNN) and employs multistage finetuning of the model on balanced benchmark datasets of real-world disaster data. The integration of DistilBERT with FFNN provides a simple and computationally efficient architecture, while the multistage finetuning enables a deeper adaptation of the model to the disaster domain, leading to demonstrably improved performance. The contributions of this study are outlined below:

- A new context-aware transfer learning approach is proposed that integrates DistilBERT with a FFNN and involves multistage finetuning of the model to adapt it to the disaster domain for enhanced detection of disaster-related informative content.
- Different model variants are designed using DistilBERT and several DL models like CNN, LSTM, and Bi-LSTM to evaluate the proposed model.
- This study addresses the issue of data imbalance, by employing random down sampling technique to balance the distribution of classes. This ensures that the model is not biased towards the majority class, thus providing unbiased results.
- Through comprehensive ablation studies, this study systematically investigates the impact of multistage finetuning, context-aware representation of tweets, and data balancing on the proposed model's performance.
- The proposed model is compared with various state-of-the-art (SOTA) baseline methods including non-contextual and context-aware transformer-based methods.
- The remaining sections of this paper are arranged as follows: The related works are covered in Section II. Section III covers the methodology. Section IV includes the experimental set up. Section V reports the experimental results. Section VI provides a discussion of the results. Lastly, Section VII concludes this study with the future work.

II. Related Work

Over the years, extensive research has been carried out and numerous methods have been proposed to extract disaster-related information useful for humanitarian organizations from SM. We group the existing studies into two categories: non-contextual methods and context-aware methods.

A. Non-contextual Methods

Preliminary studies in this field have mostly employed traditional NLP techniques like BoW along with classical ML algorithms to classify disaster-related SM content. In study [9], the authors presented a system called Tweedr to identify tweets mentioning damage or human fatalities information using a Logistic Regression classifier with BoW features. The authors in [24] extracted situational awareness information from both Hindi and English tweets using lexical and syntactic features with an SVM classifier. In another work [25], the authors employed an SVM classifier along with BoW features to detect tweets related to floods during Manila flooding.

Researchers have also focused on the detection of various information categories present in disaster tweets. In study [26], the authors trained a Naïve Bayes classifier on unigram, bigram, POS, and binary features using the Joplin 2011 tornado disaster dataset to classify tweets into “Caution and advice”, “Fatality”, “Injury”, “Offers of Help”, “Missing” and “General Population Information”. In another work, a system called AIDR [8] has been developed for identifying informative tweets during disasters. This system extracts features based on BoW model from the tweets and then classifies them into user defined categories such as ‘donations,’ ‘damage’ etc.

Additionally, deep neural networks with different word embedding models have been utilized in disaster domain utilizing SM datasets. The authors in study [13, 14] presented CNN-based systems for identifying disaster related SM posts during disasters. They utilized general and domain specific word embedding models for generating features. [15] designed a two-layer LSTM model combined with pretrained GloVe word embeddings for classifying tweet texts of seven different disaster events into informative and not-informative binary categories. In another work [16], the authors employed a Bi-LSTM with pre-trained GloVe embeddings for classifying disaster-related SM textual content.

B. Context-Aware Methods

With advancements in NLP, researchers have begun to employ pretrained transformer-based architectures in disaster domain to enhance the detection of useful disaster-related information from SM. The authors in study [27] finetuned the BERT model for adapting it to the task of identifying informative tweet texts during disasters. In another work [28], the authors provided a RoBERTa based method, an extension of BERT, for identifying disaster related informative tweet texts. They fine-tuned the architecture to adapt the model to the disaster-specific task.

From the aforementioned studies, it is evident that prevailing methods have majorly relied on traditional non-contextual language models for processing disaster-related SM texts. While a limited number of recent studies have explored context-aware transformer-based approaches, they typically employ resource-intensive models. Besides imbalanced datasets have been used, which may result in biased outcomes.

III. Methodology

Detecting disaster-related informative content on SM effectively and efficiently is crucial for effective disaster response. The problem is framed as a binary classification task: given a tweet, detect whether it is an informative or not_informative tweet. An informative tweet offers valuable details such as information about affected individuals, the extent of infrastructure damage, and resource requirements. Conversely, a non-informative tweet lacks crucial information related to humanitarian organizations or victims. We propose a new context-aware transfer learning approach leveraging the computational efficiency of the DistilBERT language model. Our methodology integrates DistilBERT with a FFNN and employs a multistage finetuning process for improved detection performance. We first provide an overview of the DistilBERT language model and subsequently present our proposed model (DistilBERT+FFNN), followed by multistage finetuning process in detail.

A. DistilBERT

DistilBERT is a distilled variant of BERT, an advanced bidirectional pretrained language model based on transformer encoded architecture. It is pretrained on a sizable dataset made up of the Wikipedia and Toronto Book Corpus. The reason for choosing DistilBERT is that it is lightweight than the BERT model in terms of parameters and runs 60% faster than BERT. The description of DistilBERT parameters is shown in Table I.

TABLE I. DistilBERT Parameters

Parameter Name	Value
Number of Layers	6
Hidden States Size	768
Attention Heads	12
Number of Parameters	66 million

DistilBERT comprises six stacked encoders, enabling it to encode the semantic and syntactic information in the tweet sequences as, shown in Fig. 1. The DistilBERT implements a multi-headed self-attention mechanism that captures information from each attention head. This enables the model to look at all the surrounding words in the input tweet sequence, allowing for a better understanding of a word in a particular context.

