

# Variational Autoencoder Framework for Multimodal Fake news Detection

*A DISSERTATION (MAJOR PROJECT – II)  
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE AWARD OF THE DEGREE  
OF*

MASTER OF TECHNOLOGY  
IN  
INFORMATION TECHNOLOGY

Submitted by:

**VIDHU TANWAR**

**2K18/ISY/14**

Under the supervision of

**DR. KAPIL SHARMA**

Head of Department (HOD), Department of Information Technology



Department of Information Technology

**Delhi Technological University**

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Bawana Road, Delhi – 110042 (India)

**July, 2020**

## **DECLARATION**

I, Vidhu Tanwar, hereby declare that the work which is being presented in the dissertation (Major Project - II) entitled “ **VARIATIONAL AUTOENCODER FRAMEWORK FOR MULTIMODAL FAKE NEWS DETECTION** ” by me in partial fulfillment of requirements for the awards of degree of Master of Technology (Information System) from Delhi Technological University, is an authentic record of my own work carried out under the supervision of Dr. Kapil Sharma, Head of Department (HOD), Information Technology Department.

The material contained in the report has not been submitted to any university or institution for the award of any degree.

**Place:** Delhi

*Vidhu Tanwar*  
**Vidhu Tanwar**

**Date:** 31/07/2020

**2K18/ISY/14**

## CERTIFICATE

This is to certify that Major Project Report - II entitled “VARIATIONAL AUTOENCODER FRAMEWORK FOR MULTIMODAL FAKE NEWS DETECTION” submitted by **VIDHU TANWAR (Roll No. 2K18/ISY/14)** for partial fulfillment of the requirement for the award of degree Master of Technology (Information System) is a record of the candidate work carried out by him under my supervision.

**Place:** Delhi

**Date:** 31/07/2020



**Dr. Kapil Sharma**

**SUPERVISOR,**

Head of Department (HOD), Department of Information Technology

Delhi Technological University, Delhi

## **ACKNOWLEDGEMENT**

I express my deep sense of gratitude towards my supervisor Dr. Kapil Sharma (Head of Department, Department of Information Technology) for his invaluable and meticulous guidance, suggesting the study, constructive criticism and encouragement during the course of research. Sincerity and perseverance are qualities that I would like to inherit from him. I am fortunate to have had an opportunity to work under his supervision. He has not only been my supervisor but also a very inspirational figure during my master's studies.

I am grateful to all the faculty members of the Department of Information Technology of Delhi Technological University for their immense support and guidance for the project. I would also like to acknowledge Delhi Technological University faculty for providing the right academic resources and environment for this work to be carried out. Last but not the least I would like to express my heartiest gratitude for the support provided to me by my parents and peer group for constantly encouraging me during the completion of work.

*VIDHU TANWAR*  
**VIDHU TANWAR**  
**2K18/ISY/14**

## ABSTRACT

Online Social media for news utilization is a have its pros and cons. If we ponder on the positives outcomes for this, it includes easy access, negligible cost, smart categorization and outreach to the very customer in seconds. But, as every coin has two sides and when we flip side of this, a series of issues come up which need immediate attention and most important among them is spreading of fake news. This has become a serious threat for the governments of countries to keep their harmony intact, keep faith of public in democracy and justice and sustenance of public trust. Subsequently detection in fake news, especially in web based platform has become a rising examination topic of interest that is pulling in colossal consideration. Current set of detection algorithms are specially indicating their powerlessness to gain proficiency with the mutual portrayal of text and visuals joined (popularly known as multimodal) information. Therefore, we present a variational auto encoder based framework, which consists of three major components encoder, decoder and fake news detector. It utilize the concatenation of visual latent features from three popular CNN architecture(VGG19,ResNet50,InceptionV3) combined with textual information to detect fake news with the help of binary classifier. We have shown the investigation on two publically available Twitter dataset and Kaggle dataset. The experimental result shows that our model improves state of the art method by the margin of ~2% in accuracy and ~3% in F1 score.

**Keywords:** *Concatenation, fake news detection, Latent features, multi-model, Variational auto encoder.*

# TABLE OF CONTENTS

<b>CANDIDATE’S DECLARATION</b>	<b>i</b>
<b>CERTIFICATE</b>	<b>ii</b>
<b>ACKNOWLEDGMENT</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>TABLE OF CONTENTS</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF TABLE</b>	<b>ix</b>
<b>LIST OF EQUATIONS</b>	<b>x</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xi</b>
<b>CHAPTER 1 INTRODUCTION</b>	<b>1</b>
1.1 WHAT IS FAKE NEWS?	1
1.1.1 Definition	1
1.1.2 Fake News Characterization	2
1.2 FEATURE EXTRACTION	2
1.2.1 News Content Features	2
1.2.2 Social Context Features	3
1.3 NEWS CONTENT MODELS	5
1.3.1 Knowledge-based models	5
1.3.2 Style-Based Model	6
1.4 SOCIAL CONTEXT MODELS	6
<b>CHAPTER 2 DEEP LEARNING MODELS</b>	<b>7</b>
2.1 NEURAL NETWORKS	7
2.1.1 Inspiration	7
2.1.2 Modal Representation	8
2.1.3 Neural network for multi-label classification	10
2.1.4 Squared Error Function	10
2.15 Cross Entropy	11
2.2 CNN	11
2.3 RNN	14
2.4 LSTM	15

<b>CHAPTER 3 LITERATURE REVIEW</b>	<b>17</b>
3.1 INTRODUCTION	17
3.2 SINGLE MODALITY-FAKE NEWS DETECTION	18
3.2.1 Textual Features	18
3.2.2 Review of existing textual based model	18
3.2.3 Visual features	22
3.3 MULTI-MODAL FAKE NEWS DETECTION	23
3.3.1 TI-CNN: Convolutional Neural Networks for Fake News Detection.	23
3.3.2 Fake News Detection on Social Media: A Data Mining Perspective.	24
3.3.3 r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection	25
3.3.4 Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs	26
3.3.5 EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection	27
3.3.6 SpotFake: A Multi-modal Framework for Fake News Detection	28
3.3.7 Multimodal variational autoencoder for fake news detection	29
<b>CHAPTER 4 DATA EXPLORATION</b>	<b>30</b>
4.1 INTRODUCTION	30
4.2 DATASET	30
4.2.1 Twitter Dataset	30
4.2.2 “all_data” Dataset	31
4.3 DATASET STATISTICS	31
4.3.1 Twitter dataset	31
4.3.1.1 Dataset Filtering	31
4.3.1.2 General Analysis	32
4.3.2 “all_data” Dataset	37
4.3.2.1 Dataset Filtering	37
4.3.2.2 General Analysis	37
4.4 VISULIZATION WITH t-SNE	42
4.5 CONCLUSION	42
<b>CHAPTER 5 THE PROPOSED WORK</b>	<b>44</b>
5.1 INTRODUCTION	44
5.1.1 Problem Statement	44

5.1.2 Motivation	44
5.2 VISUAL FEATURE EXTRACTION	44
5.2.1 VGG-19	45
5.2.2 Resnet 50	46
5.2.3 Inception V3	46
5.2.4 Image feature Extraction steps	46
5.3 TEXT FEATURE EXTRACTION	47
5.4 VARIATIONAL AUTOENCODER	48
5.5 PROPOSED MODEL	49
5.5.1 Overview	49
5.5.2 Encoder	49
5.5.3 Decoder	51
5.5.4 Fake News Detector	52
<b>CHAPTER 6 SOFTWARE REQUIREMENT AND METHODOLOGY</b>	
<b>VALIDATION</b>	<b>55</b>
6.1 SOFTWARE REQUIREMENT	55
6.1.1 Python	55
6.1.2 Hardware/Software Requirements	56
6.2 EXPERIMENTAL SETTINGS	56
6.3 PERFORMANCE MEASURE	57
<b>CHAPTER 7 RESULTS, CONCLUSIONS AND FUTURE SCOPE</b>	<b>59</b>
7.1 INTRODUCTION	59
7.2 EXPERIMENTAL RESULTS	59
7.3 CONCLUSIONS	64
7.4 FUTURE SCOPE	64
<b>References</b>	<b>66</b>
<b>LIST OF PUBLICATIONS OF CANDIDATE</b>	<b>71</b>



## LIST OF FIGURES

S.No.	Figure Name	Page No.
1.	Characterization & Detection of Fake News	2
2.	The two pictures present deforestation from two dates are taken from a single picture	4
3.	Neural network in human brain	7
4.	ANN Architecture	8
5.	Neural network with 3 layers	9
6.	Neural networks having hidden layers (2) and output layers(3 label)	10
7.	Cross-entropy Loss Function	11
8.	CNN architecture	12
9.	RNN architecture	14
10.	LSTM architecture	15
11.	Architecture of TI-CNN for Fake news Detection	24
12.	Architecture of att-RNN.	27
13.	The architecture of Event Adversarial Neural Networks (EANN)	28
14.	Twitter Dataset fields(main fields :post_text, image_id(s), label)	33
15.	Data Statistics for Twitter Dataset After Cleaning	33
16.	Test & Train Split after Pre-processing	34
17.	10 most common words in post_text with fake label	34
18.	10 most common words in post_text with real label	35
19.	Word cloud for fake and real post	35
20.	Length of words in post for fake and real	35
21.	Sentence length for fake and real post	36
22.	Box plot for Sentence length for fake and Real post	36
23.	all_data Dataset fields(main fields :title,image_url,type)	37
24.	all_data Dataset fields with relevant fields	38
25.	Data Statistics for all_data After Cleaning	38
26.	Test & Train Split after Pre-processing	38
27.	10 most common words in title with fake label	39

28	10 most common words in title with real label	39
29	Word cloud for fake and real post	40
30	Length of words in post for fake and real	40
31	Sentence length for fake and real post	41
32	Box plot for Sentence length for fake and Real post	41
33	t-SNE plot for Twitter Dataset	43
34	t-SNE plot for all_data Dataset	43
35	Example of Tweets from Twitter dataset	45
36	Architecture of Encoder of proposed model	49
37	Architecture of Decoder of proposed model	51
38	Architecture of fake news Detector of proposed model	52
39	Architecture of proposed model in Keras	53
40	Accuracy curve for Twitter dataset	59
41	Precision-Recall curve for Twitter dataset	59
42	ROC curve for Twitter dataset	60
43	Confusion Matrix for Twitter dataset	60
44	Accuracy curve for all_data dataset	61
45	Precision-Recall curve for all_data dataset	61
46	ROC curve for all_data dataset	62
47	Confusion Matrix for all_data dataset	62
48	Examples of Tweets classified by our proposed model but not by MVAE [1]	63

### **LIST OF TABLES**

<b>S.No.</b>	<b>Figure Name</b>	<b>Page No.</b>
1.	Table for Textual based Fake news Detection	18
2.	Statistics of Fakeddit	25
3.	Performance Measure of our proposed model	60

## LIST OF EQUATIONS

<b>S.No.</b>	<b>Equation Name</b>	<b>Page No.</b>
1.	Hypothesis Function	8
2.	Activation Function	8
3.	Activation of hidden layer	9
4.	Weight Matrix	9
5.	Computation of z for hidden layer 2	9
6.	Output of hidden layer 2	9
7.	Computation of z for final layer	9
8.	Output of final layer	9
9.	Squared Error Function norm 1	10
10.	Squared Error Function norm 2	10
11.	Cross Entropy Loss	11
12.	New hidden state for RNN	14
13.	Output for RNN	14
14.	new cell memory $C_t$	15
15.	Forget gate	15
16.	Memory gate	15
17.	Input gate	16
18.	Output gate	16
19.	Textual features for Encoder	49
20.	Visual features for Encoder	50
21.	Decoder	51
22.	Fake news Detector	52

### **LIST OF ABBREVIATIONS**

<b>S.No.</b>	<b>Abbreviated Names</b>	<b>Full Name</b>
1.	VAE	Variational AutoEncoder
2.	CNN	Convolutional Neural Network
3.	RNN	Recurrent Neural Network
4.	LSTM	Long Short-Term Memory
5.	VGG	Visual Geometry Group
6.	Resnet	Residual Network
7.	NLTK	Natural Language Toolkit,
8.	ANN	Artificial Neural Network
9.	BILSTM	Bidirectional Long Short-Term Memory
10.	FC	Fully Connected
11	HDF5	Hierarchical Data Formats

# CHAPTER 1

## INTRODUCTION

### 1.1 WHAT IS FAKE NEWS?

#### 1.1.1 Definition

With the arrival of internet revolution and digitization of contents, news has also gone digital. But it has own bad consequences as well. Fake news has emerged as a common problem across society that is used to disseminate the wrong information and influence the behavior of people. One recent example of fake news was seen in US presidential election [4] where it used to influence the mindset of people going into elections.

In INDIA also, with the ease of access and advancement of technology, cheaper data rates and availability even in remote villages, the reliability of people on digital news has increased multi-fold [5]. Even in case of digital news, a news with pictorial representation catches more attention of people rather than a content-based version of itself. The news companies encash this idea very well. The news with images related to content is widely read and has a greater outreach. But sadly, this idea has been used notoriously by various elements on the internet. Morphed images, misleading content and the way this false news are aimed at some range of readers help them spread the fake news.

In this thesis, we experiment the possibility of detecting fake news which has textual and visual information embedded in a single package by applying Variational Autoencoder performed on two distinct datasets containing distinct kind of news.

In the progress of detection in fake news effectively, the first and foremost part is to get the knowledge of what fake news is and how it is differentiated in categories. This thesis is based on mining perspective in the field of fake news detection [6]. The first is characterization and second is detection.

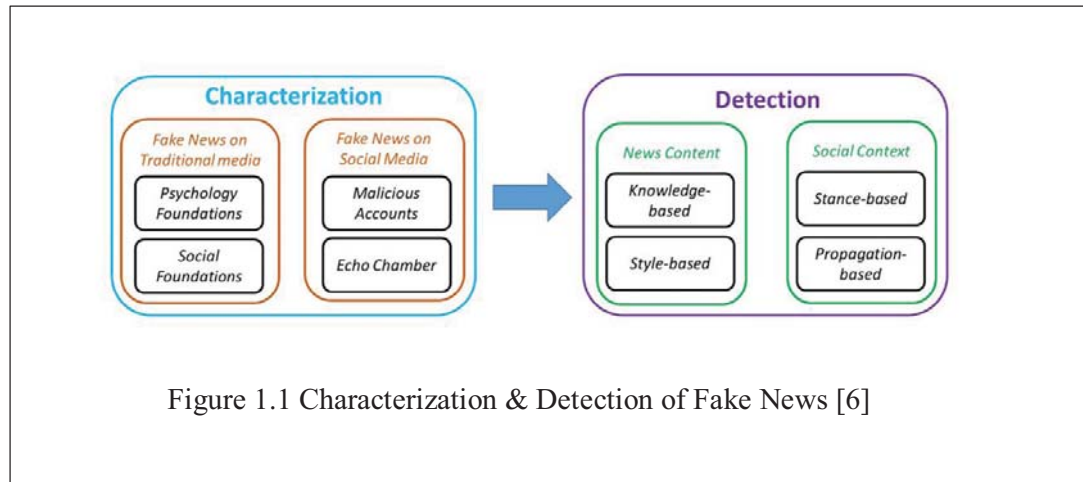


Figure 1.1 Characterization & Detection of Fake News [6]

### 1.1.2 Fake News Characterization

Fake news to be identified can be distinguished by two factors -authenticity and intent. Authenticity means verifying the contents of that particular news or article. This can be to verify like no conspiracy theory is involved and the news contents are not altered. The second factor, intent, means the motive behind misleading news. Whether the news is purposely designed to mislead the users.

## 1.2 FEATURE EXTRACTION

### 1.2.1 News Content Features

Now when we have defined a fake news and the target is set, the focus comes towards the features that can be used to classify a fake news. As per the general rules and news content, it is seen that a fake news is made up of four main components: -

- **Source:** The origin of news like author, timestamps and its reliability.
- **Headline:** The first line, a highlight of the news, used to catch the attention of the audience.
- **Body Text:** The textual content (i.e. the body of news article )of the news.
- **Image/Video:** Another part of fake news, usually the text is clubbed with pictures and videos.

Features are extracted from the above mentioned four components. The important ones from this are text and images/videos. These features are visual and lingual. As we have already understood that a fake news is used to influence a user or a group for a wrongly motivated scenario, the main difference based on lingual features is the discrimination of language. A true/valid news will be more formal and inclining towards a clear language to explain while a fake news may be well altered to weave a web of misunderstanding and presenting the facts in wrong manner or even drifting away from reality. To overcome this type of problems and identifying the fake lingual content, lexical features can be added. These may vary from a set of words, frequency of large/unique words to any of them.

On the other hand, the visual features are equally important and shall be taken into consideration. The images and videos are indeed added to the textual content to add more weight to the news. But eventually a morphed image and edited video may make it worse for the news to be considered in intended fashion. For example Fig 1.2 present the situation of deforestation in an area, still the two picture of year 2009 and 2019 are taken from a one unique picture mentioned in the bottom, however the authentic logo of WWF shows it as if it is from a confided in source [7] .

### **1.2.2 Social Context Features**

When it comes to sharing a news in social media, three important aspects are always taken into consideration. These aspects are, user perspective , post perspective and group perspective.

User aspect, for instance, can be considered in case of a news where it is feasible to analyse the behaviour of some users and the metadata(user profile data) in order to identify if a user is at risk. The risk can either be of falling into the trap of a false news or even spreading the same. The metadata plays an important role of giving information about center of interest, followers, likes etc...

Post aspect, it is used to analyze the metadata from post to provide insights about the authenticity of news. Group aspect, used to analyze the metadata from group, just like post or user, to give important information regarding that particular group.





Figure 1.2 The two pictures present deforestation from two dates are taken from a single picture [7]



## 1.3 NEWS CONTENT MODELS

### 1.3.1 Knowledge-based models

Now when we have discussed the parameters for identifying the fake news and various features for this, it is time to start explaining the models that can be created around these news content features. Considering the first model which is related to the news content is based on knowledge. The aim of the model based on news content is primarily checking the authenticity or truthfulness of the news and this can be achieved by three different methods. Those are – expert oriented, crowdsourcing oriented and computational oriented.

**Expert-oriented**, it is the method of relying on experts in the field like scientists, columnists and journalists who have a deep knowledge and vast experience in the subject related to news.

**Crowdsourcing-oriented**, relies mostly on the opinion of the crowd. It is termed using a large opinion or poll result that comes from a group of users which term the information as true or false.

**Computational-oriented**, depends on the result of mathematical computations and automatic fact checking tools. These tools have their own data base of information which is used in classifying the information. One such example is DBpedia.

The above mentioned methods have their own pros and cons. When it comes to expert oriented approach, the hiring of individual can be costly and may take its time as well. The experts may have limited information in that field and treating all the news received for them may be a tedious task. The crowd may not be right all the time and news can be designed in a way to break the computation fact checking system. The result of computation methods may not be accurate always.

