Disaster events identification from Social Media Data using Deep machine learning techniques

By

Iqra Javed

00000276288

Supervisor

Assoc Prof Dr. Hammad Afzal

Department of Computer Software Engineering

A thesis submitted in conformity with the requirements for the degree of

Masters of Science in Software Engineering (MS SE)

In

Military College of Signals

National University of Science and Technology (NUST)

Islamabad, Pakistan

September, 2022

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by ~~Mr~~/Ms **NS Iqra Javed,** Registration No. **00000276288**, of Military College of Signals has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature:________________________________

Name of Supervisor: Assoc Prof Dr. Hammad Afzal

Date: ____________________________________

Signature (HoD): ___________________________

Date: ____________________________________

Signature (Dean): ___________________________

Date: ____________________________________

Declaration

I, *Iqra Javed* declare that this thesis titled “Disaster events identification from Social Media Data using Deep machine learning techniques”, and the work presented in it are my own and has been created by me as a result of my own original research.

I confirm that;

1. This work was done wholly or mainly while in candidature for a Master of Science Degree MS at NUST.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exceptions of such quotations, this thesis is entirely my own work.
5. I have acknowledged all main sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

NS Iqra Javed,

NUST MCS MSSE 25

Copyright Notice

Dedication

"In the name of Allah, the most Beneficent, the most Merciful"
I dedicate this thesis to my parents, husband, and teachers who supported me in each step.

Acknowledgments

All praises to Allah for the strengths and His blessing in completing this thesis.

First and foremost, I want to thank my thesis supervisor Dr. Hammad Afzal, PhD, for his continuous support, monitoring and guidance. Without his assistance and dedicated involvement in every step throughout the duration, this work wouldn't have been completed. His incomparable help for constructive comments as well suggestions for all assessment tasks and thesis are major contributions to the success of this study. I would like to thank him very much for his understanding over this past duration. Also, I would like to thank the members of my committee; Assoc Prof. Naima Iltaf, and Lt Col. Khawir Mehmood for their support on this topic.

Most importantly, none of this could have happened without the unceasing support and attention of my father (Tahir Javaid), my mother (Nargis Bano) and my husband(Aamir Hassan). They have always stood by my side and supported my aspirations and have been a great source of inspiration for me. I would like to thank them for all their attention and support through my times of stress and hard work.

Abstract

Social media has seen a boom in recent years, with its users growing worldwide.The ubiquitous nature of social media has made it a useful tool to extract useful information regarding disasters. As in case of any disaster, people upload data related to the disaster on social media which can be very useful in the timely detection of disasters and preventing loss of human lives and infrastructures. This research aims to identify disaster-related tweets from bulk data, classify them to the type of event they belong to i.e. to perform multi-classification, and identify type of humanitarian aid-related information posted in a tweet during disasters, with improved performance in terms of accuracy, F1, etc. To perform this multi-class classification i.e. to detect disastrous events and further detect the type of humanitarian information present in the tweets a DistilBert+CNN+LSTM-based framework has been proposed in this paper. The framework is based on DistilBert pre-trained embeddings and for multi-class classification, CNN+LSTM with a self-attention layer has been used. After applying the proposed framework the F1 score of 98 % was achieved in disastrous events classification and an F1 score of 88% was achieved in information classification of tweets. These results obtained from the proposed framework have shown improvement over other deep learning models that were used as part of a comparative study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Problem Statement

Twitter, one of the most popular microblogging website[1], is being used by people to update posts in case of an emergency or a disaster. These quick uploads help the general public or Twitter users to get real-time updates regarding disasters[2] unlike other sources. For instance, newspapers take time to get printed, and blogs also take time to get published. Thus, timely detection of events on Twitter plays a critical role in helping organizations speed up their relief operations to minimize the effects of those disasters. Also, these posts can help to get an insight into the severity of any disaster that has happened. Hence, many disaster relief organizations are utilizing Twitter data to assess the situation during a disaster and immediately take necessary actions to combat the aftermaths of disasters.

Moreover, in the last few decades, Pakistan has faced natural hazards such as flooding, earthquakes, and landslides that have escalated into humanitarian disasters, with the loss of lives, homes, and livelihood. Natural disasters in Pakistan are likely to increase as a result of climate change and environmental degradation. More extreme weather events, coupled with poor preparedness in communities, can only increase the risks of humanitarian disasters. So, an early dissection of such disasters in Pakistan is required to prevent the losses on a larger scale. Thus here our proposed research can play a pivotal role in aiding non-governmental humanitarian organizations, government agencies, and public administration authorities to speed up the relief operations or to start them earlier before the losses increase exponentially.

With a rapidly changing environment and growing use of technology, this paper

[1] https://seosandwitch.com/microblogging-sites-list-top-10/

aims to contribute to effectively classifying tweets into different disaster events and humanitarian information tasks. Due to the informal nature of Twitter data, it is relatively difficult to perform classification on it. But with recent development in deep learning algorithms, a new dimension has been added to the field of NLP. In our study as well we have employed various deep learning algorithms to compare results with our proposed hybrid framework for multi-class classification.

1.0.1 Problem Statement

Recent studies have highlighted the importance of analyzing social media data during disaster events as it helps decision-makers to plan relief operations. Also, in the early hours of a disaster most of the actionable information related to a disaster is not available on traditional data sources rather are signaled on Social Media by everyday citizens COVID-19 pandemic being its most recent example. Thus, a timely identification of disasters helps humanitarian organizations, government agencies, and public administration authorities to make timely decisions and to launch relief efforts during emergency situations. So, the proposed research aims to identify such disasters related tweets to mitigate the effect of disaster in terms of human lives and infrastructure as much as possible.

1.0.2 Objectives

The main objectives of this thesis are listed below:

- To explore the effectiveness of Deep learning and Natural language processing methods for multi-classification of disaster related tweets.
- To propose a hybrid deep learning framework that can increase the performance of multi-classification of disaster and humanitarian information tasks in terms of accuracy and F1-Score
- To correct the data annotation and labels of datasets two source, in order to enhance the model learning

1.0.3 Thesis Contribution

In summary, the contributions made by this research are as follows:

- As part of this study, two datasets have been consolidated to get a new dataset for multiclass classification of disaster events and humanitarian information. Moreover, labels are mapped to get common classes in the resultant datasets. Also, annotation of disaster events has been done on the resultant consolidated dataset. Thus, a new consolidated dataset is formed as a part of this research.
- A novel hybrid framework has been proposed based on deep learning models using the self-attention-based mechanism in order to focus on contextual

information.

- This hybrid framework presents increased performance, on the task of multi-class classification of disaster events and humanitarian information, in terms of accuracy and F1 score.

1.0.4 Thesis Organization

1. Chapter 2 gives an overview of the problem which is targeted in this thesis work. It also describes a review of the literature and a brief description of work related to this topic in the past.
2. Chapter 3 materials and methods that have been employed over the course of research.
3. Chapter 4 explains the proposed hybrid framework used for multi-class classification
4. Chapter 5 discusses the process of formulation of problem. It also discusses the results.
5. Chapter 6 gives the conclusion of thesis work.
6. Chapter 97 lists the references used.

Chapter 2

Literature Survey and Related Work

In recent times the emergence of the latest Information and Communication Technologies(ICT) and access to high-speed internet connections have made social networking sites more accessible to people. Due to all these advancements, social networking sites such as Twitter and Facebook have been extensively used by people all across the world. People use these sites for various purposes like updating others in their circle about happenings in their day-to-day life or in case of emergencies or disasters to report about the incidents. Because of ease of usage, social media sites have proved to be a useful tool to detect disasters and emergencies happening around the globe. Moreover, other useful information including the requirement of relief operations, the number of injuries and casualties, requests for urgent supplies, etc can also be deduced from the updates uploaded by social media users.

In the past, mainstream media was the only medium to get the news or updates about various happenings in the world. However, with the increasing use of social networking sites, even disaster events are reported on these sites before the mainstream media. For instance, COVID-19 pandemic[1] that jolted the whole world was first discussed on social media platforms i.e WeChat in China days before its news came out in mainstream media. Also, Chinese people started posting on Weibo[2] about COVID-19 to seek help from their fellow citizens. Such events underline the importance of social media usage during disastrous events and in the hour of need by ordinary citizens.