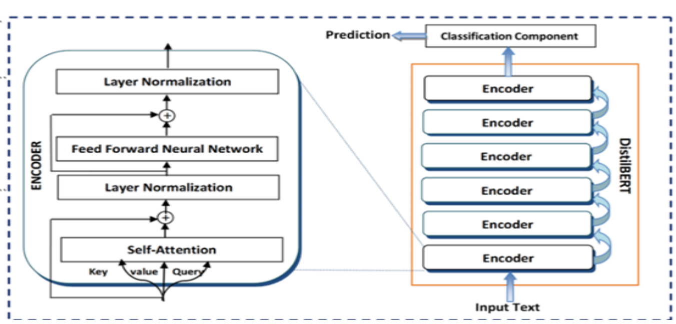

Fig. 1. DistilBERT architecture.

This stands in contrast to other text processing models like W2V and GloVe, which generates context-independent embeddings by processing the text sequence in a unidirectional manner. Thus, it can represent each word with a single vector regardless of contextual variations.

B. Proposed Model

The proposed model is an amalgamation of DistilBERT and an FFNN, leveraging their complementary capabilities to enhance the performance of disaster-related informative content detection task. The DistilBERT is utilized to extract contextual numeric representation of disaster tweet sequences, and the FFNN refines these representations into high-level abstract features for the final prediction. Importantly, FFNN introduces a layer of simplicity to the architecture, ensuring that the model retains its computational efficiency, while maintaining high performance. Fig. 2 illustrates the complete model architecture, showcasing the flow from DistilBERT to FFNN.

To obtain context-aware vector representation of the preprocessed tweets from DistilBERT model, tweets must be transformed into a format understandable by DistilBERT. For this, the tweets are passed to the batch_encode_plus method of DistilBertTokenizer class from the transformers package. It performs the following operations to transform the textual data into an appropriate format:

- Tokenization: Splitting the tweet text sequence into a list of words or sub-words called tokens.
- Padding: Making the length of the tweet sequences equal in case of the unequal sequence lengths.
- Adding special tokens: Adding special tokens such as [CLS], which stand for classification, and [SEP], which stands for separation, to indicate the beginning and the end of each sentence, respectively.
- Encoding: Substituting tokens with their corresponding IDS.
- Adding attention mask: Including an attention mask, a binary array guiding the model on which tokens to focus on it and which to ignore.

For each tweet, the *batch_encode_plus* method returns two sequences (input IDs along with attention mask), which are then input to the DistilBERT. The DistilBERT model outputs hidden states of shape (batch_size, sequence_length, hidden_size), representing the word-level/token-level embedding output of DistilBERT's layers. In this study, the output of the last hidden state is considered as it typically leads to the best empirical results [29]. Moreover, instead of using the word-level/token-level representation, the sentence-level representation of the sequence is used by taking the output for the [CLS] token, denoted as $R_{[CLS]}$. The sentence-level embedding $R_{[CLS]}$ provides the overall context of the entire sequence of tweet text. $R_{[CLS]}$ is a 2D tensor of shape (batch_size, hidden_size), which is passed to the next component of our model i.e., FFNN for predicting whether a tweet is informative or not-informative.