### 1.3.2 Style-Based Model

As we have already talked about the styles followed in fake news, these are specific styles which are used to influence the crowd or user, their behavior which in turn is used to play with the emotions at time. This method is called deception oriented stylometric methods.

Another method is objectivity oriented approach which is used to fetch the object of a news, headline or a text. These styles are mostly used by partisan articles or yellow journalism. In this the articles, news or information is attached with an attractive headline which has high probability of being seen or gazed upon by users. An example of this can be a flashy pic or a headline like “Corona vaccine got success!!!”. This kind of images/headline plays with the curiosity of user and has highest probability of being seen

## 1.4 SOCIAL CONTEXT MODELS

The last feature which we shall talk about is social context models. This model is not used yet and have a scope to be researched upon. Two approaches are used to bring these into implementation.

- Stance based
- Propagation based

Stance based methods use internal as well as external representations. For example, votes by a group on social media can be considered as explicit representation while the information extracted from metadata is termed as implicit representation.

Propagation based approaches use propagation based features related to sharing such as likes, comments on social media or number of retweets on twitter.

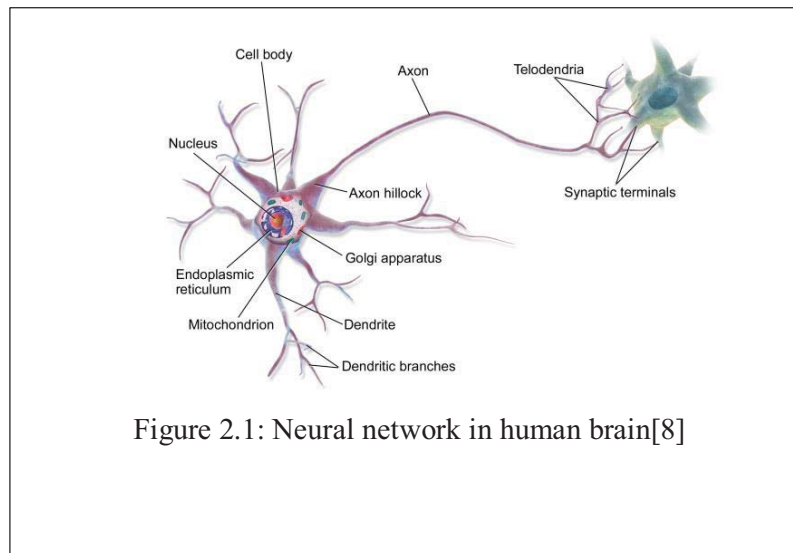
## CHAPTER 2

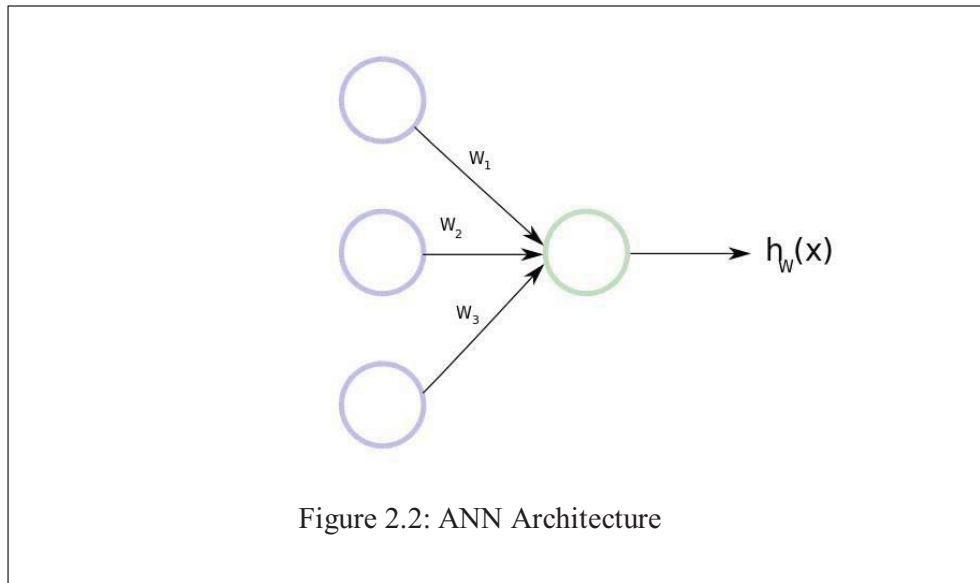
### DEEP LEARNING MODELS

#### 2.1 NEURAL NETWORKS

##### 2.1.1 Introduction

Be it, neural network or artificial neural network, these all function the same way. They try to copy the functioning of human brain (Fig 2.1) and how brain would act in those conditions if kept in. This is inspired by the neuron system of a mind. In the terms of biological words, a brain detects signals sent from different groups of dendrites and if the signal received from dendrites is powerful such that it can compel brain to act, signal then flow from an axon and goes to a dendrite to another neuron. The neurons in the human brain are unique in the system and no neuron is linked to another neuron. This phenomenon made possible by synaptic gaps and the two different neurons come in contact only when the connection of an axon from one neuron and dendrite connected together from the others are stimulated.





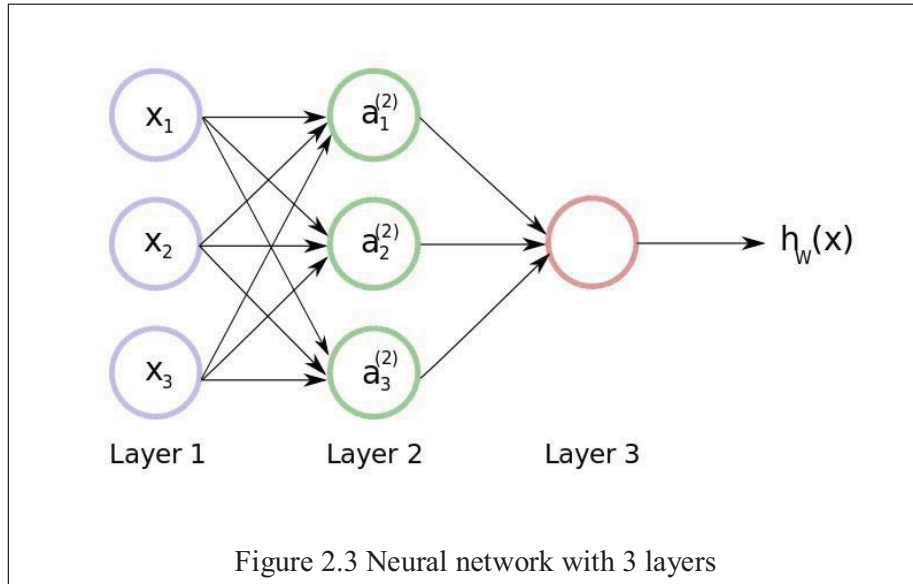
### 2.1.2 Modal Representation

For example, suppose an input  $x_1 \in (0,1)$  is given which gives information about whether the neuron is launch or not. This neuron gets accumulate by a weight  $W_1$ . As mentioned above, this part is to understand the synaptic connection where  $W_1$  is the degree of connection. Degree of connection tells about the strength of signal sent. Degree is greater if the connection is strong and smaller if the connection is weak. Being precise, it says about that decision making process in which the strength of synaptic connection is determined to decide whether the axon is stimulated or not. On the same pattern, we have  $x_2, x_3, \dots, x_n$  that get multiplied by  $W_2, W_3, \dots, W_n$  respectively. The products coming as a result of these calculations are summed up into a single unit to depict collective influence. The resultant summation of all neurons, which were given as input, are put into an activation function. The result from output functions is checked and if the value is greater than ZERO then the output unit is launched otherwise not. Fig 2.2 depicts an ANN with activation function(logistic) . In the above explained case, output of a neuron  $h_{w(x)}$  is computed as

$$h_{\theta}(x) = g(\text{Weight}^T \text{input}) \quad (2.1)$$

here,  $g(z)$  is activation function, for example logistic function

$$g(z) = \frac{1}{1+e^{-z}} \quad (2.2)$$



Similarly, different sets of weights are used to model multiple connections using different layers. Let's take an example of an ANN system which has three layers as hidden layers described in above Fig 2.3, the activation of neurons in the hidden layer (layer 2) are computed as:

$$a_0^{(2)} = g(\text{Weight}^{(1)}x_0 + \text{Weight}^{(1)}x_1 + \text{Weight}^{(1)}x_2 + \text{Weight}^{(1)}x_3) \quad (2.3)$$

In ML writing, condition (2.3) are reworked in grid documentation. Right off the bat, weight network speaking to the association between input layer and hidden layer is changed as:

$$\text{Weight of Hidden layer}(W)^{(1)} = \begin{bmatrix} W_{00}^{(1)} & W_{01}^{(1)} & W_{02}^{(1)} & W_{03}^{(1)} \\ W_{10}^{(1)} & W_{11}^{(1)} & W_{12}^{(1)} & W_{13}^{(1)} \\ W_{20}^{(1)} & W_{21}^{(1)} & W_{22}^{(1)} & W_{23}^{(1)} \\ W_{30}^{(1)} & W_{31}^{(1)} & W_{32}^{(1)} & W_{33}^{(1)} \end{bmatrix} \quad (2.4)$$

$$Z^{(2)} = W^1 x \quad (2.5)$$

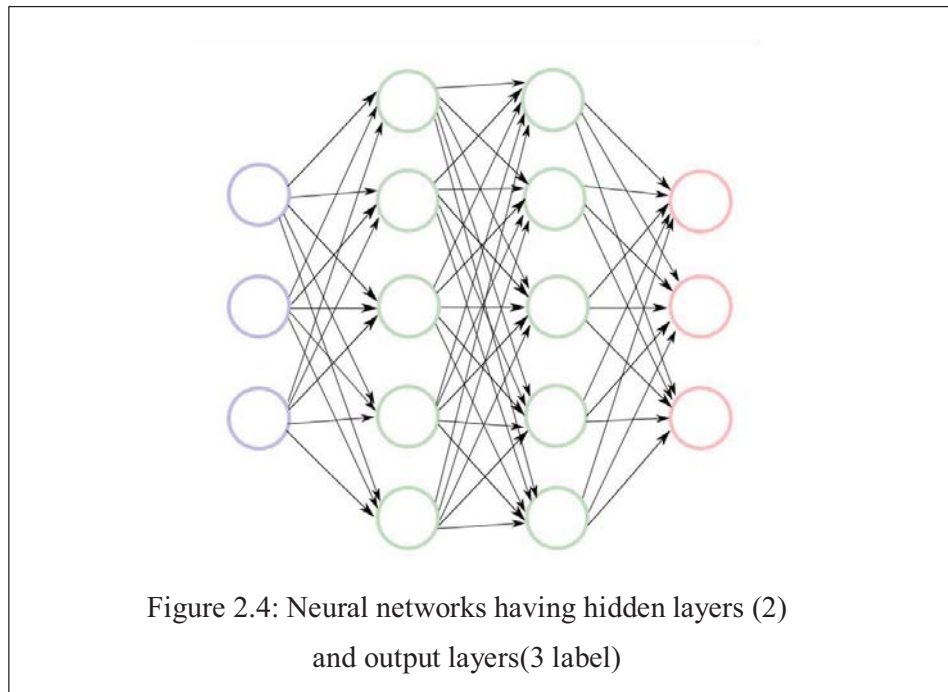
$$a^2 = \begin{bmatrix} a_0^2 \\ a_1^2 \\ a_2^2 \\ a_3^2 \end{bmatrix} = g^z \quad (2.6)$$

$$Z^3 = W^2 a^2 \quad (2.7)$$

$$h_w(x) = a^3 \quad (2.8)$$

### 2.1.3 ANN with multiple hidden layer and multiple label output

Now let's suppose that we want an output layer with a multilabel classification task (3 labels), the possible solution that comes to our mind will somewhat have the neural network structure as shown in Figure 2.4. Vector output  $hw$  is a 3-dimensional one-hot vector



### 2.1.4 Squared Error Function

Loss function indicates the distinction between anticipated field  $\hat{y}$  from the model and the actual output  $y$ . A basic methodology might be applied by taking distinction between them or standard 1:

$$\text{loss} = |\text{actual label} - \text{predicted label}| \quad (2.9)$$

For the ease of mathematical calculations, derivatives, square error, or norm 2, is applied to the loss function defined as:

$$\text{Loss} = (\text{actual label} - \text{predicted label})^2 \quad (2.10)$$

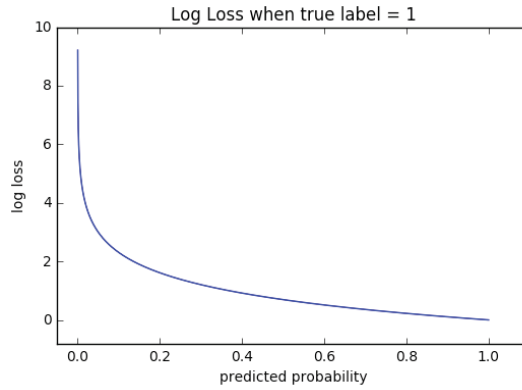


Figure 2.5 Cross-entropy Loss Function

However, Loss in equation 2.9 and 2.10 are not used frequently. On the other hand, cross entropy, has its own characteristics which are preferred in neural networks. The cross entropy is as defined in next section.

### 2.1.5 Cross Entropy

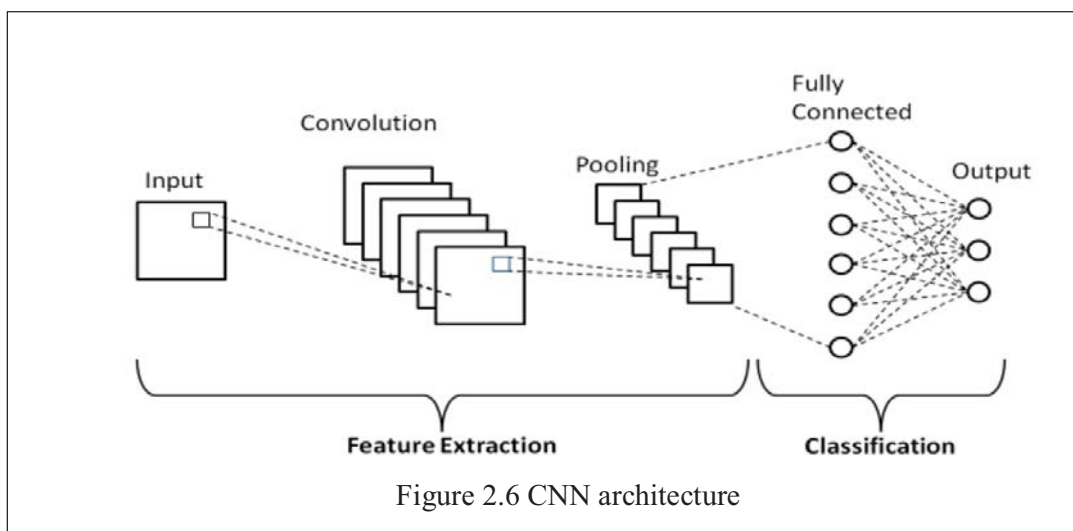
Cross entropy is defined as:

$$C(y, \hat{y}) = \sum_{i=1}^M y_i \log(\hat{y}_i) \quad (2.11)$$

Where, M is no of samples, y is actual labels and  $\hat{y}$  is predicted labels

Cross-entropy loss gauges the presentation of a grouping model whose output is a likelihood regard whose range is in between 0 and 1. Cross-entropy loss is increased when the predicted label is away from actual ground truth. So, anticipating a likelihood of .012 when the real perception mark is 1 would be awful and bring about a high misfortune esteem. An ideal model would have a log loss of 0.

The Figure 2.5 above shows the scope of conceivable misfortune esteems given a genuine perception (actual label= 1). As the predicted output from the model come near to 1, loss gradually decrease. As the predicted output from the model decreases's, loss gradually increase. Loss function mainly check errors where predicted output label is sure and wrong.



## 2.2 CNN

CNN is a well-defined subset of ANN, it is the one in which various connections are established between the neurons a one layer with the neurons of another layer. It is achieved with the combination of non-linear activation functions and weight matrix. But, a convolutional neural network (CNN) is pretty much different from the ordinary neural network. The differences are as below-

- The ordinary neural network takes vector as an input while convolutional neural network takes a matrix as an input. E.g. a 2-dimensional or 3dimensional matrix
- Ordinary neural networks take the multiplication of input vector and weight matrix into consideration while CNN calculate the convolution between them

A CNN network consists of 2 primary components named as convolutional and pooling layer. Here we take the example of a 2-dimensional matrix that is given as an input to CNN.

**Convolutional layer:** A kernel or sometimes called as filter which is actually a matrix is used ,the subparts of image are taken and kernel is slide upon them. The result coming out of this operation is called as feature map. The two main functions of this operation are input and filter matrix. In this example CNN and the filter matrix,

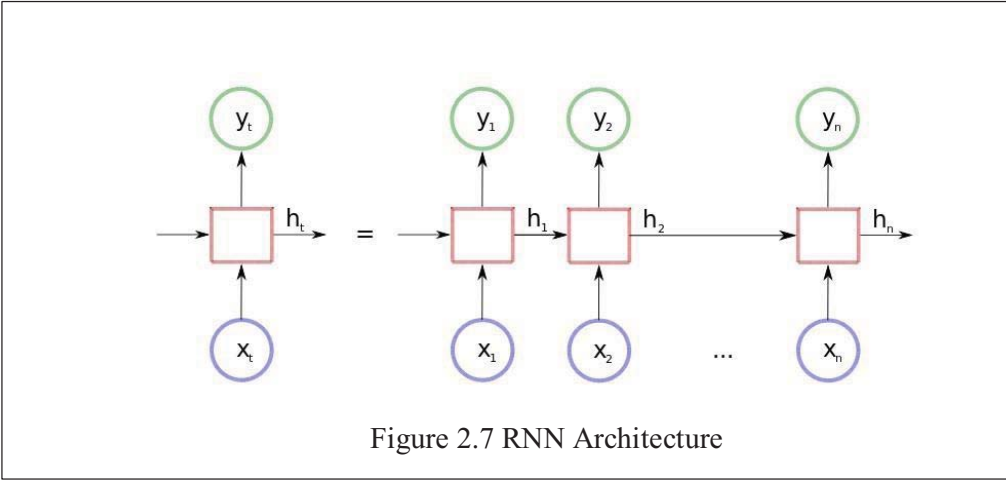


that is used to slide over input matrix, are used to find the convolutional operation between two functions. The sliding gap between the input image is determined by stride size. Filter size or kernel size is termed as size of a filter. This complete operation is executed with a motive behind that is to understand an unquestionable representation from the given input. For a definite issue, to extract distinct features from any input image distinct filters of distinct sizes are used on input image . The convolutional layer has two prime characteristics: - Sparse connectivity and share weights.

The generic neural networks are based on a phenomenon in which all the neurons from one layer are coupled to the respected neurons of another layer giving it a complete full connection. Whereas in these convolutional neural networks, the neurons are locally connected between layers, this is called “sparse connectivity”. The weight of filter that is shared because input image subparts are created then filter is slide over complete input image.

