MULTIMODAL ANALYSIS OF DISASTER MANAGEMENT

A PROJECT REPORT

Submitted By

KAVYA R.	**312215104046**
SANGEETHA V.S.	**312215104092**
SWARNALATHA N.	**312215104112**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

SSN COLLEGE OF ENGINEERING

KALAVAKKAM 603110

ANNA UNIVERSITY :: CHENNAI - 600025

April 2019

ANNA UNIVERSITY : CHENNAI 600025

BONAFIDE CERTIFICATE

Certified that this project report titled **“MULTIMODAL ANALYSIS FOR DISASTER MANAGEMENT”** is the *bonafide* work of “**KAVYA R. (312215104046)**, **SANGEETHA V.S. (312215104092)**, and **SWARNALATHA N. (312215104112)**” who carried out the project work under my supervision.

DR. CHITRA BABU
HEAD OF THE DEPARTMENT
Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

MR.B.SENTHIL KUMAR
SUPERVISOR
Assistant Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

Place:
Date:

Submitted for the examination held on. . . . .

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENTS

We thank GOD, the almighty for giving us the strength and knowledge to do this project.

We would like to thank and convey our deep sense of gratitude to our guide **Mr. B.SENTHIL KUMAR**, Assistant Professor, Department of Computer Science and Engineering, for his valuable advice and suggestions as well as his continued guidance, patience and support throughout the project.

We would like to convey our special thanks to **Dr. C. ARAVINDAN**, Professor, Department of Computer Science and Engineering for his continued guidance,encouragement and timely inputs that helped us shape and refine our work.

Our sincere thanks to **Dr. CHITRA BABU**, Professor and Head of the Department of Computer Science and Engineering, for her support and encouragement. We would like to thank our project Coordinator **Dr. S. SHEERAZUDDIN**, Professor, Department of Computer Science and Engineering for his valuable suggestions throughout the project.

We express our deep respect to the founder **Dr. SHIV NADAR**, Chairman, SSN Institutions. We also express our appreciation to **Dr. S. SALIVAHANAN**, Principal, for all the help he has rendered during this course of study.

We would like to extend our sincere thanks to all the teaching and non-teaching staffs of our department who have contributed directly and indirectly during the course of our project work. Finally, we would like to thank our parents and friends for their patience, cooperation and moral support throughout our life.

KAVYA R. **SANGEETHA V.S.** **SWARNALATHA N.**

ABSTRACT

Traditional disaster assessment of damage heavily relies on expensive GIS data, especially remote sensing image data. In recent years, social media has become a rich source of disaster information that may be useful in evaluating damage at a lower cost. Such information includes text (e.g. tweets) or images posted by eyewitnesses of disaster and emergency events such as earthquakes, typhoons , floods , tsunami etc. During the sudden onset of a critical situation, affected people resort to social media for a solution and therefore, post useful information (text and images) on Twitter that can be used for situational awareness and other humanitarian disaster response efforts, if processed timely and effectively. But the volume and velocity of tweets posted during crises today tend to be extremely high, making it hard for humanitarian organizations and professional emergency responders to process the information in a timely manner. In this project, we present an automated solution that identifies and matches the request(need) to their corresponding offer(supplies) for textual and imagery twitter data thereby accelerating the emergency relief efforts. This project stands useful for both the Government and Non-profit Organizations while undertaking Disaster Management and Relief coordination actions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

At times of natural and man-made disasters, social media platforms such as Twitter and Facebook are considered vital information sources that contain a variety of useful information such as reports of injured or dead people, infrastructure and utility damage, urgent needs of affected people, and missing or found people among others.

1.1 Motivation

Twitter has been extensively used as an active communication platform, especially during critical events such as earthquakes, floods, typhoons, etc. During the onset of such events, a variety of information is posted in real-time by affected people; by people who are in need of help (e.g., food, shelter, medical assistance, etc.) or by people who are willing to donate or offer volunteering services.

Moreover, humanitarian and formal crisis response organizations such as government agencies, public health care NGOs, and the military are tasked with responsibilities to save lives, reach people who are in need of help, etc. Situation-sensitive requirements arise during such events and formal disaster response agencies look for actionable and tactical information in real-time to effectively estimate the aftermath of a disaster, and to launch relief efforts accordingly.

At the same time, the volume and velocity of tweets posted during the crisis currently are extremely high, making it difficult for professional emergency responders to process information in a timely manner. This issue acts as one motivating factor and is resolved in this project by creating an automated tool for extraction, classification and matching of tweets.

Apart from this, another major motivation for our project is the usage of a variety of twitter data such as text, images and videos to gain insight into an event. Although many automated systems based on Artificial Intelligence and Machine Learning have been developed in this field, majority of them focus primarily on analyzing the textual content, ignoring the rich information provided by the visual content. The proposed system addresses this limitation by introducing a real-time social media text and image processing pipeline to assist organizations carry out disaster response and management operations.

Our system acts as a viable tool that matches the requests(need) to the corresponding offers(demand) in the tweets by prioritizing the similar requests based on the severity of the damage observed in the image. A typical request message involves an entity (person or organization) describing the scarcity of a certain resource or service (e.g., clothing, volunteering) and/or asking others to supply said resource. A typical offer involves an entity describing the availability and/or willingness to supply a resource.

An example scenario for request-offer mapping is illustrated in the Figure 1.1 in the form of **use-case** info graphics:

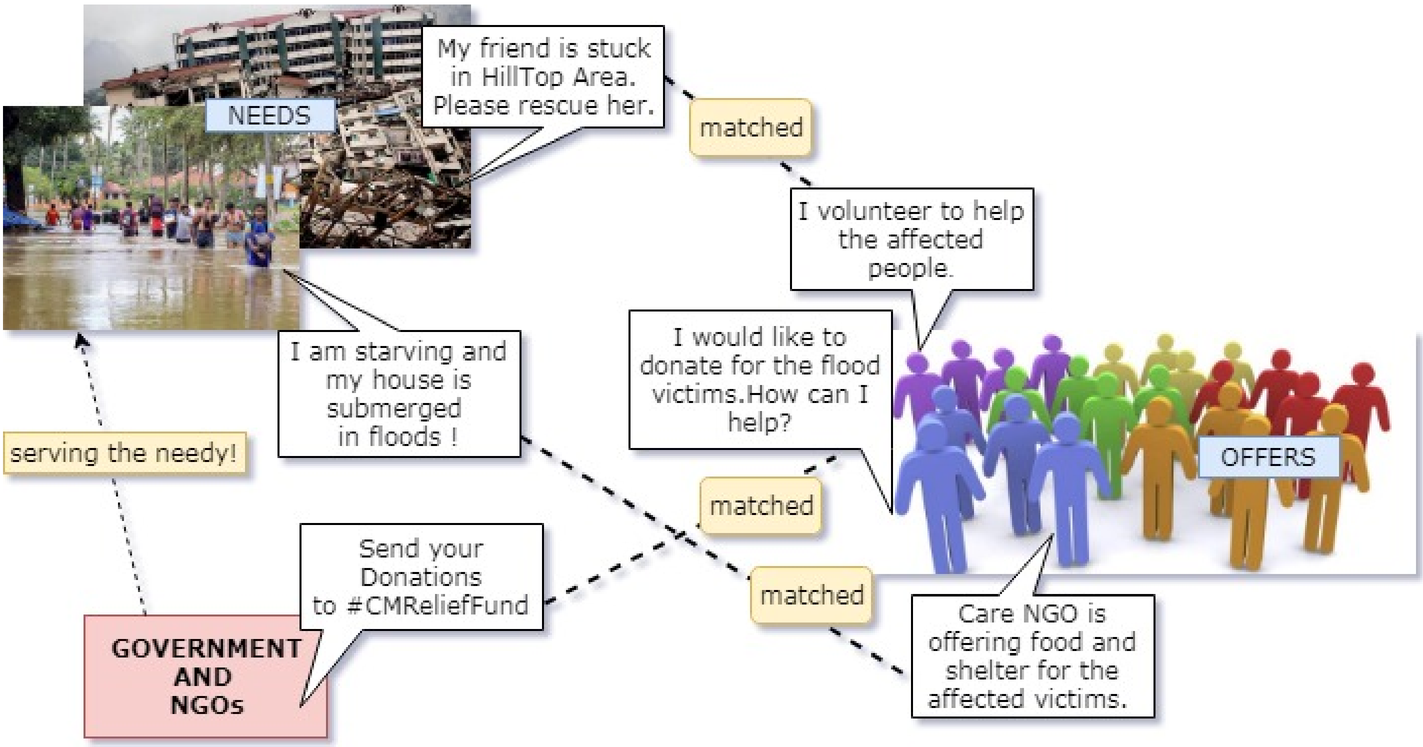

NEEDS
My friend is stuck in HillTop Area. Please rescue her.
matched
I volunteer to help the affected people.
I would like to donate for the flood victims.How can I help?
I am starving and my house is submerged in floods !
serving the needy!
matched
OFFERS
Send your Donations to #CMReliefFund
matched
GOVERNMENT AND NGOs
Care NGO is offering food and shelter for the affected victims.

FIGURE 1.1: Use-case Diagram For Request-Offer Matching

The rest of the thesis is organized as follows. The next Chapter presents the detailed Literature Survey of related works followed by Chapter 3 wherein the Problem Statement of the project is stated. Next, the proposed system methodologies, the algorithm and techniques used for the implementation of individual modules are explained in Chapter 4. The results and observations are discussed in Chapter 5. Finally, the conclusion and future work of the project is proposed in Chapter 6.

CHAPTER 2

LITERATURE SURVEY

In this chapter, a detailed description of similar and related works, the methodology used in their work, an analysis of their results and limitations are discussed.

2.1 Related Work

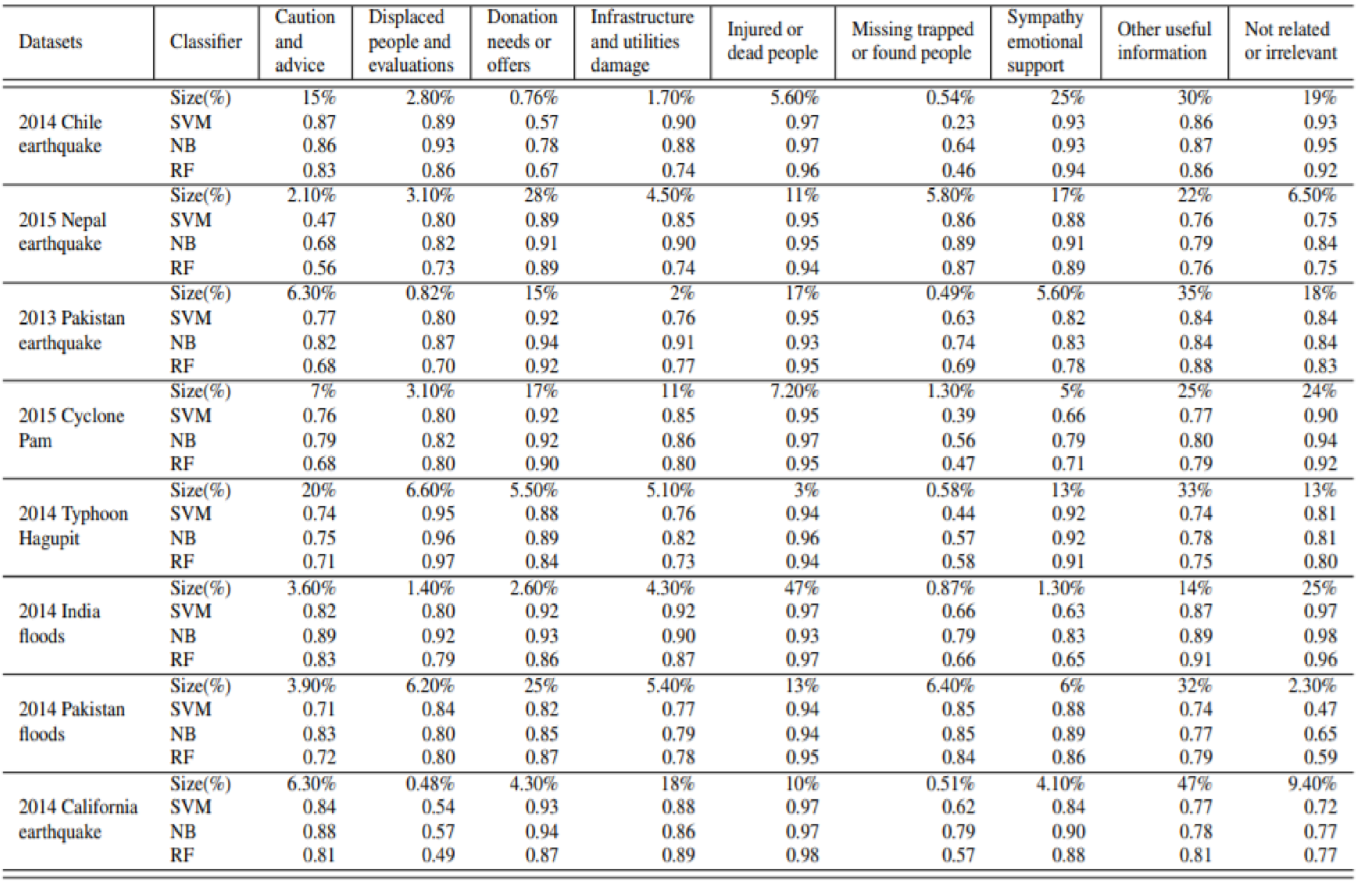
2.1.1 Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages

In Imran M et al.[11], a human-annotated Twitter corpora collected during 19 different crises that took place between 2013 and 2015 is presented. This human annotations done by volunteers and crowd-sourced workers are of two types. First, the tweets are annotated with a set of categories such as displaced people, financial needs, infrastructure, etc. Second , the tweets are annotated to identify out-of-vocabulary(OOV) terms such as slangs, place names, abbreviations, misspellings, etc. and their corrected normalized forms. This human-annotations is further used to built machine-learning classifiers such as Naive Bayes, Random Forest, and Support Vector Machines (SVM), in a multiclass classification setting, to classify messages that are useful for humanitarian efforts. Also, word2vec word embeddings trained using 52 million crisis-related messages has been furnished.

The system categorizes the data into following categories:

- Request—Disaster victims and humanitarian organizations requesting for help
- Offer—NGOs and relief measure organizations ready to offer their help
- Other useful information—Other useful information that helps understand the situation such as
  - Injured or dead people—Reports of casualties and/or injured people due to the crisis
  - Missing, trapped, or found people—Reports and/or questions about missing or found people
  - Displaced people and evacuations—People who have relocated due to the crisis, even for a short time (includes evacuations)
  - Infrastructure and utilities damage—Reports of damaged buildings, roads, bridges, or utilities/services interrupted or restored
  - Caution and advice—Reports of warnings issued or lifted, guidance and tips
  - Sympathy and emotional support—Prayers, thoughts, and emotional support

The three different kinds of classifiers are trained using the preprocessed data and the evaluation is done using the 10-folds cross-validation technique. The classification results is calculated in terms of Area Under ROC Curve for selected datasets across all classes using Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF) and are tabulated as shown in Figure 2.1.



Datasets	Classifier	Caution and advice	Displaced people and evaluations	Donation needs or offers	Infrastructure and utilities damage	Injured or dead people	Missing trapped or found people	Sympathy emotional support	Other useful information	Not related or irrelevant
2014 Chile earthquake	Size(%)	15%	2.80%	0.76%	1.70%	5.60%	0.54%	25%	30%	19%
	SVM	0.87	0.89	0.57	0.90	0.97	0.23	0.93	0.86	0.93
	NB	0.86	0.93	0.78	0.88	0.97	0.64	0.93	0.87	0.95
	RF	0.83	0.86	0.67	0.74	0.96	0.46	0.94	0.86	0.92
2015 Nepal earthquake	Size(%)	2.10%	3.10%	28%	4.50%	11%	5.80%	17%	22%	6.50%
	SVM	0.47	0.80	0.89	0.85	0.95	0.86	0.88	0.76	0.75
	NB	0.68	0.82	0.91	0.90	0.95	0.89	0.91	0.79	0.84
	RF	0.56	0.73	0.89	0.74	0.94	0.87	0.89	0.76	0.75
2013 Pakistan earthquake	Size(%)	6.30%	0.82%	15%	2%	17%	0.49%	5.60%	35%	18%
	SVM	0.77	0.80	0.92	0.76	0.95	0.63	0.82	0.84	0.84
	NB	0.82	0.87	0.94	0.91	0.93	0.74	0.83	0.84	0.84
	RF	0.68	0.70	0.92	0.77	0.95	0.69	0.78	0.88	0.83
2015 Cyclone Pam	Size(%)	7%	3.10%	17%	11%	7.20%	1.30%	5%	25%	24%
	SVM	0.76	0.80	0.92	0.85	0.95	0.39	0.66	0.77	0.90
	NB	0.79	0.82	0.92	0.86	0.97	0.56	0.79	0.80	0.94
	RF	0.68	0.80	0.90	0.80	0.95	0.47	0.71	0.79	0.92
2014 Typhoon Hagupit	Size(%)	20%	6.60%	5.50%	5.10%	3%	0.58%	13%	33%	13%
	SVM	0.74	0.95	0.88	0.76	0.94	0.44	0.92	0.74	0.81
	NB	0.75	0.96	0.89	0.82	0.96	0.57	0.92	0.78	0.81
	RF	0.71	0.97	0.84	0.73	0.94	0.58	0.91	0.75	0.80
2014 India floods	Size(%)	3.60%	1.40%	2.60%	4.30%	47%	0.87%	1.30%	14%	25%
	SVM	0.82	0.80	0.92	0.92	0.97	0.66	0.63	0.87	0.97
	NB	0.89	0.92	0.93	0.90	0.93	0.79	0.83	0.89	0.98
	RF	0.83	0.79	0.86	0.87	0.97	0.66	0.65	0.91	0.96
2014 Pakistan floods	Size(%)	3.90%	6.20%	25%	5.40%	13%	6.40%	6%	32%	2.30%
	SVM	0.71	0.84	0.82	0.77	0.94	0.85	0.88	0.74	0.47
	NB	0.83	0.80	0.85	0.79	0.94	0.85	0.89	0.77	0.65
	RF	0.72	0.80	0.87	0.78	0.95	0.84	0.86	0.79	0.59
2014 California earthquake	Size(%)	6.30%	0.48%	4.30%	18%	10%	0.51%	4.10%	47%	9.40%
	SVM	0.84	0.54	0.93	0.88	0.97	0.62	0.84	0.77	0.72
	NB	0.88	0.57	0.94	0.86	0.97	0.79	0.90	0.78	0.77
	RF	0.81	0.49	0.87	0.89	0.98	0.57	0.88	0.81	0.77

FIGURE 2.1: Classification results in terms of Area Under ROC Curve using Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF). Source: [11]

2.1.2 A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria

In Alam F et al. [2], an extensive multidimensional analysis of textual and multimedia content is conducted from millions of tweets shared on Twitter data collected during the three disasters, namely Hurricanes Harvey, Irma, and Maria. The objective of this paper includes the following tasks: It performs sentiment analysis to determine how peoples' thoughts and feelings change over time as disaster events progress. It employs topic modeling techniques to gain insight into the different topics discussed during each day in order to help concerned authorities to quickly sift through big crisis data. It classifies both textual and imagery content into several humanitarian categories as mentioned in [11]. The data collected from CrisisNLP repository[11] is made use for the analysis.