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7378494/

[2] https://thediplomat.com/2020/02/how-the-coronavirus-outbreak-played-out-on-chinas-social-media/

With the increasing climate changes, the world is facing more natural disasters[3] like floods, hurricanes, wildfires, earthquakes, etc. than before. People use social media platforms to disseminate information about such natural disasters[3] which results in a huge bulk of data over social media that may also contain redundant information as well. Event detection from such a large volume of data is a gruesome task that is humanly impossible. As quickly detecting disasters events and relevant information related to them plays a vital role in helping government and non-government organizations to lay out their plan and start immediate relief operations where needed[4]. Thus, to detect events from this data uploaded on social media sites more advanced solutions like the application of machine and deep learning algorithms are required.

2.0.1 Twitter and its role in disaster events

Our research utilizes a disaster events Twitter dataset. With growing users of Twitter, it has become an important source to collect data during emergency occasions. Twitter messages are tweets, initially limited to 140 characters (presently increased to 280). As a matter of course, using proper search mechanisms all public posts can be found by everybody. On Twitter, people can follow each other and increase their network. Also, Twitter users can utilize the functionality of hashtags to add more information related to an ongoing trend that as a result reaches more people [5]. Moreover, these users can retweet any tweet and follow each other to see posts. Consequently, sharing data has no peripheral expense for the client: Posting a tweet requires just a cell phone, a Twitter application or sign-in through the web interface, and a relatively small transfer speed (more in the case of multimedia sharing). As per Twitter use statistics, around 500 million tweets are posted each day[4].

With the increase in Twitter usage, it has emerged as one of the primary resources to break news in real-time. In this paper[6], authors have pointed out that Twitter serves as a news source too as their research concludes that over 85% of trending topics are news headlines. Also in some cases, Twitter also became a source of information for mainstream media. During emergencies, Twitter produces a huge volume of data that is more casual, unstructured, and noisy with information overload presenting a real challenge in analyzing the Twitter data[7]. Resultantly, due to the enormous amount of data uploaded during emergencies that are humanly impossible to comprehend[8], multiple challenges are being faced by researchers

[3]https://www.usgs.gov/faqs/how-can-climate-change-affect-natural-disasters
[4]http://www.internetlivestats.com/twitter-statistics/

during tweet analysis to get useful information regarding a disaster. These challenges range from extracting required information from brief and informal messages in the tweets, handling information overload as during an emergency more posts related to it are uploaded, to prioritizing different types of information that can play a vital role in reducing the damage caused by an emergency [9].

2.0.2 Extracting event and humanitarian tasks from Twitter

An event is described as a happening that occurs at a certain time and a certain place in which one or more participants are involved[10]. Event detection has received great interest from the research community where a lot of researchers have published their findings[11]. Detection of disasters from tweets has become a hot area of research in which disaster-related messages are analyzed for better disaster management and to increase situational awareness[12]. With the world increasingly facing the effects of climate change, people are faced with more natural disasters. The past years have witnessed a record-high number of disasters causing billion-dollar losses to the world economy[13]. Thus with the increasing number of disasters and use of Twitter, a great emphasis has been given to uncovering disasters and information related to them by several researchers as evident in recent works[14][15][16].

As per their job and roles, the data needs of formal disaster management authorities and other philanthropic non-governmental organizations (NGOs) differ[17]. For example, relief organizations such as police forces require information about people trapped in disaster-hit areas that need to be rescued or injured people that need urgent medical assistance, etc. to quickly execute their relief operations. In contrast to this low-level information requirement, many Humanitarian and governmental disaster management organizations need high-level information regarding a disaster such as the scale of damage caused by the disaster event, the number of citizens affected by the disaster, urgent needs of the affected people such as food, water, and shelter, etc. Thus, these different levels of data requirements can be termed as trying to understand "the big picture" versus finding "actionable insights"[18][19].

However, the posts that are uploaded during disasters not only contain information related to missing people, infrastructure damage, and injured or dead citizens rather they also contain condolences offered to people and appreciation messages for different disaster management organizations. Because of the gigantic volume of such posts, it is quite challenging for humanitarian authorities to go through all tweets manually. Hence a quick need arises to create a framework that can group tweets into various classes of humanitarian information[20]. However, this particular research has not been given due importance to date[21].

2.0.3 Classification Algorithms

In past research studies, Machine learning algorithms were the main choice for disaster-related classification. However, with the emergence of more powerful deep learning algorithms researchers started employing them to classify disaster-related tweets. Authors have also published research based on the comparative analysis between Machine (Support Vector Machines, Logistic Regression, Random Forests) and deep learning algorithms(Recurrent Neural Networks, and Convolutional Neural Networks (CNN) [22][23]. As per their experiments, CNN proved to give high performance in classifying disaster-related tweets. Moreover, quite recently various types of embedding representations, for a variety of NLP tasks, have been proposed in the research publications such as Embeddings from Language Models (ELMo) [24], Bidirectional Encoder Representation from Transformers (BERT) [25].

Authors have mentioned in [9] their survey paper in which they mention that extracting events related information involves multiple challenges ranging from parsing brief and informal messages in the tweets, handling information overload, to prioritizing different types of information. These challenges can be mapped to information processing operations such as filtering, classifying, ranking, aggregating, extracting, and summarizing information. They also mention that Convolutional Neural Networks provide considerable performance in binary as well as multi-classification of tweets over machine learning algorithms. But these deep neural network techniques require a large amount of data to train the model.

Researchers[26] have proposed a domain specific sentiment analysis approach specifically for tweets posted during hurricanes (DSSA-H) which is the most recurring and deadly disasters resulting in loss of both infrastructure and human lives. DSSA-H can retrieve hurricane-relevant tweets with a trained supervised-learning classifier, Random Forest (RF), and classify the sentiment of hurricane-relevant tweets based on a domain-adversarial neural network (DANN). Authors also compare the proposed domain-specific sentiment-classification approach with two general sentiment analysis methods i.e. SentiStrength and VADER in analyzing hurricane-relevant tweets. The proposed DANN-based sentiment-analysis approach outperforms both general methods.

In their study authors[27] have proposed an automatic data processing services have been developed that take textual messages (e.g., tweets, Facebook posts, SMS) as input to determine the disaster type the message belongs to, its relevancy, and the type of humanitarian information it conveys. They compared three machine learning algorithms namely Naïve Bayes, Random Forest, Support Vector Machines and a deep learning algorithm i.e. Convolutional Neural Network. The proposed CNN outperforms the machine learning algorithms based on data gathered from AIDR , CrisisLex , CrisisMMD , CrisisNLP , Disaster Response Data , Disasters on Social Media , SWDM .

The study [28] show that recent advances in Deep Learning and Natural Language Processing outperform prior approaches for the task of classifying informativeness and encourage the field to adopt them for their research or even deployment. They have compared Multinomial Naïve Bayes (MNB), LR, BERT, RoBERTa to classify and cluster tweets in a way that either they represent need or supply and further tweets classification into requirement of Food, shelter, health, and WASH (Water, Sanitation and Hygiene).

Researchers have presented a hybrid machine learning pipeline in their study to automatically map the evolution of disaster events across different locations using social media posts. The proposed hybrid pipeline integrates named entity recognition, location positioning and fusion, fine-tuned BERT-based classifier, and graph-based clustering. This fine-tuned BERT-based classifier performs the best amongst other baseline deep learning models such as BiGRU, BiLSTM, CNN GRU, CNN LSTM, DPCNN, KMax CNN, and R-CNN[29].

2.0.4 Word Embeddings

Authors in their research have proposed a word-embedding, generated by famous Word2vec tool, based Ad Hoc Information Retrieval system that outperforms conventional term matching based IR model[30]. Conventional IR perform well only when queries and documents use the same vocabulary, and both are sufficiently detailed. However, in case of twitter, limited characters restriction means that many of the necessary words will not be present in a tweet. Thus, the authors show that the proposed semantic matching techniques i.e. using word embedding based method on the disaster-specific SMERP 2017 dataset, is more effective for this task than word embedding trained on the large social media collection provided for the TREC 2011 Microblog track dataset.

Usage of social media to identify disasters has become a hot research area over the years. Researchers in the study [31] have recommended an automatically-generated and human-curated list of 380 terms, that were frequently found in tweets related to disasters, for Twitter querying to extract disaster related tweets. The list was formed by studying terms frequency distribution in the CrisisLexT6 collection; looking for terms that were discriminative and frequent in disaster related tweets, and common across various disasters; and performing a crowd-sourced curation step to curate the list. This list can also help in multi-classification

of disaster event tweets rather than just classifying them into a tweet that is relevant or irrelevant.
Researchers have proposed CrisisBERT, an end-to-end transformer-based model for two crisis classification tasks, namely crisis detection and crisis recognition, which shows promising results across accuracy and f1 scores. They have also proposed Crisis2Vec, an attention-based, document-level contextual embedding architecture for crisis embedding, which achieves better performance than conventional crisis embedding methods such as Word2Vec and GloVe[12].