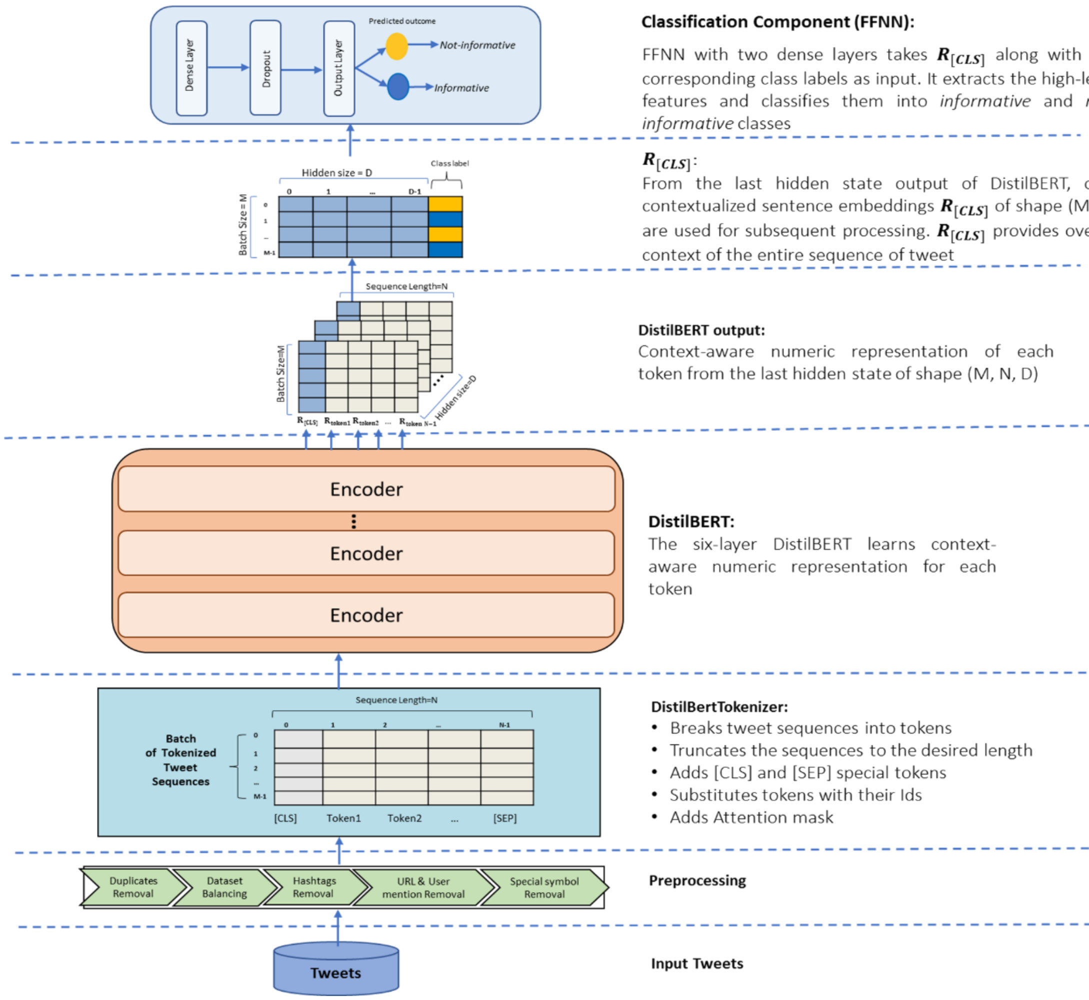

Fig. 2. Proposed model (DistilBERT+FFNN) for the detection of disaster-related informative tweets.

The FFNN is comprised of two dense layers: a hidden layer and an output layer. The hidden layer uses a popular Rectified Linear Unit (ReLU) activation function. The ReLU function is computed as shown in Eq. (1) and Eq. (2):

$$(x) = \max(0, x) \quad (1)$$

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x > 0 \end{cases} \quad (2)$$

It outputs 0 for every value of x < 0 and x itself for all values of x>0. The output of the FFNN is computed as shown in the formula in Eq. (3) and Eq. (4):

$$Z = \sum_{i=0}^{n} W_i \times X_i + B \quad (3)$$

$$f(z) = f(\sum_{i=0}^{n} W_i \times X_i + B) \quad (4)$$

where, f is an activation function (i.e., sigmoid for the output layer), X is the previous layer output, W is the weight matrix, n is the number of inputs from the incoming layer, and B is the bias.

The output of the sigmoid function is always between 0 and 1. The sigmoid activation function is computed in Eq. (5).

$$y' = \frac{1}{1+e^{-z}} \quad (5)$$

where, y' is the model's predicted value, z is the output generated by the last layer of FFNN as computed in (3). Since, the number of classes is two in our case, Binary Cross Entropy (BCE) loss function is employed which computes how far the model deviates from the correct prediction i.e., error as illustrated in Eq. (6).

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \cdot log y'_i + (1-y_i) \cdot \log(1-y'_i)) \quad (6)$$

where N is the training set size, y_i and y_i' is the actual class label and predicted value for the i^{th} sample in the dataset. BCE outputs a loss value that tells how wrong the models' predictions are. The lower the loss value, the higher the model's performance in making accurate predictions.

C. Multistage Finetuning Procedure

This study employs a multistage finetuning approach to finetune the proposed model. In this approach, the proposed model undergoes finetuning in a series of two stages. In the first stage (stage-1), the DistilBERT model is frozen and only the FFNN undergoes finetuning. This allows the FFNN to learn the task-specific knowledge without altering the general knowledge acquired by DistilBERT during pretraining. The model is trained for six epochs using Adam optimizer with a learning rate of 5e-5 and batch-size of 64. During the error back-propagation, the pretrained weights of DistilBERT will remain unchanged; only the FFNN will learn. Once, the weights of FFNN are learned in the subsequent stages (stage-2), the DistilBERT layers are unfrozen, and the entire architecture undergoes further finetuning for six additional epochs. The pretrained DistilBERT weights also get updated during finetuning, implying that the error gets back-propagated through the entire architecture. A lower learning rate is set to prevent significant updates to the gradient. A callback mechanism is used to stop the training process when the model has stopped improving. The DistilBERT attention dropout and DistilBERT dropout are also slightly increased from their default values. This two stage finetuning process allows the model to further adapt to our task and improve overall performance.

IV. Experimental Setup

This section presents the setup for experiments, including datasets used, evaluation metrics, model variants, baseline methods, and training details.

A. Dataset Description and Pre-Processing

The current study evaluates the proposed model on a real-world disaster tweet dataset called CrisisMMD [30]. This dataset encompasses tweets from seven disaster events broadly annotated into informative and not_informative classes. A few examples of informative and not_informative tweets from the CrisisMMD dataset are shown in Fig. 3. An overview of the number of tweets present in each disaster dataset is provided in Table II and the class distribution of each disaster dataset can be seen in Fig. 4.

Before conducting any experiments, the datasets are cleaned by eliminating duplicate tweets, hashtags, URLs, user mentions and various symbols such as “@,” “!”,“#”, “&,” and “%”. As can be observed from Fig. 4, the dataset is imbalanced, to address the class imbalance and ensure equal representation of *informative* and *not_informative* tweets, a balanced sample is obtained using random down-sampling. This simple and effective technique reduces the majority class to match the size of the minority class. Finally, a total count of about 8.6k tweet samples from the CrisisMMD dataset are used for subsequent processing. From each event dataset 70% tweet samples are used for training, 10% are used for validation, and the remaining 20% are used for testing the model’s performance.

