Fake News Detection Using Deep Learning

Waleed Asghar

Hood College

ABSTRACT

Today's internet is primarily a layered cake of data. The layers never get ridiculed rather they are growing exponentially. A significant proportion of that cake is covered by the news. Considering fake news, it has been there since the advent of the internet. We can define fake news as "fictitious articles that are fabricated deliberately to deceive the readers". Many social media and news outlets publish fake news to get attention or to get views from public.

The goal is to generate profit by spreading rumor or fake news among people which can sometime cause chaos or unrest among the general public. So there is a need of identifying such article or news and prevent those. The purpose of the work is to come up with solution that can be utilized by the users to detect and filter out accurately the fake news and false information posted by articles or some other outlets

In this paper we highlight an overview of the application of DL techniques to fake news detection dataset. We classify and validate the dataset, in order to produce best fake news detection results. Although a good number of research papers have been published ,the application of DL to fake news detection is still an important part considering the growing the hybrid network globally: to stimulate further work in this areas, we conclude the paper proposing the most accurate model for the fake news detector.

INTRODUCTION

As defined by S Marsland, Machine learning is a branch of Artificial Intelligence that pushes forward the idea that, by giving access to the right data, machines can learn by themselves how to solve a specific problem[1]. By leveraging complex mathematical and statistical tools, DL renders machines capable of performing independently intellectual tasks that have been traditionally solved by human beings[2].

In today's digital era ,the widespread problem of fake news is very difficult to tackle. Thousands of information sharing platforms propagating fake news or misinformation.

In this paper, our focus will be on DL techniques. The practical implementation of these techniques will be solely in python and pytorch library. We will code DL techniques in python on our fake news dataset in order to elect best algorithm to be used in detecting fake news.
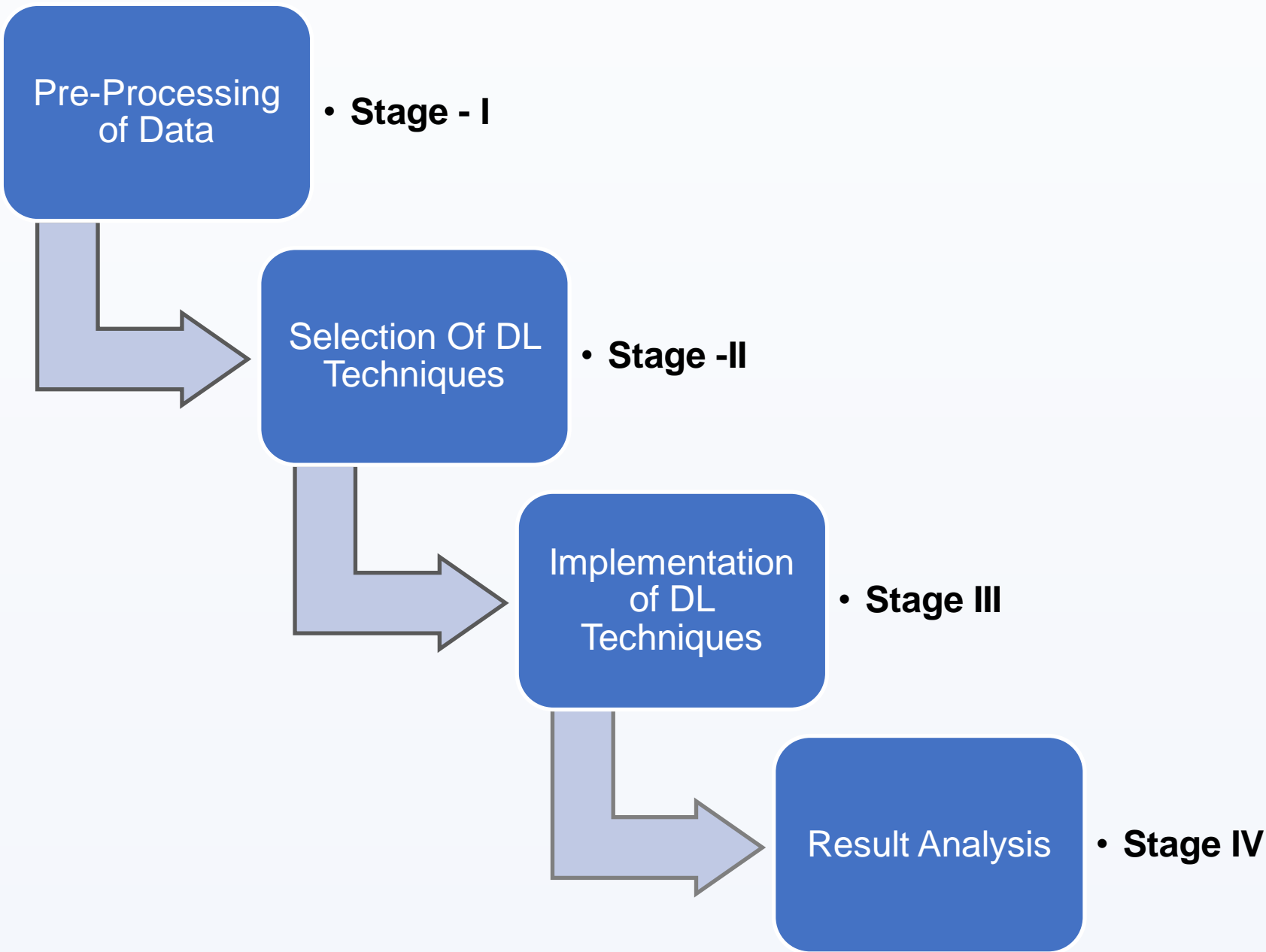
PROPOSED METHODOLOGY