To perform the sentiment analysis, it makes use of the Stanford sentiment analysis classifier that classifies the text into 5 categorical labels such as Very Negative, Negative, Neutral, Positive and Very Positive. The accuracy of the classifier for fine-grained sentiment labels is 80.7%. The classification of humanitarian categories are performed using a decision tree based learning scheme known as Random Forest. The performance obtained using the test set in terms of the F-measure is F1 = 0.64 and accuracy of 0.66.

The topic modelling is evaluated using LDA (Latent Drichlet Allocation) method to obtain the top 30 frequently used terms. Standford NER tagger is used for the identification of most mentioned named entities that helps in discovering important stories related to actual local emergency needs.The reported F-measure of this NER system is 86.72% to 92.28% for different datasets.

The images are classified as severe, mild and none and are trained based on the human annotated ground truth values using their relevancy filtering model. The overall accuracy of the resulting damage assessment models varied from 76% to 90%.

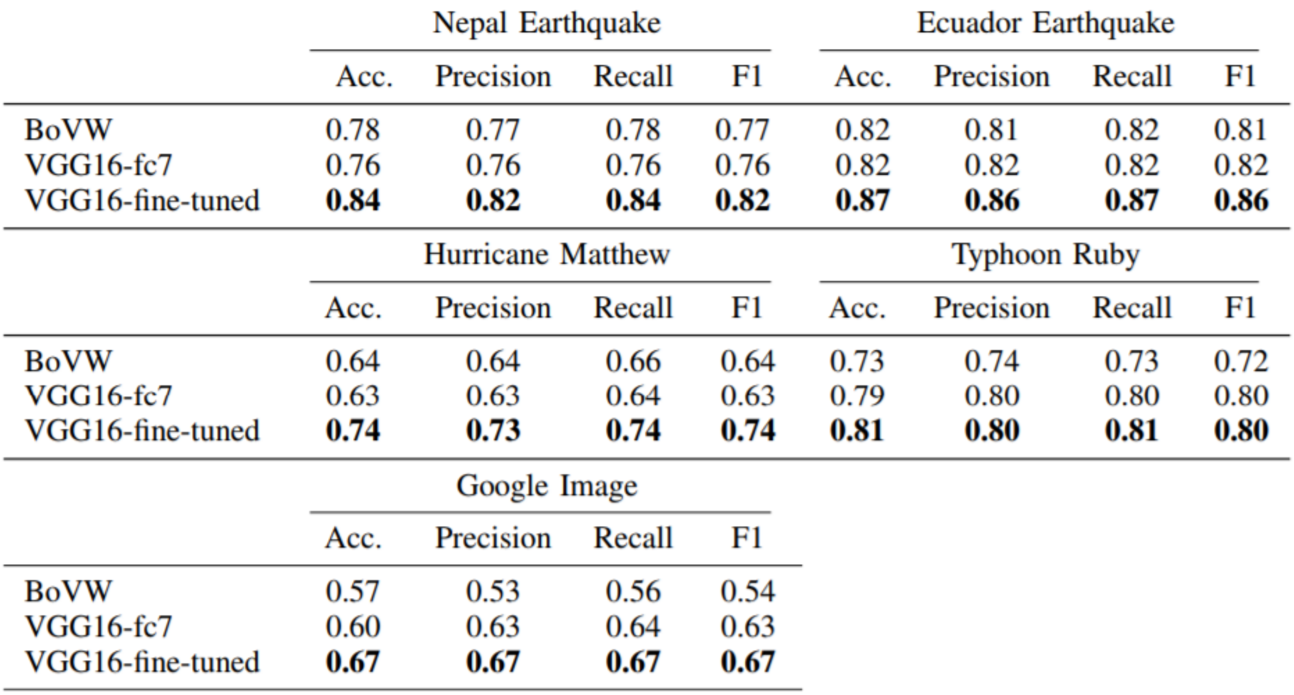
The major limitation of this work is that it is semi-automated and requires human in the loop for machine training. Further, it does not address the scalability issues that arise owing to the volume of data being processed.

.

2.1.3 Damage Assessment from Social Media Imagery Data During Disasters

In Nguyen D T et al.[5], determination of the level of destruction caused by the disasters has been performed. In this study, labeled data from past disaster events as well as data collected from the Web have been leveraged. These data resources are further annotated using the Crowdflower (Crowdsourcing platform) based on fixed guidelines. For damage assessment, three levels namely : severe damage, mild damage, and little-to-no damage are taken into consideration. This work employs both traditional computer vision techniques such as Bag-of-Visual-Words (BoVW) model as well as state-of-the-art deep learning techniques such as Convolutional Neural Networks (CNN) to assess the level of destruction in disaster images.

The event specific experiments using the three learning techniques (i.e.BoVW, VGG16-fc7, and VGG16-fine-tuned) are evaluated in terms of precision, recall, F1 score and overall accuracy as shown in Figure 2.2. The experimental analysis of this work shows that domain-specific fine-tuning of deep CNNs outperforms the traditional BoVW models by a considerable margin. However, there are two main challenges, i.e. low prevalence of the training data and non-trivial human-labeling tasks in this work.



	Nepal Earthquake				Ecuador Earthquake			
	Acc.	Precision	Recall	F1	Acc.	Precision	Recall	F1
BoVW	0.78	0.77	0.78	0.77	0.82	0.81	0.82	0.81
VGG16-fc7	0.76	0.76	0.76	0.76	0.82	0.82	0.82	0.82
VGG16-fine-tuned	**0.84**	**0.82**	**0.84**	**0.82**	**0.87**	**0.86**	**0.87**	**0.86**

	Hurricane Matthew				Typhoon Ruby			
	Acc.	Precision	Recall	F1	Acc.	Precision	Recall	F1
BoVW	0.64	0.64	0.66	0.64	0.73	0.74	0.73	0.72
VGG16-fc7	0.63	0.63	0.64	0.63	0.79	0.80	0.80	0.80
VGG16-fine-tuned	**0.74**	**0.73**	**0.74**	**0.74**	**0.81**	**0.80**	**0.81**	**0.80**

	Google Image			
	Acc.	Precision	Recall	F1
BoVW	0.57	0.53	0.56	0.54
VGG16-fc7	0.60	0.63	0.64	0.63
VGG16-fine-tuned	**0.67**	**0.67**	**0.67**	**0.67**

FIGURE 2.2: Event-specific results for all events in terms of precision, recall, f1 score, and overall accuracy using three different learning schemes.
Source:[5]

2.1.4 Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster

In Varga I et al.[3], a method for discovering matches between problem reports and aid messages is proposed. Their system contributes to problem-solving in a large scale disaster situation by facilitating communication between victims and humanitarian organizations. A machine learning based system is developed to recognize problem reports, aid messages and problem-aid tweet matches.

First, location names in tweets are identified by matching tweets against their location dictionary. Then, each tweet is paired with each dependency relation in the tweet, which is referred to as the nuclei and given to the problem report and aid message recognizer. A tweet-nucleus-candidate pair judged as problem report

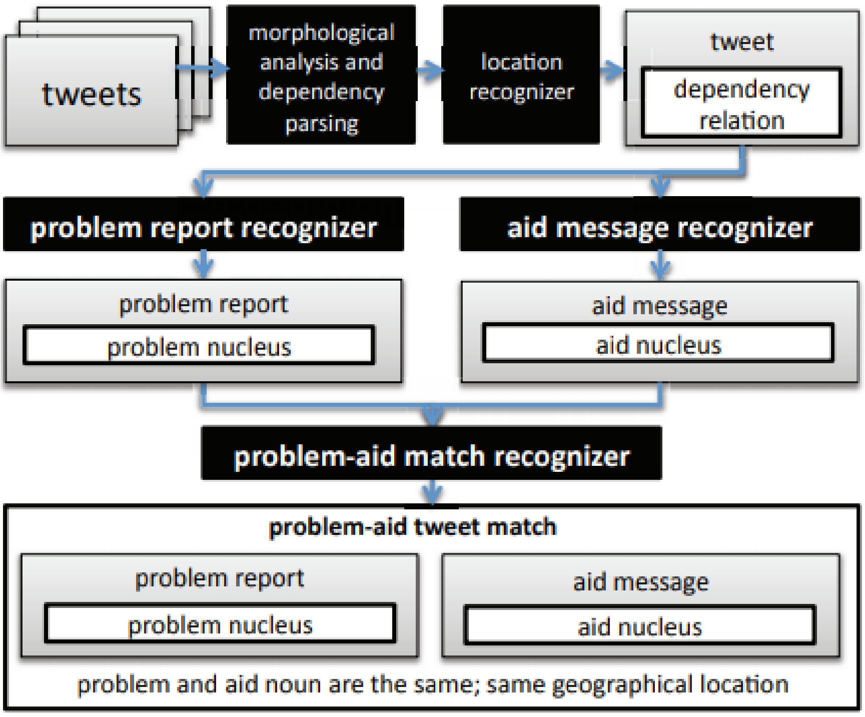


FIGURE 2.3: Problem-aid matching system overview.
Source:[3]

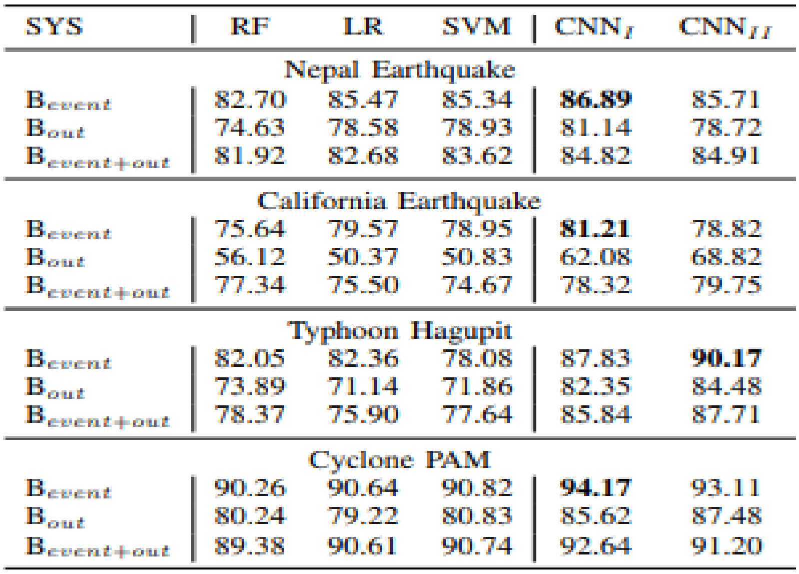
is combined with another tweet-nucleus-candidate pair recognized as an aid message if the two nuclei share the same noun and the tweets share the same location name.In this work,it does not employ structural characteristics of tweets as restrictions (e.g. a problem report and its aid message need to be in the same tweet chain).

2.1.5 CrisisMMD: Multimodal Twitter Datasets from Natural Disasters

In Alam F et al.[4], human-labeled multimodal datasets collected from Twitter during seven recent natural disasters including earthquakes, hurricanes, wildfires, and floods is presented. The tweets that do not contain at least one image URL is filtered out. The data is further annotated by paid workers from a well-known crowdsourcing platform based on three humanitarian tasks .The first task aims to categorize the data into two high-level categories called Informative or Not informative. The second task, on the other hand, aims to identify critical and potentially actionable information such as reports of injured or dead people, infrastructure damage, etc. from the tweets. For this purpose, seven humanitarian categories is used. The final task aims at assessing the severity of damage shown in the images. The severity of damage is based on the extent of physical destruction to a build-structure. First the task one is executed by human annotators, and only the informative tweets are passed to task two. In second task, the images are annotated manually and only those included in the 'infrastructure and utility damage' category are given to task three. In task three, the images are classified as severe, mild and no damage categories. The labelled dataset is available as CrisisMMD which is published at the CrisisNLP site. The major limitation of this work is that it is semi-automated and requires human annotators.

2.1.6 Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks

In Nyugen D T et al.[9], neural network based classification methods for binary and multi-class tweet classification task is presented. Data from use data from multiple sources: (1) CrisisNLP, (2) CrisisLex, and (3) AIDR is used. The data is preprocessed and the tweets are tokenized using the CMU TweetNLP tool. The unigram, bigram and trigram features are then extracted from the tweets as features. The features are converted to TF-IDF vectors by considering each tweet



SYS	RF	LR	SVM	CNN$_I$	CNN$_{II}$
		Nepal Earthquake			
B$_{event}$	82.70	85.47	85.34	**86.89**	85.71
B$_{out}$	74.63	78.58	78.93	81.14	78.72
B$_{event+out}$	81.92	82.68	83.62	84.82	84.91
		California Earthquake			
B$_{event}$	75.64	79.57	78.95	**81.21**	78.82
B$_{out}$	56.12	50.37	50.83	62.08	68.82
B$_{event+out}$	77.34	75.50	74.67	78.32	79.75
		Typhoon Hagupit			
B$_{event}$	82.05	82.36	78.08	87.83	**90.17**
B$_{out}$	73.89	71.14	71.86	82.35	84.48
B$_{event+out}$	78.37	75.90	77.64	85.84	87.71
		Cyclone PAM			
B$_{event}$	90.26	90.64	90.82	**94.17**	93.11
B$_{out}$	80.24	79.22	80.83	85.62	87.48
B$_{event+out}$	89.38	90.61	90.74	92.64	91.20

FIGURE 2.4: Performance Comparison based on AUC.
Source:[9]

as a document. These features are used by the non-neural models. Some of non-neural existing models including: (i) Support Vector Machine (SVM), (ii) Logistic Regression (LR), and (iii) Random Forest (RF) are experimented with for comparision.

A CNN model is trained by optimizing the cross entropy and maximum number of epochs as 25. Various dropout rates and minibatch sizes are experimented with and rectified linear units (ReLU) is used for the activation function. The CNN model is initialized using two types of pretrained word embeddings. (i) Crisis Embeddings (CNN1) : trained on all crisis tweets data (ii) Google Embeddings (CNN2) trained on the Google News dataset.

The results of classification by several non-neural classifiers is compared with the CNN based classifier using the AUC (Area under the ROC) score. CNN outperforms the traditional non-neural methods. The drawback in this work is that CNN does not consider the semantic relatedness of a tweet under several classes which brings inconsistency in labeling.

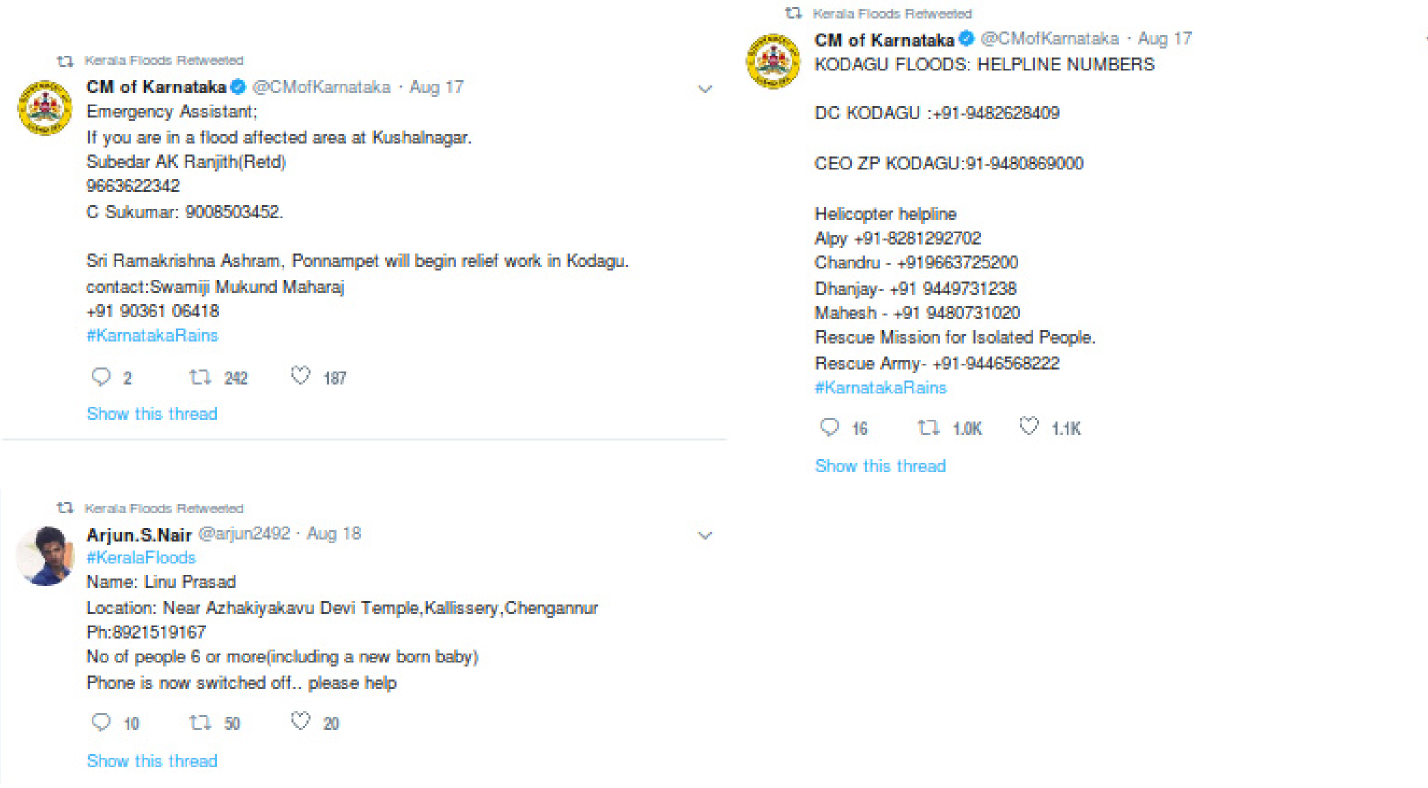
CHAPTER 3

PROBLEM STATEMENT

The proposed system takes real-time twitter data and maps the request tweets to their corresponding offer tweets by taking both the text and image into consideration.