Chapter 3

Materials and Methods

As part of this research,we took two different datasets of tweets gathered during disasters that have happened in the world.After performing some preprocessing steps on these datasets a new consolidated dataset was formed for disaster events and humanitarian information classification.The newly formed dataset has been employed in the process of training the models be it benchmark algorithm chosen to do a comparative study or our proposed hybrid framework.During process of model training dataset has been divided in 80 to 20 ratio for training and evaluation part each.
After initial phase of data cleaning to convert our datasets into deep learning models readable format this data has to be converted into word embeddings.For this purpose,different contextual word embeddings were explored as part of this study.These word embeddings include BERT embeddings,DistilBert embeddings and Glove embeddings.These word embeddings convert the input data into a matrix form where each word is given a meaningful value.Thus,they play a vital role in classification of different labels.
As embeddings are given as input to some classification algorithms,here comes the role of deep learning algorithms.Quite recently,deep learning algorithms have jolted the tectonic plates of NLP research with their impressive performance.So keeping in view the encouraging results shown by these deep learning models as part of this research we have employed various deep learning models from different categories including Recurrent Neural Networks, Artificial Neural Networks,Transformers etc.The details of all of these are explained in the subsequent subsections.

3.0.1 Dataset Preparation

Firstly,most important step before training a model is to prepare data for the models to get higher accuracies.Because data can have a lot of discrepancies which can reduce the performance of training models.For this reason as part of our classification task we perform a number of preprocessing steps that are explained in detail in the following sections.

3.0.1.1 Dataset Collection

tweet_id	tweet_text	class_label
7.33E+17	.@GreenABEnergy How can @AirworksCanada assist in the cleanup? #AlbertaStrong	rescue_volunteering_or_donation_effort
7.3E+17	RT @katvondawn: Thoughts & prayers going to all those being affected by the wildfire in Canada. Truly, truly heartbreaking. #Pr	sympathy_and_support
7.31E+17	Glacier Farm Media pledges $50K in support for Fort McMurray wildfire disaster relief.	rescue_volunteering_or_donation_effort
7.34E+17	Beatton Airport Road wildfire in northern B.C. leaves a patchwork of damage - #VernonNews	infrastructure_and_utility_damage
7.33E+17	RT @dana_balsor: @InsuranceBureau will Insur. professionals be entering our homes without us present? #ymmfire listening to Albe	other_relevant_information
7.29E+17	RT @reporterchris: The @NHL is donating $100,000 to the Canadian Red Cross in support of the wildfire relief efforts in Fort McMurr	rescue_volunteering_or_donation_effort
7.29E+17	Foothills family, coming to the PDL game today? Were collecting cash donations for the Red Cross for #FortMcMurray	rescue_volunteering_or_donation_effort
7.35E+17	U.S. #wildfires have already burned 5 times as much acreage as last year at this time:	infrastructure_and_utility_damage
7.31E+17	Ad Hoc Committee To Coordinate Federal Wildfire Efforts #Saskatoon #YXE	rescue_volunteering_or_donation_effort
7.29E+17	RT @good_archer: #usanews New homes spring up for Fort McMurray evacuees - Temporary homes for evacuees fleeing the wildfir	displaced_people_and_evacuations
7.33E+17	Nicely done by @iancbates for @NatGeoPhotos of Fort McMurray evacuees	displaced_people_and_evacuations
7.33E+17	Sam is doing my hair, proceeds today at @hairkix #yyc goes towards wildfire relief efforts	rescue_volunteering_or_donation_effort
7.29E+17	My thoughts are for the people of Fort McMurray fleeing the wildfire. My respect to the firefighters battling the blaze.	other_relevant_information
7.29E+17	The wildfire is the latest in a lengthening lineage of early wildfires in the northern reaches of the globe that	other_relevant_information
7.29E+17	Such an awful natural disaster, thoughts are with the families #canadianfamily	sympathy_and_support
7.36E+17	Giveaway at 300 followers+Wildfire Fundraiser Continuing Overwatch #PS4 #Overwatch #Twitch	rescue_volunteering_or_donation_effort
7.3E+17	The @WaterlooMasjid is raising money for people affected by the Fort McMurray wildfire. At 810, @CBCKW891 hears why they deci	rescue_volunteering_or_donation_effort
7.31E+17	RT @TravelLeisure: Air Canada blames algorithm for fares of $4K during wildfire evacuation:	displaced_people_and_evacuations
7.29E+17	and people are asking why should i donate .	rescue_volunteering_or_donation_effort
7.33E+17	Evacuation order lifted for wildfire in northeast B.C.: An evacuation order has been rescinded for a rural ar	displaced_people_and_evacuations

FIGURE 3.1: HumAID Dataset Structure

id	event	source	text	lang	lang_conf	class_label
4.08E+17	2013_ny_train-	crisislext2	4 Dead in Train Crash Reflected NY's Diversity: 4 dead in train crash reflec	en	1	affected_individual
3.48E+17	2013_alberta_	crisislext2	Floods kill 3, 75,000 forced from Calgary homes http://t.co/c6OPFjFLkA |	en	1	affected_individual
4.07E+17	2013_glasgow_	crisislext2	@orgasmicgomez Glasgow helicopter victim named @onedirslaytion	en	0.715936	affected_individual
3.33E+17	2013_banglade	crisislext2	Death toll from factory collapse in Bangladesh passes 1,000 as recovery o	en	1	affected_individual
2.95E+17	2013_brazil_ni	crisislext2	Local media: Brazil nightclub fire kills at least 90: SAO PAULO (Reuters) - A	en	1	affected_individual
4.00E+17	2013_phillipine	crisislext2	#YolandaPH #Biliran: 10,000 Filipinos left homeless in Biliran. (Source: AN	en	0.967107	affected_individual
4.10E+17	2013_glasgow_	crisislext2	Glasgow helicopter crash funerals to be held today. FUNERALS for two of	en	1	affected_individual
3.03E+17	2013_russia_m	crisislext2	Russia cleans up after meteor blast injures more than 1,000: CHELYABINS	en	1	affected_individual
3.02E+17	2013_russia_m	crisislext2	A meteor streaked across de sky above Russia's Ural Mountains this morn	en	1	affected_individual
2.78E+17	2012_philippin	crisislext2	RT @ANCALERTS: NDRRMC: 'Pablo' death toll now at 540 http://t.co/5jD	en	0.967897	affected_individual
9.12E+17	hurricane_mari	crisismmd	Survivors of Hurricane Maria need assistance https://t.co/nOwU2JCaNy	en	NA	affected_individual
4.07E+17	2013_ny_train-	crisislext2	#tcot MTA identifies 4 people killed in NYC train crash: NEW YORK (AP) â€	en	1	affected_individual
4.09E+17	2013_glasgow_	crisislext2	RT @TieDyeTheSea: A helicopter crashed into a pub in Glasgow, killing pe	en	1	affected_individual
4.00E+17	2013_phillipine	crisislext2	RT @denverpost: Typhoon deaths climb into thousands in Philippines, one	en	1	affected_individual
3.60E+17	2013_spain_tr	crisislext2	At least 35 killed as train derails in Spain: reports - Yahoo! News http://t.c	en	1	affected_individual

FIGURE 3.2: CrisiBench Dataset Structure

The datasets that have been used in this research have been taken from two different sources. One is HumAID dataset, this dataset contains human annotated tweets collected during different disasters that happened between 2016-2019. The dataset has been annotated into 10 different classes of humanitarian information tasks. Thus, this dataset was a strong candidate to perform multi-class classification on it. The other dataset taken for this study is CrisisBench data, this dataset has a collection of tweets from eight different published datasets. This dataset is annotated into humanitarian information classes and informativeness.the structure of both the datasets are provided in Figure 3.1 and Figure 3.2 respectively.

Disaster Event labels	
S. No.	Labels
1	Earthquake
2	Cyclone
3	Hurricane
4	Wildfires
5	Floods
Humanitarian Task labels	
S. No.	Labels
1	caution_and_advice
2	sympathy_and_support
3	requests_or_urgent_needs
4	displaced_people_and_evacuations
5	injured_or_dead_people
6	missing_or_found_people
7	infrastructure_and_utility_damage
8	rescue_volunteering_or_donation_effort

Table 3.1: Classes used in multi-class classification

3.0.1.2 Dataset Annotation

The resultant dataset formed from the combination of datasets was further annotated to prepare it for multi-class classification of disaster events. Tweets in the two source datasets were not labeled according to disaster event classification. In HumAID dataset tweets were in separate files according to a specific event they were collected from i.e Hurricane Maria 2017, Maryland floods 2018, etc. Thus, in the first step tweets from all these separate files were gathered and annotated manually into disaster event labels. In addition to this, in CrisisBench data tweets were labeled according to the occurrence of events like 2013_pakistan_earthquake,2012_philipinnes_floods, etc. which were labeled into disaster events labels as well. As a result, the new dataset formed from the combination of these datasets and manual annotation contains five different disaster events as shown in Table 3.1.