**Pooling layer:** Pooling layer is also known as subsampling. In max pooling ,Pooling layer extract the maximum value of input image based on kernel size and average of value of input image in case of average pooling

An engrossing attribute of a convolutional neural network is its ability to composition called as compositionality. To create the learning of non -activation function pooling and convolution layer are paired together. Compositionality, sparse connectivity and shared weight permit to learn the distinct features of an input that was provided. Notwithstanding with being popular for tasks such as captioning of images or recognizing the objects which come under the category of computer vision tasks, a convolutional neural network has also been inducted in the task of languages processing in natural scenarios and surprisingly it is yielding the astonishing results. The input sequence which is represented as a matrix is applied. The filters which may vary in numbers are slid over it.



### 2.3 RNN

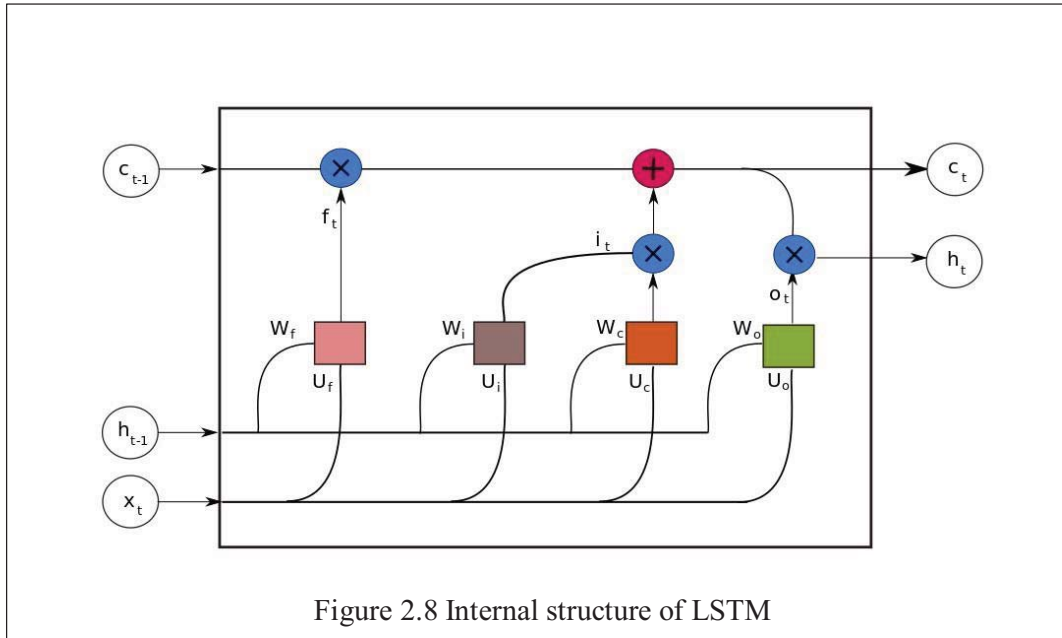
Another subset of ANN is RNN. In this type of neural networks, a subnetwork or a cell is repeated numerous times in the system to read the different inputs. This structure is as illustrated as in Figure 2.7

The output of the hidden state  $h_t$  is calculated as follows:

$$h_t = g_h(W_h x_t + U_h h_{t-1} + b_h) \tag{2.12}$$

$$y_t = W_y h_t + b_y \tag{2.13}$$

This type of systems is specifically architecture to handle the sequential data where the inputs are not provided to neural network system all at once. The sequential data is fragmented into smaller pieces of same or different sizes to be fed into a network cell in sequence of their respective numbers. Although it is designed to deal efficiently and act accordingly on the sequence of data provided, it has been observed very often that Recurrent neural networks have their own limitations in capturing long stream of data and its dependencies. Resultantly, a LSTM network has been devised to overcome the short comings of this system. This LSTM model is a doctored version of existing RNNs with extra features such as gating mechanism.



## 2.4 LSTM

When the existing RNN systems had their own limitations in processing the data, this LSTM model has made advances into neural network systems owing its ability to take care of limitation of RNN. To provides the essence of memory three new gates are added into a network cell. In reality, memory is stored as buffer and updated as soon as the network cell detects and read the input stream at every time stamp. Figure 2.8 provide an overview of LSTMs with four gates: memory, input, output and forget gate

The memory cell value  $C_t$  is calculated as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.14)$$

**Forget gate:** chooses what data to be removed from the current state of memory.

Given input  $x$  at time stamp  $t$ , forget gate  $f_t$  is calculated as:

$$f_t = g_f(W_f x_t + U_f h_{t-1} + b_f) \quad (2.15)$$

**Memory gate:** It create the new memory at time stamp  $t$ . Given input  $x$  at time stamp  $t$ , memory gate  $C_t$  is calculated as:

$$C_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2.16)$$

**Input gate:** This gate gives the information of how much amount of data of the memory cell is used with the updated memory. Given an input  $x$  at time stamp  $t$ , it is computed as

$$i_t = g_i(W_i x_t + U_i h_{t-1} + b_i) \quad (2.17)$$

**Output gate:** It provide the information of how much part of the data in memory cell is extracted as output. It is computed as:

$$o_t = g_o(W_o x_t + U_o h_{t-1} + b_o) \quad (2.18)$$

Hidden layer output is updated as:

$$h_t = o_t * g(c_t) \quad (2.19)$$

The long reliance issue is tended, with the nearness of inward memory and its capacity to get update consecutively.

# CHAPTER 3

## LITERATURE REVIEW

### 3.1 INTRODUCTION

This section presents a review of the different studies that have developed various techniques for detecting fake news. “Fake news” [9], this term has been taking rounds on internet for decades now but there is no formal definition of it. Social media recently has been used as a tool to disseminate the fake information in lesser time because of the masses available on it. As the issue was presented to the owners of these social media platforms [10] by the governments, they have come into action. Fake news can be defined [11] conventionally as one that consists of false claims, passed on statements, speech and posts in the textual or multimedia form. Precisely it states that deceptive news by some news outlet is more harmful and makes it harder to distinguish than the news in first definition. Fake news can be either rumor, misinformation, disinformation, hoax or a biased propaganda by some person or the organization. The broad categorization in [12] is Rumour, Hoax, Misinformation, clickbait, satire and propaganda.

In [9] the authors have surveyed different detection methods and opportunities. The study was covered by four perspectives, how the false knowledge proceeds, writing styles, propagation patterns and the credibility of the creators and the spreaders. Most of the fake news is not analysed in the aspects of complex patterns and the network traffic data patterns. Classification and utilization of data has always been a challenge in fake news detection. Here also, the utilization of information from different modalities has to be taken care of. As most of the existing approaches are old and highly dependent on text Deep learning technique [13], [14], [15].Lately, visual got started to be considered as another important factor [16] Due to increase in demand of multi-media, people have started to rope in visual information too in detection of fake news. On the other hand, the process of validating the content of multimedia went under relatively less scrutiny.

### 3.2 SINGLE MODALITY-FAKE NEWS DETECTION

#### 3.2.1 Textual Features

Literary highlights is factual or textual highlights extricated from text content of posts, tweets or news article which are investigated in numerous writings of phony news identification [17, 18, 19, 6]. Shockingly, phonetic examples are definitely not however surely knowing, since they are profoundly reliant on explicit occasions and comparing area information [20]. Hence, it is hard to configuration hand-created printed highlights for customary AI based phony news location models. To survive this impediment, In paper [21] proposes a profound learning model to distinguish counterfeit news. In particular, it conveys repetitive neural systems to get familiar with the portrayals of posts in a period arrangement as literary highlights. Trials results show the adequacy of profound learning based models.

#### 3.2.2 Review of existing textual based model

Table 3.1 Table for Textual based Fake news Detection

Refs	Objectives	Techniques	Obtained results with merits	Demerits
[22]	To detect fake news of social media by ensuring the verification process via multi-voting model.	Term frequency-inverse document frequency; count-vectorizer and hash-vectorizer were used as a feature extraction process. Then,	Suggested models were applied on three datasets, namely, News Trends, Kaggle and Reuters. Performance metrics such as accuracy, precision, recall, F1 score and	Though the system has improved the detection accuracy, the efficiency of the detection classifier is not explored. If the input size increases, then the

		<p>Passive Aggressive (PA), Logistic Regression (LR), Linear support vector (LSV) and Linear SVM were used for classification purpose.</p>	<p>specificity. Multi-voting model is the novel approach employed. The news trends datasets have achieved an accuracy of 94.5 (Tf-IDF); 93.6 (CV); 87.1(HV). Kaggle datasets have achieved an accuracy of 98.9 (Tf-IDF); 98.7 (CV); 95.8(HV). Likewise, Reuters datasets have achieved an accuracy of 97.2 (Tf-IDF); 96.5 (CV); 90.2(HV).</p>	<p>system lowered the efficiency of the classifier.</p>
[23]	To develop a user behavior model on detecting the false news on Twitter.	An unsupervised approach was employed here. Clustering and frequent itemset mining were used for	The suggested classifier was studied in Military airstrikes in Syria in Sep. 2017. For 8 clusters, the system has	Lack of geolocation prediction and analysis

		constructing the classifiers.	achieved 100% precision.	
[24]	To detect automatic fake news by improving pre-training classifiers.	Bidirectional Encoder Representations from Transformers (BERT)	CNN and Daily Mail datasets were used for analytic purpose. Performance measures analyzed are precision, recall and f1 score. Compared to prior algorithms, 0.14 F1 score was improved.	Data imbalance issue arises, when the authenticity of the data is altered.
[25]	To detect fake news earlier by theory-based approaches.	News article content is analyzed at four distinct levels, namely, syntax-level, semantic-level, lexicon-level, and discourse-level. Then, a supervised approach was framed to classify the contents.	Semi-Supervised classifiers such as SVM, Random forest & XG Boost were used for study purpose. PolitiFact & Buzzfeed datasets were for experimental purpose. The suggested model has achieved 0.892(accuracy); 0.877 (precision); 0.908 (recall) &	Interpretation of the data and its relationships are not effectively approached. Some complex news data are ignored for study purpose.



			0.892 (F1-score) for PolitiFact dataset. Likewise, Buzzfeed dataset has helped for achieving 0.879(accuracy); 0.85 (precision); 0.902 (recall) & 0.879 (F1-score)	
[26]	To detect the fake news of different sources of social media.	Improved Part of Speech (POS) Bidirectional LSTM and Convolutional Neural Networks.	Liar -Liar datasets were used on this hybrid model LSTM and CNN and achieved an accuracy of 42.2% with gain 3.3%.	Some learning patterns of the news content are difficult to formalize the hidden layers.
[27]	To detect fake news on different multi-modals deceptive systems.	Different neural networks architecture was used for study purpose. AdaBoost and NN models were explored.	The class with the highest incorrect prediction in this manner is disinformation (40.08% of tweets) followed by the conspiracy (39.13%) and propaganda (37.45%). The	The false-positive rate is higher in the combination of disinformation and propaganda posts.

			<p>least incorrectly predicted class is satire (0.72%), then hoax (2.19%), verified (5.55%), and clickbait (11.26%).</p> <p>Between all collections, about 31.5% of tweets fool all of our models in this way.</p>	
[28]	To study about the media-rich fake news detection models.	Surveyed about the characterization of a news story of different content types.	This paper has provided better insights into fake news detection systems.	

### 3.2.3 Visual features

Visual highlights have been demonstrated to be a significant marker for counterfeit news identification [16, 6]. Be that as it may, extremely restricted examinations are led on checking the believability of sight and sound substance on online networking. The fundamental highlights of connected pictures in the posts are investigated in the work [30, 31, 32]. Be that as it may, these highlights are till hand-created and can barely speak to complex disseminations of visual substance.

### 3.3 MULTI-MODAL FAKE NEWS DETECTION

To take in include portrayals from different viewpoints, profound neural systems was effectively experimented in different undertakings, including however not constrained to image captioning [34, 35], visual question answering [33] and counterfeit news location [36]. [36] proposes a fake news detection system where textual , social information of user as well as visual features are extracted and combine them in a complete system.

To beat the impediments of existing work and limited work on visual features, we propose a variational autoencoder based fake news detection system, which altogether enhances the exhibition on counterfeit news location on various occasions. The proposed model in this thesis not just consequently learns multi-modular component portrayals, but help us to detect fake news

#### 3.3.1 TI-CNN: Convolutional Neural Networks for Fake News Detection [37]

In this paper, [37] propose to examine the "fake news discovery" issue. To build a system which can automatically distinguish fake news is hard to create because system based on fact checking is still an open issue, very few current models can be applied to overcome the current issue of counterfeit news .The first step was an intensive examination of information used in fake news, heaps of helpful unequivocal highlights is distinguished from the content articles and pictures utilized for the issue of counterfeit news. Other than express highlights, their internal features are also extracted from articles and pictures utilized in counterfeit news, which is captured by lot of inert highlights separated by means of the different convolutional and pooling layers in our model.[37] proposes this model named as TI-CNN (Text and Image data based Convolutional Neural Network) was suggested by this paper. The last step was combining the latent and external features together in same element space, TI-CNN is prepared with both the content what's more, picture data at the same time. Broad trials carried on this present reality counterfeit news datasets have illustrated the viability of TI-CNN in fathoming the fake new recognition issue.

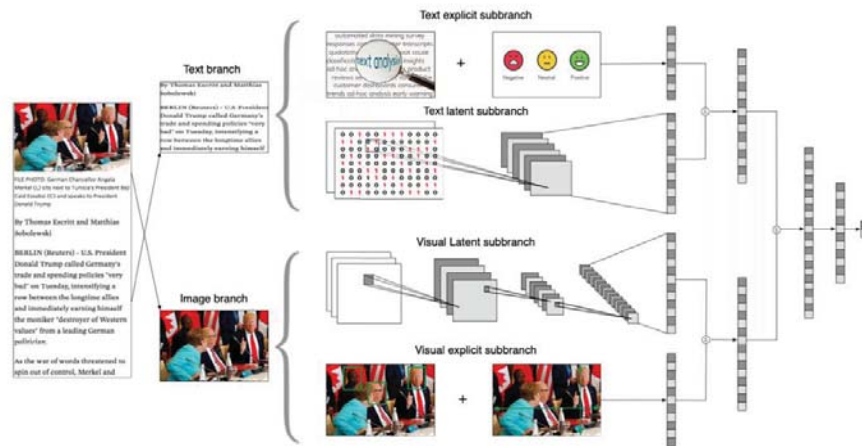


Figure 3.1 Architecture of TI-CNN for Fake news Detection [37]

### 3.3.2 Fake News Detection on Social Media: A Data Mining Perspective [6]

[6] provided a detailed review on Fake news Detection which states that today's life which is mostly dependent on online resources is a double edged sword having pros and cons of itself. On one side, its ease of access to almost all kind of data with least efforts and quick scattering leads an individual to read, propagate and forward the news with utmost ease. But on the other hand, these offerings have their dark side as well through which fake or low calibre news with purposefully irrelevant data are spread easily. These news are generally called as "counterfeit news". This widespread of low quality news may have adverse impacts on society and may be targeted to harm the communal harmony as well. In this manner, counterfeit news discovery on web based life has as of late become a developing exploration that is drawing in colossal consideration. The identification of this counterfeit news through some most popular web based models and networking media poses greater challenge and difficulties that make existing discovery calculations from customary news media insufficient or not appropriate. The First point mentioned as, fake news article are written in a manner so that they seems to like real news article. Therefore, information other than news article is also explored for Fake News Discovery Second, abusing this assistant data is trying in and of itself as clients' social commitment with counterfeit news produce information that is huge, inadequate,

Table 3.2 Statistics of Fakeddit [38]

Statistics of Fakeddit	No. of samples
Total no. of samples	1,063,106
No. of Fake samples	628,501
No. of Real samples	527,049
Multimodal samples	682,996

unstructured, and boisterous. Throughout this model, they came up with a survey of vast reach in which they went on recognizing fake or counterfeit news on internet. This includes some extra ordinary topics like news portrayals on brain science and social speculations, existing calculations from an information mining point of view, assessment measurements and agent datasets. They additionally talk about related exploration zones, open issues, and future examination bearings for counterfeit news location via web-based networking media.

**3.3.3 r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection [38]**

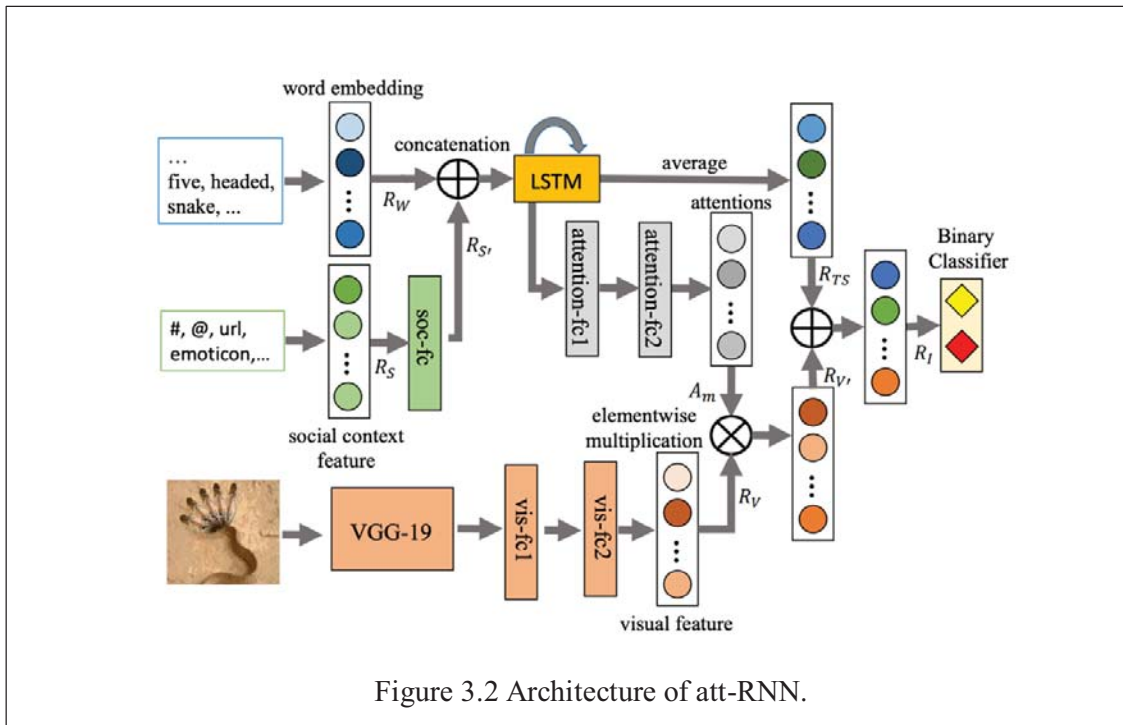
In any case, an absence of viable, thorough datasets has been an issue for counterfeit news examination and recognition model. Earlier fake news datasets don't give multimodal text and picture information, metadata, remark information, and clear lables for classification of fake news identification. In this paper [38], a new multimodal dataset named as Fakeddit comprising of more than a million of examples with distinct labels (not just fake and real). In the wake of being handled through a few phases of audit, the tests are marked by three different lables(2,3,6) arrangement classes through inaccessible management. They develop half and half text+image models and perform broad examinations for various varieties of

arrangement, showing the significance of the novel part of multimodality and fine-grained grouping one of a kind to Fakeddit. This dataset consists a large quantity of multimedia contents coming from very diverse resources. Data set is resourced from Reddit, a social news and discussion forum on various issues. Each issue is called a subreddit, which has its own theme. It consists of more than 8 lac submission from 21 different subreddits, which consists of image, text, comments, submission by other users on same subreddit, score of subreddit, source domain, up votes and down votes. Almost two third of the samples had multimedia contents while the remaining only had textual information

### **3.3.4 Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs [36]**

[36] proposes a novel architecture of Recurrent Neural Network with a consideration instrument (attention-RNN) to meld multiple modals highlights to resolve powerful rumor discovery. To create the model they have taken the textual data and pass it though the LSTM to get the features whereas visual data is passed though VGG19 architecture and social setting are also consider and passed though the RNN Broad tests were directed on 2 interactive media talk datasets gathered from Weibo and Twitter.