- Pre-Processing of Data • **Stage - I**
- Selection Of DL Techniques • **Stage -II**
- Implementation of DL Techniques • **Stage III**
- Result Analysis • **Stage IV**

Fig 1: Stages of Methodology

DATASET OVERVIEW

The whole fake news detection will be based on word clouds. There will be two-word clouds.

  – A reliable word cloud and a non-reliable one (real & fake words).

Hence, the algorithm will be classifying the news as real if it finds words in it that present in the below-mentioned word cloud.
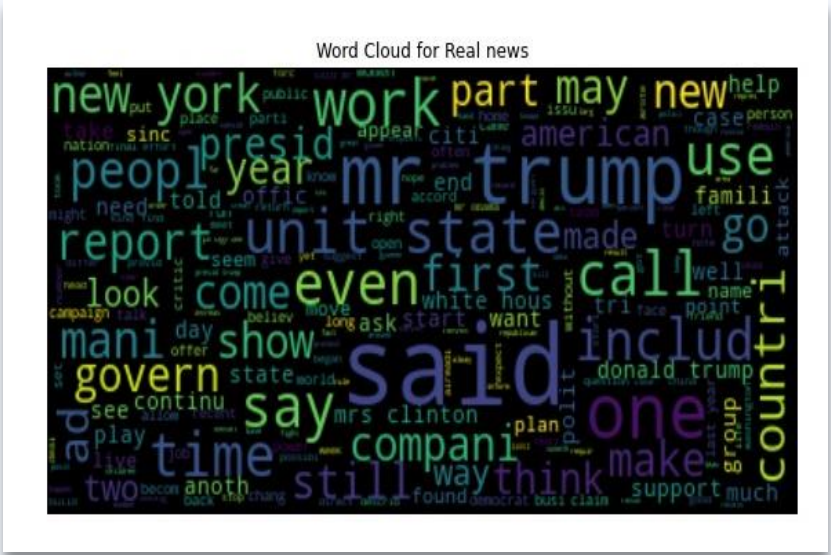

Fig 2: Word Cloud for Real news

Fig 3: Word Cloud for Fake news

  Classification of fake news will be based on these clouds. These word clouds have been first preprocessed. Processes like stop words, stemmization and punctuation removal have been applied.