For multi-class classification of humanitarian information classes relevant to disasters were taken from both datasets. In both of the source datasets, the number of humanitarian task classes was different. Also, both datasets had different names for classes thus they were labeled manually to get the same class name in the resulting dataset. Thus in order to form a uniform dataset common classes are taken from both datasets which are mentioned in Table 3.1. Moreover, name mapping was performed on class names to get uniform names in the new dataset. For example, in CrisisBench affected_people was a class name that has a similar context to the class of injured_or_dead_people from HumAID dataset so they were combined under a single class name. After performing these steps eight common humanitarian task classes are extracted to perform multi-class classification.

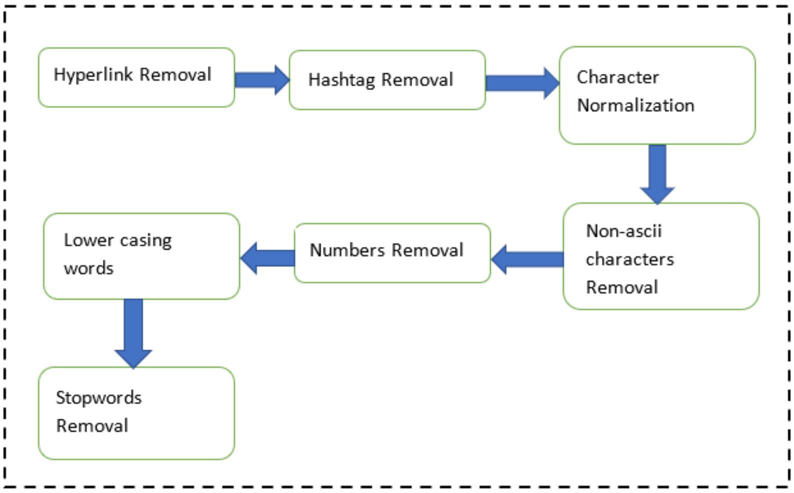

FIGURE 3.3: Dataset preprocessing steps

3.0.1.3 Dataset preprocessing

In order to perform classification on the newly formed dataset, it has to be cleaned and processed before training the framework. This involves a number of steps that are shown in Figure 3.3 and explained here:

1. **Hyperlink Removal** As a first step of preprocessing, hyperlinks are removed from tweets as they can make the learning process imprecise because they can add noise to the data and their unstructured nature carries no significant information.

2. **Hashtag Removal** Hashtag symbol '#' has also been removed from the dataset but the words succeeding to hashtag are kept in order not to miss any useful information.

3. **Character Normalization** As Twitter is an informal source of data, there are a lot of unusual writings in tweets. In this step, extra characters that are repeated more than two times are removed e.g todaaaaaay will be changed to today.

4. **Non-ascii Character Removal** Special characters give no useful information rather they add noise to data thus they are also removed.

5. **Numbers Removal** Numbers do not represent a sentiment thus they are also removed with the exception of '0' because it can show a negative sentiment.

6. **Lower casing words** In order to achieve uniformity, all the tweets in the dataset have been lower-cased.

7. **Stop words removal** In the last step of data preprocessing stopwords are removed to reduce the feature vector space [32] and noise as well.

3.0.2 Extracting Features from Dataset

For this research feature matrices of tweets have been used to classify data into desired classes of disaster events and humanitarian information.Word embeddings have been utilized to perfrom the task of feature extraction that include Bert,DistilBert and Glove embeddings.With excepption of Glove both embeddings take into account the contxet in which the word has been used.The details of these word embeddings have been given below:

- **Glove Embeddings - Global Vector for Word Representation** Glove embeddings have been developed by Standford University developers.It is an unsupervised learning model that is employed to form meaningful cetor for words given as input.In this research,as a part of our comparative study we have used Glove that are trained on huge corpus of data.During its development around 2 Billion tweets with 27 Billion tokens and 200-dimension vectors were taken.Thus,with huge data corpus Glove helps in creating matrices with extracted features from the tweets. These matrices with feature representation enhances the overall performance of deep learning models.

- **BERT Embeddings - Bidirectional Encoder Representation from Transformers** One drawback that is being faced while using Glove embeddings is that they do not produce contextual feature representations rather they give context-free matrices to be used in classification algorithms.However, these problems diminish if BERT is used to as a word embedding.Because BERT embeddings generate embedding matrix taken into consideration the context of the sentences.Unlike other embeddings that create embeddings of fixed length for each word,BERT embeddings create matrices of the words by considering other words around it.For instance,by looking at two sentences i.e 'A man went fishing to the bank' and 'Bank was robbed by a man' in both these sentences bank has been used in entirely different context.BERT embeddings will create different matrix for bank in both sentences unlike other context-free embeddings.

- **DistilBERT Embeddings - Bidirectional Encoder Representation from Transformers** DistilBert is the distilled version of BERT that improves performance with minimal computational cost as well[33].So, keeping in mind the effectiveness of DistilBERT over BERT we have used DIstilBERT embeddings for our research. Same as BERT, DistilBert also captures the context information in the text as well which gives it an extra edge as well. Because, as in our study, we have utilized disaster tweets in which consideration of context plays a vital role i.e sentences like 'He put the stage on ablaze' and 'Jungle is fully ablaze' use the word ablaze in two different contexts. Thus, to cater to such problems and effectively classify tweets DistilBert embeddings make a huge difference.

3.0.3 Classification Task

Th word embeddings are given as input to machine learning and deep learning models like CNN,LSTM,BERT,DistilBert etc.These models require input to be given as numerical features which are generated as output of employing word embeddings.Various classification models have been used during the course of this research which are summarized in the subsections below:

- **BERT Model** Ever since its entry, BERT has caused a stir in the Machine Learning community.BERT has been developed by developers at Google which was first published as paper[34].The most distinguish feature of BERT is its ability to train from both left and right side of input matrix.Previous models used to process text sequence from left to right or combined left to right and right to left for training the models.However,the published research in which BERT was proposed of show that a language model which is bidirectionally trained can have a deeper sense of language context and flow than single-direction language models. In the paper[34], the researchers have explained that Masked Language Model (MLM) technique has been used in BERT models.This technique covers 15% of the input sequences are taken as predicted and replaced with symbols.

- **DistilBert Model** DistilBert is another variation of BERT having same bidirectional transformer architecture.However,DistilBert is a distilled version of BERT which works on less resources than BERT model.This distilled version of BERT has forty percent less parameters than BERT itself.Also,coming to the speed it is 60% faster than BERT as well.However,DistilBert still manages to keep 97% language understanding capabilities of BERT while utilizing less resources which makes it a competitive candidate in the machine learning research community.This distillation process is done over

BERT layers,after distillation the number of layers in DistilBert have reduced than in BERT.

- **RoBerta Model** RoBERTa stands for Robustly Optimized BERT Pretraining Approach which was presented by researchers at Facebook and Washington University[35].RoBERTa is built on BERT's language masking strategy, in which the model learns to expect deliberately hidden sections of textual content inside any unannotated language examples. RoBERTa, which was implemented in PyTorch, modifies key hyperparameters in BERT, consisting of getting rid of BERT's next-sentence pretraining goal, and training with tons large mini-batches and learning rates. This lets in RoBERTa to enhance at the masked language modeling goal in comparison with BERT and ends in higher downstream task performance.

- **XLM-Roberta Model** This model is based on Roberta model was presented by authors in their research[36].XLM-Roberta was developed to work on multilingual languages.It has shown improved performance in terms of accuracy and various other matrices over multi lingual BERT as shown by researchers in their publication.XLM-R is a multilingual model trained on 100 different languages. Unlike some XLM multilingual models, it does not require language tensors to understand which language is used, and should be able to determine the correct language from the input ids.