3.1 Input

Real-time social media data containing either only text or both text and image.



Kerala Floods Retweeted
CM of Karnataka @CMofKarnataka · Aug 17
Emergency Assistant;
If you are in a flood affected area at Kushalnagar.
Subedar AK Ranjith(Retd)
9663622342
C Sukumar: 9008503452.

Sri Ramakrishna Ashram, Ponnampet will begin relief work in Kodagu.
contact:Swamiji Mukund Maharaj
+91 90361 06418
#KarnatakaRains
2 242 187
Show this thread

Kerala Floods Retweeted
Arjun.S.Nair @arjun2492 · Aug 18
#KeralaFloods
Name: Linu Prasad
Location: Near Azhakiyakavu Devi Temple,Kallissery,Chengannur
Ph:8921519167
No of people 6 or more(including a new born baby)
Phone is now switched off.. please help
10 50 20
Show this thread

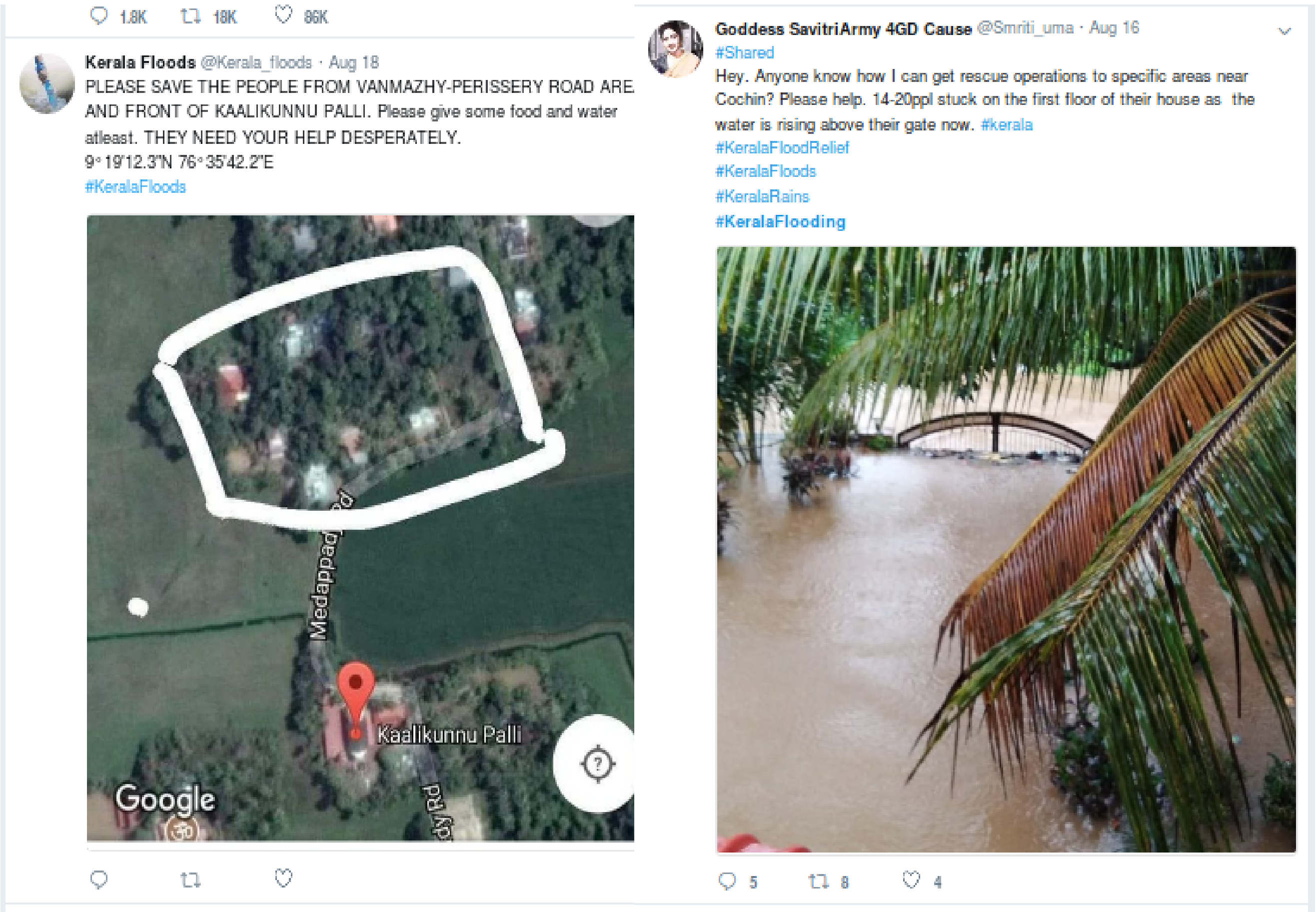
Kerala Floods Retweeted
CM of Karnataka @CMofKarnataka · Aug 17
KODAGU FLOODS: HELPLINE NUMBERS

DC KODAGU :+91-9482628409

CEO ZP KODAGU:91-9480869000

Helicopter helpline
Alpy +91-8281292702
Chandru - +919663725200
Dhanjay- +91 9449731238
Mahesh - +91 9480731020
Rescue Mission for Isolated People.
Rescue Army- +91-9446568222
#KarnatakaRains
16 1.0K 1.1K
Show this thread

FIGURE 3.1: Tweets containing textual contents only.
Source: https://twitter.com/



1.8K 18K 86K

Kerala Floods @Kerala_floods · Aug 18

PLEASE SAVE THE PEOPLE FROM VANMAZHY-PERISSERY ROAD ARE
AND FRONT OF KAALIKUNNU PALLI. Please give some food and water
atleast. THEY NEED YOUR HELP DESPERATELY.
9° 19'12.3"N 76° 35'42.2"E
#KeralaFloods

Goddess SavitriArmy 4GD Cause @Smriti_uma · Aug 16

#Shared
Hey. Anyone know how I can get rescue operations to specific areas near
Cochin? Please help. 14-20ppl stuck on the first floor of their house as the
water is rising above their gate now. #kerala
#KeralaFloodRelief
#KeralaFloods
#KeralaRains
#KeralaFlooding

5 8 4

FIGURE 3.2: Tweets containing both text and image.
Source:https://twitter.com/

3.2 Output

Matched (request,offer) pairs based on requirements and severity of the damage that helps in relief operations

3.3 Use Case

The proposed system can be employed to assist the government and humanitarian organizations to launch relief operations immediately in disaster-hit regions.

CHAPTER 4

PROPOSED SYSTEM METHODOLOGY

The proposed system aims at multimodal analysis for disaster management which involves analysis of both text and image from social media in order to help the government as well as humanitarian organizations to perform relief operations immediately.

The pipeline of the system consists of 4 phases: **Collection, Classification, Prioritization and Matching.**

The proposed system architecture is illustrated in Fig 4.1.

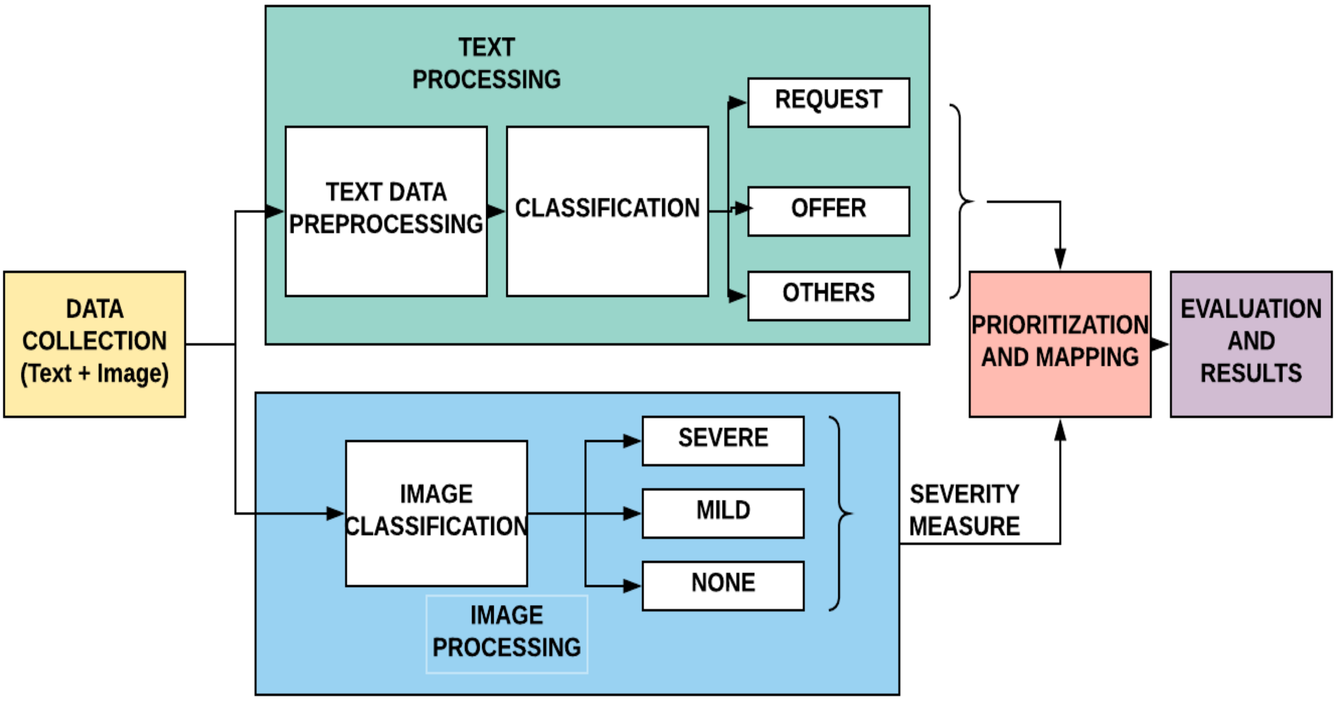


FIGURE 4.1: Proposed System Architecture

```
Result: matched_tweets
while true do
    $hashtag = trending\_hashtags(\text{APIcredentials}, location)$;
    if check($hashtag$) /* Disaster based hashtag */ then
        $tweets = extract\_tweets(\text{APIcredentials}, hashtag)$;
        $tweets = preprocess(tweets)$;
        for $t$ ←tweets do
            if contains($t$,$onlytext$) then
                $class\_label \leftarrow classify\_text(t)$;
            else
                /* contains both text and image */
                $class\_label \leftarrow classify\_text(t)$;
                $priority \ast_{\leftarrow} classify\_image(t)$;
            end
        end
        $matching \leftarrow match$(request,offer)
    else
        continue;
    end
end
*Priority is computed only for the request tweets.
```

Algorithm 1: Proposed System Algorithm

4.1 Data Collection

The Collection phase includes real-time extraction and analysis of both textual and imagery tweets that are relevant to a particular ongoing disastrous event. The live tweets pertaining to disasters are fetched based on hashtag. The **Tweepy** API is imported for retrieval of twitter data.Tweepy is open-sourced, hosted on

GitHub and enables Python to communicate with Twitter platform. The flowchart for live tweets extraction is shown in Fig 4.2.

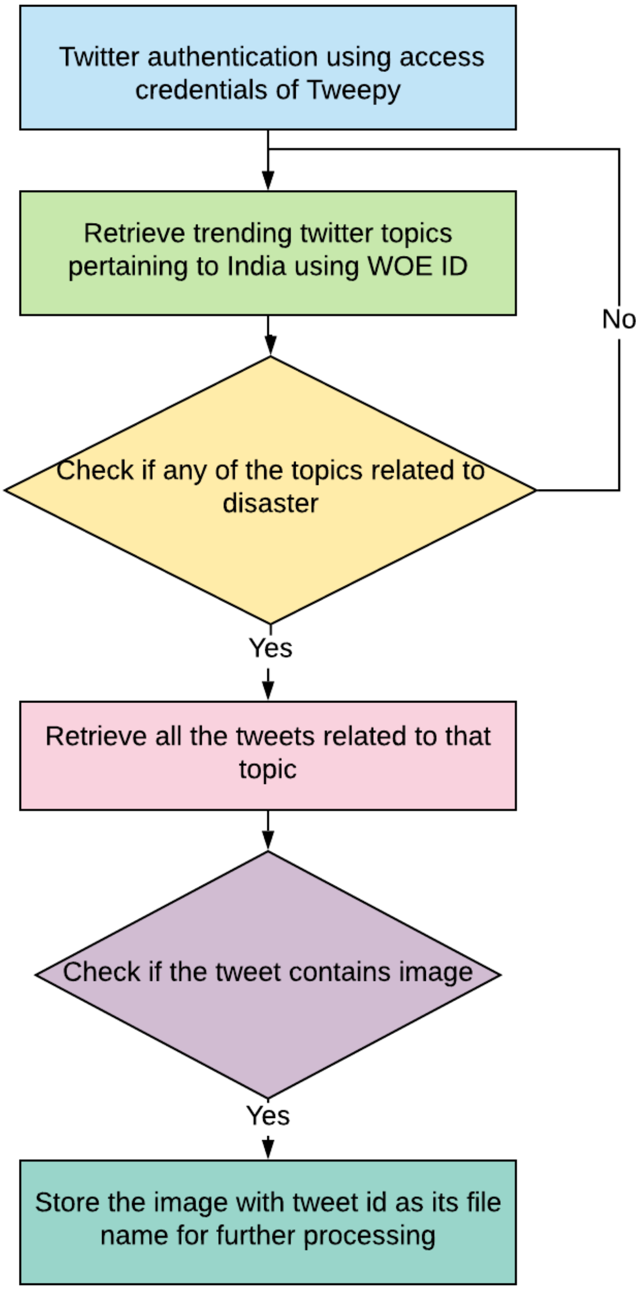

Twitter authentication using access credentials of Tweepy
Retrieve trending twitter topics pertaining to India using WOE ID
Check if any of the topics related to disaster
No
Yes
Retrieve all the tweets related to that topic
Check if the tweet contains image
Yes
Store the image with tweet id as its file name for further processing

FIGURE 4.2: Flowchart Depicting Live Tweet Extraction

The stepwise implementation of the algorithm is stated as follows:

1. The necessary headers like tweepy,json,csv are imported.
2. The tweet access is authenticated by providing the consumer_key , consumer_secret , access_token , access_token_secret of the registered twitter user.
3. The trending hashtags of India is collected by providing its WOE_ID i.e. Where On Earth Identifier.
4. The list of hashtags are analysed for the presence of a disaster using keywords like avalanche, volcano, earthquake, disaster, tsunami, tornado, drought, storm, flood, tremor.
5. If a disaster is found, the corresponding tweets (both text and image) are extracted by specifying the hashtag. The images are detected using media URL stored with tweet ids as filename.

Algorithm 2: Live Tweets Extraction

This collected data is given as input to next module.

4.1.1 Data Pre-Processing

The stored tweets in the previous module are then preprocessed. Several preprocessing tasks were performed on the input data using the tweet preprocessor API.

The tweet text were converted to uniform lower case characters. The URLs, SMILEYs/EMOJIs and USER MENTIONS (@) were cleaned off from the text.

Repetition of tweets were avoided using by removing retweets (RT). The word from hashtags were retained by the removing the ’#’ prefix from #word.

The HTMLParser module was used to parse the text file formatted in HTML and XHTML which includes changing of escape sequences like &amp; to &. Further, the tweet texts were normalised changing the short forms like ’re, ’nt, ’ve to are, not, have, etc.

Finally, the non-ascii and special characters were removing from the text. This preprocessed text is provided as input for the LSTM module for text classification as specified in the forthcoming section.

4.2 Classification

In this module, the text classification is performed using LSTM model and the image classification is done using Convolutional Neural Network.

4.2.1 Text Classification Using LSTM

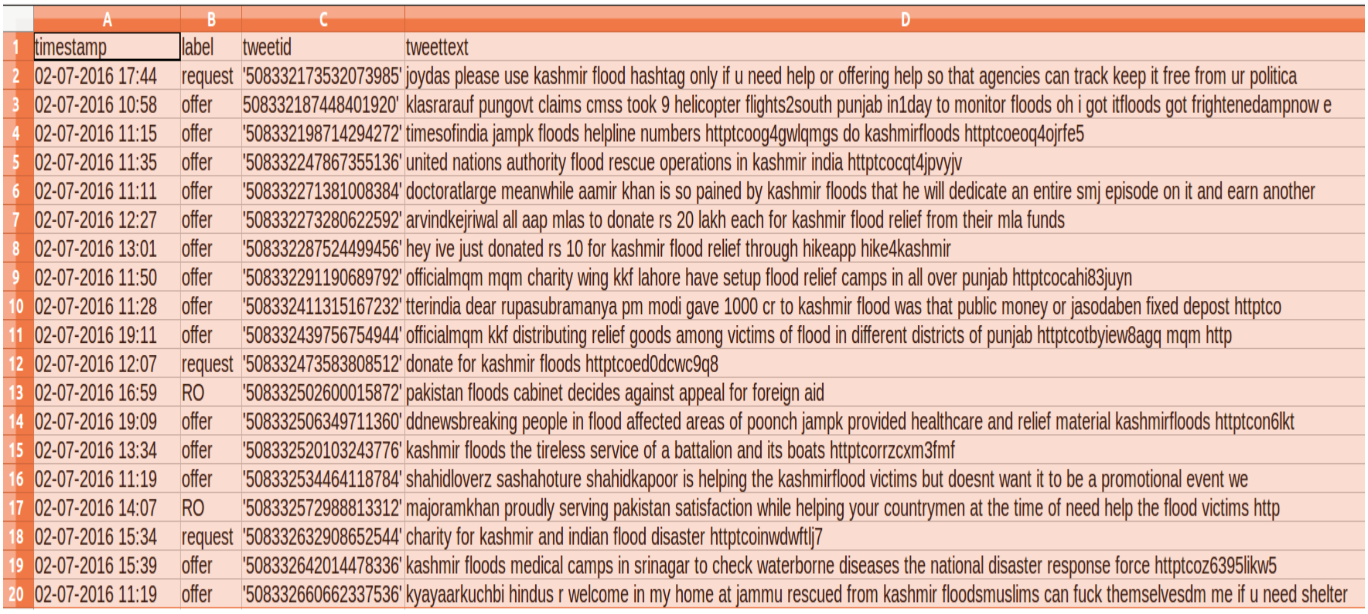
The Text Classification phase segregates the extracted data into three categories namely:

- Request—Disaster victims and humanitarian organizations requesting for help
- Offer—NGOs and relief measure organizations ready to offer their help
- None — The tweet is neither a request nor an offer.