TABLE II. CrisisMMD Dataset Details

Disaster event	#Tweets
Hurricane Harvey	4434
Hurricane Maria	4556
Hurricane Irma	4504
Sri Lanka Floods	1022
Iraq-Iran Earthquake	597
Mexico Earthquake	1380
California Wildfires	1588

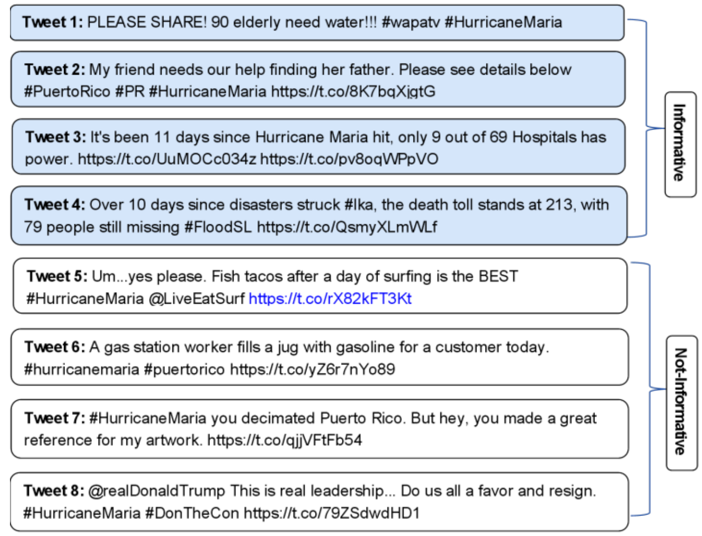

Fig. 3. Examples of informative and not_informative tweets from the CrisisMMD dataset.

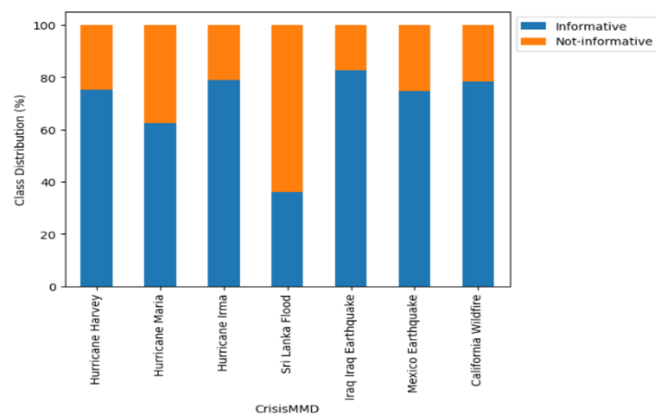

Fig. 4. CrisisMMD class distribution.

B. Evaluation Metrics

To assess our proposed model’s performance, several metrics are used:

1) Precision (P): shows the proportion of correctly predicted informative tweets to the total predicted informative tweets and is equal to:

$$P = \frac{TP}{TP+FP} \tag{7}$$

2) Recall (R): shows the proportion of correctly predicted informative tweets to the total actual informative tweets and is equal to:

$$R = \frac{TP}{TP+FN} \tag{8}$$

3) F1-Score (F1): merges the precision and recall by computing their harmonic mean, as:

$$F1 = 2 \times \frac{P \times R}{P+R} \tag{9}$$

where, TP indicates "True Positive", which means an informative tweet is correctly predicted as informative. TN indicates "True Negative", which means a not_informative tweet is correctly predicted as not_informative. FP indicates "False Positive", which means a not_informative tweet is incorrectly predicted as informative, and FN indicates "False Negative", which means an informative tweet is incorrectly predicted as not_informative.

C. Model Variants

To evaluate the effectiveness of the proposed model (DistilBERT+FFNN), we design three different model variants. These variants are formed by integrating DistilBERT with various popular DL architectures specifically CNN, LSTM, and Bi-LSTM. The description of each model variant is provided below.

1) DistilBERT+CNN: CNN is combined on top of DistilBERT as a classification component and the input to CNN is a 2-dimensional tensor X obtained from DistilBERT. This tensor undergoes a convolution operation with a filter matrix, generating a new feature map through element-wise multiplication. The resulting feature map is then subjected to a pooling layer, extracting maximum values to form a pooled output. This output is subsequently fed into the output layer, to determine the tweet's class as shown in Eq. (4).

2) DistilBERT+LSTM: An LSTM is used on top of DistilBERT and the input to the LSTM is DistilBERT output. An LSTM comprises recurrently connected memory units. Each unit comprises of a cell state, an input gate, an output gate and a forget gate. The cell state keeps information over arbitrary periods, and the information flow into and out of the cell state is governed by the gates. This allows the model to retain important information. The output from the LSTM layer is fed to an output layer for classification as shown in Eq. (4).

3) DistilBERT+Bi-LSTM: Another variant is designed by using a Bi-LSTM on top of DistilBERT. A Bi-LSTM layer trains two separate LSTM layers of opposite directions (forward and backward) simultaneously on the input generated from DistilBERT and then concatenates the outputs from both layers. This output is fed to a final output dense layer for classifying tweets into informative and not-informative classes as shown in Eq. (4).