  The figure below shows the balanced distribution of reliable and non-reliable word clouds after processing.
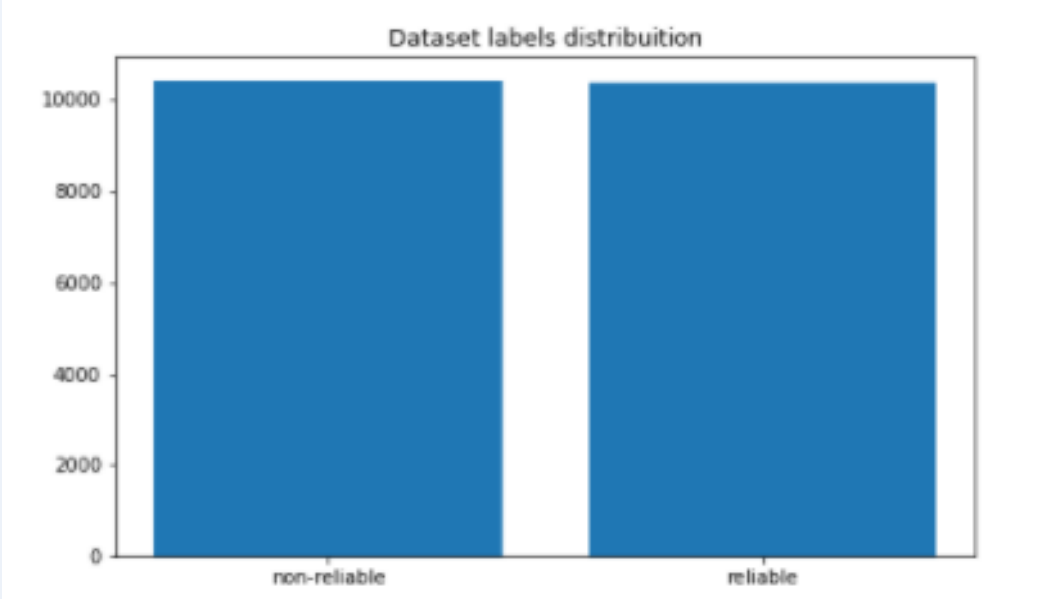

Fig 4: Reliable & Non-Reliable Word Clouds Graphic Representation

The training dataset is taken from Kaggle[3]. Particulars of the dataset have been stated below:
Id: unique id for a news article
Title: the title of a news article
Author: author of the news article
Text: the text of the article; could be incomplete
Label: label that marks the article as potentially unreliable

SELECTED NEURAL NETWORKS

Selected Neural Network Algorithms:

1. RCNN (Recurrent Based Convolutional Neural Network)
2. LSTM (Long Short-Term Memory Neural Network)
3. Bi_LSTM (Bi-directional Long Short-Term Memory)

RESULTS ANALYSIS

LSTM		
Learning Rate	0.005	0.0005
Validation Accuracy	94	84
Training Accuracy	97	85
Valid Loss	0.029	0.39
Train Loss	0.091	0.2999

TABLE 1: LSTM Results

Bi – LSTM		
Learning Rate	0.005	0.0005
Validation Accuracy	94.5	85
Training Accuracy	98	86
Valid Loss	0.023	0.379
Train Loss	0.098	0.362