- **LSTM-Long Short Term Memory** LSTMs are a type of Recurrent Neural Networks but a better and improved one.They have more memory than other RNNs due to which it is possible for them to store more information and provide more accurate results.Going into details of LSTM,it has basic three components on which it is based that are Forget Gate,Input Gate and Output Gate.The Forget gate is reponsible for deciding which information is kept for calculating the cell state and which is not relevant and can be discarded.Next is Input Gate that updates the cell state and decides which information is important to be kept. As forget gate helps to drop unimportant information, the input gate helps to find out important information and store certain data in the memory.In the last gate that is Output Gate,classification is performed by applying different functions on it like sigmoid.

- **CNN - Convolutional Neural Networks** A convolutional neural network (CNN) is one of the most used neural networks which is formed from multiple layers of fully connected neurons. In CNN each neuron is connected to all neurons in the layer prior to it. CNN is also capable of learning hidden features in word embeddings of the input text to efficiently capture the semantics in a sentence [37]. The CNN consists of three main layers, a convolution layer, a max pooling layer, and a softmax layer. The main functionality of these layers is to extract features from the input matrix, reduce convolutional matrix size to reduce computational costs, and classify the resulting output. As per previous studies, CNNs give more precise results than other machine learning algorithms because they learn features automatically[38]

Chapter 4

Proposed Methodology:A framework for Disaster and Humanitarian Information Classification

4.1 Proposed Methodology

In this study,firstly two different datasets were combined HumAID [39] and CrisisBench[40] datasets.The resultant dataset is used to perform two types of classification i.e multi-class classification on events and multi-class classification on humanitarian tasks. These classifications are performed by the proposed framework of DistilBERT embeddings, CNN,LSTM, and self-attention layers. All of these steps will be elaborated on in this section.

4.1.1 Feature Extraction

In our proposed multi-class classification framework contextualized DistilBert embeddings have been used. Word embeddings have been utilized in Natural Language Processing tasks because of their effectiveness in representing sentences in low dimension space and effectively capturing their semantic meanings as well. Among a lot of word embeddings, BERT embeddings have gained success because of their nature to form embeddings based on the context of the word[41]. However, BERT utilizes more resources thus DistilBert has been studied in this research which gave promising results, architecture of both are portrayed in Figure4.1.DistilBert is the distilled version of BERT that improves performance with minimal computational cost as well[33].

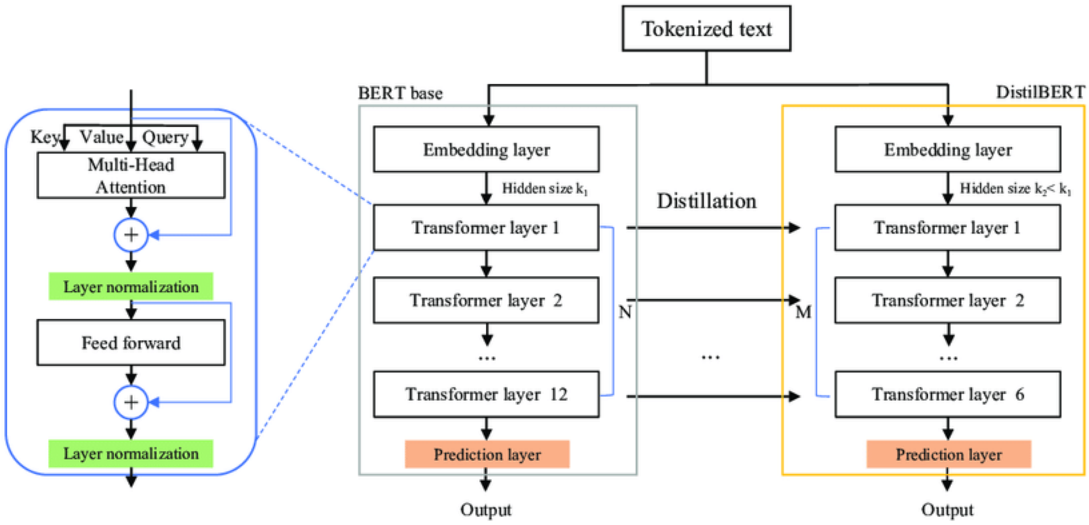

FIGURE 4.1: BERT vs DistilBert architecture [1]

So, keeping in mind the effectiveness of DistilBERT over BERT we have used DIstilBERT embeddings for our research. Same as BERT, DistilBert also captures the context information in the text as well which gives it an extra edge as well. Because, as in our study, we have utilized disaster tweets in which consideration of context plays a vital role i.e sentences like 'He put the stage on ablaze' and 'Jungle is fully ablaze' use the word ablaze in two different contexts. Thus, to cater to such problems and effectively classify tweets DistilBert embeddings make a huge difference.

As DistilBERT has the same functioning as that of BERT so similar to BERT, Masked Language Modeling(MLM) mechanism has been employed in DistilBERT as well which randomly 15% of the words in the sentence are taken as predicted and replaced with symbols. These replaced words are obtained through self-learning by the DistilBERT model. In the process of training, instead of sequentially processing the texts, DistilBERT adopts the Transformer mechanism. A transformer is an encoder-decoder architecture model which uses attention mechanisms[42] to forward a more complete picture of the whole sequence to the decoder at once rather than sequentially. This attention mechanism used by Transformer calculates the relationship between each word in a sentence and portrays the relevance and importance of various words in a sentence. Resultantly the representation that we get captures the context of words in a sentence making it more useful than a simple word vector.

4.1.2 Deep Learning Models

In our proposed framework DistilBERT embeddings are given as input to our proposed hybrid framework that consists of ANNs and self-attention layers to perform multi-class classification of disaster events and humanitarian tasks that are elaborated below.

4.1.2.1 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is one of the most used neural networks which is formed from multiple layers of fully connected neurons. In CNN each neuron is connected to all neurons in the layer prior to it. CNN is also capable of learning hidden features in word embeddings of the input text to efficiently capture the semantics in a sentence [37]. The CNN consists of three main layers, a convolution layer, a max pooling layer, and a softmax layer that are illustrated in Figure 4.3. The main functionality of these layers is to extract features from the input matrix, reduce convolutional matrix size to reduce computational costs, and classify the resulting output. As per previous studies, CNNs give more precise results than other machine learning algorithms because they learn features automatically[38]. Thus, the output embedding from DistilBert is given as input to a 1D Convolutional Neural network (CNN). Moreover, CNN is capable of extracting semantic information from a sequence of sentences and it can learn effective and suitable feature representations. Thus, CNN along with DistilBert plays a vital role in the multi-class classification of tweets.

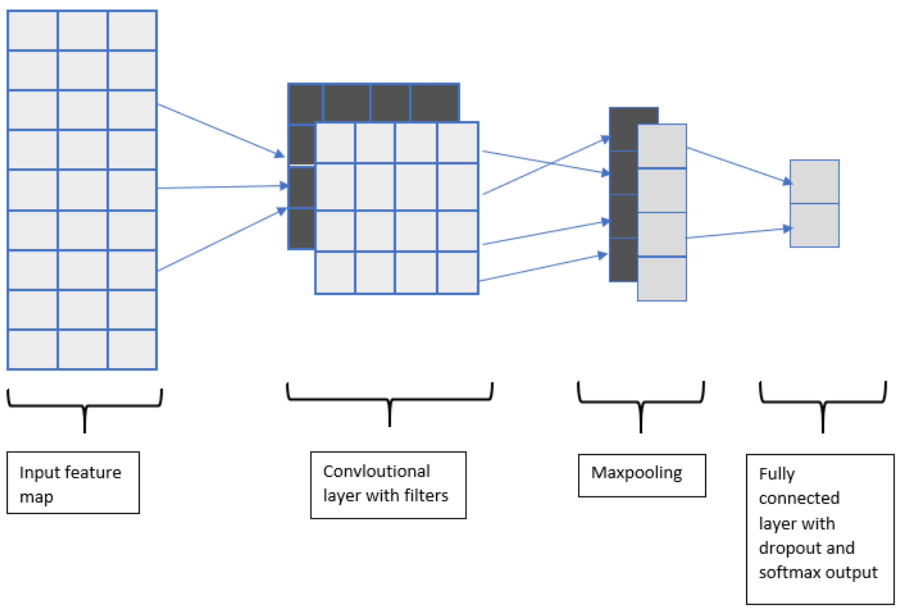

FIGURE 4.2: 1D-CNN architecture

4.1.2.2 Long-Short Term Memory-LSTM

LSTMs are a type of Recurrent Neural Networks but a better and improved one.They have more memory than other RNNs due to which it is possible for them to store more information and provide more accurate results.Going into details of LSTM,it has basic three components on which it is based that are Forget Gate,Input Gate and Output Gate.The Forget gate is reponsible for deciding which information is kept for calculating the cell state and which is not relevant and can be discarded.Next is Input Gate that updates the cell state and decides which information is important to be kept. As forget gate helps to drop unimportant information, the input gate helps to find out important information and store certain data in the memory.In the last gate that is Output Gate,classification is performed by applying different functions on it like sigmoid as illustrated in Figure4.3.So, the output from CNN is passed through LSTM which also retains the context of the input due to its higher memory cells.