D. Baseline Methods

The proposed model is further evaluated by comparing it against the SOTA baseline methods for disaster-related informative content detection. These methods are grouped into two categories: non-contextual methods and context-aware methods. A brief description of each baseline method is provided below:

1) Non-contextual methods:

a) W2V-CNN [13]: The authors in [13] employed a CNN model with filters of different sizes for identifying disaster related SM posts. They used general pretrained W2V embeddings with the CNN model.

b) CW2V-CNN [14]: The authors in [14] used a pretrained crisis embedding model (CW2V) [31] and trained a custom CNN with different filters to identify disaster related SM posts during a disaster.

c) GloVe-LSTM [15]: The authors in [15] used a pretrained GloVe word embeddings of 100 dimension with a 2-layer LSTM model for identifying informative textual content from SM during disasters.

d) GloVe-Bi-LSTM [16]: The authors in [16] used a pretrained GloVe word embeddings of 300 dimension along with a Bi-LSTM neural network for detecting informative textual content from SM during disasters.

2) Context-aware methods:

a) Finetuned-BERT [27]: The authors in [27] finetuned BERT for classifying SM textual posts into *informative* and *not_informative* classes during disasters.

b) Finetuned-RoBERTa [28]: The authors in [28] used a variant of BERT called RoBERTa and finetuned it for identifying informative SM textual posts during disasters.

All of these baseline methods have used the CrisisMMD dataset, with the exception of W2V-CNN [13], CW2V-CNN [14]. For these specific methods, we rely on results computed in [32] on the same CrisisMMD dataset, to facilitate a comprehensive and consistent comparison in this study.

E. Training Details

All the experiments are executed using TensorFlow framework and Google Colab cloud platform with Python programming language. The optimal hyperparameters for the stage-1 and stage-2 of our model finetuning approach are listed in Table III and IV, respectively.

TABLE III. PARAMETERS FOR STAGE-1 OF FINETUNING

Hyperparameter	Value
Learning rate	5e-5
Number of epochs	6
Batch Size	64
Dropout	0.2

TABLE IV. PARAMETERS FOR STAGE-2 OF FINETUNING

Hyperparameter	Value
Learning rate	2e-5
Number of epochs	6
Batch Size	64
DistilBERT Dropout	0.2
DistilBERT Attention Dropout	0.2

V. EXPERIMENTAL RESULTS

In this section, we present a comprehensive evaluation of the performance of the proposed model and its various variants employed in this study. Additionally, a thorough comparison is conducted between the proposed model and SOTA baseline methods. Finally, this section covers detailed ablation studies.

A. Proposed Model vs. Different Model Variants

To demonstrate that the proposed model (DistilBERT+FFNN) is an effective model for the informativeness classification task, it is compared against different model variants, as discussed in Section C. The P, R, and F1 of the proposed model and the model variants on CrisisMMD dataset are reported in Table V, with bold values indicating the best results. As per the table, all the model variants exhibit good performance across seven disasters. The proposed model exhibits the highest P ranging from 73.12 to 96.20, R in the range of 73.11 to 96.04, and F1 in the range of 72.20 to 96.04. These results render the proposed model an effective choice for detecting disaster related informative content.

B. Proposed Model vs. SOTA Baseline Methods

In this subsection, we first present a performance comparison of the proposed model against non-contextual baseline methods. Next, we analyze the comparative performance of the proposed model with respect to context-aware baseline methods.

1) Proposed model vs. non-contextual baseline methods: The comparison of results in terms of F1 of the proposed model against non-contextual methods: W2V-CNN [13], CW2V-CNN [14], GloVe-LSTM [15] and GloVe-Bi-LSTM [16] are reported in Table VI. As per the results, the proposed model consistently outperforms the non-contextual methods across all disasters, with the exception of Hurricane Irma, where GloVe-LSTM [15] outperforms in terms of F1. On average, the proposed model significantly enhances F1, with an improvement ranging from 2.75% to 19.86%.

2) Proposed model vs context-aware baseline methods: The efficacy of the proposed model is further validated through comparisons with recent transformer-based context-aware methods: Finetuned-BERT [27] and Finetuned-RoBERTa [28]. The corresponding results, expressed in terms of F1, are detailed in Table VII. A comprehensive examination of the table reveals that the proposed model outperforms Finetuned-BERT [27] across six out of seven disasters and surpasses Finetuned-RoBERTa [28] on five out of seven disasters from the CrisisMMD dataset. The proposed model exhibits an average F1 improvement of 5.47% over Finetuned-BERT [27] and 1.6% over Finetuned-RoBERTa [28].

C. Ablation Studies

In this subsection, we conduct ablation experiments to investigate the effect of multistage finetuning, context-aware representation of tweets, and data balancing on the proposed model performance.

1) Effect of multistage finetuning: To understand the effect of multistage finetuning on the model performance, we conduct a comparative analysis with a single stage (stage-1) finetuned model, where only the FFNN is finetuned while keeping DistilBERT parameters frozen. The results depicted in Fig. 5 report the F1 achieved with and without multistage finetuning on CrisisMMD dataset. From the results, consistent improvement is observed with multistage finetuning, where the model undergoes a series of two finetuning stages. Specifically, multistage finetuning boosts the performance by 3.89% to 6.85% in terms of F1. This emphasizes the efficacy of multistage finetuning in enhancing the model's ability to capture and adapt to the nuances present in disaster-related tweets, resulting in improved performance.

2) Effect of DistilBERT context-aware tweet representation: To evaluate the impact of context-aware tweet representation derived from DistilBERT on the informativeness classification task, additional experiments are performed, comparing DistilBERT against two non-contextual word embedding models. Experiments are designed where the FFNN classifier uses CW2V [33] and GloVe [34] embedding models.