TABLE 2: Bi - LSTM Results

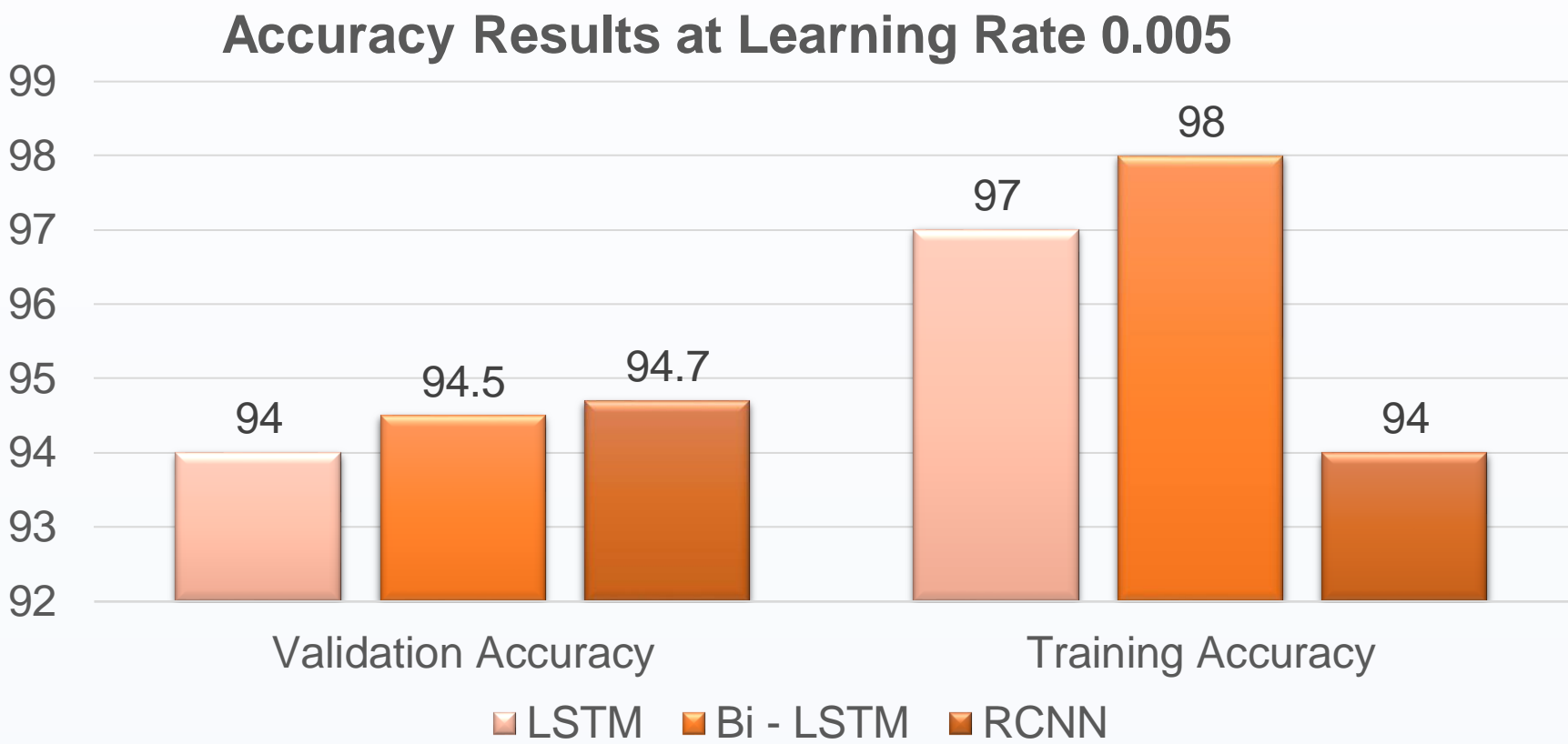| RCNN		
Learning Rate	0.05	0.005
Validation Accuracy	95.6	94.7
Training Accuracy	97	94
Valid Loss	0.05	0.09
Train Loss	0.088	0.109