4.2.1.1 Dataset

The dataset for training was obtained from a resource published by Quatar Computing Research Institute(QCRI) group called CRISISNLP under the title Human-annotated Twitter Corpora for NLP of Crisis-related Messages[11].

This dataset[11] contains human annotated data for 10 different disasters that took place between 2013 and 2015 having more than 2000 labelled tweets each created by crowdsource workers. It was created for the purpose of classifying disaster data using machine learning classifiers like SVM, Random forest(RF), Naive Bayes(NB).



	A	B	C	D
1	timestamp	label	tweetid	tweettext
2	02-07-2016 17:44	request	'508332173532073985'	joydas please use kashmir flood hashtag only if u need help or offering help so that agencies can track keep it free from ur politica
3	02-07-2016 10:58	offer	508332187448401920'	klasrarauf pungovt claims cmss took 9 helicopter flights2south punjab in1day to monitor floods oh i got itfloods got frightenedampnow e
4	02-07-2016 11:15	offer	'508332198714294272'	timesofindia jampk floods helpline numbers httptcoog4gwlqmgs do kashmirfloods httptcoeoq4ojrfe5
5	02-07-2016 11:35	offer	'508332247867355136'	united nations authority flood rescue operations in kashmir india httptcocqt4jpvyjv
6	02-07-2016 11:11	offer	'508332271381008384'	doctoratlarge meanwhile aamir khan is so pained by kashmir floods that he will dedicate an entire smj episode on it and earn another
7	02-07-2016 12:27	offer	'508332273280622592'	arvindkejriwal all aap mlas to donate rs 20 lakh each for kashmir flood relief from their mla funds
8	02-07-2016 13:01	offer	'508332287524499456'	hey ive just donated rs 10 for kashmir flood relief through hikeapp hike4kashmir
9	02-07-2016 11:50	offer	'508332291190689792'	officialmqm mqm charity wing kkf lahore have setup flood relief camps in all over punjab httptcocahi83juyn
10	02-07-2016 11:28	offer	'508332411315167232'	tterindia dear rupasubramanya pm modi gave 1000 cr to kashmir flood was that public money or jasodaben fixed depost httptco
11	02-07-2016 19:11	offer	'508332439756754944'	officialmqm kkf distributing relief goods among victims of flood in different districts of punjab httptcotbyiew8agq mqm http
12	02-07-2016 12:07	request	'508332473583808512'	donate for kashmir floods httptcoed0dcwc9q8
13	02-07-2016 16:59	RO	'508332502600015872'	pakistan floods cabinet decides against appeal for foreign aid
14	02-07-2016 19:09	offer	'508332506349711360'	ddnewsbreaking people in flood affected areas of poonch jampk provided healthcare and relief material kashmirfloods httptcon6lkt
15	02-07-2016 13:34	offer	'508332520103243776'	kashmir floods the tireless service of a battalion and its boats httptcorrzcxm3fmf
16	02-07-2016 11:19	offer	'508332534464118784'	shahidloverz sashahoture shahidkapoor is helping the kashmirflood victims but doesnt want it to be a promotional event we
17	02-07-2016 14:07	RO	'508332572988813312'	majoramkhan proudly serving pakistan satisfaction while helping your countrymen at the time of need help the flood victims http
18	02-07-2016 15:34	request	'508332632908652544'	charity for kashmir and indian flood disaster httptcoinwdwftlj7
19	02-07-2016 15:39	offer	'508332642014478336'	kashmir floods medical camps in srinagar to check waterborne diseases the national disaster response force httptcoz6395likw5
20	02-07-2016 11:19	offer	'508332660662337536'	kyayaarkuchbi hindus r welcome in my home at jammu rescued from kashmir floodsmuslims can fuck themselvesdm me if u need shelter

FIGURE 4.3: Sample text dataset

It classified the data into 9 categories:

1. Injured or dead people
2. Missing, trapped, or found people
3. Displaced people and evacuations
4. Infrastructure and utilities damage

5. Donation needs or offers or volunteering services
6. Caution and advice
7. Sympathy and emotional support
8. Other useful information
9. Not related or irrelevant

Out of these nine categories, ‘Donation needs or offers or volunteering services’ alone were scrapped out for our purpose.

In order to handle sequential data, recurrent neural network (RNN) is implemented here. Due to the problem of vanishing gradients, Long Short Term Memory(LSTM) model is constructed. Around 6000 human annotated request and offer tweets were utilized to build and test the model. 80% of the dataset was used for training (precisely 4425 tweets) and the remaining 20% for testing.

4.2.1.2 Text Classification Model Architecture using LSTM (Variation 1)

An initial LSTM architecture model was built as shown in Fig 4.4 with a one embedding Layer followed by a uni-directional LSTM layer and then a fully connected Dense layer.

4.2.1.3 Text Classification Model Architecture using LSTM (Variation 2)

The performance of the initial model was improved by making the certain changes to the architecture. These changes are tabulated in Table 4.1.

This improved model contains an intial pretrained glove embedding layer, two stacked bidirectional LSTM layers followed by a dense fully connected layer.

The model plot of our LSTM model II is illustrated as shown in Fig.4.5.

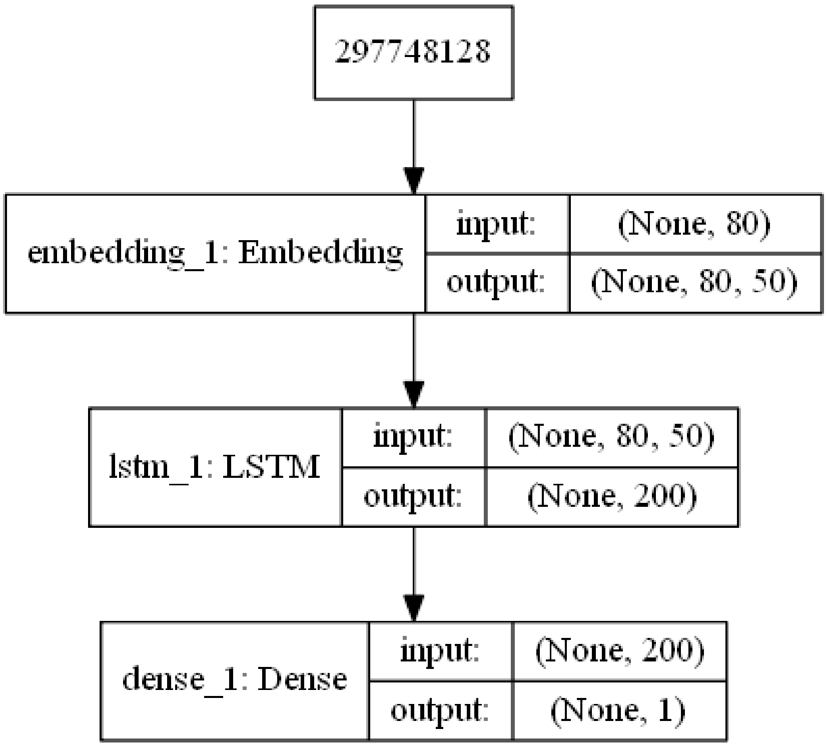


FIGURE 4.4: Text Classification Model Architecture using LSTM (Variation 1)

TABLE 4.1: Improvement in the LSTM model

	VARIATION 1	**VARIATION 2**
1.Preprocessing	Basic text preprocessing	Improved preprocessing using tweet preprocessor package.
2.Vocabulary Size	Fixed as a constant value-20000	Set as the number of unique words in the vocabulary
3.Embedding	A seperate layer for creating the word embeddings	Word vectors are created using PRETRAINED GLOVE EMBEDDINGS
4.Padding	A constant padding length as 50 was set.	Fixed as the maximum number of words appearing in a tweet in the dataset(=30).
5.LSTM	Single layer Unidirectional LSTM - Learning from past to future	stacked Bidirectional LSTM

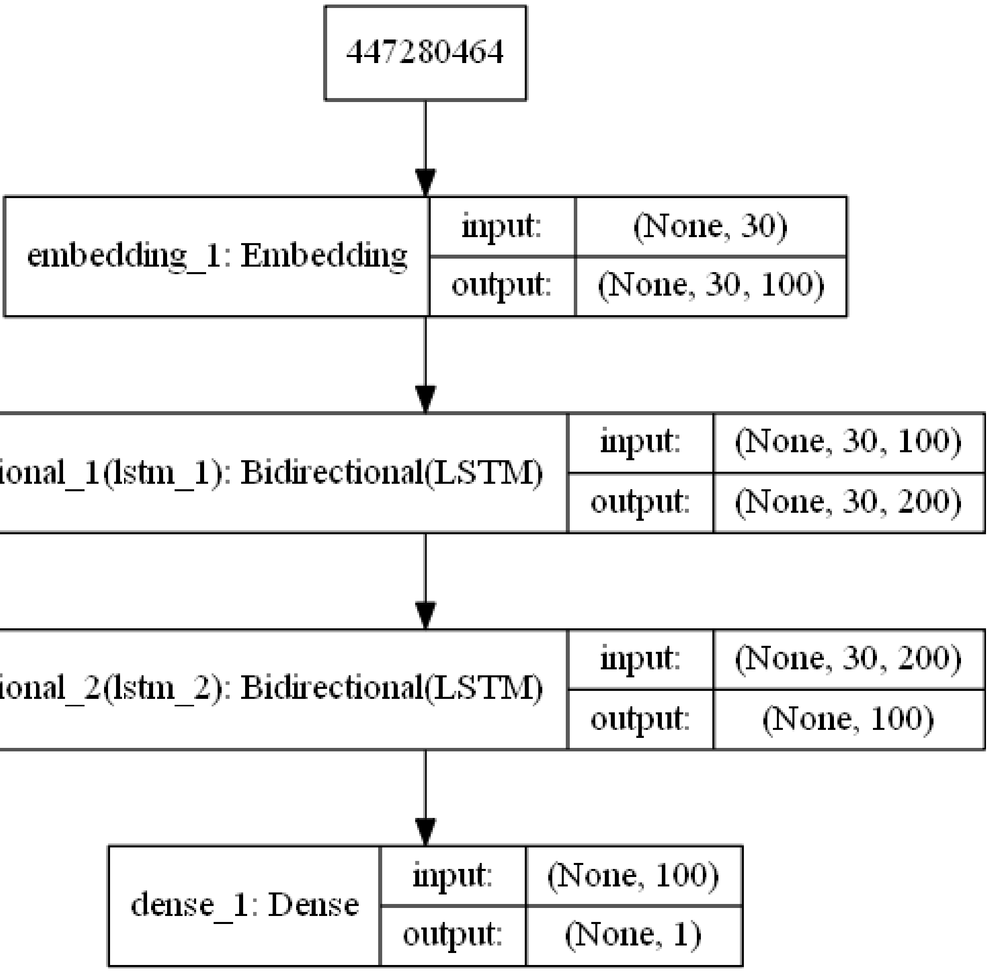


FIGURE 4.5: Text Classification Model Architecture using LSTM (Variation 2)

4.2.2 Image Classification Using CNN

The image classification phase classifies the input image into three categories, providing insight into the gravity of the damage inflicted by the disaster. The three classes are : 1. Mild, 2. Severe, 3. None

4.2.2.1 Dataset

The dataset for training the images for classifying as Severe, Mild or None was obtained from a resource cited in Damage Assessment from Social Media

Imagery data During Disasters. This dataset[2] is developed for the purpose of assessing the damage during disasters using social media images collected during 4 natural disasters like Typhoon Ruby(2014), Nepal Earthquake(2015), Eucador Earthquake(2016), Hurricane Mathew(2016) and google images. The infrastructure and utility damage images(relevant) are classified as severe,mild or little or no damage categories. It has a total of 25820 such images. The image dataset comprises of 19703 images out of which 15258 images were used for training and the remaining 3815 images were used for testing i.e. 80% training data and 20% testing data.

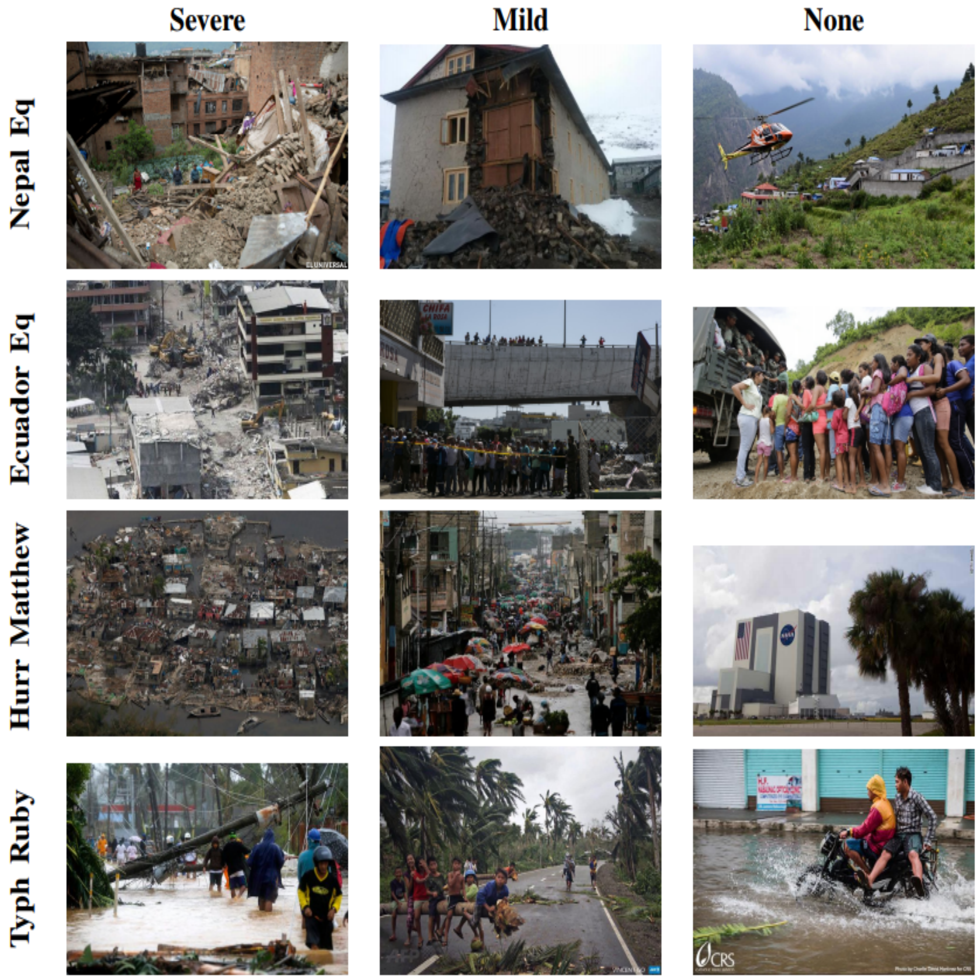
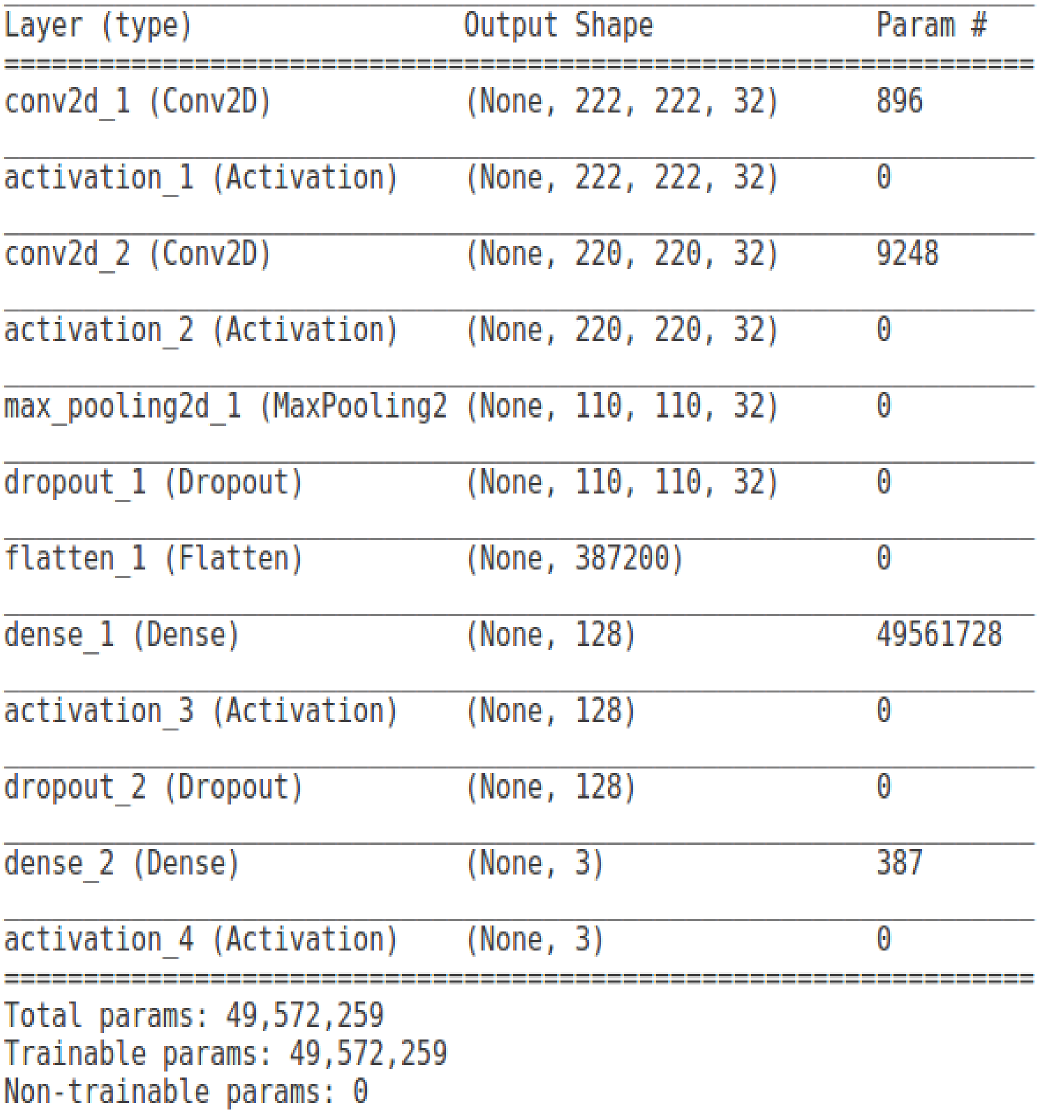


FIGURE 4.6: Image Dataset Sample

4.2.2.2 Initial model

The input images were resized to a common, fixed size (224x224). An initial basic CNN model was designed with two layers of convolution. The architectural summary of this model is shown in Figure 4.7.



```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 222, 222, 32)      896
_________________________________________________________________
activation_1 (Activation)    (None, 222, 222, 32)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 220, 220, 32)      9248
_________________________________________________________________
activation_2 (Activation)    (None, 220, 220, 32)      0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 110, 110, 32)      0
_________________________________________________________________
dropout_1 (Dropout)          (None, 110, 110, 32)      0
_________________________________________________________________
flatten_1 (Flatten)          (None, 387200)            0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               49561728
_________________________________________________________________
activation_3 (Activation)    (None, 128)               0
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 387
_________________________________________________________________
activation_4 (Activation)    (None, 3)                 0
=================================================================
Total params: 49,572,259
Trainable params: 49,572,259
Non-trainable params: 0
```

FIGURE 4.7: Image Classification Model Architecture using CNN

In an attempt to improve the performance of the model, an alternate pre-trained VGG-16 model was fine tuned using transfer learning technique.