TABLE V. COMPARISON OF THE PROPOSED MODEL VS. DIFFERENT VARIANTS

Disasters	Models											
	DistilBERT+CNN			*DistilBERT+LSTM*			*DistilBERT+Bi-LSTM*			*Proposed Model*		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Hurricane Harvey	81.23	81.23	81.23	81.10	81.10	81.10	83.24	83.21	83.00	**84.57**	**84.56**	**84.56**
Hurricane Maria	83.03	83.03	83.02	82.21	82.20	82.20	81.31	81.30	81.25	**84.03**	**84.01**	**83.11**
Hurricane Irma	82.00	81.22	81.19	80.05	80.05	80.04	82.04	82.04	82.01	**83.12**	**83.07**	**82.12**
Sri Lanka Floods	94.09	94.09	94.08	93.01	93.00	93.00	94.09	94.04	94.00	**96.20**	**96.04**	**96.04**
Mexico Earthquake	71.10	71.09	71.09	73.06	72.31	71.40	**73.19**	73.01	72.24	**73.19**	**73.17**	**73.17**
Iraq-Iran Earthquake	81.04	81.03	81.03	80.12	80.10	80.10	81.11	81.10	81.09	**82.09**	**82.05**	**82.02**
California Wildfires	72.09	71.30	71.30	72.03	72.01	72.01	72.06	72.06	72.03	**73.12**	**73.11**	**72.20**

TABLE VI. COMPARISON OF THE PROPOSED MODEL VS. NON-CONTEXTUAL BASELINE METHODS IN TERMS OF F1

Disasters	**Methods**				
	W2V-CNN [13]	***CW2V-CNN [14]***	***GloVE-LSTM [15]***	***GloVe -Bi-LSTM [16]***	***Proposed Model***
Hurricane Harvey	74.20	75.00	81.00	70.93	**84.56**
Hurricane Maria	69.10	70.00	82.00	69.51	**83.11**
Hurricane Irma	69.60	70.00	**83.00**	57.17	82.12
Sri Lanka Floods	90.02	91.00	92.00	72.57	**96.04**
Mexico Earthquake	64.20	64.00	71.00	51.83	**73.17**
Iraq-Iran Earthquake	59.10	60.00	81.00	63.43	**82.02**
California Wildfires	57.40	58.00	64.00	48.81	**72.20**
Average F1	69.09	69.71	79.14	62.03	**81.89**

TABLE VII. COMPARISON OF THE PROPOSED MODEL VS. CONTEXT-AWARE BASELINE METHODS IN TERMS OF F1

Disasters	**Methods**		
	Finetuned-BERT [27]	***Finetuned-RoBERTa [28]***	***Proposed Model***
Hurricane Harvey	83.00	84.50	**84.56**
Hurricane Maria	79.00	**88.00**	83.11
Hurricane Irma	73.00	82.00	**82.12**
Sri Lanka Floods	96.00	95.50	**96.04**
Mexico Earthquake	**78.00**	74.00	73.17
Iraq-Iran Earthquake	63.00	72.50	**82.02**
California Wildfires	63.00	65.50	**72.20**
Average F1	76.42	80.29	**81.89**

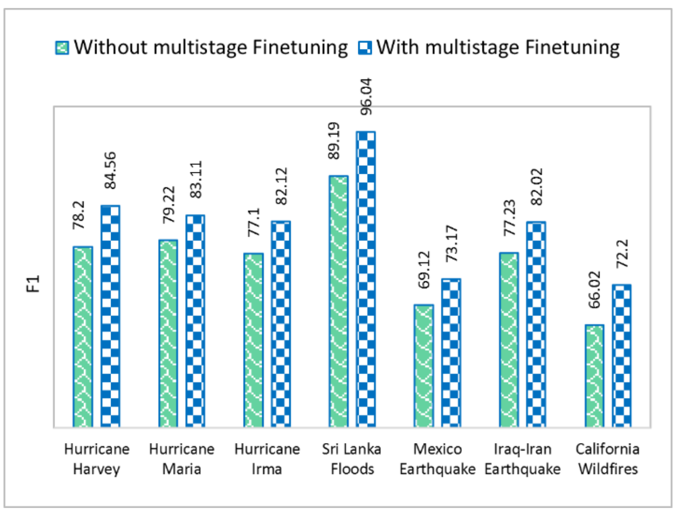

Fig. 5. F1 of the proposed model with and without multistage finetuning.

The CW2V is a 300-dimensional embedding, pretrained using W2V model on disaster related Twitter corpus. GloVe is a pretrained 100-dimensional embedding trained on Wikipedia and web text words. Fig. 6 depicts the F1 of the proposed model and non-contextual models. Our investigation reveals that the context-aware representation generated by DistilBERT model improves the results across all disasters compared to the non-contextual models. Overall, with DistilBERT, a significant improvement in the range of 8.11% to 19.56% compared to GloVe and 9.75% to 21.2% compared to CW2V in F1 is achieved across seven disasters. These results clearly highlight that DistilBERT model captures better contextual information from tweet sequences leading to better detection performance than GloVe and CW2V models.