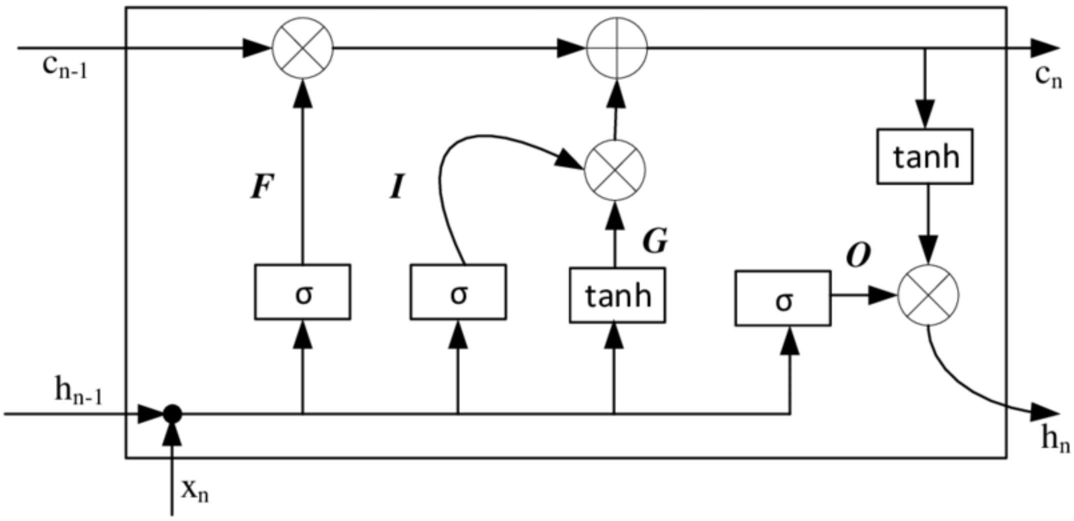

FIGURE 4.3: LSTM architecture

4.1.2.3 Self-Attention Layer

Lastly, in order to pay more attention to contextual information, we use CNN with a self-attention layer. CNN outputs the input variables fed to it with a focus on all of the input variables, so an attention layer is deployed to focus more on relevant words' contexts. A self-attention layer is chosen instead of an attention layer because the attention mechanism only lets output to focus attention on a particular input, whereas, self-attention lets inputs interact with each other hence calculating the attention of all other inputs with respect to one input. The self-attention phenomenon is capable of drawing global dependencies between inputs and outputs. So, a self-attention mechanism has been deployed to focus more on relevant features from input.

Chapter 5

Experiment and Results

5.1 Experiment and results

In this section, we discuss the experiments performed and their results in order to propose a highly effective and efficient approach for multi-class classification.

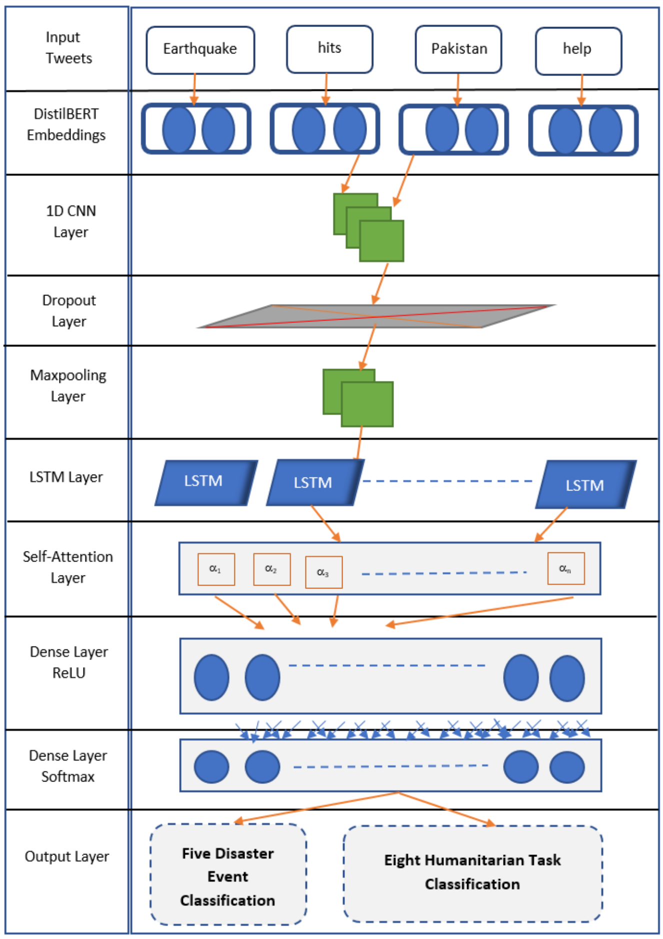

FIGURE 5.1: Proposed Framework Architecture

5.1.1 Proposed Framework Environment Setting

Our Final Proposed hybrid deep learning model is based on a self-attention mechanism. In our proposed model DistilBERT word embedding has been used for feature extraction along with a hybrid model consisting of a Convolutional Neural Network (CNN), and a Self-attention Layer. The proposed model gave the best performance in terms of accuracy and F1 score. Figure 5.1 illustrates the architectural model of the proposed methodology.

A newly formed dataset obtained from a combination of two datasets, as mentioned earlier, is split into training, development, and test data respectively. This data is converted into embeddings using the DistilBERT base uncased model. In order to achieve higher performance from this proposed framework, DistilBERT embeddings with an embedding size of 128 are given as an input matrix to the CNN layer.
CNN is capable of extracting semantic information from a sequence of sentences and it can learn effective and suitable feature representations. In this experimentation, 1D-CNN has been used with a filter size of 2064 and kernel size of 24. Also in the 1D-CNN layer, ReLU activation has been used that decides the relevancy of neurons in the process of classification. Next to the 1D-CNN layer is another dropout layer with a learning rate of 0.5 as well. The max pooling layer with a filter size of 24 comes next to the dropout layer. This layer selects the most prominent features from the matrix.
After CNN layer,another deep learning algorithm namely LSTM has been used to improve performance of the proposed hybrid framework.The output of CNN is given to LSTM layer which is having 1024 memory cells.LSTM layer has been added keeping in mind its greater memory ability to retain smenatic information of the input given to it.Thus,after passing through this layer output is generated on which classification function has been applied.
The output of this layer is given to the self-attention layer which is deployed to focus more on relevant features from the input matrix. The self-attention layer with the softmax function is applied to learn the contribution weights of various words in a tweet. This weight vector is multiplied with different relevant words to learn the contribution of these words in prediction and pay the attention to them according to their weights. Softmax function in the self-attention layer also increases the rate of learning semantic features. In the end, two Fully-Connected-Layers (FCL) are applied with different Activation Functions (AF). First FCL is constituted by Dense Layer with Relu Activation Function and 128 learning units. After that last FCL with Softmax Activation Function has been used to classify

tweets into disaster events and humanitarian tasks respectively.
For optimization during the training process, Adam optimizer has been used in our experimentation over the proposed hybrid framework. Adam optimizer has been used with a learning rate of 3e-05, epsilon of 1e-08, decay of 0.0,1 and clip-norm of 1.0 as settings. Also, sparse categorical cross entropy has been as the loss function.

5.1.2 Metrics

Dimensions	Explanation
True Positive(TP)	Disaster events and humanitarian information instances that are classified correctly.
True Negative(TN)	Tweets with label 1 are predicted as class 2,3 etc.
False Positive(FP)	Tweets with class 2,3 etc are predicted as class 1
False Negative(FN)	Tweets with label 2,3 are predicted as 2,3 etc.

Table 5.1: Elaboration of TP, TN, FP, FN

The evaluation metrics that have been used in this study to gauge the performance of our benchmark classifiers and proposed hybrid framework are Accuracy and F1-score. All the four significant dimensions i.e. True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are catered in the chosen metrics (refer to Table5.1). These valuation metrics are summarized below:

Accuracy: It is the measure of all the cases that have been correctly identified, which means it takes into account the number of correctly classified cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

F1-Score: F1 score takes Precision and Recall into account and it is a result of the harmonic mean of precision and recall. F1 score is an improved measure than the accuracy metric because F1-score measures the cases that have been incorrectly classified as well.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.3}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.4}$$

5.1.3 Benchmark Algorithms

As part of our research, we have studied and implemented various deep learning models,and word embeddings that have been proposed in recent works, to serve as benchmarks for evaluating our proposed hybrid framework.