4.2.2.3 Transfer learning using VGG-16

CNNs are rarely trained from scratch for new datasets because state-of-the-art CNNs (i) are getting deeper everyday, and (ii) require larger datasets to train on. However, collecting large datasets for the particular problem at hand is usually hard in practice (as in the current study). Therefore, it is common to devise new techniques based on pre-trained networks.

A popular approach is to use the existing weights of a pre-trained CNN as an initialization for the new dataset, which is often referred to as fine-tuning. In this transfer-learning approach, the last layer of the network is adapted to the task at hand (i.e., number of categories in the softmax layer and sometimes even the loss function) and the pre-trained network is fine-tuned according to the training images from the new dataset. This approach allows us to transfer the features and the parameters of the network from the broad domain (i.e., large scale image classification) to the specific one (i.e., disaster image analysis).

We adapted the VGG-16 network pre-trained on the ImageNet dataset to classify the images in our disaster image dataset into one of the three damage categories namely severe, mild, and none. The VGG-16 network trained on the ImageNet dataset using over 1.2M images and 1000 categories. The VGG-16 network consists of 16 layers and around 140 million weight parameters.

The weights until the block5_pool layer were retained and these layers were flagged False for training. A Flatten layer followed by two dense fully connected layers were added and finally a dense Softmax layer with three outputs was included. Features were computed by forward propagating a 224224 RGB image through thirteen convolutional, five max pooling layer, two fully connected dense layers and finally the softmax layer.

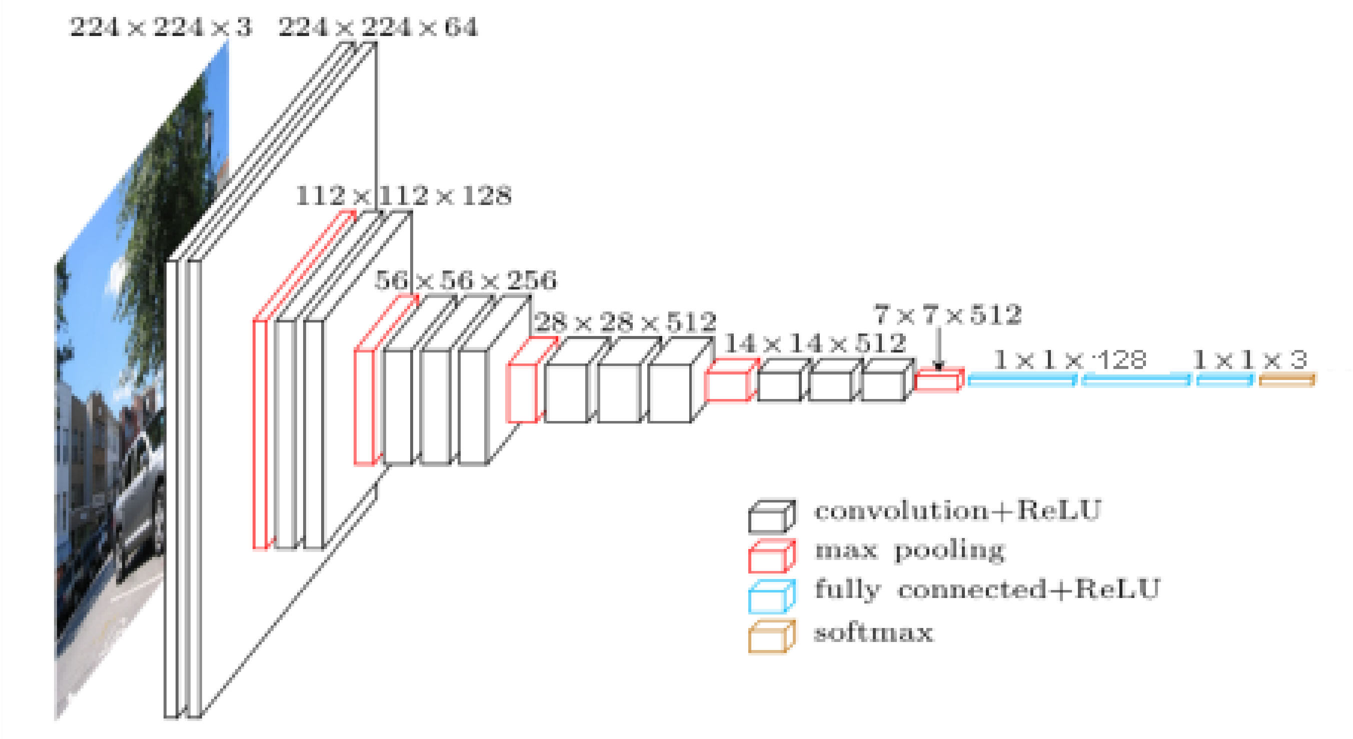
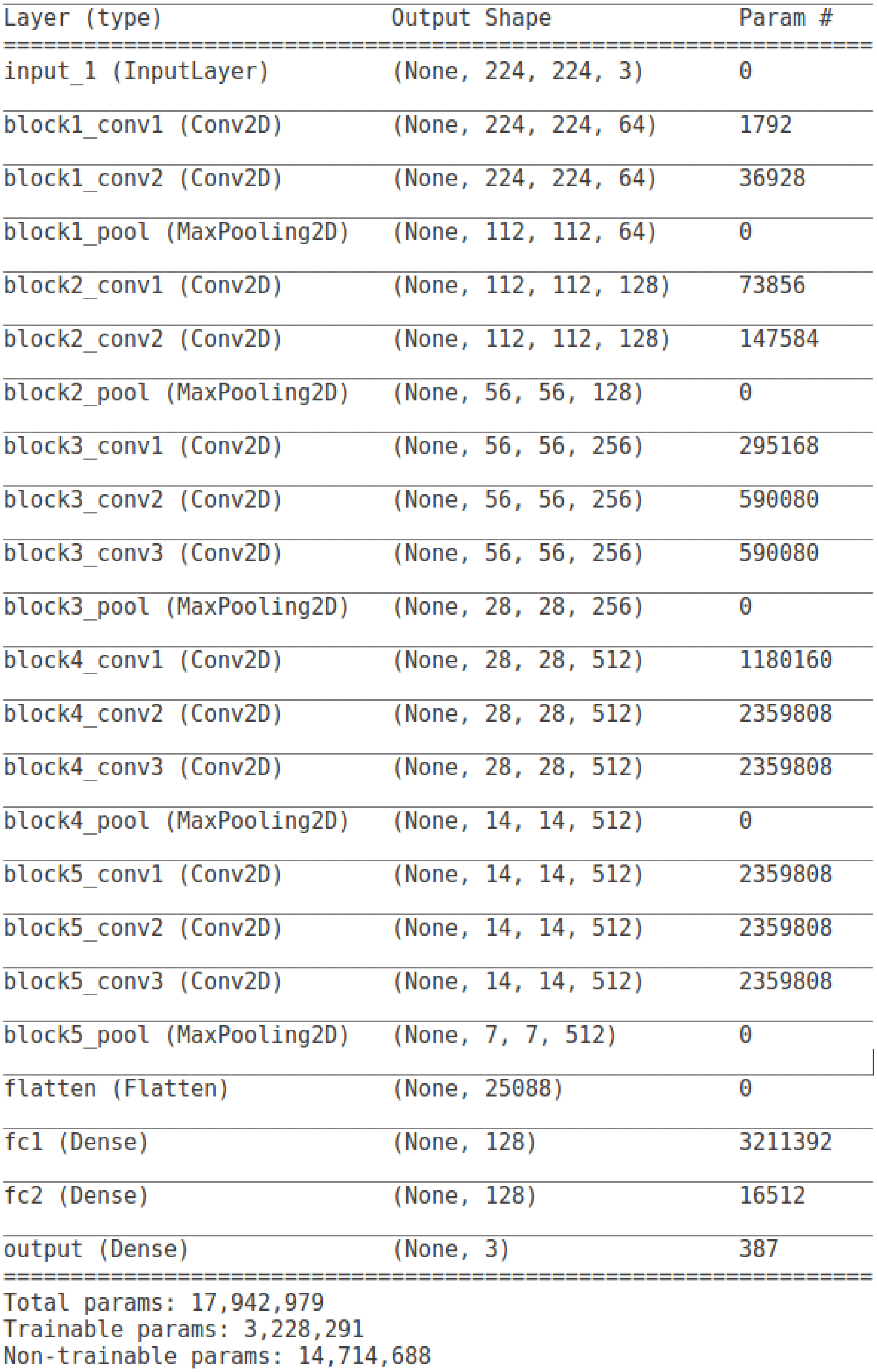


FIGURE 4.8: Architecture Diagram of Modified VGG-16 Source:https://www.cs.toronto.edu/ frossard/post/vgg16/

The Image classification model architecture using CNN is shown in Fig.4.8. The Architectural summary of the model is given in Fig. 4.9.



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 224, 224, 3)       0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, 224, 224, 64)      1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, 224, 224, 64)      36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, 112, 112, 64)      0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, 112, 112, 128)     73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, 112, 112, 128)     147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, 56, 56, 128)       0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, 56, 56, 256)       295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, 56, 56, 256)       590080
_________________________________________________________________
block3_conv3 (Conv2D)        (None, 56, 56, 256)       590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, 28, 28, 256)       0
_________________________________________________________________
block4_conv1 (Conv2D)        (None, 28, 28, 512)       1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 14, 14, 512)       0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 7, 7, 512)         0
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
fc1 (Dense)                  (None, 128)               3211392
_________________________________________________________________
fc2 (Dense)                  (None, 128)               16512
_________________________________________________________________
output (Dense)               (None, 3)                 387
=================================================================
Total params: 17,942,979
Trainable params: 3,228,291
Non-trainable params: 14,714,688
```

FIGURE 4.9: Image Classification Model Architecture using Modified VGG-16

4.3 Matching Request To Offer

The tweets and severity measure obtained as the result of the previous text and image classification modules respectively, along with the predicted label are fed as input to this module. The severity measure of image associated with requests are used for prioritization. The mapping is done one offer to one request and one offer to many requests.

4.3.1 One to One Mapping

The algorithm of the one to one matching process is illustrated as shown:

```
Input: requests,offers,image_severity
Output: matched_tweets
maxm = threshold_value
for o ←offers do
    for r ←requests do
        measure = similarity(r,o);
        if measure $\geq$maxm then
            if abs(measure-maxm) $\leq$0.01 then
                matched_request ← max_severity(r, o);
            else
                no updation of matched request;
            end
        else
            continue;
        end
    end
    display(matched_request,o);
    remove_from_list(matched_request);
end
```

Algorithm 3: One to One Matching Algorithm

4.3.2 One to Many Mapping

Since there could arise a situation wherein a single offer can satisfy multiple requests, one offer to many requests mapping is performed. This mapping is done by considering top five most similar request for the current offer tweet based on similarity measure. These are further sorted based on the severity measure of the images. This prioritizes similar tweets.

```
Input: requests,offers,image_severity
Output: matched_tweets
maxm = threshold_value
for o ←offers do
    for r ←requests do
        measure = similarity(r,o);
        if measure ≥maxm then
            request_list ←r
        end
        top_5 = sort_desc(request_list,measure);
        matched_request = prioritize(top_5,image_severity);
    end
    display(matched_request,o);
end
```

Algorithm 4: One to Many Matching Algorithm

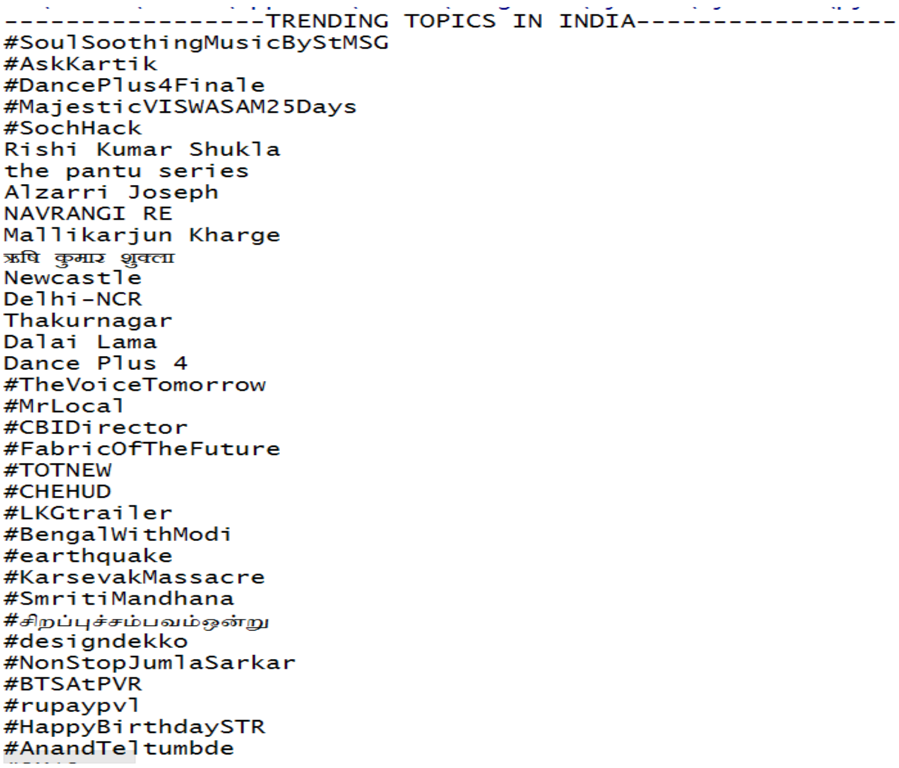
CHAPTER 5

IMPLEMENTATION RESULTS

In this section, the implementation results and their performance comparisons are presented in detail.

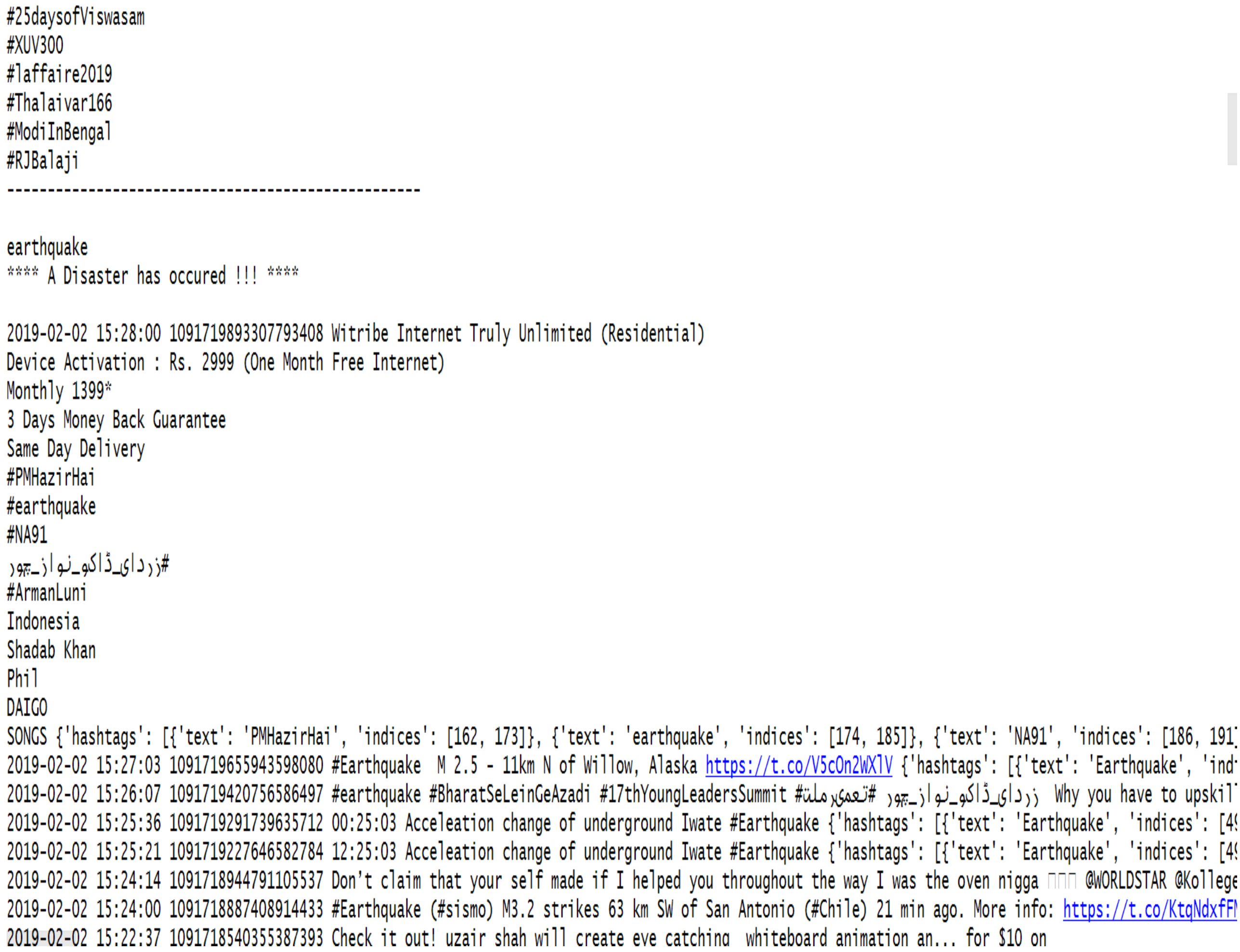
5.1 Live Tweet Extraction

Since the system operates on real-time tweets posted on twitter by the authenticated user, the trending tweets are extracted based on geographic location specified by WOE_ID i.e Where On Earth Identifier.