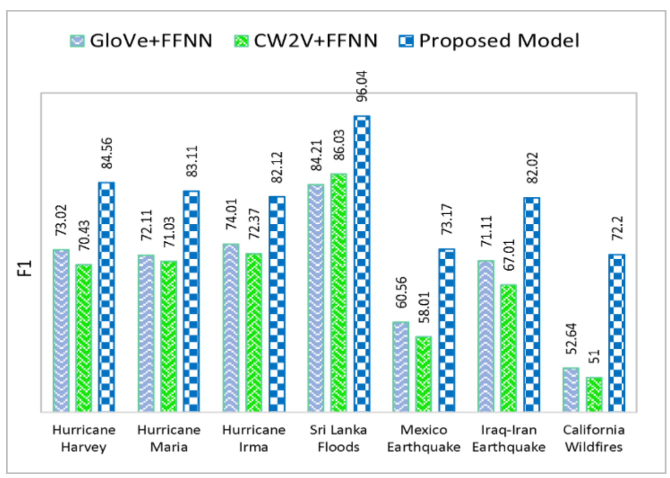

Fig. 6. F1 of the proposed model and non-contextual models.

3) Effect of data balancing: As mentioned earlier, the datasets are imbalanced, and random down sampling is employed to ensure a balanced representation. Table VIII presents a comparative analysis of the proposed model's performance in terms of F1 when trained on imbalanced and balanced dataset. From the table, it can be concluded that the model's performance is notably enhanced, showcasing improved and unbiased results when trained on balanced datasets. On the other hand, training on imbalanced datasets leads the model to disproportionately learn from the majority class, thereby introducing bias towards that class.

TABLE VIII. F1 ANALYSIS OF THE PROPOSED MODEL ON IMBALANCED VS. BALANCED DATASET

Disasters	Class 0 (*not_informative*) 1 (*informative*)	Imbalanced	Balanced
Hurricane Harvey	0	61.51	**84.69**
	1	86.24	**84.43**
Hurricane Maria	0	73.89	**82.92**
	1	89.33	**83.30**
Hurricane Irma	0	61.45	**82.03**
	1	80.24	**82.21**
Sri Lanka Floods	0	94.42	**95.59**
	1	71.09	**96.49**
Mexico Earthquake	0	48.80	**73.10**
	1	79.02	**73.24**
Iraq-Iran Earthquake	0	42.87	**82.00**
	1	79.86	**82.04**
California Wildfires	0	36.58	**72.10**
	1	87.45	**72.30**

VI. DISCUSSION

Upon a thorough evaluation of our proposed model on seven real-world disasters from a benchmark CrisisMMD dataset, it stands out as both effective and computationally efficient for the informativeness classification task. In the performance comparison between the proposed model and its variants, it is noteworthy that the integration of CNN, LSTM, and Bi-LSTM DL models with DistilBERT does not enhance the classification performance. Conversely, a simple FFNN proves sufficient for extracting high-level abstract features from the contextualized representations generated from DistilBERT. The results confirm the effectiveness of the proposed model for the informativeness classification task. The outperformance of the proposed model against non-contextual methods based on W2V, CW2V and GloVe models underscores the importance of context-aware representation of tweets in the effective detection of disaster-related informative content. The comparison of the proposed model with SOTA context-aware methods based on BERT and RoBERTa, demonstrates clear advantages of the proposed model in the detection performance. While BERT and RoBERTa are powerful transformer architectures renowned for their contextual understanding, require significant computational resources due to their large size and complexity. The proposed model built on DistilBERT known for its enhanced computational efficiency capitalizes on the efficiency gains through the multi-stage fine-tuning process. The ablation experimental results provide deeper insights, emphasizing the advantages of our multistage finetuning approach for significantly enhancing overall model performance. Additionally, the combination of DistilBERT with an FFNN, and data balancing collectively contributes to the effective and computationally efficient detection of disaster-related informative content on SM. This positions the model as well-suited for time-critical disaster response applications. Nevertheless, one of the limitations of this study is that it considers only English language tweets, while people often communicate in their native language during disasters. So, multilingual transformer-based models are worth investigating to tackle multi-linguality issues.

VII. CONCLUSION AND FUTURE WORK

In the realm of disaster management, this study introduces a novel context-aware transfer learning approach harnessing the computational efficiency of DistilBERT for the detection of disaster-related informative content on SM. Our methodology integrates DistilBERT with an FFNN, providing a simple yet effective architecture. A key feature of our approach involves multistage finetuning of the model on seven real-world disasters, resulting in improved detection performance.

This work contributes to the broader goal of improving the effectiveness and efficiency of disaster response systems using NLP and AI technologies. By developing a specialized model for disaster-related informative content detection, we aim to provide a valuable tool for disaster response organizations to better identify critical information during disasters and provide situational awareness to decision-makers. The model can be implemented in any system for filtering actionable informative content during disasters from irrelevant content, enabling disaster responders to improve their ability to deliver help and make well-informed decisions. Furthermore, the proposed model holds promise for diverse domains such as epidemics and civil unrest monitoring on SM. With domain-specific finetuning, it can be readily adapted to identify informative content during outbreaks, protests, or other volatile situations enabling real-time interventions.

In the future, we plan to extend this work for identifying and categorizing multimodal informative content into distinct humanitarian information categories, including "affected individuals", "infrastructural damage", “help and rescue”, “resource needs” etc. to enable a more targeted and efficient disaster response. Moreover, we recognize the need to explore cross-domain informative content detection, particularly in situations where labeled data for the ongoing disaster might be scarce or unavailable.