They fuse multimodal substance on informal communities to take care of the difficult rumor discovery issue. Rather than conventional physically made features, textual, visual and social setting substance are spoken to by means of profound impartial systems. It proposes an inventive attention with RNN instrument (attention-RNN) for viable multiple modals highlight combination. The system wires highlights from three modalities and uses the consideration component for highlight arrangement. For validation of the model against competitive algorithms, the paper evaluated attention-RNN on two popular datasets named as Weibo and Twitter, respectively. Outcome of their modal have shown that it achieves the simplest performance on both datasets, as compared with exiting feature-based methods and state-of-the-art neural network models. Weibo and Twitter datasets have analysed for all social context features. Weibo dataset helped to achieve 0.788 accuracies whereas, 0.682 achieved by Twitter datasets.



### 3.3.5 EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection [39]

As news perusing via web-based networking media turns out to be increasingly well known, counterfeit news turns into a significant issue concerning general society and government. The fake news can exploit media content to misdirect peruses which help it to spread among people consuming this news, it causes negative impacts and control the opinion of people regarding some subject. The major exceptional challenges for counterfeit news discovery via web-based networking media is the means by which to distinguish counterfeit news on recently developed occasions. Shockingly, a large portion of the existing methodologies can scarcely deal with this test, since they will in general learn occasion explicit highlights that cannot be moved to inconspicuous occasions. To provide solution to the above mentioned problem, [39] propose an start to finish system named Event Adversarial Neural Network (EANN), that can infer occasion invariant highlights and in this way advantage the recognition of fake news on recently showed up occasions. It comprises of three primary parts: the fake news finder, the multi-modular component extractor and the occasion discriminator. For removing the printed and visual highlights from posts multi-modular component extractor is used. It helps out the fake news locator to

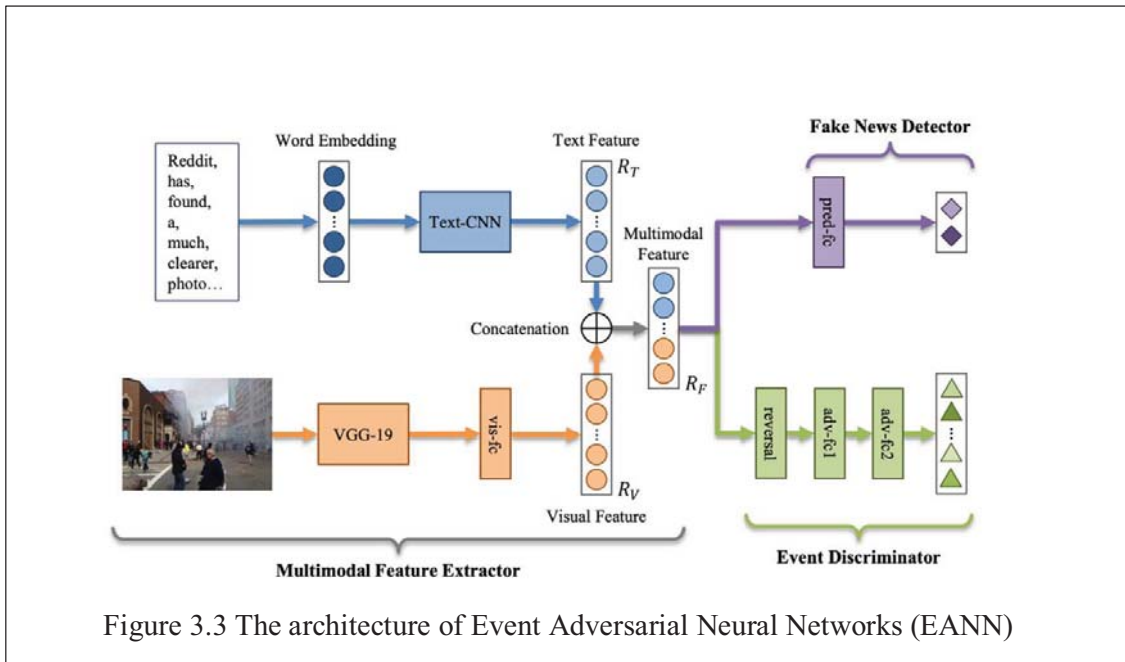


Figure 3.3 The architecture of Event Adversarial Neural Networks (EANN)

become familiar with the discriminable portrayal for the location of fake news. The job of occasion discriminator is to evacuate the occasion explicit highlights what's more, keep shared highlights among occasions. Broad analyses are led on sight and sound datasets gathered from Weibo and Twitter. It learns occasion invariant highlights utilizing an adversarial system alongside a multimodal include extractor. However, both these models don't have any express target to find relationships over the modalities. Twitter and Weibo datasets were used for experimental purpose. performance measures such as accuracy, precision and fl- score. Twitter datasets have achieved 0.715 (accuracy); 0.822(precision); 0.638 (recall) and 0.719 (F1 measure).

### 3.3.6 SpotFake: A Multi-modal Framework for Fake News Detection [40]

A multi-modal framework was developed by [40] which exploited textual and visual features. The architecture plans to recognize whether a given news article is genuine or counterfeit. Other subtask like event discriminator [39] or Decoder-Encoder model [1] is not consider here. The novel oddity of Spot Fake model was to join the intensity of linguistic models, for example Bidirectional Encoder Representations from Transformers (BERT) to join logical data. Picture highlights were found out from VGG-19 pre-prepared on ImageNet dataset. The Textual and the visual model is combined together through concatenation and further used for



classification. Hence, Different language models like BERT was combined with VGG-19 pre-trained architecture on ImageNet datasets. The suggested model has achieved an accuracy of 77.77% (Twitter) and 89.23% (Weibo). Limited hidden layers are taken for analytic purpose.

### **3.3.7 Multimodal variational autoencoder for fake news detection [1]**

Multimodal Variational Autoencoder (MVAE) was suggested by [1] that detect the fake news via learning the probabilistic latent variable models. Most of the complex patterns are ignored, which was improved by multimodal representations. Experimental results have shown that the improvement of 6% in accuracy and 5% in F1 score. Some characteristics of the users are not explored under neural network architectures.

Our model is inspired from [1] which is based on variational auto encoder, which consist of Bi-LSTM, for text feature extraction and VGG-19 for image feature extraction. The latent vectors produced by concatenation of these two vectors were fed into a decoder for reconstructing the original samples. The same latent vectors were also used for a secondary task of fake news detection.

## Chapter 4

### DATA EXPLORATION

#### 4.1 INTRODUCTION

The decent beginning stage in the examination is to provide a few information investigations of the dataset. The primary thing to be done is factual examination, for example, checking the quantity of words per labels or tallying the quantity of sentence. At that point it is conceivable to attempt to get an understanding of the information dissemination by proving dimensionality decrease and create plot of information in 2D.

#### 4.2 DATASET

##### 4.2.1 Twitter Dataset

The scarcity or a limited access to properly managed multimedia information is a concern. To avoid this, we use standard dataset which helps in testing our model to detect fake news. The dataset consists of factual information on social media obtained from Twitter. The Twitter dataset was released as part of MediaEval [2] to verify multimedia usage with primary focus as to identify the fake multimedia contents including images and text on social media. This data set consists textual contents, videos, images and context details for every tweet related to dataset which was posted on social media site Twitter. It is a large dataset containing approximately 17000 different tweets spanning over years related to 17 events across globe. Two parts of this data set include – the development set and the test set. The development set is considered bigger than the test set and contains around 9000 post label as fake and 6000 post label are real while the testing set contains only 2000 tweets. As the emphasis of our model is on textual and image information, we have omitted the tweets in which either the image was not available or the tweets which had videos associated with them. The dataset is available online [2].

#### **4.2.2 “all\_data” Dataset**

Another dataset used in this thesis consist of 20,015 total news, i.e., 11,941 news label as fake and 8,074 news label as real. It is accessible online [3]. For counterfeit news, it consist of text and text related information scratched from 240 distinct sites from the Megan Risdal on Kaggle . The news labelled as real is slithered from notable definitive news sites, i.e., the New York Times, Washington Post, and so forth. The Kaggle dataset contains numerous data for example, the title, text, picture, creator and site. To uncover the natural contrasts between genuine and counterfeit news, we exclusively utilize the title, type and picture data.

### **4.3 DATASET STATISTICS**

#### **4.3.1 Twitter dataset**

The main dataset used for creating the proposed model is the twitter dataset. So, data exploration will start from the twitter dataset. Twitter dataset is divided into two parts devset and testset. Devset is furnished along with ground truth and is utilized by us to build up our methodology to detect fake news. For the primary errand, it contains post related on the 17 occasions , involving altogether 193 instances of genuine and 220 instances of abused pictures/recordings, related with 6,225 genuine and 9,596 fake posts. Whereas the testset, is utilized for assessment. For the fundamental undertaking, it involves 104 instances of genuine and abused pictures and 25 instances of genuine and abused recordings, altogether connected with 1,107 and 1,121 posts, separately.

##### **4.3.1.1 Dataset Filtering**

To start with the analysis, dataset need to be clean. Firstly, file containing image URL is read line by line and images are downloaded from the mentioned URL. All the images are stored in the database and further used for image feature extraction.

Now, post file is read line by line and image id associated with particular post is retrieved. Each image id is checked in images database whether images are available or not. If, all the image id corresponding to the post are present in database then the post is valid otherwise the post is discarded.

Since the dataset has been refined, data statistics given by the dataset makers and data statistics processed in the wake of cleaning are changed. Result in change of number of samples actually used for proposed model is given in Figure 4.1-4.3.

#### **4.3.1.2 General Analysis**

In addition to cleaning, linguistic analysis is also performed on the dataset. Most frequent words are taken out from the post present in the dataset based on the labels fake and real. Before finding the most frequent words, posts are pre-processed by removing the URL, removing the alphanumeric characters and tokenizing the post with the help of NLTK [41] library. The subsequent advance comprises of extracting words from sentence, removal of stop words, (for example, 'an', 'a', 'the'), accentuation, words with length less than size are removed, removal of non-alphanumeric words, numeric qualities and labels, (for example, html labels) are expelled. At long last, the quantity of words despite everything present is utilized.

Figure 4.4-4.6 illustrate the most frequent words as well as most frequent trigrams in the post i.e. the textual content of our dataset.

Another important highlight to take a gander at is the dispersion of the quantity of words in the content. To be sure, sooner or later it is expected to fix the size of consistent length of writings and when the post is less than the consistent length it is cushioned with zero otherwise long sentences are trimmed. It is hence expected to examine the length of the writings so as to pick the correct one. Next task in the data exploration is analysis of number of words used in the title as well as sentence length distribution on the basis of label of post. Figure 4.7-4.9 provide the boxplot for the same. It can be analysed from the fig that the

sentence length is under twenty and the sentence length below two can be discarded

	0	1	2	3	4	5	6
0	post_id	post_text	user_id	image_id(s)	username	timestamp	label
1	324597532548276224	Don't need feds to solve the #bostonbombing wh...	886672620	boston_fake_03,boston_fake_35	SantaCruzShred	Wed Apr 17 18:57:37 +0000 2013	fake
2	325145334739267584	PIC: Comparison of #Boston suspect Sunil Tripa...	21992286	boston_fake_23	Oscar_Wang	Fri Apr 19 07:14:23 +0000 2013	fake
3	325152091423248385	I'm not completely convinced that it's this Su...	16428755	boston_fake_34	jamwil	Fri Apr 19 07:41:14 +0000 2013	fake
4	324554646976868352	Brutal lo que se puede conseguir en colaboraci...	303138574	boston_fake_03,boston_fake_35	rubenson80	Wed Apr 17 16:07:12 +0000 2013	fake
5	324315545572896768	4chan and the bombing. just throwing it out th...	180460772	boston_fake_15	Slimlenny	Wed Apr 17 00:17:06 +0000 2013	fake
6	324581777614180352	4chan thinks they found pictures of the bomber...	46224814	boston_fake_08	iamyadvinder	Wed Apr 17 17:55:00 +0000 2013	fake
7	324665423956176896	Ola ke ase, investigando las bombas de Boston ...	90735851	boston_fake_35	rcr866	Wed Apr 17 23:27:23 +0000 2013	fake
8	325464125868216321	4chan ThinkTank - Imgur <a href="http://t.co/hQt2fmxE48">http://t.co/hQt2fmxE48</a>	142785938	boston_fake_03,boston_fake_35	GlebGgs	Sat Apr 20 04:21:09 +0000 2013	fake
9	325099014355820544	@DLoesch have you seen this? Bomber #2 looks ...	21769179	boston_fake_13	larrygloverlive	Fri Apr 19 04:10:19 +0000 2013	fake

Figure 4.1 Twitter Dataset fields(main fields :post\_text, image\_id(s), label)

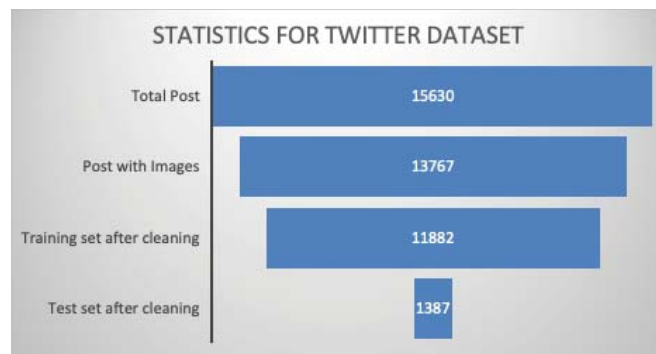
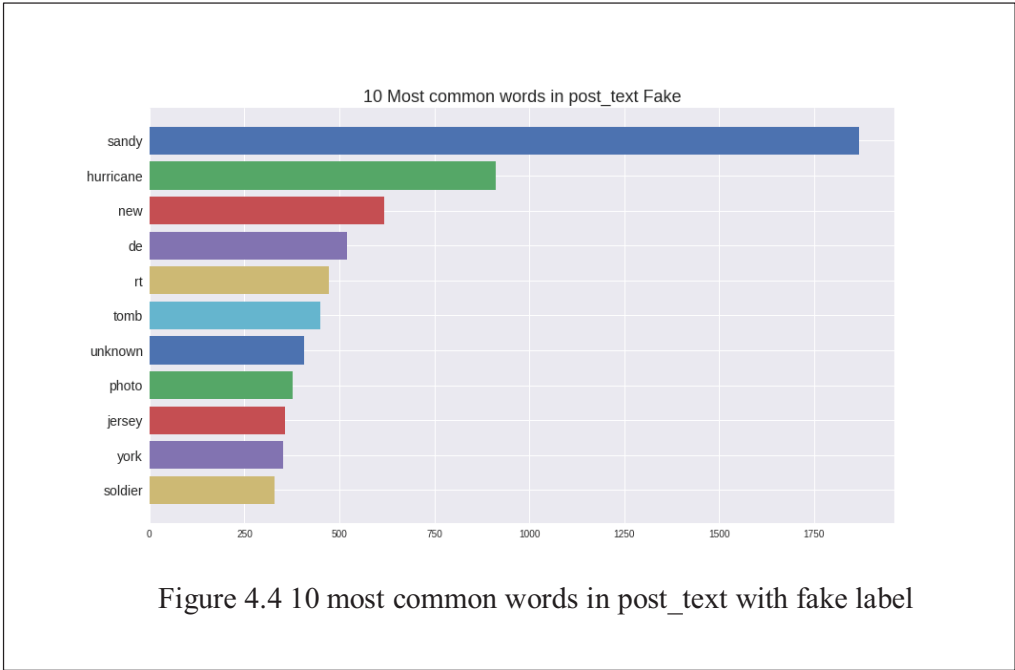
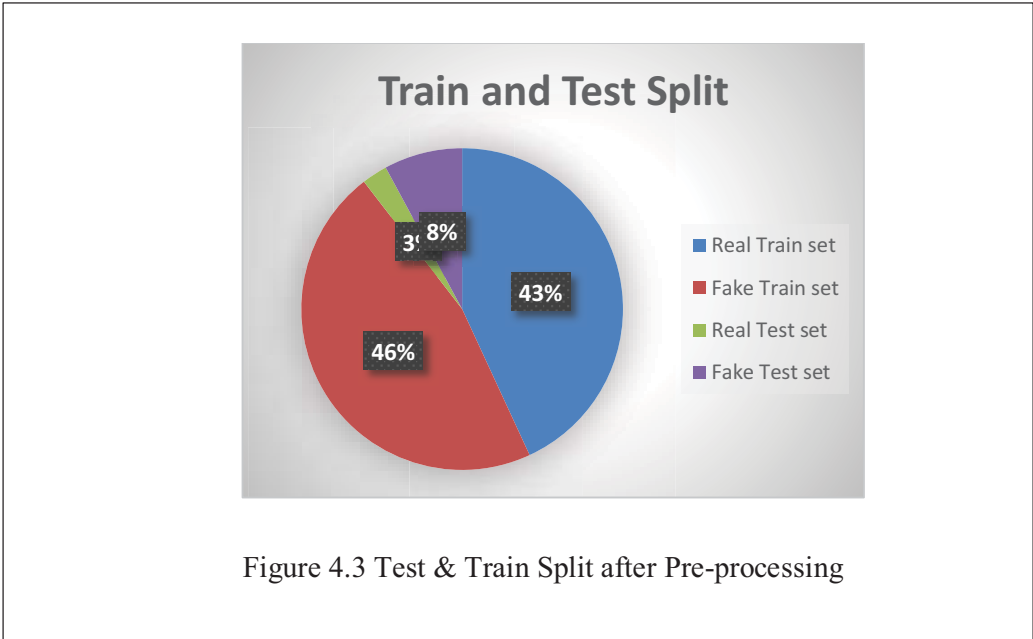