```
-----------------TRENDING TOPICS IN INDIA-----------------
#SoulSoothingMusicByStMSG
#AskKartik
#DancePlus4Finale
#MajesticVISWASAM25Days
#SochHack
Rishi Kumar Shukla
the pantu series
Alzarri Joseph
NAVRANGI RE
Mallikarjun Kharge
ऋषि कुमार शुक्ला
Newcastle
Delhi-NCR
Thakurnagar
Dalai Lama
Dance Plus 4
#TheVoiceTomorrow
#MrLocal
#CBIDirector
#FabricOfTheFuture
#TOTNEW
#CHEHUD
#LKGtrailer
#BengalWithModi
#earthquake
#KarsevakMassacre
#SmritiMandhana
#சிறப்புச்சம்பவம்ஒன்று
#designdekko
#NonStopJumlaSarkar
#BTSAtPVR
#rupaypvl
#HappyBirthdaySTR
#AnandTeltumbde
```

FIGURE 5.1: Trending Hashtags In India

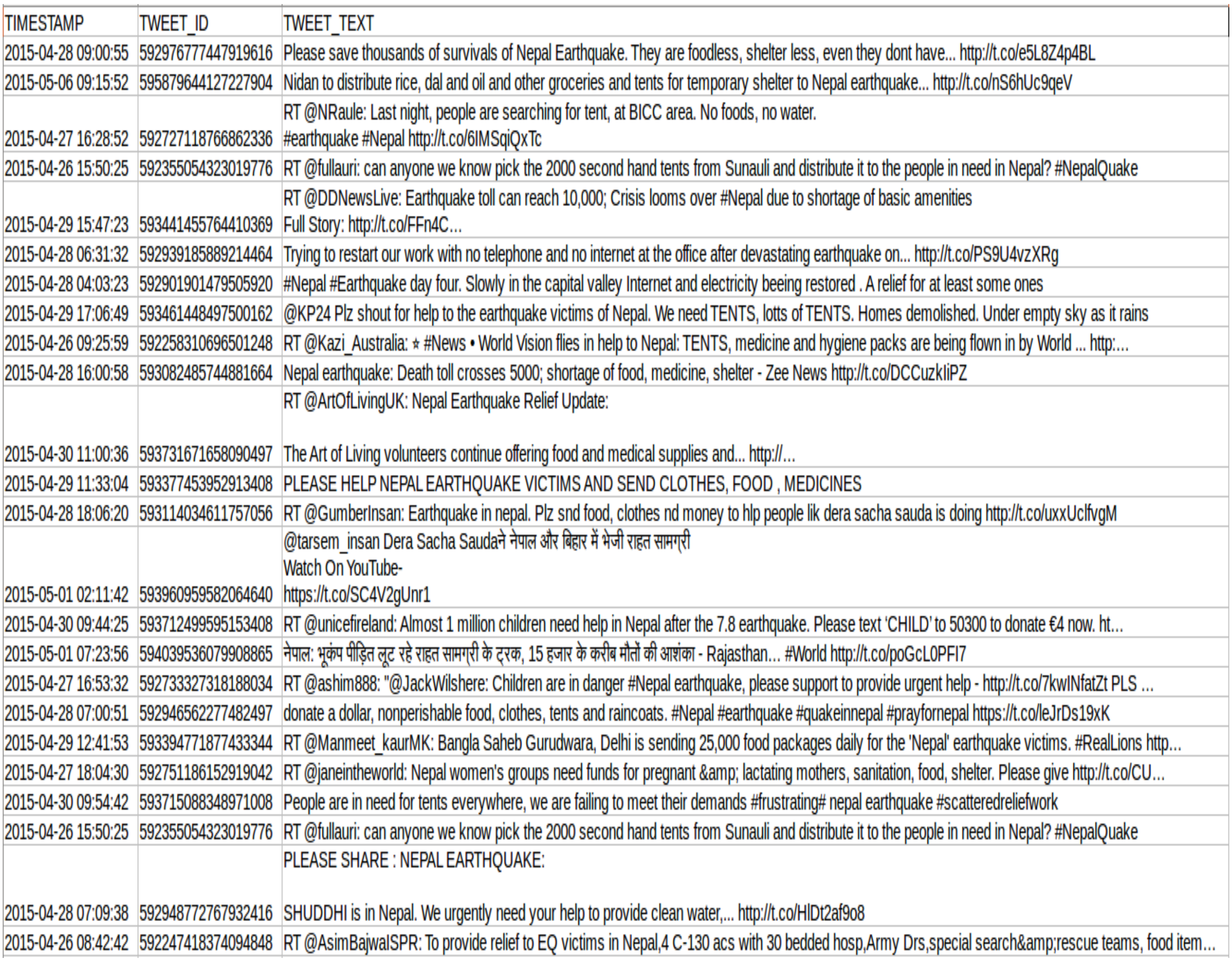


```
#25daysofViswasam
#XUV300
#laffaire2019
#Thalaivar166
#ModiInBengal
#RJBalaji
---------------------------------------------------

earthquake
**** A Disaster has occured !!! ****

2019-02-02 15:28:00 1091719893307793408 Witribe Internet Truly Unlimited (Residential)
Device Activation : Rs. 2999 (One Month Free Internet)
Monthly 1399*
3 Days Money Back Guarantee
Same Day Delivery
#PMHazirHai
#earthquake
#NA91
#زردای_ڈاکو_نواز_چور
#ArmanLuni
Indonesia
Shadab Khan
Phil
DAIGO
SONGS {'hashtags': [{'text': 'PMHazirHai', 'indices': [162, 173]}, {'text': 'earthquake', 'indices': [174, 185]}, {'text': 'NA91', 'indices': [186, 191
2019-02-02 15:27:03 1091719655943598080 #Earthquake  M 2.5 - 11km N of Willow, Alaska https://t.co/V5cOn2WXlV {'hashtags': [{'text': 'Earthquake', 'ind
2019-02-02 15:26:07 1091719420756586497 #earthquake #BharatSeLeinGeAzadi #17thYoungLeadersSummit #تعمیرملت #زردای_ڈاکو_نواز_چور  Why you have to upskil
2019-02-02 15:25:36 1091719291739635712 00:25:03 Acceleation change of underground Iwate #Earthquake {'hashtags': [{'text': 'Earthquake', 'indices': [4
2019-02-02 15:25:21 1091719227646582784 12:25:03 Acceleation change of underground Iwate #Earthquake {'hashtags': [{'text': 'Earthquake', 'indices': [4
2019-02-02 15:24:14 1091718944791105537 Don't claim that your self made if I helped you throughout the way I was the oven nigga □□□ @WORLDSTAR @Kollege
2019-02-02 15:24:00 1091718887408914433 #Earthquake (#sismo) M3.2 strikes 63 km SW of San Antonio (#Chile) 21 min ago. More info: https://t.co/KtqNdxfF
2019-02-02 15:22:37 1091718540355387393 Check it out! uzair shah will create eve catching  whiteboard animation an... for $10 on
```

FIGURE 5.2: Extracted Tweets During Disaster

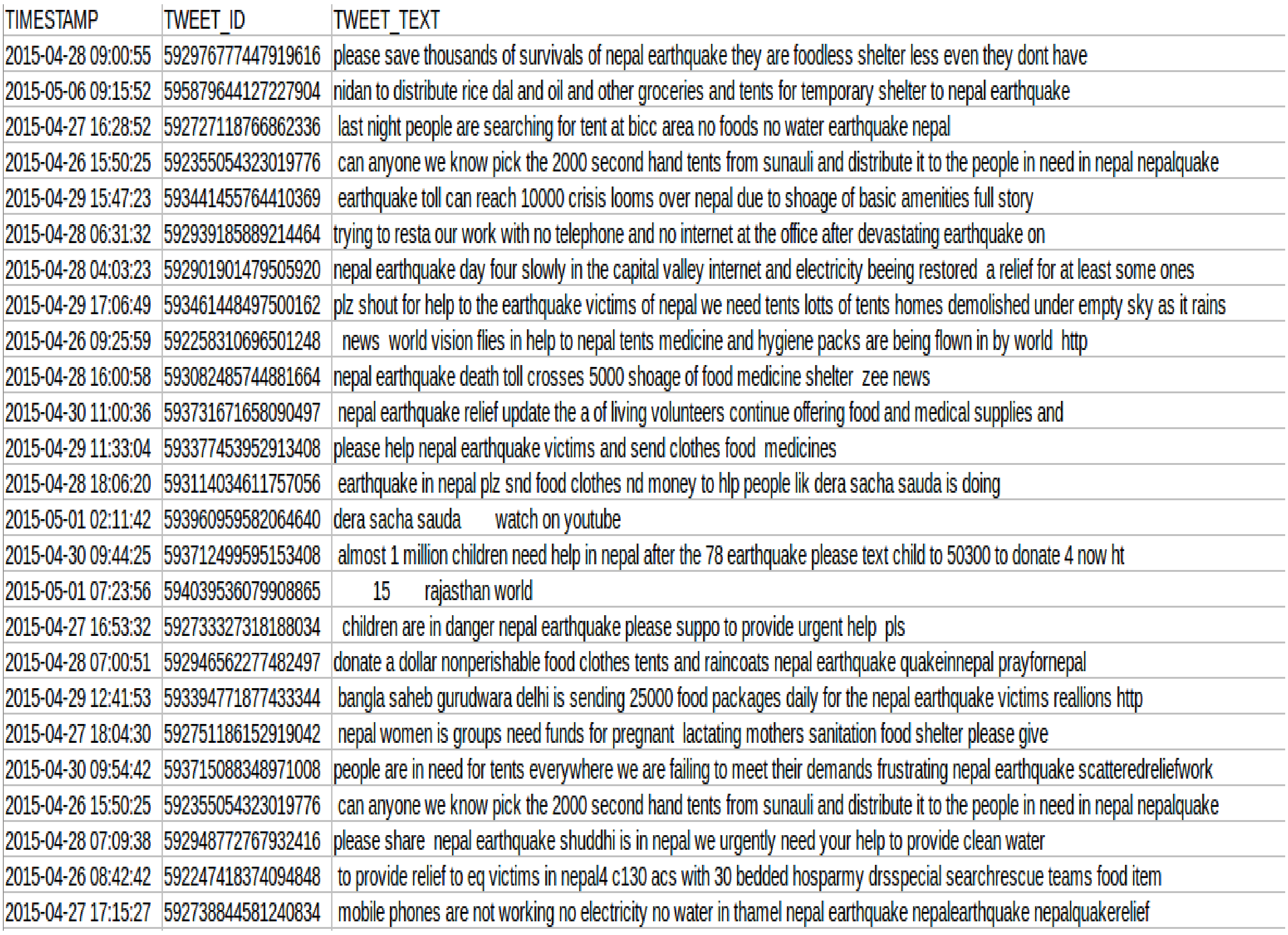
These trending hashtags are analysed to check for the presence of a disaster as shown in Fig 5.2.

5.2 Data Preprocessing



TIMESTAMP	TWEET_ID	TWEET_TEXT
2015-04-28 09:00:55	592976777447919616	Please save thousands of survivals of Nepal Earthquake. They are foodless, shelter less, even they dont have... http://t.co/e5L8Z4p4BL
2015-05-06 09:15:52	595879644127227904	Nidan to distribute rice, dal and oil and other groceries and tents for temporary shelter to Nepal earthquake... http://t.co/nS6hUc9qeV
2015-04-27 16:28:52	592727118766862336	RT @NRaule: Last night, people are searching for tent, at BICC area. No foods, no water. #earthquake #Nepal http://t.co/6IMSqiQxTc
2015-04-26 15:50:25	592355054323019776	RT @fullauri: can anyone we know pick the 2000 second hand tents from Sunauli and distribute it to the people in need in Nepal? #NepalQuake
2015-04-29 15:47:23	593441455764410369	RT @DDNewsLive: Earthquake toll can reach 10,000; Crisis looms over #Nepal due to shortage of basic amenities Full Story: http://t.co/FFn4C…
2015-04-28 06:31:32	592939185889214464	Trying to restart our work with no telephone and no internet at the office after devastating earthquake on... http://t.co/PS9U4vzXRg
2015-04-28 04:03:23	592901901479505920	#Nepal #Earthquake day four. Slowly in the capital valley Internet and electricity beeing restored . A relief for at least some ones
2015-04-29 17:06:49	593461448497500162	@KP24 Plz shout for help to the earthquake victims of Nepal. We need TENTS, lotts of TENTS. Homes demolished. Under empty sky as it rains
2015-04-26 09:25:59	592258310696501248	RT @Kazi_Australia: ⭑ #News • World Vision flies in help to Nepal: TENTS, medicine and hygiene packs are being flown in by World ... http:…
2015-04-28 16:00:58	593082485744881664	Nepal earthquake: Death toll crosses 5000; shortage of food, medicine, shelter - Zee News http://t.co/DCCuzkIiPZ
2015-04-30 11:00:36	593731671658090497	RT @ArtOfLivingUK: Nepal Earthquake Relief Update: The Art of Living volunteers continue offering food and medical supplies and... http://…
2015-04-29 11:33:04	593377453952913408	PLEASE HELP NEPAL EARTHQUAKE VICTIMS AND SEND CLOTHES, FOOD , MEDICINES
2015-04-28 18:06:20	593114034611757056	RT @GumberInsan: Earthquake in nepal. Plz snd food, clothes nd money to hlp people lik dera sacha sauda is doing http://t.co/uxxUclfvgM
2015-05-01 02:11:42	593960959582064640	@tarsem_insan Dera Sacha Saudaने नेपाल और बिहार में भेजी राहत सामग्री Watch On YouTube- https://t.co/SC4V2gUnr1
2015-04-30 09:44:25	593712499595153408	RT @unicefireland: Almost 1 million children need help in Nepal after the 7.8 earthquake. Please text 'CHILD' to 50300 to donate €4 now. ht…
2015-05-01 07:23:56	594039536079908865	नेपाल: भूकंप पीड़ित लूट रहे राहत सामग्री के ट्रक, 15 हजार के करीब मौतों की आशंका - Rajasthan... #World http://t.co/poGcL0PFI7
2015-04-27 16:53:32	592733327318188034	RT @ashim888: "@JackWilshere: Children are in danger #Nepal earthquake, please support to provide urgent help - http://t.co/7kwINfatZt PLS …
2015-04-28 07:00:51	592946562277482497	donate a dollar, nonperishable food, clothes, tents and raincoats. #Nepal #earthquake #quakeinnepal #prayfornepal https://t.co/leJrDs19xK
2015-04-29 12:41:53	593394771877433344	RT @Manmeet_kaurMK: Bangla Saheb Gurudwara, Delhi is sending 25,000 food packages daily for the 'Nepal' earthquake victims. #RealLions http…
2015-04-27 18:04:30	592751186152919042	RT @janeintheworld: Nepal women's groups need funds for pregnant &amp; lactating mothers, sanitation, food, shelter. Please give http://t.co/CU…
2015-04-30 09:54:42	593715088348971008	People are in need for tents everywhere, we are failing to meet their demands #frustrating# nepal earthquake #scatteredreliefwork
2015-04-26 15:50:25	592355054323019776	RT @fullauri: can anyone we know pick the 2000 second hand tents from Sunauli and distribute it to the people in need in Nepal? #NepalQuake
2015-04-28 07:09:38	592948772767932416	PLEASE SHARE : NEPAL EARTHQUAKE: SHUDDHI is in Nepal. We urgently need your help to provide clean water,... http://t.co/HlDt2af9o8
2015-04-26 08:42:42	592247418374094848	RT @AsimBajwalSPR: To provide relief to EQ victims in Nepal,4 C-130 acs with 30 bedded hosp,Army Drs,special search&amp;rescue teams, food item…

FIGURE 5.3: Tweets Before Preprocessing



TIMESTAMP	TWEET_ID	TWEET_TEXT
2015-04-28 09:00:55	592976777447919616	please save thousands of survivals of nepal earthquake they are foodless shelter less even they dont have
2015-05-06 09:15:52	595879644127227904	nidan to distribute rice dal and oil and other groceries and tents for temporary shelter to nepal earthquake
2015-04-27 16:28:52	592727118766862336	last night people are searching for tent at bicc area no foods no water earthquake nepal
2015-04-26 15:50:25	592355054323019776	can anyone we know pick the 2000 second hand tents from sunauli and distribute it to the people in need in nepal nepalquake
2015-04-29 15:47:23	593441455764410369	earthquake toll can reach 10000 crisis looms over nepal due to shoage of basic amenities full story
2015-04-28 06:31:32	592939185889214464	trying to resta our work with no telephone and no internet at the office after devastating earthquake on
2015-04-28 04:03:23	592901901479505920	nepal earthquake day four slowly in the capital valley internet and electricity beeing restored a relief for at least some ones
2015-04-29 17:06:49	593461448497500162	plz shout for help to the earthquake victims of nepal we need tents lotts of tents homes demolished under empty sky as it rains
2015-04-26 09:25:59	592258310696501248	news world vision flies in help to nepal tents medicine and hygiene packs are being flown in by world http
2015-04-28 16:00:58	593082485744881664	nepal earthquake death toll crosses 5000 shoage of food medicine shelter zee news
2015-04-30 11:00:36	593731671658090497	nepal earthquake relief update the a of living volunteers continue offering food and medical supplies and
2015-04-29 11:33:04	593377453952913408	please help nepal earthquake victims and send clothes food medicines
2015-04-28 18:06:20	593114034611757056	earthquake in nepal plz snd food clothes nd money to hlp people lik dera sacha sauda is doing
2015-05-01 02:11:42	593960959582064640	dera sacha sauda watch on youtube
2015-04-30 09:44:25	593712499595153408	almost 1 million children need help in nepal after the 78 earthquake please text child to 50300 to donate 4 now ht
2015-05-01 07:23:56	594039536079908865	15 rajasthan world
2015-04-27 16:53:32	592733327318188034	children are in danger nepal earthquake please suppo to provide urgent help pls
2015-04-28 07:00:51	592946562277482497	donate a dollar nonperishable food clothes tents and raincoats nepal earthquake quakeinnepal prayfornepal
2015-04-29 12:41:53	593394771877433344	bangla saheb gurudwara delhi is sending 25000 food packages daily for the nepal earthquake victims reallions http
2015-04-27 18:04:30	592751186152919042	nepal women is groups need funds for pregnant lactating mothers sanitation food shelter please give
2015-04-30 09:54:42	593715088348971008	people are in need for tents everywhere we are failing to meet their demands frustrating nepal earthquake scatteredreliefwork
2015-04-26 15:50:25	592355054323019776	can anyone we know pick the 2000 second hand tents from sunauli and distribute it to the people in need in nepal nepalquake
2015-04-28 07:09:38	592948772767932416	please share nepal earthquake shuddhi is in nepal we urgently need your help to provide clean water
2015-04-26 08:42:42	592247418374094848	to provide relief to eq victims in nepal4 c130 acs with 30 bedded hosparmy drsspecial searchrescue teams food item
2015-04-27 17:15:27	592738844581240834	mobile phones are not working no electricity no water in thamel nepal earthquake nepalearthquake nepalquakerelief

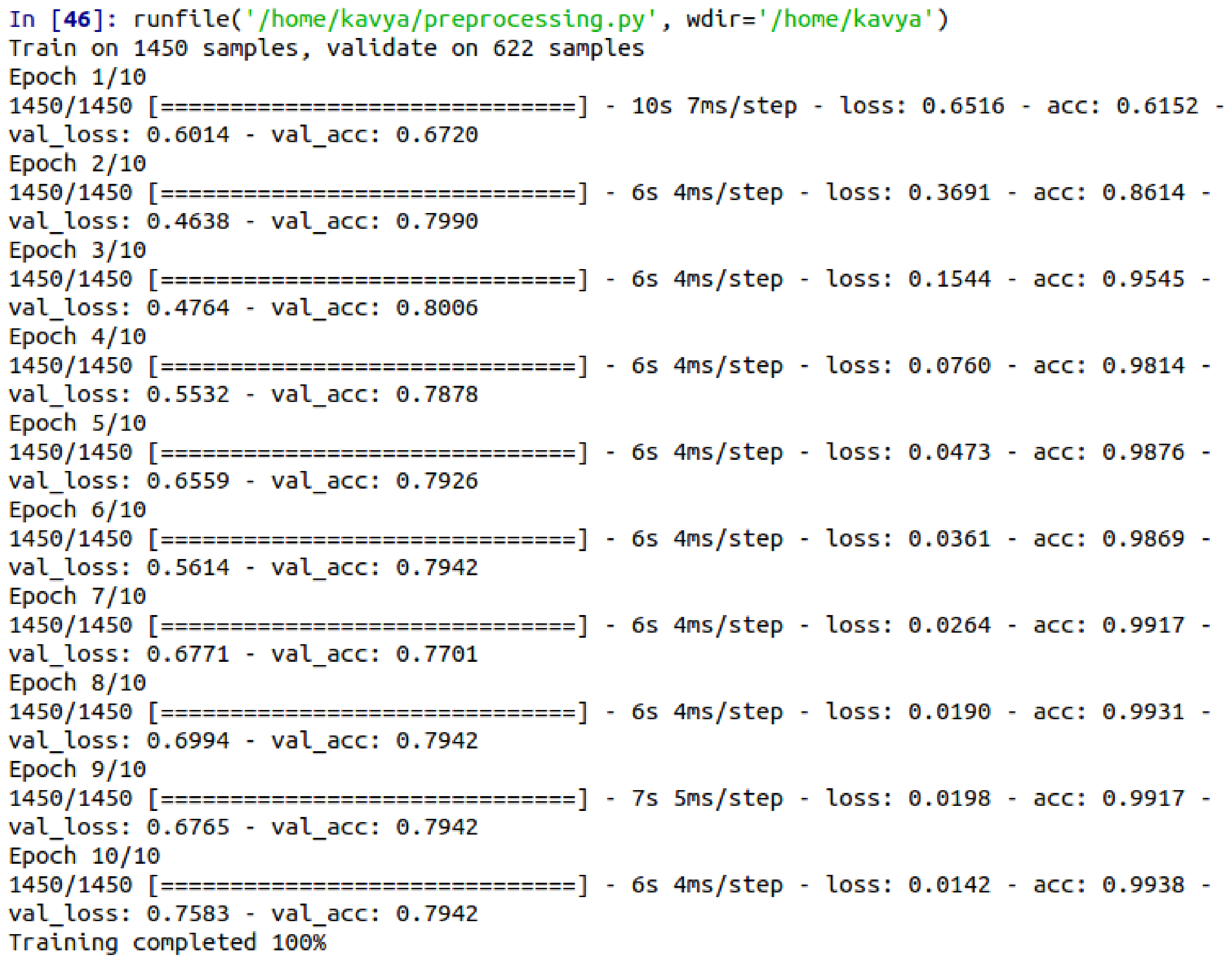
FIGURE 5.4: Tweets After Preprocessing

5.3 Classification

5.3.1 Text Classification

5.3.1.1 Training and Testing

The text classification model is trained as given in section 4. The training results for LSTM Architecture I and LSTM Architecture II are given in Fig 5.5 and Fig 5.6 respectively.



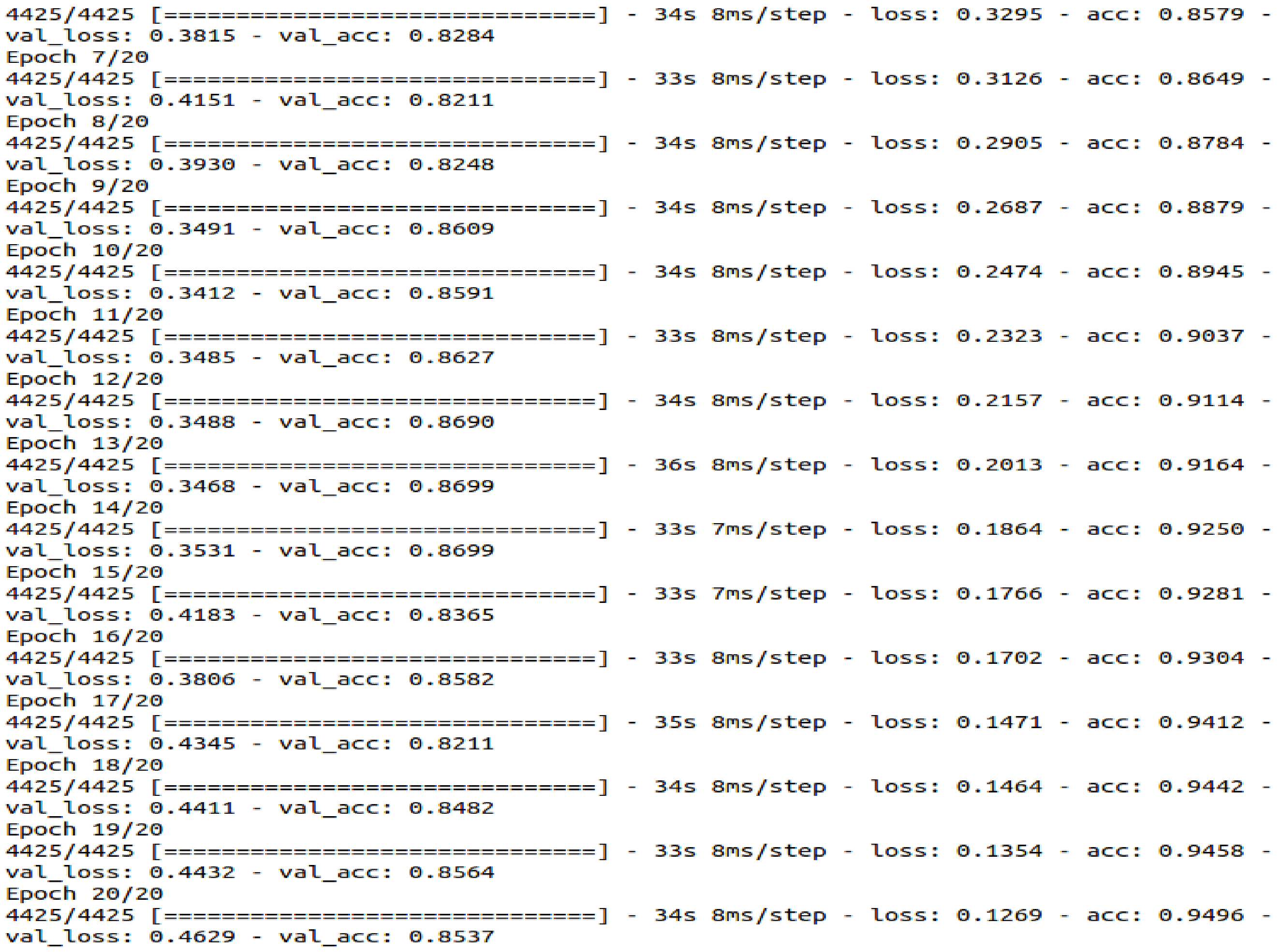
```
In [46]: runfile('/home/kavya/preprocessing.py', wdir='/home/kavya')
Train on 1450 samples, validate on 622 samples
Epoch 1/10
1450/1450 [==============================] - 10s 7ms/step - loss: 0.6516 - acc: 0.6152 -
val_loss: 0.6014 - val_acc: 0.6720
Epoch 2/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.3691 - acc: 0.8614 -
val_loss: 0.4638 - val_acc: 0.7990
Epoch 3/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.1544 - acc: 0.9545 -
val_loss: 0.4764 - val_acc: 0.8006
Epoch 4/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0760 - acc: 0.9814 -
val_loss: 0.5532 - val_acc: 0.7878
Epoch 5/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0473 - acc: 0.9876 -
val_loss: 0.6559 - val_acc: 0.7926
Epoch 6/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0361 - acc: 0.9869 -
val_loss: 0.5614 - val_acc: 0.7942
Epoch 7/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0264 - acc: 0.9917 -
val_loss: 0.6771 - val_acc: 0.7701
Epoch 8/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0190 - acc: 0.9931 -
val_loss: 0.6994 - val_acc: 0.7942
Epoch 9/10
1450/1450 [==============================] - 7s 5ms/step - loss: 0.0198 - acc: 0.9917 -
val_loss: 0.6765 - val_acc: 0.7942
Epoch 10/10
1450/1450 [==============================] - 6s 4ms/step - loss: 0.0142 - acc: 0.9938 -
val_loss: 0.7583 - val_acc: 0.7942
Training completed 100%
```

FIGURE 5.5: Training Result for Text Classification Model (Variation 1)

5.3.1.2 Accuracy Comparison

The testing accuracy of Model I is 79.42%. The changes that were made as per Table 4.1 led to the improvement in the overall accuracy of the model reaching as high as 85.37% testing accuracy.

The training vs testing accuracy of the two models are graphical represented as shown in Fig 5.7.