REFERENCES

[1] Wang, Z., & Ye, X. (2018). Social media analytics for natural disaster management. *International Journal of Geographical Information Science* 32(1), 49-72.

[2] Saleem, S., & Mehrotra, M. (2022). Emergent Use of Artificial Intelligence and Social Media for Disaster Management. In *Proceedings of International Conference on Data Science and Applications: ICDSA 2021, Volume 2* (pp. 195-210). Springer Singapore.

[3] Saleem, S., & Mehrotra, M. (2023, July). An Analytical Framework for Analyzing Tweets for Disaster Management: Case Study of Turkey Earthquake 2023. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

[4] Vieweg, S. (2010) Microblogged contributions to the emergency arena: Discovery, interpretation and implications. In *Proceedings of Computer Supported Collaborative Work* (pp. 515-516).

[5] Alam, F., Alam, T., Hasan, M. A., Hasnat, A., Imran, M., & Ofli, F. (2023). MEDIC: a multi-task learning dataset for disaster image classification. *Neural Computing and Applications*, 5(3), 2609-32.

[6] Caragea, C., Silvescu, A., & Tapia, A. H. (2016, May). Identifying informative messages in disaster events using convolutional neural networks. In *International conference on information systems for crisis response and management* (pp. 137-147).

[7] Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions. *Information Processing & Management*, 57(5), 102261.

[8] Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 159-162).

[9] Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. In *Proceedings of the 11th International ISCRAM Conference* (pp. 269-272).

[10] Tunio, M. H., Jianping, L., Butt, M. H. F., Memon, I., & Magsi, Y. (2022). Fruit Detection and Segmentation Using Customized Deep Learning Techniques. In *2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* (pp. 1-5). IEEE.

[11] Dhaka, D., Saleem, S., & Mehrotra, M. (2023). USE-Based Deep Learning Approach to Detect Spammers on Twitter. In *International Conference on Smart Trends in Computing and Communications* (pp. 407-416). Singapore: Springer Nature Singapore.

[12] Ajmal, S., Sarfraz, M. S., Memon, I., Bilal, M., & Alam, K. A. (2023). PUB-VEN: a personalized recommendation system for suggesting publication venues. Multimedia Tools and Applications, 1-22.

[13] Burel, G., & Alani, H. (2018). Crisis event extraction service (crees)-automatic detection and classification of crisis-related content on social media. In *Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.

[14] Alam, F., Ofli, F., & Imran, M. (2019). CrisisDPS: Crisis Data Processing Services. In *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.

[15] Kumar, A., Singh, J., Dwivedi, Y., & Rana, N. (2020). A deep multi-modal neural network for informative Twitter content classification during emergencies. *Annals of Operations Research*, 1-32.

[16] Khattar, A., Quadri, S. M. K (2022). CAMM: Cross-Attention Multimodal Classification of Disaster-Related Tweets. *IEEE Access*, 10, 92889-902.

[17] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[18] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[19] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

[20] Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). Ammus: A survey of transformer-based pretrained models in natural language processing. *arXiv preprint arXiv:2108.05542*.

[21] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*, 80(8), 11765-88.

[22] Chang, J. W., Yen, N., & Hung, J. C. (2022). Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classification as an instance. *Journal of Ambient Intelligence and Humanized Computing*, 4663-79.

[23] Shin, C. Y., Park, J. T., Baek, U. J., & Kim, M. S. (2023). A Feasible and Explainable Network Traffic Classifier Utilizing DistilBERT. *IEEE Access*.

[24] Rudra, K., Ganguly, N., Goyal, P., & Ghosh, S. (2018). Extracting and summarizing situational information from the twitter social media during disasters. *ACM Transactions on the Web (TWEB)*, 12(3), 1-35.

[25] Parilla-Ferrer, B. E., Fernandez, P. L., Ballena, J. T. (2014). Automatic classification of disaster-related tweets. In *Proceedings of the International Conference on innovative engineering technologies (ICIET)* (pp 62-69).

[26] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Extracting information nuggets from disaster-Related messages in social media. In *International conference on information systems for crisis response and management* (pp. 791-801).

[27] Madichetty, S., Muthukumarasamy, S., & Jayadev, P. (2021). Multi-modal classification of Twitter data during disasters for humanitarian response. *Journal of ambient intelligence and humanized computing*, 10223-37.

[28] Madichetty, S., & Madisetty, S. (2023). A RoBERTa based model for identifying the multi-modal informative tweets during disaster. *Multimedia Tools and Applications*, 1-19.

[29] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune bert for text classification?. In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18* (pp. 194-206). Springer International Publishing.

[30] Alam, F., Ofli, F., & Imran, M. (2018, June). Crisismmd: Multimodal twitter datasets from natural disasters. In *Proceedings of the Twelfth international AAAI conference on web and social media* (Vol. 12, No. 1).

[31] Alam, F., Joty, S., & Imran, M. (2018, June). Graph based semi-supervised learning with convolution neural networks to classify crisis related tweets. In *Proceedings of the international AAAI conference on web and social media* (Vol. 12, No. 1).

[32] Madichetty, S. (2020). Classifying informative and non-informative tweets from the twitter by adapting image features during disaster. *Multimedia Tools and Applications*, 79, 28901-28923.

[33] Nguyen, D. T., Joty, S., Imran, M., Sajjad, H., & Mitra, P. (2016). Applications of online deep learning for crisis response using social media information. *arXiv preprint arXiv:1610.01030*.

[34] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).