Figure 4.2 Data Statistics for Twitter Dataset After Cleaning



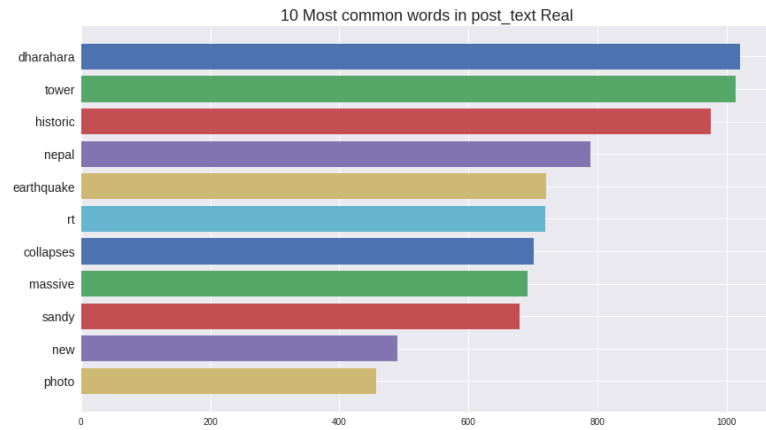


Figure 4.5 10 most common words in post\_text with real label



Figure 4.6 Word cloud for fake and real post

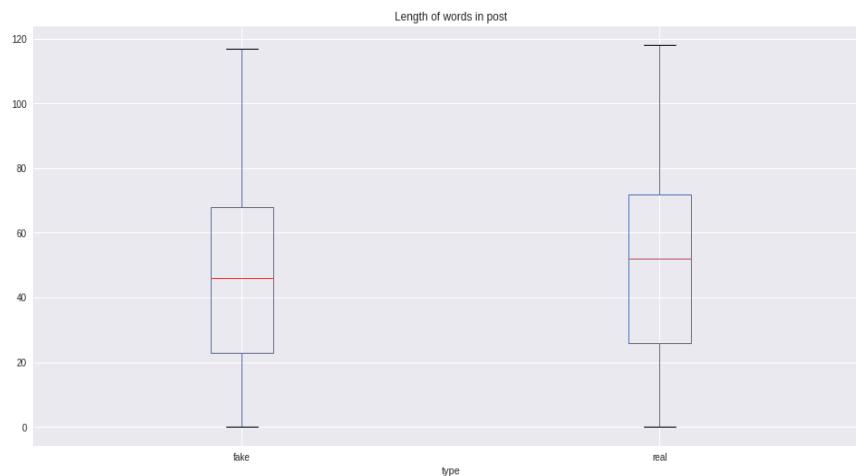
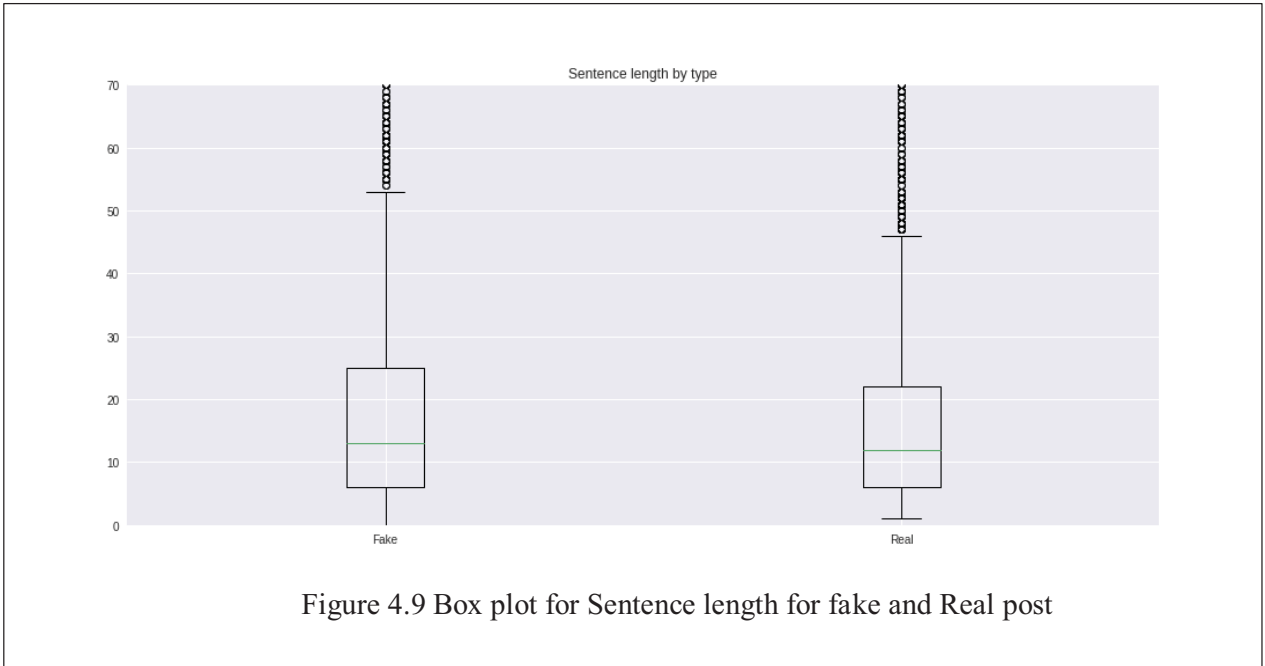
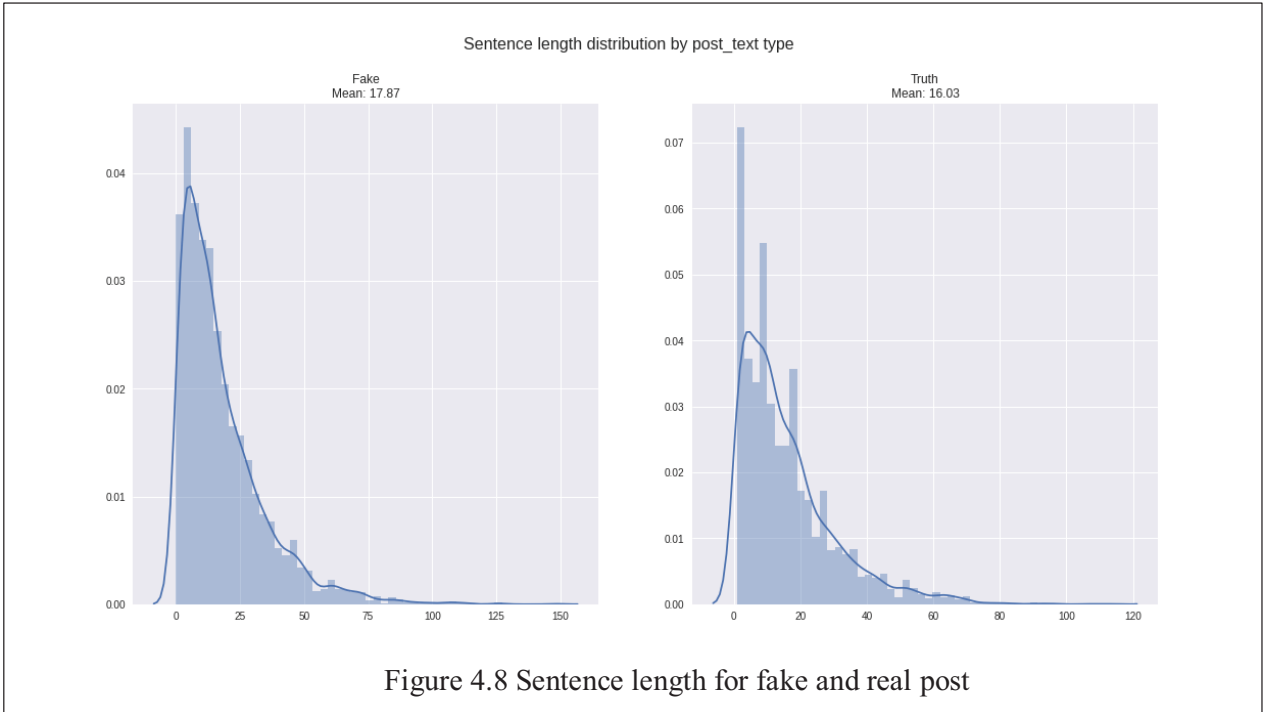


Figure 4.7 Length of words in post for fake and real





### 4.3.2 “all\_data” dataset

#### 4.3.2.1 Dataset Filtering

Similar procedure is used for all\_data dataset as mentioned in section 4.3.1.1. Firstly, the csv file of size 131.1MB is downloaded and saved to database. Since proposed model requires multi-modal data, the columns or features used from csv are title, image\_url and type. Fig 4.10-4.13 illustrate the dataset with main columns required for proposed model.

Images are downloaded from the features “image\_url” columns and stored in database. All the images are stored in the database and further used for image feature extraction.

Each title is verified whether corresponding image is present in the database. If no, title is discarded. After filtering post with images following is the stats

#### 4.3.2.2 General Analysis

Linguistic analysis is performed with the similar procedure mentioned in 4.3.1.2. The figure 4.13-4.19 plots showed the findings on all\_data.csv

Unnamed: 0	Unnamed: 0.1	author	comments	country	crawled	domain_rank	id	language	likes	...	fear	joy	sadness	surprise	trust	negative
0	1	NaN	JEREMY W. PETERS	0.0	US	2017-03-14 08:25:04	0 3.0	english	0.0 ...	6	20	5	14	30	14	
1	2	NaN	STEVE EDER	0.0	US	2017-03-14 08:25:36	0 4.0	english	0.0 ...	4	4	4	5	9	8	
2	3	NaN	MAGGIE HABERMAN ASHLEY PARKER	0.0	US	2017-03-14 08:25:36	0 5.0	english	0.0 ...	8	15	8	6	26	15	
3	4	NaN	NELSON D. SCHWARTZ SUI-LEE WEE	0.0	US	2017-03-14 08:25:36	0 6.0	english	0.0 ...	10	10	10	6	32	24	
4	5	NaN	MAGGIE HABERMAN	0.0	US	2017-03-14 08:25:37	0 7.0	english	0.0 ...	3	6	2	4	14	4	

5 rows x 54 columns

Figure 4.10 all\_data Dataset fields(main fields :title,image\_url,type)

	uuid	title	main_img_url	type
0	f182f05dc3191ba4cb741e22f75fb43b	At Donald Trump<U+2019>s Properties, a Showcas...	https://static01.nyt.com/images/2016/11/23/us/...	real
1	220b87845a5eb01509b66c8008bf3728	Trump Foundation Tells New York It Has Stopped...	https://static01.nyt.com/images/2016/10/18/us/...	real
2	247e97e1da2dc67fcb31e20b84b2d960	Donald Trump Prepares for White House Move, bu...	https://static01.nyt.com/images/2016/11/12/us/...	real
3	e1f572512a36071cbca6056a31577389	Luring Chinese Investors With Trump<U+2019>s N...	https://static01.nyt.com/images/2016/10/21/bus...	real
4	584700e476e0d3c20731cb3d28e6ce2b	Melania and Barron Trump Won<U+2019>t Immediat...	https://static01.nyt.com/images/2016/11/21/us/...	real