```
4425/4425 [==============================] - 34s 8ms/step - loss: 0.3295 - acc: 0.8579 -
val_loss: 0.3815 - val_acc: 0.8284
Epoch 7/20
4425/4425 [==============================] - 33s 8ms/step - loss: 0.3126 - acc: 0.8649 -
val_loss: 0.4151 - val_acc: 0.8211
Epoch 8/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.2905 - acc: 0.8784 -
val_loss: 0.3930 - val_acc: 0.8248
Epoch 9/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.2687 - acc: 0.8879 -
val_loss: 0.3491 - val_acc: 0.8609
Epoch 10/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.2474 - acc: 0.8945 -
val_loss: 0.3412 - val_acc: 0.8591
Epoch 11/20
4425/4425 [==============================] - 33s 8ms/step - loss: 0.2323 - acc: 0.9037 -
val_loss: 0.3485 - val_acc: 0.8627
Epoch 12/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.2157 - acc: 0.9114 -
val_loss: 0.3488 - val_acc: 0.8690
Epoch 13/20
4425/4425 [==============================] - 36s 8ms/step - loss: 0.2013 - acc: 0.9164 -
val_loss: 0.3468 - val_acc: 0.8699
Epoch 14/20
4425/4425 [==============================] - 33s 7ms/step - loss: 0.1864 - acc: 0.9250 -
val_loss: 0.3531 - val_acc: 0.8699
Epoch 15/20
4425/4425 [==============================] - 33s 7ms/step - loss: 0.1766 - acc: 0.9281 -
val_loss: 0.4183 - val_acc: 0.8365
Epoch 16/20
4425/4425 [==============================] - 33s 8ms/step - loss: 0.1702 - acc: 0.9304 -
val_loss: 0.3806 - val_acc: 0.8582
Epoch 17/20
4425/4425 [==============================] - 35s 8ms/step - loss: 0.1471 - acc: 0.9412 -
val_loss: 0.4345 - val_acc: 0.8211
Epoch 18/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.1464 - acc: 0.9442 -
val_loss: 0.4411 - val_acc: 0.8482
Epoch 19/20
4425/4425 [==============================] - 33s 8ms/step - loss: 0.1354 - acc: 0.9458 -
val_loss: 0.4432 - val_acc: 0.8564
Epoch 20/20
4425/4425 [==============================] - 34s 8ms/step - loss: 0.1269 - acc: 0.9496 -
val_loss: 0.4629 - val_acc: 0.8537
```

FIGURE 5.6: Training Result for Text Classification Model (Variation 2)

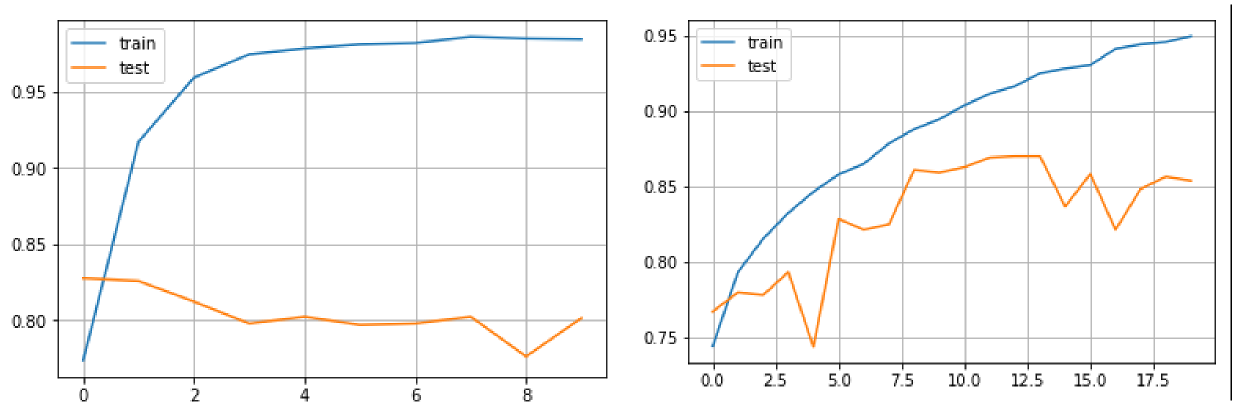
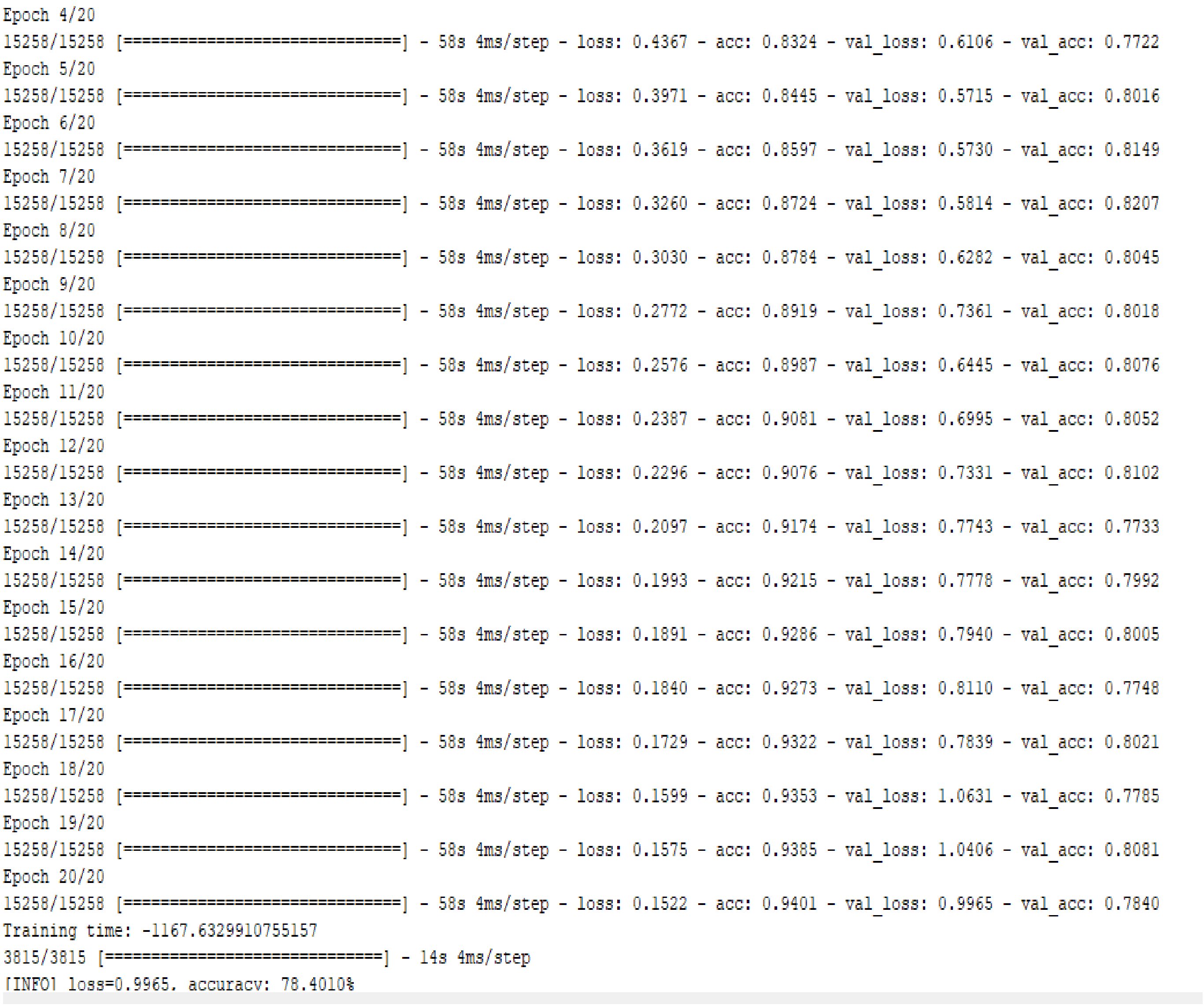


FIGURE 5.7: Comparison Of Training And Testing Accuracy Of LSTM Variations 1 And 2 Respectively.

5.3.2 Image Classification

5.3.2.1 Training and Testing

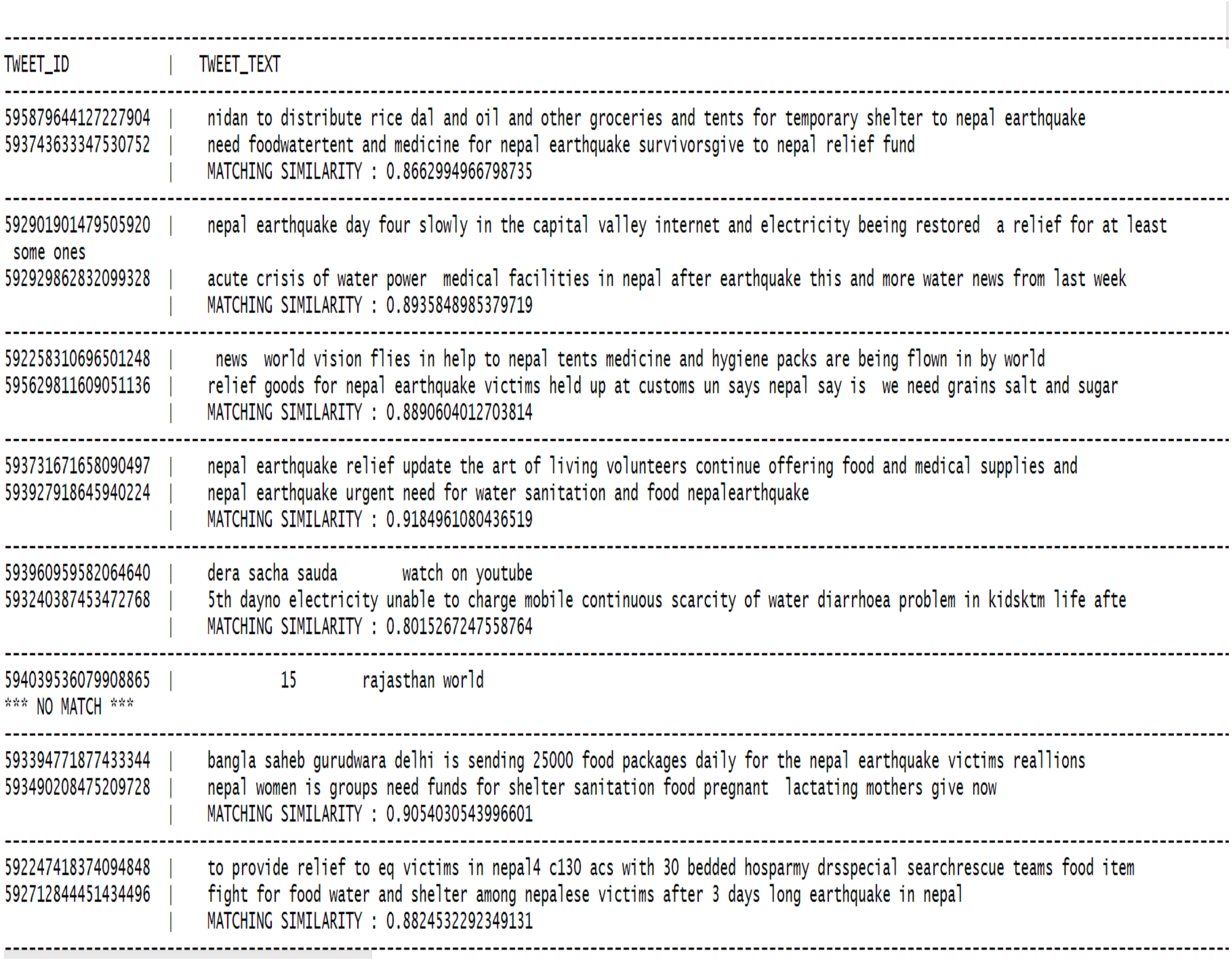