- **BERT** BERT-based uncased model with L=12, H=768, A=12 where L represents the number of transformer blocks, H represents hidden layers size, and A represents the number of attention heads respectively.
- **DistilBert** It is a distilled version of BERT which has been used with a batch size of 64 and a learning rate of 2e-5.
- **RoBERTa** Another variant of BERT, RoBERTa which is a robustly optimized method has been used as another benchmark algorithm.
- **XLM-RoBERTa** XLM-RoBERTa has also been taken into consideration as one of the baseline algorithms.
- **BERT+LSTM+CNN** In this hybrid model, BERT has been used as word embeddings and LSTM along with CNN for multi-class classification respectively.
- **BERT and CNN** BERT word embeddings are given as input to 1D-CNN in this hybrid model.
- **Glove with CNN** Glove word embeddings are given as input to 1D-CNN in this model.
- **Glove with Bi-LSTM** Glove word embeddings are given as input to Bi-LSTM in this model.

  Both the word embeddings results have been summarized in tabular form in Table 5.2

Models	Disaster Event		Humanitarian Information	
	F1	Accuracy	F1	Accuracy
Proposed Hybrid Framework	**0.981**	**0.981**	**0.88**	**0.88**
Glove +CNN	0.91	0.91	0.81	0.81
Glove +BiLSTM	0.93	0.93	0.84	0.84

TABLE 5.2: Glove and DistilBERT word Embeddings

5.1.4 Results

Overall, the experiment results of benchmark algorithms, hybrid models, and proposed hybrid framework shows that our proposed framework has given promising results in the multi-class classification of five disaster events and eight humanitarian information classes.

5.1.4.1 Disaster Event Classification

In our newly consolidated dataset, there are five disaster events that need to be classified. Overall, for this task almost all the algorithms, that we have experimented with, have shown consistent performance. But our proposed hybrid framework has surpassed all these algorithms and gave the highest performance of 98.10 % as shown in Table 5.3.

5.1.4.2 Humanitarian Information Classification

Eight humanitarian information classes have been classified as present in newly consolidated data. All the benchmark algorithms have shown different accuracy and F1 scores over this multi-class classification. Nonetheless, our proposed hybrid gives the highest performance of 88% over the new dataset as shown in Table 5.3. Also, the most important point to take into consideration is the fall in performance over the different number of classes.

Models	Disaster Event		Humanitarian Information	
	F1	Accuracy	F1	Accuracy
Proposed Hybrid Framework	**0.981**	**0.981**	**0.88**	**0.88**
DistilBert	0.97	0.97	0.83	0.83
RoBERTa	0.97	0.97	0.78	0.78
XLM-RoBERTa	0.97	0.97	0.77	0.77
BERT+LSTM+CNN	0.97	0.97	0.862	0.862
BERT & CNN	0.976	0.976	0.72	0.72

Table 5.3: Experiment Results

Results of the proposed hybrid framework have been given in graphical form in Figures 5.2,5.3,5.4,5.5