Figure 4.11 all\_data Dataset fields with relevant fields

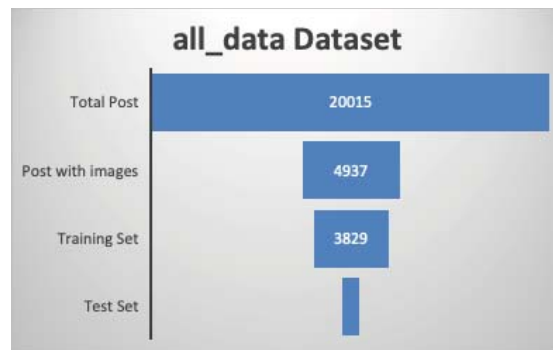


Figure 4.12 Data Statistics for all\_data Dataset After Cleaning

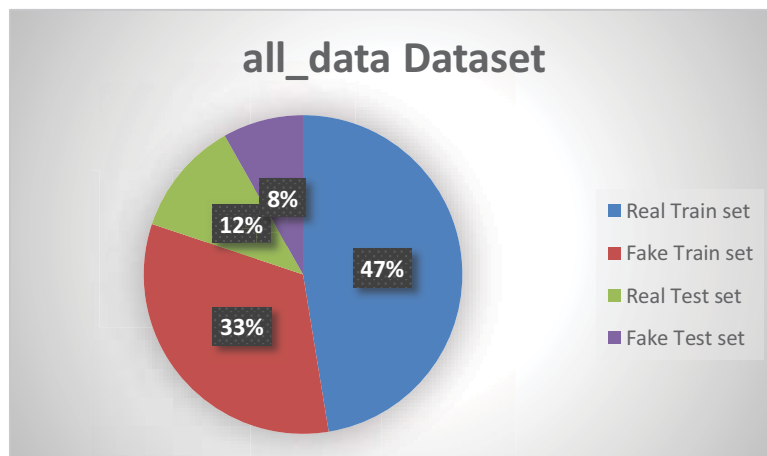


Figure 4.13 Test & Train Split after Pre-processing

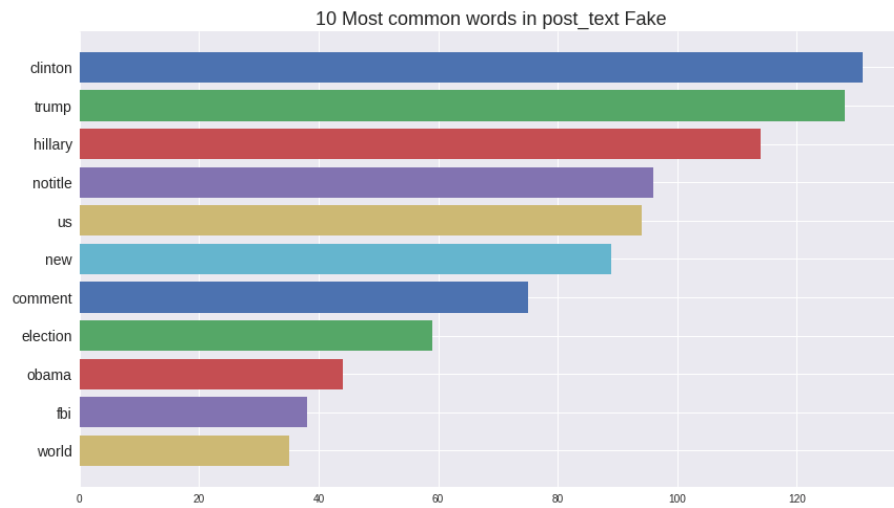


Figure 4.14 10 most common words in post\_text with fake label

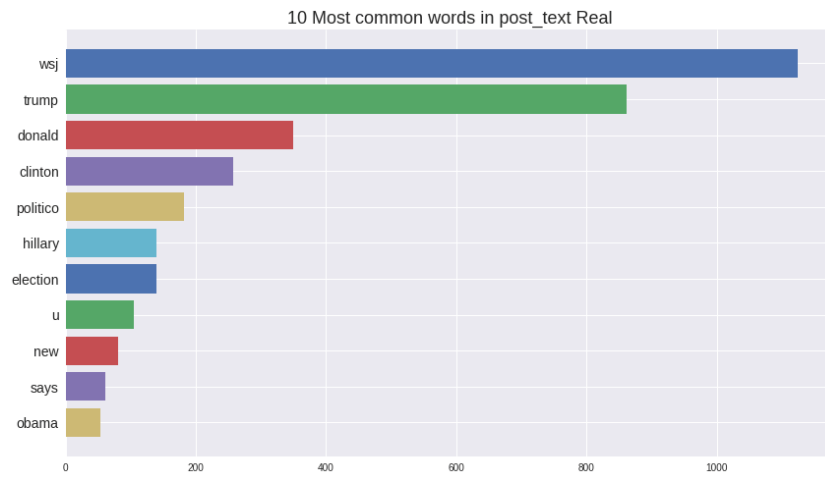


Figure 4.15 10 most common words in post\_text with real label



Figure 4.16 Word cloud for fake and real post

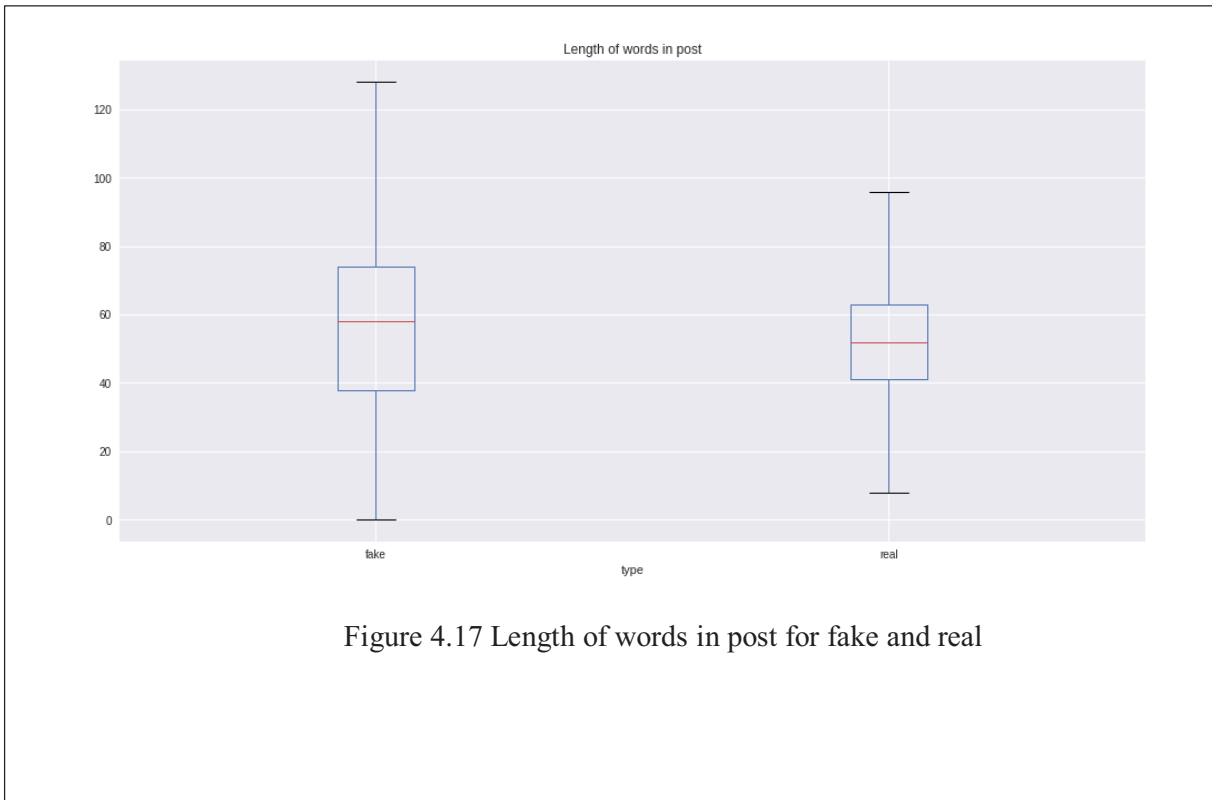
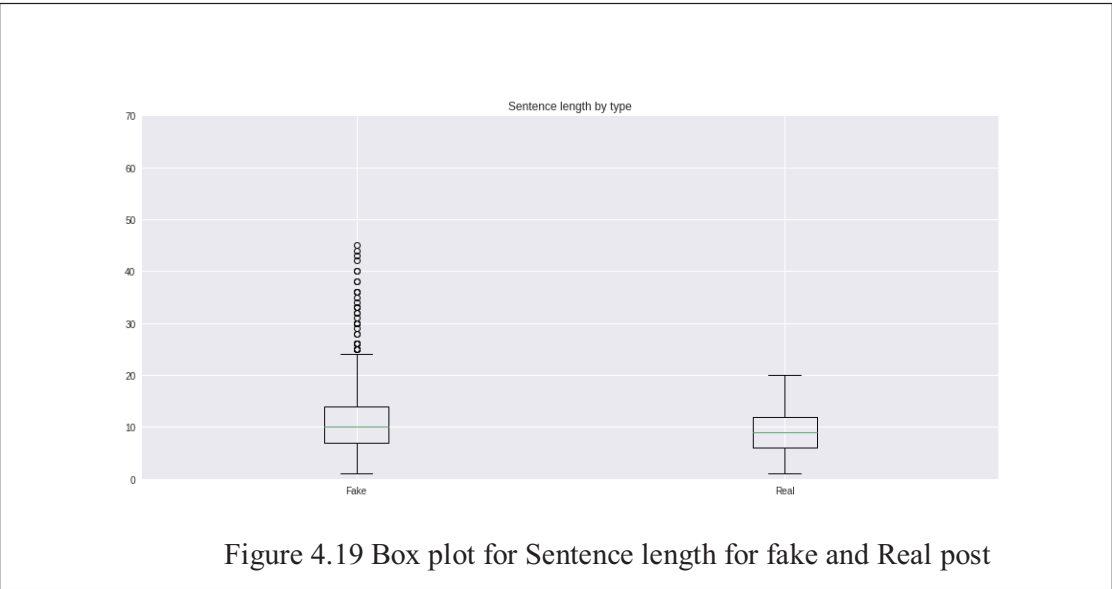
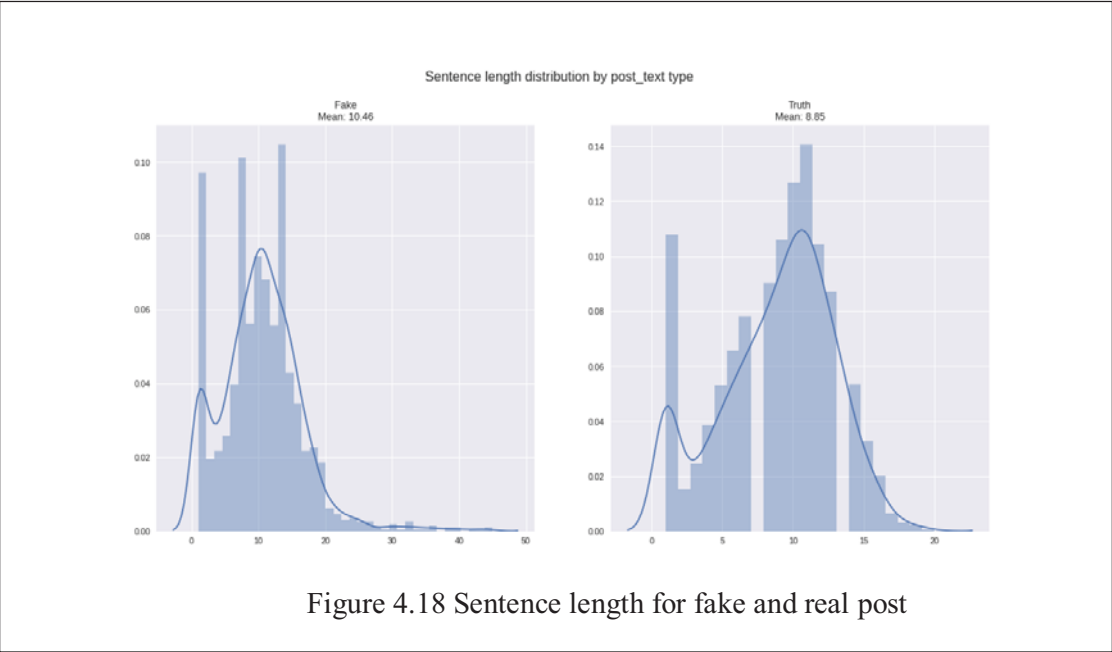


Figure 4.17 Length of words in post for fake and real



## 4.4 VISULIZATION WITH t-SNE

So to create a 2D picturization of the dataset, it is expected to change text into numerical manner and eventually lessen the measurement so as to permit it to be plotted on a 2D or 3D plot. Here word2Vec using the Gensim model is utilized. This produces a matrix of size number of samples \* word embedding size (200 dimension). All the post or title are read line by line and tokenize into words. Stop words are removed and text is lowered. Then each word from the sentence of post is embedded from word2Vec and store in a data frame. This will create a corpus of large array. For instance, a corpus of 12000 samples would create a grid as a 12000×200 scanty grid. As said previously, plot-chime in 200 measurement is beyond the realm of imagination. So as to do as such, the quantity of measurements should be diminished. Here, head segment investigation before t-SNE [42] will be utilized together. Figure 4.20 and 4.21 illustrate t-sne plot for the above mentioned procedure

## 4.5 CONCLUSION

Data exploration has helped to decide maximum length of sentences used as a input to our encoder network.t-sne plot plotted for both dataset have given and insight that classification can be easily performed on all\_data as compared to Twitter data because of scattered result.

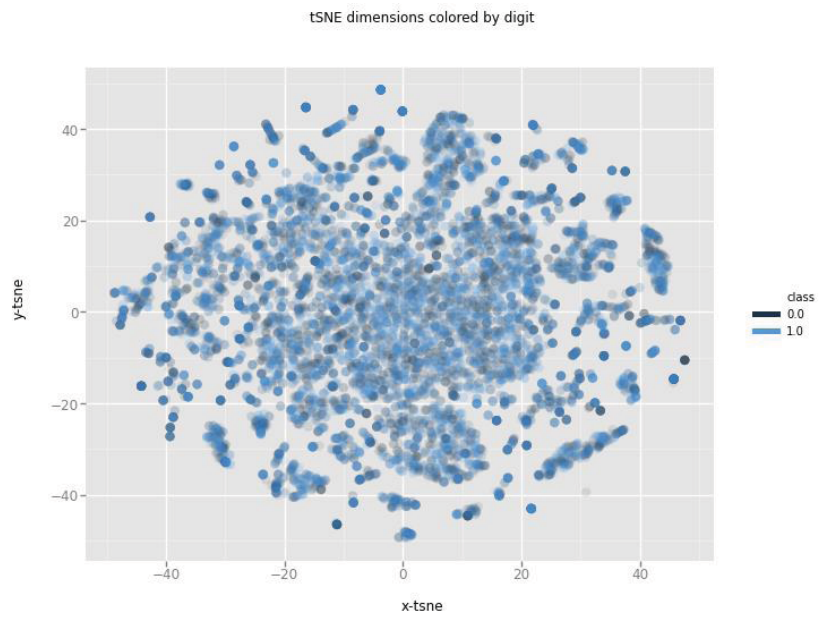


Figure 4.20 t-SNE plot for Twitter Dataset(1-real & 0-fake)

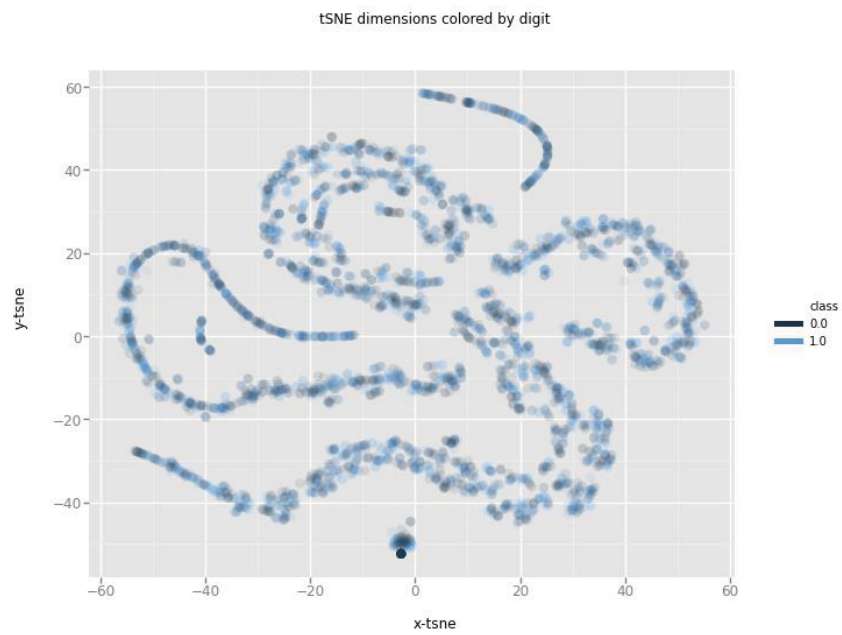


Figure 4.21 t-SNE plot for all\_data Dataset

# **CHAPTER 5**

## **THE PROPOSED WORK**

### **5.1 INTRODUCTION**

#### **5.1.1 Problem Statement**

As number of individuals utilizing on the web based life is expanding so is the substance of fake news on it. Utilizing it to control or change the perspectives on individuals to ones benefit. Past occasions have indicated fake phony news can make decimating changes. The point of the task is to distinguish counterfeit news in online web based life utilizing text and picture based highlights and attempt to think of a superior classifier which can be utilized as an instrument to sift through fake news from genuine news.

#### **5.1.2 Motivation**

Recognizing Fake news is a difficult that is been long open and there is no clear answer for it. Specialists have recently analyzed highlights like content substance, text style, client commitment, validity of client distributed it. A common new article on online web based life comprises of title, text content, top picture and client responses and remarks. We will concentrate on making the classifier multimodal wherein it can likewise utilize picture highlights to more readily recognize counterfeit news.

### **5.2 VISUAL FEATURE EXTRACTION**

Visual signs have been demonstrated to be a significant controller in detecting fake news[43]. As it is mentioned , detection in fake news abuses the personal exposure of individuals and accordingly frequently depends on electrifying or even counterfeit pictures to incite outrage or enthusiastic reaction of buyers. Image-based highlights is





Text: Nepal Quake: Death Toll hits 3700  
#NepalEarthquake #death

Figure 5.1 Example of Fake news from Twitter dataset

extricated from image components (for example pictures and recordings) to catch the unique attributes for fake news. Figure 5.1 gives us the insight of how fake image and text can be different and can be utilized for checking fake news. Faking pictures were recognized in light of different client level and tweet-level hand-created highlights utilizing arrangement system [44]. As of late, different visual and factual highlights has been extricated for news confirmation [16]. Visual highlights incorporate lucidity score, soundness score, comparability circulation histogram, decent variety score, furthermore, grouping score. Measurable highlights incorporate check, picture proportion, multi-picture proportion, hot picture proportion, long picture proportion, and so on. Various Deep CNN image model used in proposed model:

### 5.2.1 VGG-19 [45]

VGG-19 is a 19 layers deep CNN. The system contains layers of convolution pooling layers and fully connected layers pretrained on millions of images of ImageNet database. It classify the images into 1000 classes for example like cat, dog and other numerous creatures. Hence VGG-19 is a complete system which has learned from a million of images and we can use this model to extract image features of fake news detection. The input size of VGG-19 is  $224 * 224$ .

### 5.2.2 Resnet50 [46]

Resnet50 is a 50 layers deep CNN. The system contains layers of convolution pooling layers and fully connected layers pretrained on millions of images of ImageNet database. It classify the images into 1000 classes for example like cat, dog and other numerous creatures. Hence Resnet50 is a complete system which has learned from a million of images and we can use this model to extract image features of fake news detection. The input size of Resnet50 is  $224 * 224$ .

### 5.2.3 Inception-v3 [47]

Inception-v3 is a 48 layers deep CNN. The system contains layers of convolution pooling layers and fully connected layers pretrained on millions of images of ImageNet database. It classify the images into 1000 classes for example like cat, dog and other numerous creatures. Hence Inception-v3 is a complete system which has learned from a million of images and we can use this model to extract image features of fake news detection. The input size of Inception-v3 is  $229 * 229$ .

### 5.2.4 Image feature Extraction steps:-

- Download all the images from the image\_url mentioned in file set\_images.txt in Twitter dataset and image\_url of all\_data.csv .
- Split the dataset into training and test split for all\_data dataset.
- For twitter dataset splitting was already provided as devset and testset
- Load all the images as list using os.listdir().Read the images one by one from list.
- Convert the image size using image.load\_img(image\_path, target\_size=(224, 224)) to the desired input size for various deep CNN model
- preprocess\_input() of vgg-19 ,resnet50 and inception v3 is used to preprocess all the images.It subtracts the mean RGB value from each pixel
- Keras model for vgg19 ,resent50 and inception v3 is created:

- `tf.keras.applications.VGG19(include_top=True,weights="imagenet",input_tensor=None,input_shape=None,pooling=None,classes=1000,classifier_activation="softmax",)` [49]
  - `tf.keras.applications.ResNet50(include_top=True,weights="imagenet",input_tensor=None,input_shape=None,pooling=None,classes=1000,**kwargs)` [49]
  - `tf.keras.applications.InceptionV3(include_top=True,weights="imagenet",input_tensor=None,input_shape=None,pooling=None,classes=1000,classifier_activation="softmax",)` [49]
- Last 2 layers are removed from each deep cnn model.
  - Concatenate vgg-19,resnet-50 and inception v3 extracted features
  - Store the image features in pickle file

## 5.3 TEXT FEATURE EXTRACTION

- Filter post and title based on image pickle. Create `train_post-with_img.txt`
- Extract the `post_text` from `post.txt` of dev and test set from Twitter dataset and title from `all_data.csv`.
- Convert all upper bound post to lower bound,remove urls and junk charcter.
- Check whether the the post is in English language or not. If no, convert every post to English with the help of `googletrans` .
- Store the translated tweets in a pickle file.
- Read each post one by one, remove the stops words like (is,a,the,have...) from the post. Tokenize the post text.
- After preprocessing, skip the post with length less than 2.
- Create a `Word2Vec` model using `genism`. Create the embeddings of each and every word present in the post. Save the embedding matrix in numpy array.
- Set the sequence length to 20. Post having length less than 20 is appended worth zero.

- Read the file line by line. Construct the input vector of text of size no of samples \* 20. For image #samples \* 8192(concatenated vector of vgg19, resnet50, inceptionv3), embedding matrix of size #samples \* 32 and label #samples \* 1.

## 5.4 VARIATIONAL AUTOENCODER

VAE consists of an encoder, decoder and loss function as its three main components. The encoder is a neural system. Its information is a datapoint  $x$ , its yield is a concealed portrayal  $z$ , and it has weight and bias as  $\theta$ . For example, suppose  $x$  is a 28 \* 28-pixel photograph of a written by hand number. The encoder 'encodes' the information which is 784-dimensional into an idle (covered up) portrayal space  $z$ , which is considerably less than 784 measurements. This is commonly alluded to as a 'bottleneck' in light of the fact that the encoder must get familiar with an effective pressure of the information into this lower-dimensional space. How about we indicate the encoder  $q_{\theta} = (z|x)$ .

The decoder is another neural net. Its information is the portrayal  $z$ , it yields the boundaries to the likelihood conveyance of the information, and has weights and bias  $\phi$ . The decoder is meant by  $p_{\phi} = (x|z)$ . Running with the manually written digit model, suppose the photographs are high contrast and speak to every pixel as 0 or 1. The decoder gets as information the inert portrayal of a digit  $z$  and yields 784 Bernoulli boundaries, one for every one of the 784 pixels in the picture. The decoder 'interprets' the genuine esteemed numbers in  $z$  into 784 genuine esteemed numbers somewhere in the range of 0 and 1. Data from the first 784-dimensional vector can't be completely transmitted, in light of the fact that the decoder just approaches a synopsis of the data (as an under 784-dimensional vector  $z$ ). What amount of data is lost? We measure this utilizing the reconstruction lost

In the VAE, our misfortune work is made out of two sections: Reconstruction Loss: This misfortune contrasts the model yield and the model info. This can be the misfortunes we utilized in the autoencoders, for example, L2 Loss. KL Divergence Loss: This misfortune contrasts the idle vector and a zero mean, unit difference Gaussian circulation.

The misfortune we use here will be the KL disparity misfortune. This misfortune term punishes the VAE in the event that it begins to create idle vectors that are not from the ideal appropriation.

## 5.5 PROPOSED MODEL

### 5.5.1 Overview

Our model depicted in Fig 5.1 is based on Variational Auto Encoder to communicate the detection problem in fake news. The main intention around this model is to combine the two modalities (text and images) of a tweet into single form. It comprises of three components –

- **Encoder:** To encode data from images and text into a feature vector.
- **Decoder:** To construct data of images and text into original form
- **Fake news detector:** It uses the encoded information to categorize the fake or real information

### 5.5.2 Encoder

The text and image from a post are passed into this encoder and subsequently a combine vector of both the text and image features is the output from encoder. This encoder itself is divided into two smaller components as – textual encoder and visual encoder.

1. **Textual Encoder:** The feed in to this sub part of encoder, textual encoder, is a stream of words present in the post represented as a vector  $W$ , where  $W = [W_1 W_2 \dots W_n]$ , where  $n$  is the number of words in the post. Each word, a part of vector, is pre-processed by removing the URL, removing junk character from the post. Each word is embedded using “Word2Vec” modal.

To bring out the relative features from a text, we use Bidirectional LSTM (Long Short Term Memory) cells. LSTM memory cells are used for extracting the textual features.

The final state of LSTM output can be attained by concatenating the data from different layers and corresponding states. And, to retrieve the textual features from this a formula can be used –

$$TF = Fn(W_{FC} OP_{LSTM}) \quad (5.1)$$

Where –

TF = Textual features,  $W_{FC}$  = Weight matrix of fully connected layers,  $OP_{LSTM}$  = Output from LSTM cells and Fn is the function to calculate the final output.

2. **Visual Encoder:** If the text is sent as input to textual encoder then images attached with a post are sent as input to this visual encoder. We use the pre-trained VGG-19 [15], pre-trained resnet50 [20] and pre-trained inceptionV3 [21] architecture network pre-trained over the ImageNet database, and use the last full-connected layer as output. Concatenation of latent feature vector from ResNet50, VGG-19 and Inceptionv3 is used to train the network. Concatenation of these three CNN architecture features is performed and to get the similar sized representation of text and image the concatenated vector passed through fully connected layers. The function for the same is-

$$VF = Fn(W_{FC} OP_{VGG-19 \oplus Resnet50 \oplus InceptionV3}) \quad (5.2)$$

Where-

VF = Visual features,  $W_{FC}$  = Weight matrix of fully connected layers,  $OP_{(VGG-19 \oplus Resnet50 \oplus InceptionV3)}$  feature representation from VGG-19, resnet50, inceptionV3 and Fn is the function to calculate the final output.