TABLE 3: RCNN Results
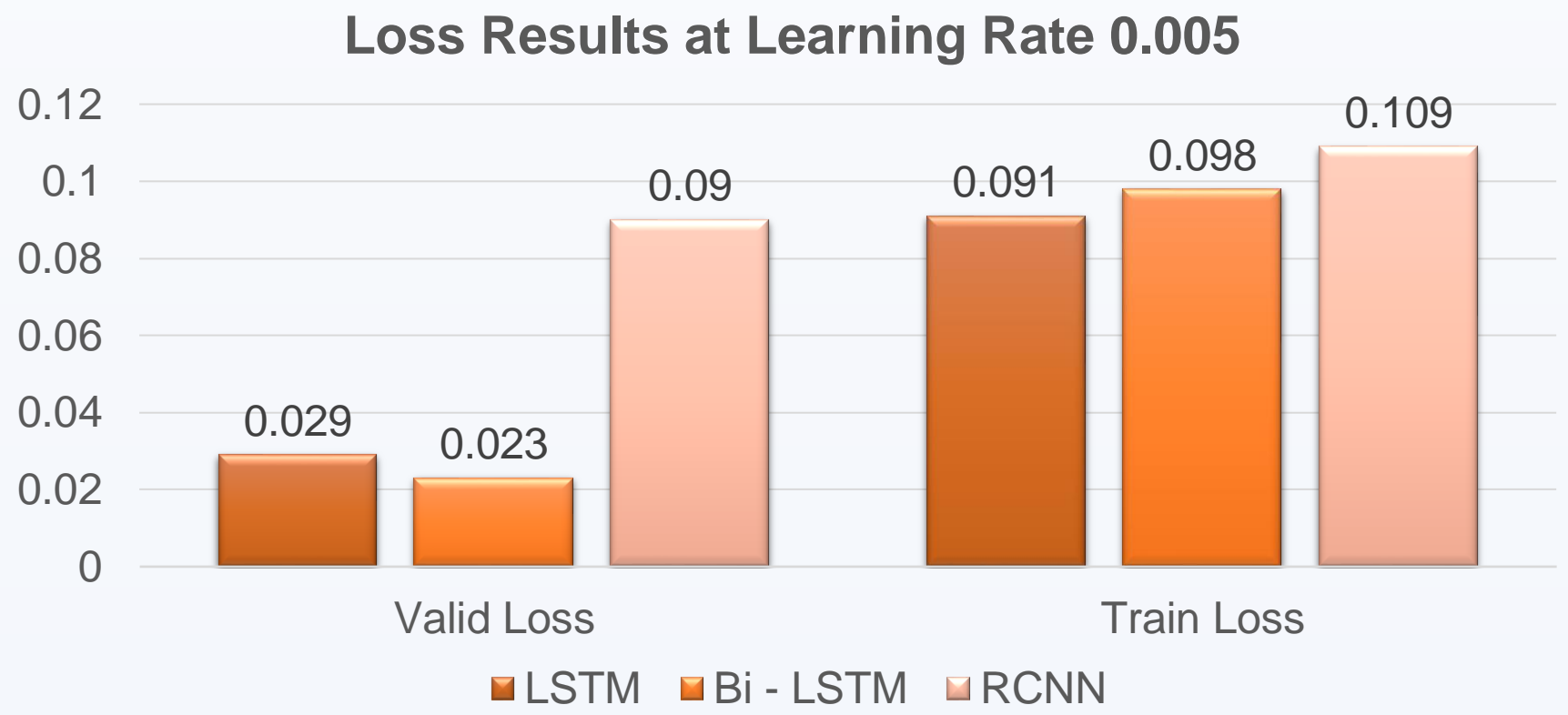

Fig 5: Accuracy Results

Fig 6: Loss Results

ANALYSIS

The validation confusion matrix results also are very much in alignment with the conclusion results. The true positive result makes it pretty much clear.