```
Epoch 4/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.4367 - acc: 0.8324 - val_loss: 0.6106 - val_acc: 0.7722
Epoch 5/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.3971 - acc: 0.8445 - val_loss: 0.5715 - val_acc: 0.8016
Epoch 6/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.3619 - acc: 0.8597 - val_loss: 0.5730 - val_acc: 0.8149
Epoch 7/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.3260 - acc: 0.8724 - val_loss: 0.5814 - val_acc: 0.8207
Epoch 8/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.3030 - acc: 0.8784 - val_loss: 0.6282 - val_acc: 0.8045
Epoch 9/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.2772 - acc: 0.8919 - val_loss: 0.7361 - val_acc: 0.8018
Epoch 10/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.2576 - acc: 0.8987 - val_loss: 0.6445 - val_acc: 0.8076
Epoch 11/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.2387 - acc: 0.9081 - val_loss: 0.6995 - val_acc: 0.8052
Epoch 12/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.2296 - acc: 0.9076 - val_loss: 0.7331 - val_acc: 0.8102
Epoch 13/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.2097 - acc: 0.9174 - val_loss: 0.7743 - val_acc: 0.7733
Epoch 14/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1993 - acc: 0.9215 - val_loss: 0.7778 - val_acc: 0.7992
Epoch 15/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1891 - acc: 0.9286 - val_loss: 0.7940 - val_acc: 0.8005
Epoch 16/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1840 - acc: 0.9273 - val_loss: 0.8110 - val_acc: 0.7748
Epoch 17/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1729 - acc: 0.9322 - val_loss: 0.7839 - val_acc: 0.8021
Epoch 18/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1599 - acc: 0.9353 - val_loss: 1.0631 - val_acc: 0.7785
Epoch 19/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1575 - acc: 0.9385 - val_loss: 1.0406 - val_acc: 0.8081
Epoch 20/20
15258/15258 [==============================] - 58s 4ms/step - loss: 0.1522 - acc: 0.9401 - val_loss: 0.9965 - val_acc: 0.7840
Training time: -1167.6329910755157
3815/3815 [==============================] - 14s 4ms/step
[INFO] loss=0.9965, accuracy: 78.4010%
```

FIGURE 5.8: Training Result for Image Classification Model using Modified VGG-16

5.4 Matching

The output of the matched (offer,request) pair is stated as below :

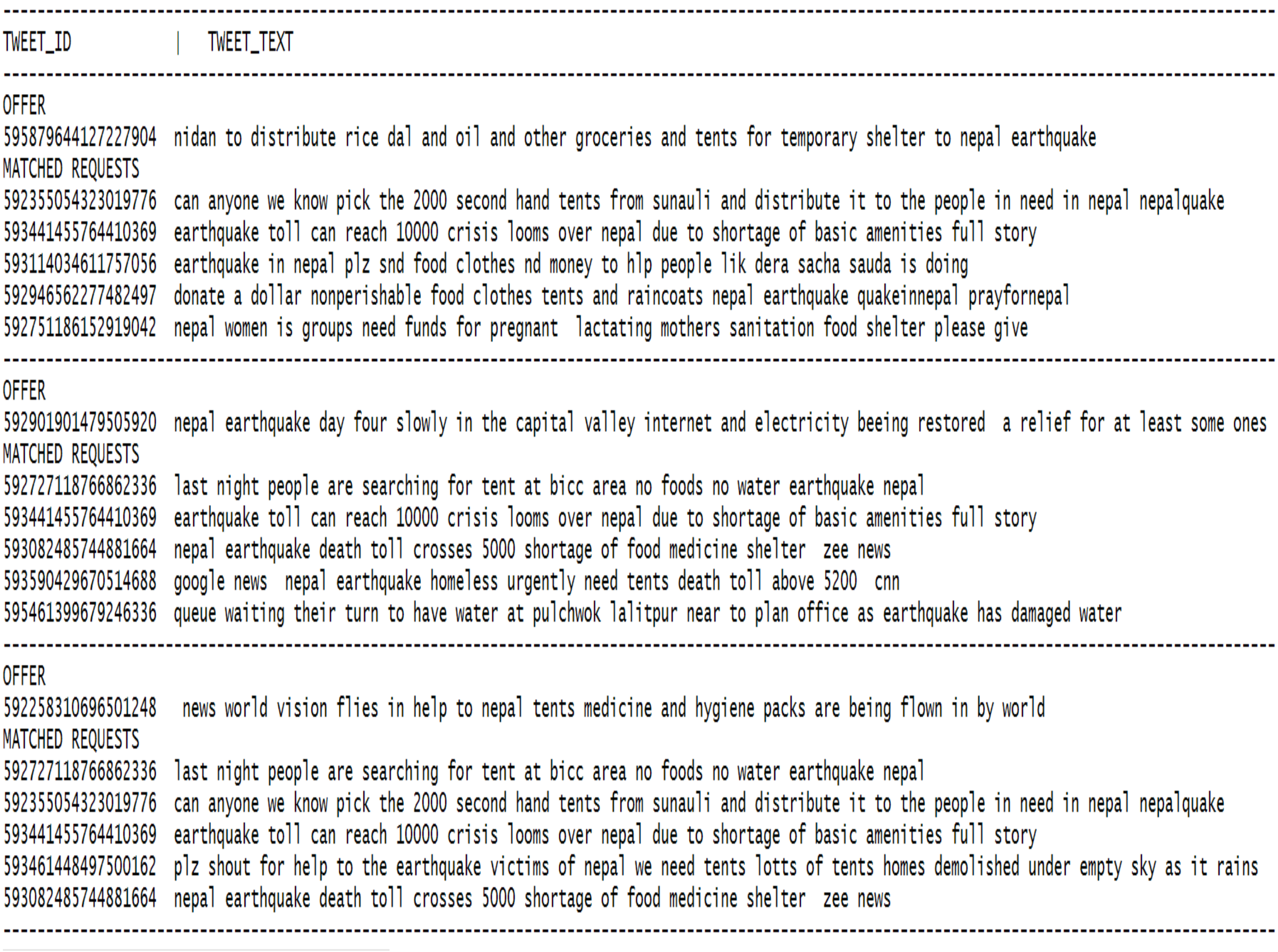


```
-------------------------------------------------------------------------------------------------------------------------------
TWEET_ID            |   TWEET_TEXT
-------------------------------------------------------------------------------------------------------------------------------
595879644127227904  |    nidan to distribute rice dal and oil and other groceries and tents for temporary shelter to nepal earthquake
593743633347530752  |    need foodwatertent and medicine for nepal earthquake survivorsgive to nepal relief fund
                    |    MATCHING SIMILARITY : 0.8662994966798735
-------------------------------------------------------------------------------------------------------------------------------
592901901479505920  |    nepal earthquake day four slowly in the capital valley internet and electricity beeing restored  a relief for at least
 some ones
592929862832099328  |    acute crisis of water power  medical facilities in nepal after earthquake this and more water news from last week
                    |    MATCHING SIMILARITY : 0.8935848985379719
-------------------------------------------------------------------------------------------------------------------------------
592258310696501248  |     news  world vision flies in help to nepal tents medicine and hygiene packs are being flown in by world
595629811609051136  |    relief goods for nepal earthquake victims held up at customs un says nepal say is  we need grains salt and sugar
                    |    MATCHING SIMILARITY : 0.8890604012703814
-------------------------------------------------------------------------------------------------------------------------------
593731671658090497  |    nepal earthquake relief update the art of living volunteers continue offering food and medical supplies and
593927918645940224  |    nepal earthquake urgent need for water sanitation and food nepalearthquake
                    |    MATCHING SIMILARITY : 0.9184961080436519
-------------------------------------------------------------------------------------------------------------------------------
593960959582064640  |    dera sacha sauda       watch on youtube
593240387453472768  |    5th dayno electricity unable to charge mobile continuous scarcity of water diarrhoea problem in kidsktm life afte
                    |    MATCHING SIMILARITY : 0.8015267247558764
-------------------------------------------------------------------------------------------------------------------------------
594039536079908865  |           15       rajasthan world
*** NO MATCH ***
-------------------------------------------------------------------------------------------------------------------------------
593394771877433344  |    bangla saheb gurudwara delhi is sending 25000 food packages daily for the nepal earthquake victims reallions
593490208475209728  |    nepal women is groups need funds for shelter sanitation food pregnant  lactating mothers give now
                    |    MATCHING SIMILARITY : 0.9054030543996601
-------------------------------------------------------------------------------------------------------------------------------
592247418374094848  |    to provide relief to eq victims in nepal4 c130 acs with 30 bedded hosparmy drsspecial searchrescue teams food item
592712844451434496  |    fight for food water and shelter among nepalese victims after 3 days long earthquake in nepal
                    |    MATCHING SIMILARITY : 0.8824532292349131
-------------------------------------------------------------------------------------------------------------------------------
```

FIGURE 5.9: One-to-One Mapping of Offer to Request

NOTE: In each matched set,the first line indicates the offer along with tweet_id and the second one represents the request.

The output of the matched (offer,[request1,request2,...]) pair is stated as below :



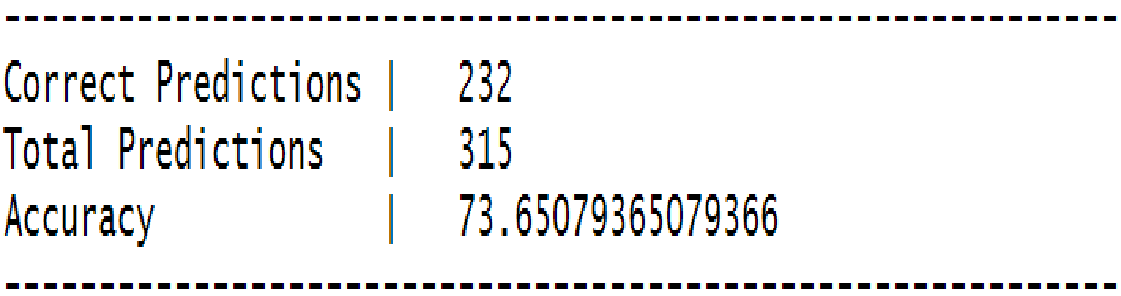
```
-----------------------------------------------------------------------------------------------------------------------
TWEET_ID            |  TWEET_TEXT
-----------------------------------------------------------------------------------------------------------------------
OFFER
595879644127227904  nidan to distribute rice dal and oil and other groceries and tents for temporary shelter to nepal earthquake
MATCHED REQUESTS
592355054323019776  can anyone we know pick the 2000 second hand tents from sunauli and distribute it to the people in need in nepal nepalquake
593441455764410369  earthquake toll can reach 10000 crisis looms over nepal due to shortage of basic amenities full story
593114034611757056  earthquake in nepal plz snd food clothes nd money to hlp people lik dera sacha sauda is doing
592946562277482497  donate a dollar nonperishable food clothes tents and raincoats nepal earthquake quakeinnepal prayfornepal
592751186152919042  nepal women is groups need funds for pregnant  lactating mothers sanitation food shelter please give
-----------------------------------------------------------------------------------------------------------------------
OFFER
592901901479505920  nepal earthquake day four slowly in the capital valley internet and electricity beeing restored  a relief for at least some ones
MATCHED REQUESTS
592727118766862336  last night people are searching for tent at bicc area no foods no water earthquake nepal
593441455764410369  earthquake toll can reach 10000 crisis looms over nepal due to shortage of basic amenities full story
593082485744881664  nepal earthquake death toll crosses 5000 shortage of food medicine shelter  zee news
593590429670514688  google news  nepal earthquake homeless urgently need tents death toll above 5200  cnn
595461399679246336  queue waiting their turn to have water at pulchwok lalitpur near to plan office as earthquake has damaged water
-----------------------------------------------------------------------------------------------------------------------
OFFER
592258310696501248   news world vision flies in help to nepal tents medicine and hygiene packs are being flown in by world
MATCHED REQUESTS
592727118766862336  last night people are searching for tent at bicc area no foods no water earthquake nepal
592355054323019776  can anyone we know pick the 2000 second hand tents from sunauli and distribute it to the people in need in nepal nepalquake
593441455764410369  earthquake toll can reach 10000 crisis looms over nepal due to shortage of basic amenities full story
593461448497500162  plz shout for help to the earthquake victims of nepal we need tents lotts of tents homes demolished under empty sky as it rains
593082485744881664  nepal earthquake death toll crosses 5000 shortage of food medicine shelter  zee news
-----------------------------------------------------------------------------------------------------------------------
```

FIGURE 5.10: One-to-Many Mapping of Offer to Requests

NOTE: In each matched set,the first line indicates the offer along with tweet_id and the following line represents the requests.

5.5 Testing

The proposed system was tested for the FIRE 2017 IRMiDis dataset and an accuracy of around 75% was attained :



```
------------------------------------------------------------
Correct Predictions |   232
Total Predictions   |   315
Accuracy            |   73.65079365079366
------------------------------------------------------------
```

FIGURE 5.11: Testing of the Proposed System

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this thesis, we presented an automatic tool for matching request and offer tweets for a real-world multimodal disaster-related dataset.Since only a small proportion of the complete dataset falls in the category of request and offer, it was quite a challenge to obtain quality data that is annotated as well. The use of state-of-art deep learning techniques have proved to be useful for task of classification.

In the future, we will redefine the matching algorithm to be flexible enough to handle multiple disasters at the same time and also take into account the whether the entity that demands or supplies is an individual or organization. We will be dealing with issues of quantity/capacity in our future work, which would require us to know more information attributes such as how many food rations, or how many shelter beds are supplied or demanded. This is a relevant aspect of this problem that is not supported by our dataset currently. We will also utilize the information present in audio and video formats to enhance the matching.

Finally, we will expand our system to handle multiple languages and not just be limited to English language tweets, though it contributes to more than one third of worldwide tweets.

REFERENCES

[1] Porto de Albuquerque J, Herfort B, Brenning A and Zipf A (2015) 'A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management', International Journal of Geographical Information Science, 29:4, 667-689.

[2] Alam F, Ofli F, Imran M and Aupetit M (2018) 'A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria', 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM).

[3] Varga I, Sano M, Torisawa K, Hashimoto C, Ohtake K,Kawai T, Jong-Hoon Oh and De Saeger S (2013) 'Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster', Association for Computational Linguistics (ACL) Conference.

[4] Alam F, Ofli F and Imran M (2018) 'CrisisMMD: Multimodal Twitter Datasets from Natural Disasters', 12th International AAAI Conference on Web and Social Media (ICWSM).

[5] Nguyen D T, Ofli F, Imran M and Mitra P (2017) 'Damage Assessment from Social Media Imagery Data During Disasters', IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[6] Imran M, Elbassuoni S, Castillo C, Diaz F and Meier P (2013) 'Extracting Information Nuggets from Disaster-Related Messages in Social Media', 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM).

[7] Alam, F, Imran, M and Ofli F (2017) ‘Image4Act: Online Social Media Image Processing for Disaster Response’, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[8] Imran M, Elbassuoni S, Castillo C, Diaz F and Meier P (2013) ‘Practical Extraction of Disaster-Relevant Information from Social Media’, 22nd International Conference on World Wide Web companion.

[9] Nguyen D T, Al-Mannai K A, Joty S, Sajjad H, Imran M and Mitra P (2017) ‘Robust Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks’, 11th International AAAI Conference on Web and Social Media (ICWSM).

[10] Starbird K and Stamberger J (2010) ‘Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting’, 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM).

[11] Imran M, Mitra P and Castillo C (2016) ‘Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages’, 10th Language Resources and Evaluation Conference (LREC), pp. 1638-1643.