In the later stage, the textual and image representations are concatenated together and passed into a fully connected layer to get a shared representation of image and

text. The TF and VF are concatenated and passed through a fully connected layer to form the final shared representation named as RF.

### 5.5.3 Decoder

A decoder is the inverse of encoder. This is used to recreate information from inspected multimodal portrayal. Same as encoder, this decoder too has two components as textual decoder and visual decoder.

- **Textual Decoder:** The multimodal representation is passed as input to this decoder and it recreate the words back. The multimodal is gone through a fully connected layer which generates input for bi-directional LSTMs. The output from LSTM is again passed into fully connected layers and an activation function by which the final output is constructed as a word.
- **Visual Decoder:** The task of visual decoder is to construct back the images from given input. It reconstructs the Concatenated output of VGG-19, Resnet50 and InceptionV3 features from multimodal. The multi-modal representation is passed through multiple fully connected layers to get the expected output.

The decoder decodes the visual and textual features from a shared representation and presents the final output.

$$(WF, IF) = Decoder(RF, P_{DEC}) \quad (5.3)$$

Where-

WF is final textual word, IF is final visual representation from multimedia, RF is final output from encoder and  $P_{(DEC)}$  is all the corresponding parameters.

### 5.5.4 Fake News Detector

This part of the model takes multi-modal representation as an input and classifies the same as fake or a real post. This too comprises multiple fully connected layers having their own activation functions. Fake news detector denoted by  $F_{FND}$  is a function that gives the score in range between 0 and 1 to all the multimedia posts.

$$OP_{FND} = F_{FND} (RF, \phi FND) \quad (5.4)$$

Where-

$OP_{FND}$  is the probability of a multimedia post being fake or real,  $RF$  is final output from encoder and  $\phi FND$  denotes all the parameters in fake news detector.

A score '1' means news is fake while a score of '0' stamps the news as real. We use the sigmoid logistic function to constrain the values between 0 and 1. Therefore, we use cross-entropy to measure the loss of classification

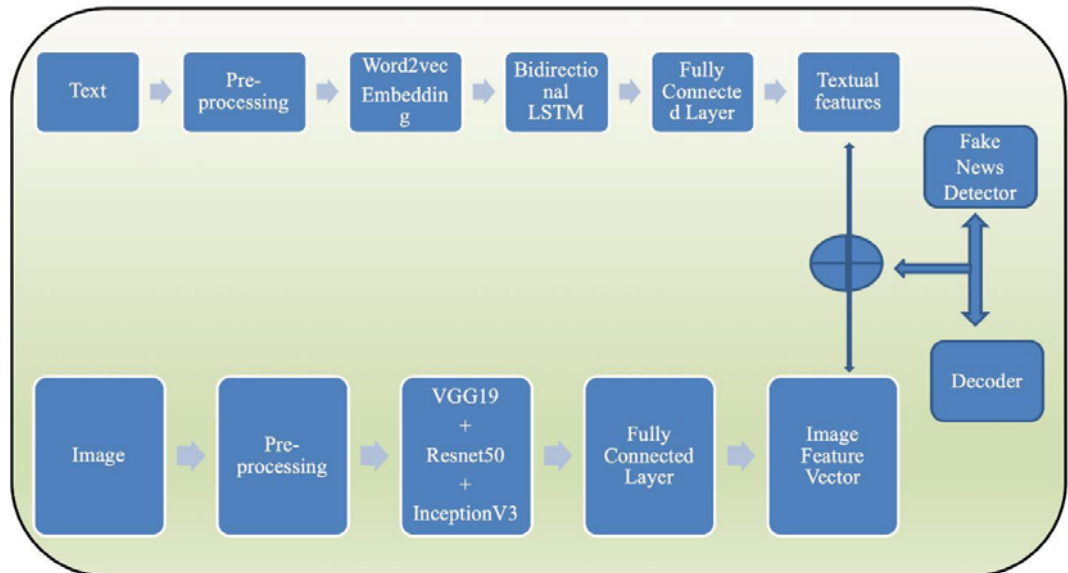


Figure 5.2 Architecture of Encoder of Proposed Model



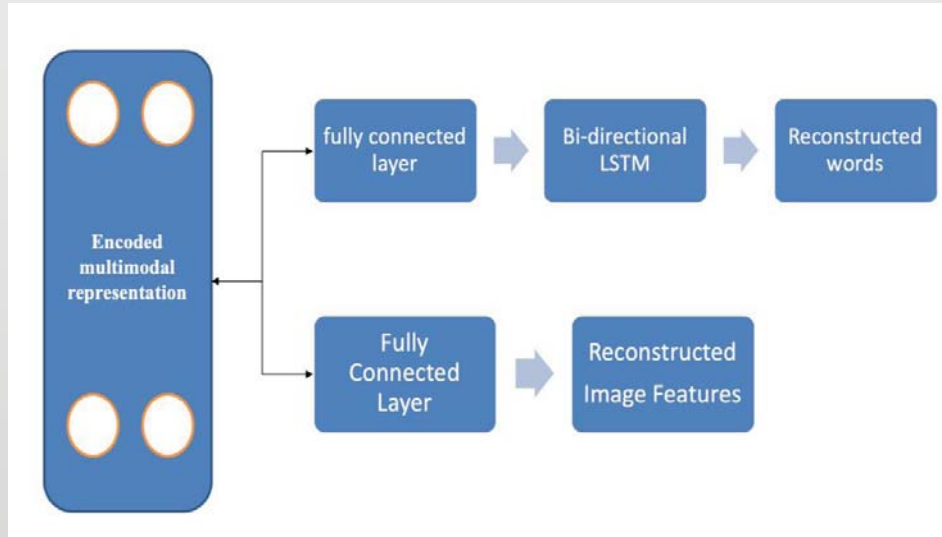


Figure 5.3 Architecture of Decoder of Proposed Model

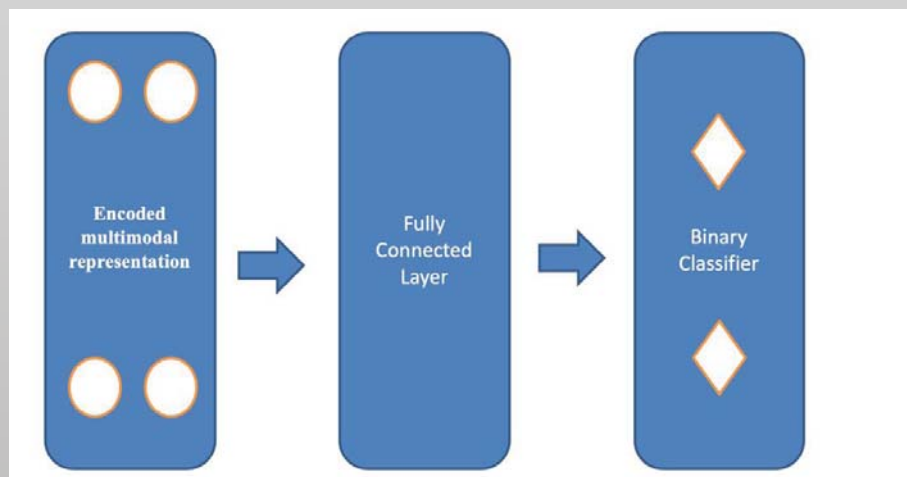


Figure 5.4 Architecture of Fake news Detector of Proposed Model

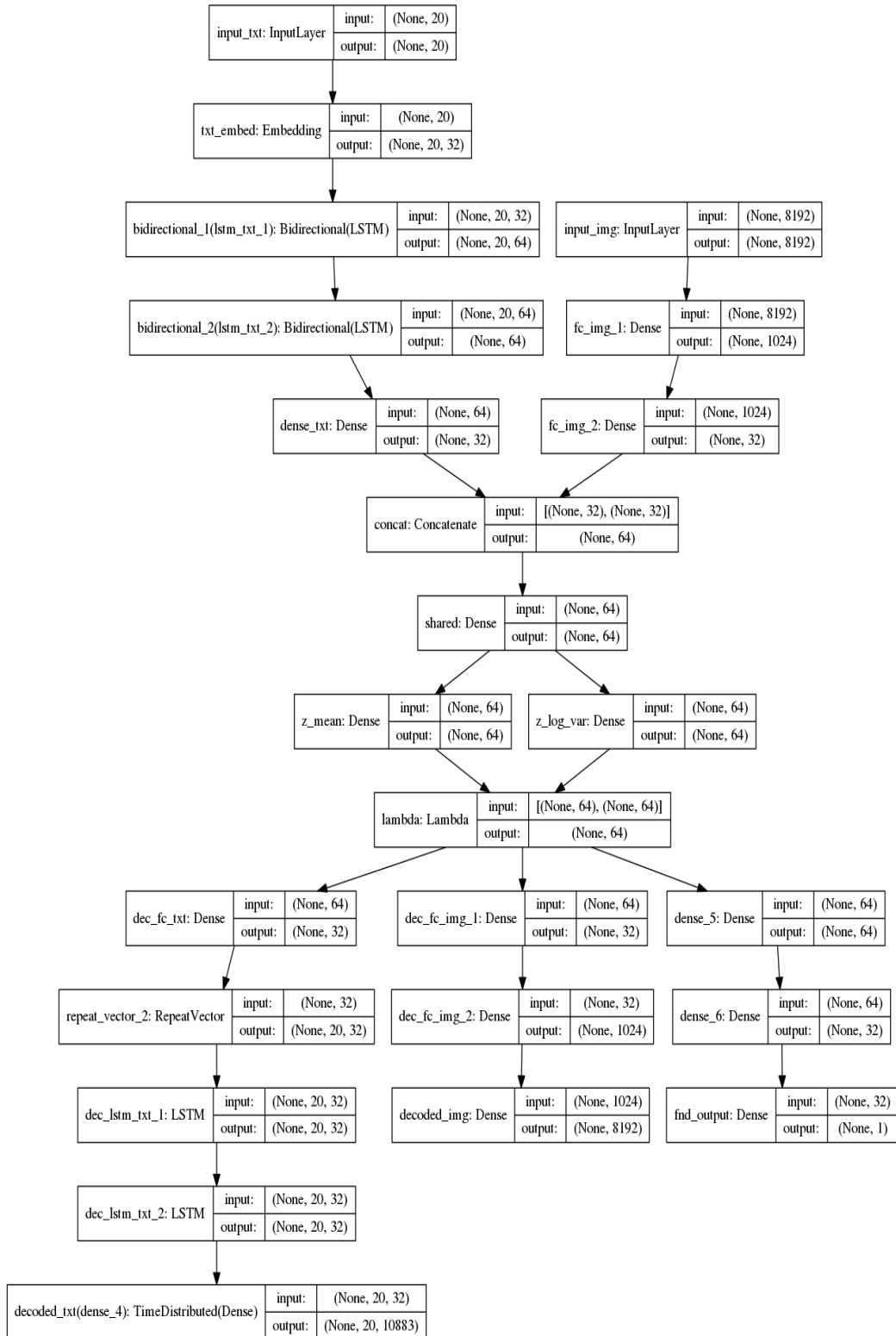


Figure 5.5 The proposed Model in Keras

# CHAPTER 6

## SOFTWARE REQUIREMENT AND METHODOLOGY VALIDATION

### 6.1 SOFTWARE REQUIREMENT

In order to design the Fake News Detection System based on Variational Autoencoder, Python is used as language. Anaconda tool is the basic requirement for creating the model. The model can run on Windows/Mac/Linux.

#### 6.1.1 Python

Python is a high level programming language which have the capabilities of object oriented programming language Its syntax is like English language that's the reason python is used for various machine learning and deep learning models. It's freely available as it's a open source programming language. Due to the libraries' available for machine learning and deep learning it has become a very popular language as it save lot of time .we can easily create algorithm and mathematical notation in Python .Following are the libraries which we have used to create our proposed model.

- Numpy
- Scikit-learn
- Keras
- Mathplotlib
- Pandas

To create Neural Network Model or any Deep Learning model Keras is the popular library of python. It runs on TensorFlow or Theano as its background. Keras was made to be easy to use, measured, simple to broaden, and to work with Python. The API was "intended for people, not machines," and "follows best practices for decreasing subjective burden." .New modules can be easily include, as new classes

and capacities. Models are characterized in Python code, not independent model design documents.

### **6.1.2 Hardware/Software Requirements**

Following is the specification of hardware required to create out proposed model:

1. Central Processing Unit (CPU): Intel core i5 or above, Quad core or higher microprocessor based system can be utilized.
2. GPU: GTX 1050 or above
3. RAM:8GB of RAM is required to run the model
4. Monitor - A 17" or larger VGA or better quality monitor/TFT/LCD.
5. Memory - 8GB of RAM is recommended.
6. Disk space – Disk space with 256 GB at least required to run the model.
7. Software: Anaconda with Python 3.5 and above is required for model Execution

## **6.2 EXPERIMENTAL SETTINGS**

Creating word to vector embedding , we use the distributed representation of Word2Vec for words. The data set contains tweets which had language other than english also. Those tweets have firstly been translated into english to keep the information intact and in machine readable format. We implement pre-processing standard text for Twitter dataset.

For visual contents, the concatenation of final layer output of a pre-trained VGG-19, ResNet5 and InceptionV3 on ImageNet set has been extinguishly used. The characteristic dimensions obtained from concatenation of latent features is 8192.

For text, we use LSTM and fully connected layer with dimension 32 each. On one end, the visual encoder comprises of two fully connected layers of dimension 1024 and 32 The third component of our model, the fake news detector, is made of two layers in size as 64 and 32. These two layers are fully connected

## 6.3 PERFORMANCE MEASURE

There are 4 important terms to measure performance:

**True Positives:** When the predicted and actual labels are fake

**True Negatives:** When the predicted and ground truth labels are real

**False Positives:** When the ground truth label is real but we predicted fake .

**False Negatives:** When the ground truth label is fake but we predicted real.

Following are the measure which we have employed for performance evaluation:

### 1. Accuracy:

- `sklearn.metrics.accuracy_score(y_true, y_pred, *, normalize=True, sample_weight=None)` [48]
- It provide us accuracy level of the model using test label and predicted output.
- It's a proportion of exactness ,it match home much our predicted label matched with actual ground truth label
- $Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Samples}$

### 2. Precision Score:

- `sklearn.metrics.precision_score(y_true, y_pred, *, labels=None, pos_label=1, average='binary', sample_weight=None, zero_division='warn')` [48]
- Calculated as  $\frac{True\ Positive}{True\ Positive + False\ Positive}$
- The precision is naturally the capacity of the classifier not to name as positive an example that is negative

### 3. Recall

- `sklearn.metrics.recall_score(y_true, y_pred, *, labels=None, pos_label=1, average='binary', sample_weight=None, zero_division='warn')` [48]
- $\frac{True\ Positive}{True\ Positive + False\ Negative}$

- The recall is instinctively the capacity of the classifier to discover all the positive examples.

#### 4. F1 Score

- `sklearn.metrics.f1_score(y_true, y_pred, *, labels=None, pos_label=1, average='binary', sample_weight=None, zero_division='warn')` [48]
- The F1 score can be deciphered as a weighted normal of the exactness and review, where a F1 score arrives at its best an incentive even from a pessimistic standpoint score at 0. The general commitment of exactness and review to the F1 score are equivalent
- Calculated as  $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$

#### 5. Classification Report

- `sklearn.metrics.classification_report(y_true, y_pred, *, labels=None, target_names=None, sample_weight=None, digits=2, output_dict=False, zero_division='warn')` [48]
- Shows the report of Precision recall f1-score and accuracy

Also, Receiver Operating characteristic curves (ROC curves) have been plotted for each dataset and model used. ROC curves help in performance measurement of binary classifier system at different threshold settings. In this curve, true positive rate is plotted on Y axis and false positive rate is plotted on X axis. These two are plotted with 100 specificity and various cut off points. Also, the area under the curve measures discrimination, that is, classifying. More the area, better the classification.

## CHAPTER 7

### RESULTS, CONCLUSION and FUTURE SCOPE

#### 7.1 INTRODUCTION

This chapter highlights the major experimental results, conclusion, and the possible future work based on Variational Autoencoder based Fake News Detection. The main aim of this chapter is to provide all the experimental findings of the research work and to conclude the research work.

#### 7.2 EXPERIMENT RESULTS

The data exploration of twitter dataset and all\_data.csv is done in section 4.2.1 and respectively. Table 7.1 displays both the MVAE [1] findings and our suggested approach on two datasets. The accuracy of our fake news detector in case of fake and real news is reported through our model and it significantly improves by adding more visual latent features to the model. Accuracy graph is also plotted in figure 7.1 and 7.5

A Precision Recall curve is essentially a diagram with Precision esteems on the y-hub and Recall esteems on the x-hub. As it were, the PR bend contains  $TP/(TP+FN)$  on the y-hub and  $TP/(TP+FP)$  on the x-pivot. Note, that Precision is additionally called the Positive Predictive Value (PPV). Review is additionally called Sensitivity, Hit Rate or True Positive Rate (TPR). Curve showed a high precision and high recall.

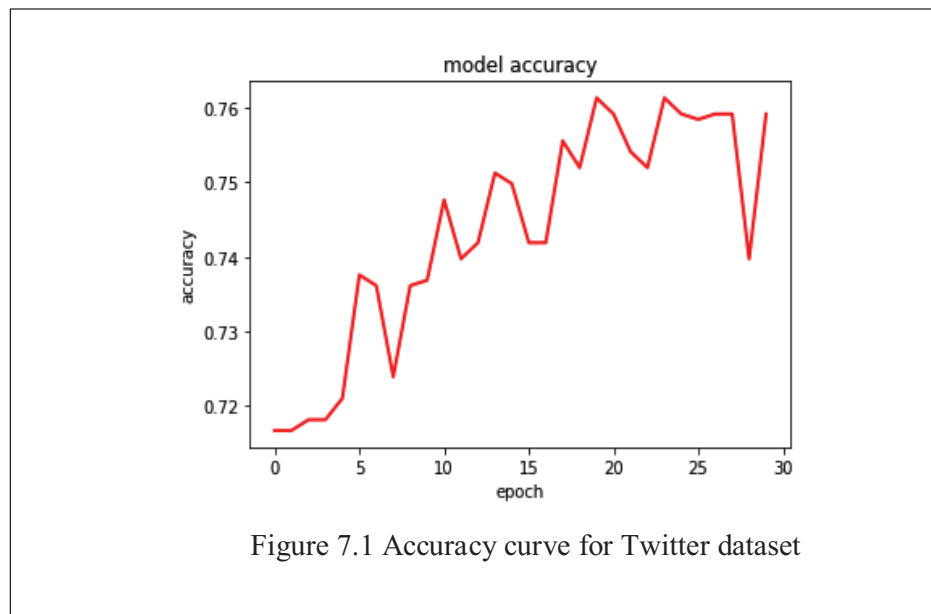
ROC curve is a 2-D bend parametrized by one boundary of the characterization calculation. AUC is consistently somewhere in the range of 0 and 1. ROC bend can be acquired plotting TPR on y-pivot and TNR on x-hub. AUC gives exactness of the proposed model.