Considering Bi – LSTM is the appropriate choice since it provides almost same validation accuracy but in addition to that it results in giving least validity loss which is also an extremely important aspect to investigate.
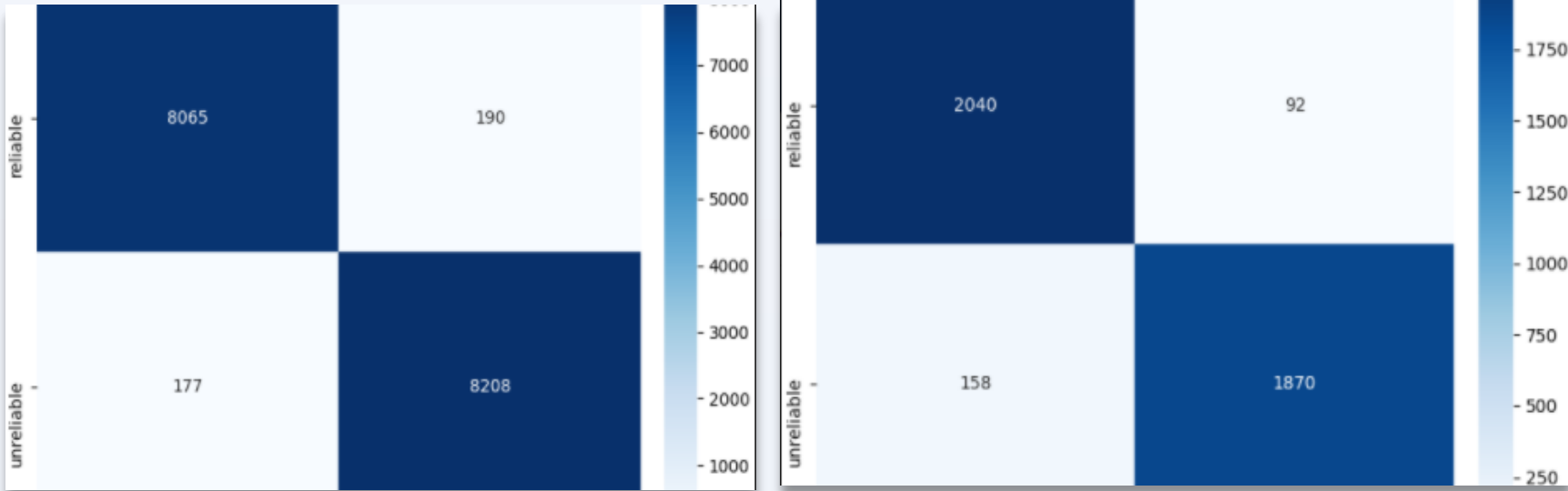
Confusion Matrix of Validation (Bi-LSTM)   Confusion Matrix of Training (Bi-LSTM)

CONCLUSION

Considering all of aspects, we have concluded that since validation accuracy is prime suspect to declare the best DL technique for the fake news detection,

  – After analysis of results we come to conclusion that RCNN provides the most validation accuracy results but similarly it also provide most validation loss.

  – On the other hand, Bi-LSTM provides almost same validation accuracy as RCNN with less validation loss as compared to RCNN so Bi-LSTM would be appropriate choice.

The curve plot for Bi-LSTM at learning rate 0.005 shows significant decrease in the validation loss and similarly significant increase in the valid accuracy that justifies our conclusion.
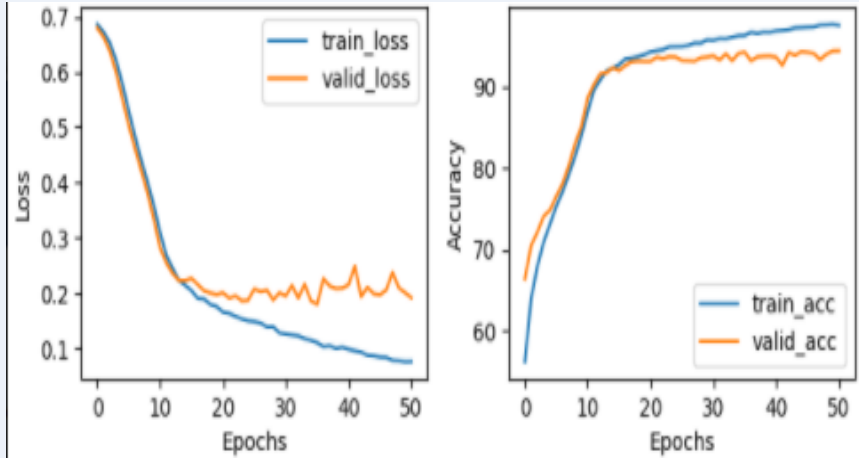

Fig 7: Curve Plot For Bi-LSTM Accuracy And Loss results (Epochs)

REFERENCES

[1] S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.

[2] An Overview on Application of Machine Learning Techniques in Optical Networks Francesco Musumeci, Member, IEEE, Cristina Rottondi, Member, IEEE.

[3] https://www.kaggle.com/c/fake-news/data