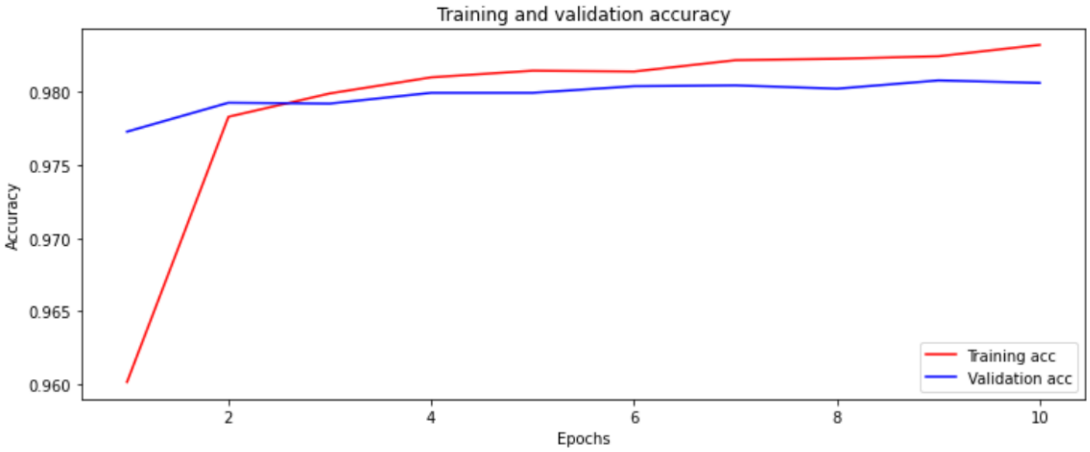

FIGURE 5.2: Training and Validation Accuracy of Disaster events

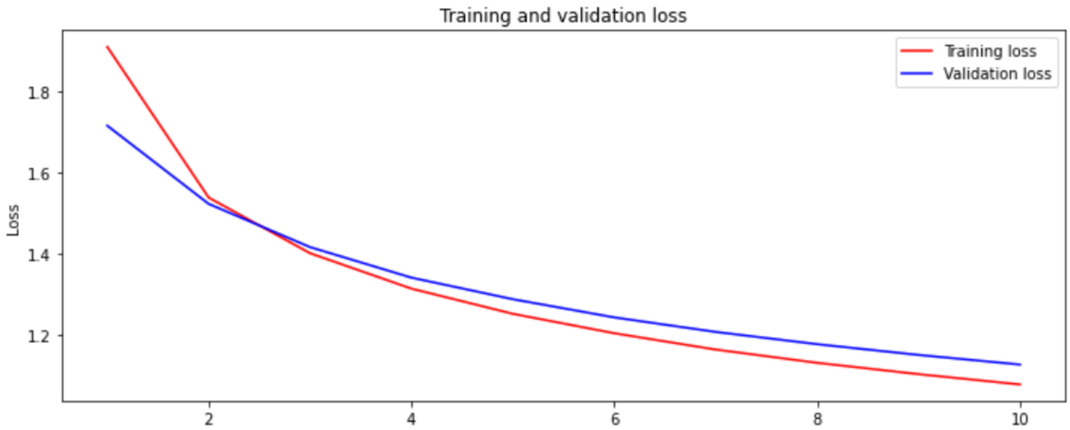

FIGURE 5.3: Training and Validation Loss of Disaster events

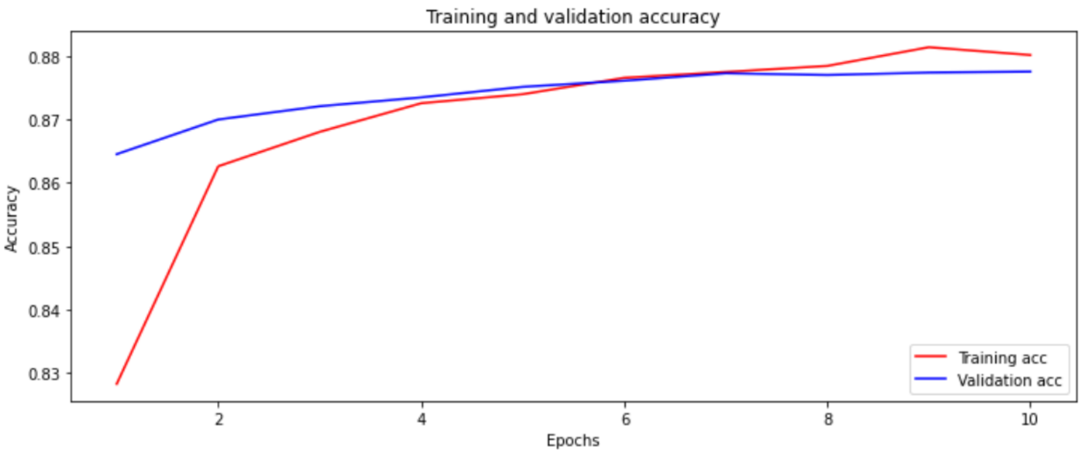

FIGURE 5.4: Training and Validation Accuracy of Humanitarian classes

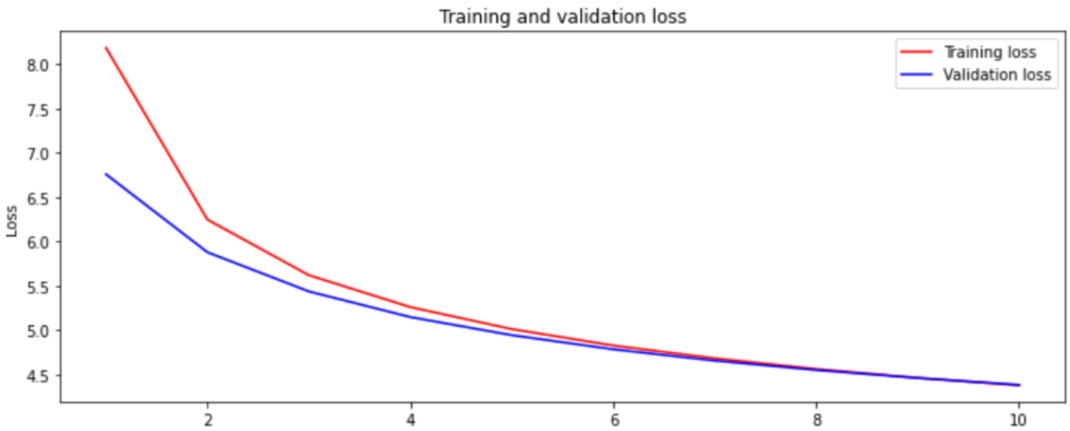

Figure 5.5: Training and Validation Loss of Humanitarian classes

Chapter 6

Conclusion

Social media, mainly Twitter has become a source of public-generated data from which important information can be extracted using deep learning algorithms. In case of disaster happening as well, people post a lot of data related to it on Twitter. However, classification of this data in order to gain insights into disaster events is a gruesome task due to the large number of posts and the ambiguous nature of natural language as well. Thus for our research, we have employed deep learning models to classify tweets into relative disaster events and humanitarian information classes. We have proposed a new hybrid framework to perform multi-class classification on tweets which gave promising results when trained over dataset formed from consolidation, mapping, and label correction of HumAid and Crisis-Bench datasets. Thus a new consolidated dataset has also been produced as part of this research. Finally, our research work also shows that DistilBert used in our hybrid framework shows improved performance over simple BERT embeddings.

Bibliography

[1] H. Adel, A. Dahou, A. Mabrouk, M. Elsayed Abd Elaziz, M. Kayed, I. Elhenawy, S. Alshathri, and A. Ali, "Improving crisis events detection using distilbert with hunger games search algorithm," *Mathematics*, vol. 10, p. 447, 01 2022.

[2] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," pp. 675–684, 01 2011.

[3] H. Purohit, A. Hampton, S. Bhatt, V. Shalin, A. Sheth, and J. Flach, "Identifying seekers and suppliers in social media communities to support crisis coordination," *Computer Supported Cooperative Work*, vol. 23, 12 2014.

[4] W. Xiang and B. Wang, "A survey of event extraction from text," *IEEE Access*, vol. 7, pp. 173111–173137, 2019.

[5] M. Eriksson and E.-K. Olsson Gardell, "Facebook and twitter in crisis communication: A comparative study of crisis communication professionals and citizens," *Journal of Contingencies and Crisis Management*, vol. 24, 06 2016.

[6] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," WWW '10, (New York, NY, USA), p. 591–600, Association for Computing Machinery, 2010.

[7] H. To, S. Agrawal, S. Kim, and C. Shahabi, "On identifying disaster-related tweets: Matching-based or learning-based?," 04 2017.

[8] S. Hiltz, J. Kushma, and L. Plotnick, "Use of social media by u.s. public sector emergency managers: Barriers and wish lists," 01 2014.

[9] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: Survey summary," pp. 507–511, 04 2018.

[10] *ACE (Automatic Content Extraction) English Annotation Guidelines for Events*, 5.4.3 2005.07.01 ed., 2005.

[11] X. Chen, S. Wang, Y. Tang, and T. Hao, “A bibliometric analysis of event detection in social media,” *Online Information Review*, vol. 43, 10 2018.

[12] J. Liu, T. Singhal, L. T. M. Blessing, K. L. Wood, and K. H. Lim, “Crisisbert: a robust transformer for crisis classification and contextual crisis embedding,” *CoRR*, vol. abs/2005.06627, 2020.

[13] A. Smith, “2018 u.s. billion-dollar weather and climate disasters - in context,” 02 2019.

[14] N. Said, K. Ahmad, M. Regular, K. Pogorelov, L. Hassan, N. Ahmad, and N. Conci, “Natural disasters detection in social media and satellite imagery: a survey (to be appeared in multimedia tools and applications),” 01 2019.

[15] L. S. Snyder, M. Karimzadeh, C. Stober, and D. S. Ebert, “Situational awareness enhanced through social media analytics: A survey of first responders,” in *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–8, 2019.

[16] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes twitter users: Real-time event detection by social sensors,” pp. 851–860, 01 2010.

[17] F. Alam, F. Ofli, and M. Imran, “Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes harvey, irma, and maria,” *Behaviour & Information Technology*, vol. 39, pp. 1–31, 05 2019.

[18] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, “Microblogging during two natural hazards events: What twitter may contribute to situational awareness,” vol. 2, pp. 1079–1088, 01 2010.

[19] H. Zade, K. Shah, V. Rangarajan, P. Kshirsagar, M. Imran, and K. Starbird, “From situational awareness to actionability: Towards improving the utility of social media data for crisis response,” *Proc. ACM Hum.-Comput. Interact.*, vol. 2, nov 2018.

[20] A. Kumar, J. P. Singh, and S. Saumya, “A comparative analysis of machine learning techniques for disaster-related tweet classification,” *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*, pp. 222–227, 2019.

[21] C. Wang and D. Lillis, “Classification for crisis-related tweets leveraging word embeddings and data augmentation,” 03 2020.

[22] D. Nguyen, K. Mannai, S. Joty, H. Sajjad, M. Imran, and P. Mitra, "Rapid classification of crisis-related data on social networks using convolutional neural networks," 08 2016.

[23] V. K. Neppalli, C. Caragea, and D. Caragea, "Deep neural networks versus naive bayes classifiers for identifying informative tweets during disasters," in *ISCRAM*, 2018.

[24] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," 02 2018.

[25] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018.

[26] F. Yao and Y. Wang, "Domain-specific sentiment analysis for tweets during hurricanes (dssa-h): A domain-adversarial neural-network-based approach," *Computers, Environment and Urban Systems*, vol. 83, p. 101522, 2020.

[27] F. Alam, M. Imran, and F. Ofli, "Crisisdps: Crisis data processing services," 05 2019.

[28] S. Padhee, T. K. Saha, J. R. Tetreault, and A. Jaimes, "Clustering of social media messages for humanitarian aid response during crisis," *CoRR*, vol. abs/2007.11756, 2020.

[29] C. Fan, F. Wu, and A. Mostafavi, "A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters," *IEEE Access*, vol. PP, pp. 1–1, 01 2020.

[30] A. Bandyopadhyay, D. Ganguly, M. Mitra, S. K. Saha, and G. J. Jones, "An embedding based ir model for disaster situations," *Information Systems Frontiers*, vol. 20, p. 925–932, oct 2018.

[31] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "Crisislex: A lexicon for collecting and filtering microblogged communications in crises," *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, pp. 376–385, 01 2014.

[32] H. Saif, M. Fernandez, Y. He, and H. Alani, "On stopwords, filtering and data sparsity for sentiment analysis of Twitter," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*,

(Reykjavik, Iceland), pp. 810–817, European Language Resources Association (ELRA), May 2014.

[33] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter," *CoRR*, vol. abs/1910.01108, 2019.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *ArXiv*, vol. abs/1810.04805, 2019.

[35] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019.

[36] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," *CoRR*, vol. abs/1911.02116, 2019.

[37] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Doha, Qatar), pp. 1746–1751, Association for Computational Linguistics, Oct. 2014.

[38] D. T. Nguyen, K. Al-Mannai, S. R. Joty, H. Sajjad, M. Imran, and P. Mitra, "Rapid classification of crisis-related data on social networks using convolutional neural networks," *CoRR*, vol. abs/1608.03902, 2016.

[39] F. Alam, U. Qazi, M. Imran, and F. Ofli, "Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks," *CoRR*, vol. abs/2104.03090, 2021.

[40] F. Alam, H. Sajjad, M. Imran, and F. Ofli, "Standardizing and benchmarking crisis-related social media datasets for humanitarian information processing," *CoRR*, vol. abs/2004.06774, 2020.

[41] A. K. Chanda, "Efficacy of BERT embeddings on predicting disaster from Twitter data," *arXiv e-prints*, p. arXiv:2108.10698, Aug. 2021.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017.