The main highlight of ROC curve is the specificity and sensitivity trade-off, if there is a increase in sensitivity then specificity decrease's. As we can see in Figure

7.3 and 7.7 the curve is closer to left hand border. Hence the model is more accurate. Performance of classifier is calculated by AUC, having higher area i.e. .89 and .90. To get the trade-off of precision and recall, precision-recall curve is used .Figure 7.2 & 7.6 shows high area of the curve which result in high precision and recall.

Table 7.1 Performance Measure of out proposed model

Model	Accuracy	Precision	Recall	F1-Score
MVAE-Textual [1]	.52	.58	.55	.56
MVAE-Visual [1]	.59	.69	.51	.59
MVAE [1]	.74	.80	.71	.75
EANN [39]	.64	.81	.49	.61
Proposed Model - Twitter Dataset	<b>.76</b>	.83	.73	.79
Proposed Model - all_data	<b>.84</b>	.86	.70	.80





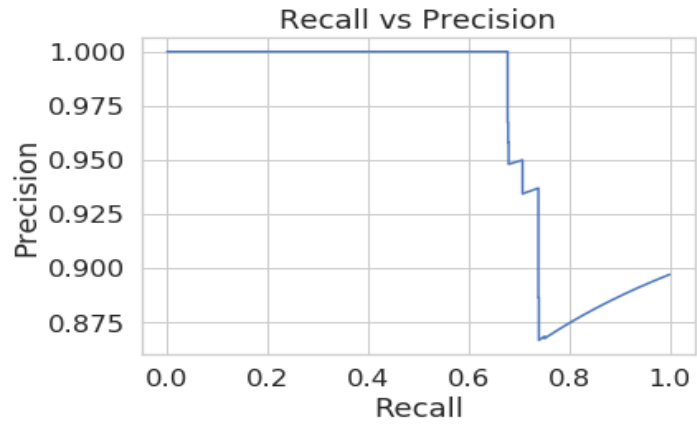


Figure 7.2 Precision-Recall curve for Twitter dataset

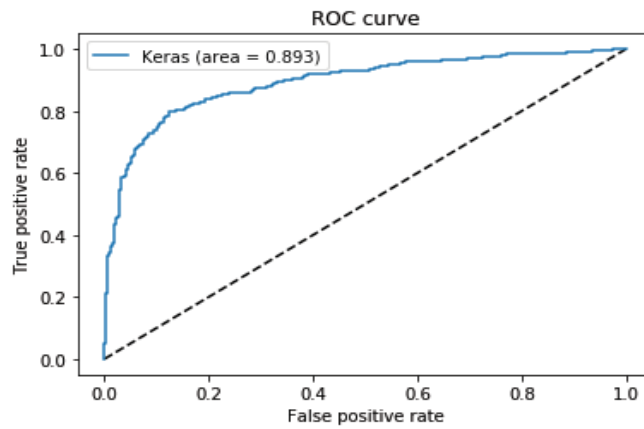


Figure 7.3 ROC curve for Twitter dataset

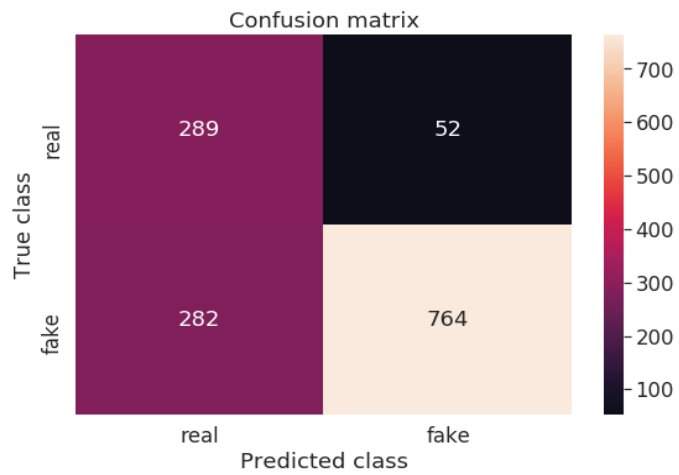


Figure 7.4 Confusion Matrix for Twitter dataset

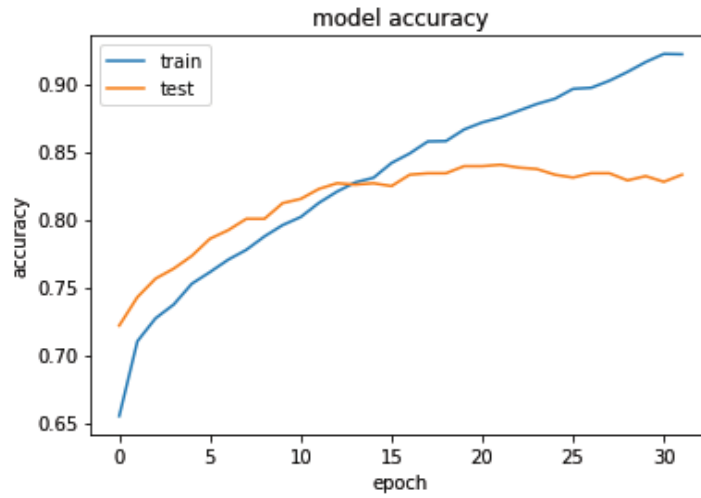


Figure 7.5 Accuracy Curve for alldata dataset

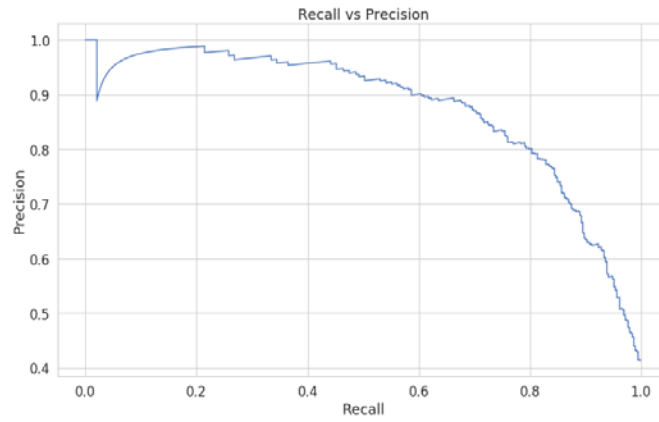


Figure 7.6 Precision -Recall Curve for alldata dataset

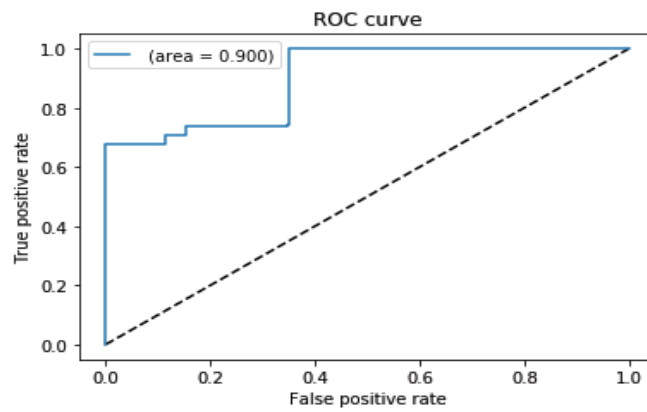


Figure 7.7 ROC Curve for alldata dataset

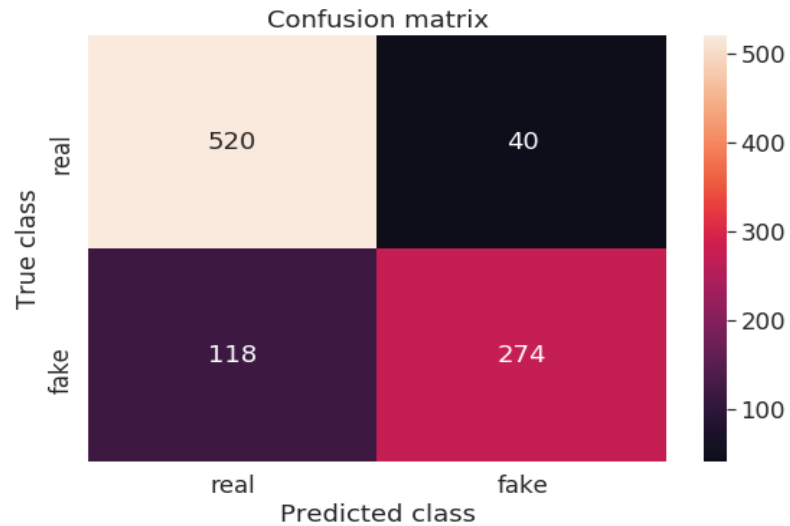




Figure 7.8 Confusion Matrix for alldata dataset

We have analysed the prediction labels of MVAE[1] and our proposed model. After training our variational encoder based model with the help of keras .The output of the trained models were hdf5 which store large amount of complex data. We load the weights stored in hdf5 file of our trained model and test the model on the training set.

The next step we did is comparison of actual labels , predicted label and predicted labels for MVAE.As illustrated in Figure 7.9 following are examples of some images which were not classified by MVAE[1] but correctly classified on our proposed model.



**Text: Not sure it real or fake .  
Pray for mh370**



**Text: Panic over, guys – Courtney Love has found missing flight MH370**

Figure 7.9 Examples of Tweets classified by our proposed model but not by MVAE [1]

## 7.3 CONCLUSION

In our thesis, we have undergone the exploration of multimodal (combination of textual and visual features) fake news detection. Overcoming the limitations of the current textual models, which uses machine learning or deep learning on textual features, we tackle the challenge of learning combination of latent features of image using three popular CNN architecture VGG-19, ResNet50 and InceptionV3 as well as textual feature.

There are three modules of our proposed model, an encoder for extracting the textual and visual features vectors, a decoder for reconstructing text and image vector and a fake news detector for classification of news post or tweets into fake or real. Our proposed model gets trained by continuous evolution and learning about the encoder, decoder and the fake news detector. The presentation assessment of our proposed design is assessed on two genuine world datasets.

Due to social media and internet fake news is all around across any social media account. This thesis tried to focus into false news characteristics and its techniques designed are reviewed. Given the challenges related to detecting the false news have made the researchers, to understand the fundamentals of those origins of fake news. The comparative analysis will help the upcoming researchers of open challenges in this field.

## 7.4 FUTURE SCOPE

The open research challenges are:

- Datasets in multi-modal: Most of the public repositories contain variants of fake news. It is also an open challenge for focussing the research objects that covers all news data types.
- Verification models on multi-modal data: Different linguistic models were designed for detecting the false news. Visual based features are difficult to recognize false news.

- Source verification: Origin of the fake news is not explored by any researchers.
- Credibility assessment: Chain of false news under the same (or) different authors are not studied from the aspects of propagation and knowledge-based features.

In future, we will add more explicit feature based on textual information or user profile data to improve our accuracy. We will also explore our model on other publically available fake news dataset, which contain images.

## References

- [1] Khattar, Dhruv, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. "Mvae: Multimodal variational autoencoder for fake news detection." In The World Wide Web Conference, pp. 2915-2921. 2019.
- [2] Boididou, Christina, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, and Yiannis Kompatsiaris. "Verifying Multimedia Use at MediaEval 2015." *MediaEval* 3, no. 3 (2015): 7.
- [3] <https://drive.google.com/open?id=0B3e3qZpPtccsMFo5bk9Ib3VCc2c>
- [4] Allcott, Hunt, and Matthew Gentzkow. "Social media and fake news in the 2016 election." *Journal of economic perspectives* 31, no. 2 (2017): 211-36.
- [5] Aneez, Zeenab, Tabereh Ahmed Neyazi, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. "Reuters Institute India digital news report." Reuters Institute for the Study of Journalism/India Digital News Report. Retrieved from [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-03/India\\_DNR\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-03/India_DNR_FINAL.pdf) on 26 (2019): 19.
- [6] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." *ACM SIGKDD explorations newsletter* 19, no. 1 (2017): 22-36.
- [7] Wwf 10yearschallenge, 2019
- [8] [https://en.wikipedia.org/wiki/Neural\\_circuit](https://en.wikipedia.org/wiki/Neural_circuit)
- [9] Zhou, Xinyi, and Reza Zafarani. "Fake news: A survey of research, detection methods, and opportunities." *arXiv preprint arXiv:1812.00315* (2018).
- [10] "Global social media ranking, 2019"
- [11] Tandoc Jr, Edson C., Zheng Wei Lim, and Richard Ling. "Defining "fake news" A typology of scholarly definitions." *Digital journalism* 6, no. 2 (2018): 137-153.
- [12] Meel, Priyanka, and Dinesh Kumar Vishwakarma. "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities." *Expert Systems with Applications* (2019): 112986.
- [13] Ajao, Oluwaseun, Deepayan Bhowmik, and Shahrzad Zargari. "Fake news identification on twitter with hybrid cnn and rnn models." In *Proceedings of the 9th international conference on social media and society*, pp. 226-230. 2018.

- [14] Roy, Arjun, Kingshuk Basak, Asif Ekbal, and Pushpak Bhattacharyya. "A deep ensemble framework for fake news detection and classification." arXiv preprint arXiv:1811.04670 (2018).
- [15] Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumor and stance jointly by neural multi-task learning." In Companion Proceedings of the The Web Conference 2018, pp. 585-593. 2018.
- [16] Jin, Zhiwei, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. "Novel visual and statistical image features for microblogs news verification." IEEE transactions on multimedia 19, no. 3 (2016): 598-608.
- [17] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." In Proceedings of the 20th international conference on World wide web, pp. 675-684. 2011.
- [18] Gupta, Aditi, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. "Tweetcred: Real-time credibility assessment of content on twitter." In International Conference on Social Informatics, pp. 228-243. Springer, Cham, 2014.
- [19] Kwon, Sejeong, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. "Prominent features of rumor propagation in online social media." In 2013 IEEE 13th International Conference on Data Mining, pp. 1103-1108. IEEE, 2013.
- [20] Ruchansky, Natali, Sungyong Seo, and Yan Liu. "Csi: A hybrid deep model for fake news detection." In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797-806. 2017.
- [21] Ma, Jing, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. "Detecting rumors from microblogs with recurrent neural networks." (2016): 3818.
- [22] Kaur, Sawinder, Parteek Kumar, and Ponnurangam Kumaraguru. "Automating fake news detection system using multi-level voting model." Soft Computing (2019): 1-21.
- [23] Orlov, Michael, and Marina Litvak. "Using behavior and text analysis to detect propagandists and misinformers on twitter." In Annual International Symposium on Information Management and Big Data, pp. 67-74. Springer, Cham, 2018.
- [24] Jwa, Heejung, Dongsuk Oh, Kinam Park, Jang Mook Kang, and Heuseok Lim. "exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder

- Representations from Transformers (BERT)." *Applied Sciences* 9, no. 19 (2019): 4062.
- [25] Zhou, Xinyi, Atishay Jain, Vir V. Phoha, and Reza Zafarani. "Fake news early detection: A theory-driven model." *Digital Threats: Research and Practice* 1, no. 2 (2020): 1-25.
- [26] Balwant, Manoj Kumar. "Bidirectional LSTM Based on POS tags and CNN Architecture for Fake News Detection." In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2019.
- [27] Volkova, Svitlana, Ellyn Ayton, Dustin L. Arendt, Zhuanyi Huang, and Brian Hutchinson. "Explaining multimodal deceptive news prediction models." In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, pp. 659-662. 2019.
- [28] Parikh, Shivam B., and Pradeep K. Atrey. "Media-rich fake news detection: A survey." In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 436-441. IEEE, 2018.
- [29] Zhang, Jiawei, Limeng Cui, Yanjie Fu, and Fisher B. Gouza. "Fake news detection with deep diffusive network model." *arXiv preprint arXiv:1805.08751* (2018).
- [30] Gupta, Manish, Peixiang Zhao, and Jiawei Han. "Evaluating event credibility on twitter." In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pp. 153-164. Society for Industrial and Applied Mathematics, 2012.
- [31] ping Tian, Dong. "A review on image feature extraction and representation techniques." *International Journal of Multimedia and Ubiquitous Engineering* 8, no. 4 (2013): 385-396.
- [32] Wu, Ke, Song Yang, and Kenny Q. Zhu. "False rumors detection on sina weibo by propagation structures." In *2015 IEEE 31st international conference on data engineering*, pp. 651-662. IEEE, 2015.
- [33] Antol, Stanislaw, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. "Vqa: Visual question answering." In *Proceedings of the IEEE international conference on computer vision*, pp. 2425-2433. 2015.
- [34] Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.



- [35] Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: A neural image caption generator." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156-3164. 2015.
- [36] Jin, Zhiwei, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. "Multimodal fusion with recurrent neural networks for rumor detection on microblogs." In Proceedings of the 25th ACM international conference on Multimedia, pp. 795-816. 2017.
- [37] Yang, Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S. Yu. "TI-CNN: Convolutional neural networks for fake news detection." arXiv preprint arXiv:1806.00749 (2018).
- [38] Nakamura, Kai, Sharon Levy, and William Yang Wang. "r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection." arXiv preprint arXiv:1911.03854 (2019).
- [39] Wang, Yaqing, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. "Eann: Event adversarial neural networks for multi-modal fake news detection." In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, pp. 849-857. 2018.
- [40] Singhal, Shivangi, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. "SpotFake: A Multi-modal Framework for Fake News Detection." In 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), pp. 39-47. IEEE, 2019.
- [41] Loper, Edward, and Steven Bird. "NLTK: the natural language toolkit." arXiv preprint cs/0205028 (2002).
- [42] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9, no. Nov (2008): 2579-2605.
- [43] <https://www.wired.com/2016/12/photos-fuel-spread-fakenews/>
- [44] Gupta, Aditi, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy." In Proceedings of the 22nd international conference on World Wide Web, pp. 729-736. 2013.
- [45] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

- [46] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [47] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826. 2016.
- [48] <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>
- [49] <https://keras.io/api/applications/>

## LIST OF PUBLICATIONS BY CANDIDATE

- [1] Vidhu Tanwar and Kapil Sharma “**Multi-Model Fake News Detection Based on Concatenation of Visual Latent Features**” The 9th International Conference on Communication and Signal Processing (ICCSP) held at Department of Electronics and Communication Engineering, Adhiparasakthi Engineering College in association with IEEE.
- [2] Vidhu tanwar, Kapil Sharma. “**A Review on Enhanced Techniques for Multi-Modal Fake News Detection** ” *International Conference on Recent Innovations in Computing (ICRIC-2020)*. Central University of Jammu, J & K. ICRIC-2020.