*Name Tanishq Pal*

*Roll No 2101320100172*

*Group 43*

*Reseach Paper : Stratification of Depressed and Non-Depressed Texts from Social Media using LSTM and its Variants*

*BY : Keerthan Kumar T G1, Anoop R2, Shashidhar G Koolagudi3, Trupthi Rao2, Ashwini Kodipalli2.*

*Department of Information Science and Engineering, Siddaganga Institute of Technology, Tumkur, India.*

*Department of artificial intelligence and Data Science, Global academy of Technology, Bengaluru, India.*

*Department of Computer science and Engineering National Institute of Technology, Karnataka, Surathkal, India.*

**Research Overview :**

This research was done to identiy the best LSTM model that can predict or segregate the depressive and depressive text from social media . The models tested were Classical LSTM , Bidirectional LSTM , GRU (gated recurrent unit) , bidirectional GRU , and Stacked LSTM . They was tested by F1 score . recall , precision and accuracy . Results shows that the classical LSTM and GRU have better accuracy but bidirectional models has better precision .

**Litreature Review :\**

Many NLP applications have used LSTM models for better accuracy and precision .

The data used by the authors to train the model is twitter data .

The dataset was classified in negative and positive tweets .

The dataset was processed using ML algo like SVM , k-nearest algo , decsion tree and LSTM , out which LSTM shows better results with (83%) .

They used their own dataset and out of all models the word based GRU was best performing with accuracy of (98%) .

Them some other autor used reddit data with NLP and text classification approaches to detect depressive attitude .

Later ML classifiers are used for model features .(91%)

Other used a text + audio dataset with two models for each data . In both CNN and LSZTM are used .(f1-0.8 ,0.75)

Some other used LSTM and RNN to model the dataset for more accuracy .(99%)

Then other combined CNN and BiLSTM on a benchmark depression dataset. (94.28%)

Some other used LSTM deep neural network .

**Dataset :**

From : kaggle

Involves : Reddit + Twitter dataset

Classification : Deperessed (1) And Non Depressed (0)

Rows : 12570

The dataset was collected in a way that the user data that they interact with rather than the data they share is collected .

Example : Depression Quotes ,depressive quotes etc (username ) was considered as depressive and one sharing about sports , news and cars etc were labelled as non depressive .

**Metheodalogy :**

1. **Data Pre-processing** : The data collected was hjaving URLs ,special characters and numeric values that were omitted. The data was formed in a way to decrese the count of unique words to improve model performance. They used SMOT(Synthetic Minority Over-sampling Technique ) technique to oversample the depressed to match the no - depressed . Then using one hot encoding the text data was converted into numerical data.

60 : 20 : 20 - training set , test set and validation set

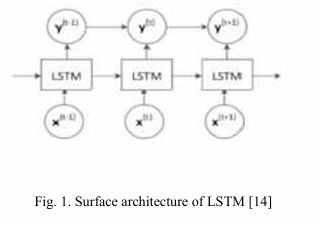
Trained using training set , tested using test set and assesment using validation set .

1. **Model Selection** : It contains selection from : (i). Classic LSTM (ii). Stacked LSTM (iii). Bidirectional LSTM (iv). GRU (v). Bidirectional GRU

All these models were implemented using keras Deep learning lib .

1. Classical LSTM :

LSTM has cell that can store data and there is forget and recall information based on input entered a each step .



This architecture of LSTM model comprises of cells in sequential order that process input from previous cell .They each have three gated mechanisms (forget gate , input gate and output gate ).

The forget gate makes sure the irrelevant data is ignored .The cell predicts a value between 0 and 1 , such that near 1 means data is kept and near 0 meaning it is ignored .

Input : previous hidden state + current input

The input gate determines the new info that needs to be added .

Input : previous concealed state + current input

Output : 0 - not added 1 - added

The output gate determines what data to be given as output .

Input: previous concealed state and current input

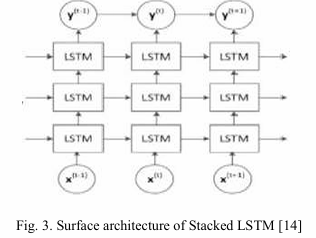
Output : 1 - shown , 0 - not shown

+ a memory cell is there to store info overtime based on input + forget gate .

1. Stacked LSTM :

In this is has multiple LSTM layers stacked on top of eachother to provide more sophesticated representation of input .

Each LSTM layer take input from previous layer and final layer is he output layer.

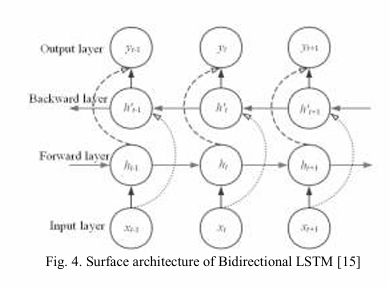


1. Bi-directional LSTM :

This model can sore both the past and future context of the input sequences .It has 2 LSTM models one for forward direction and one for backward direction processing the input .

The output of each LSTM network is concatenated at each step for final output .

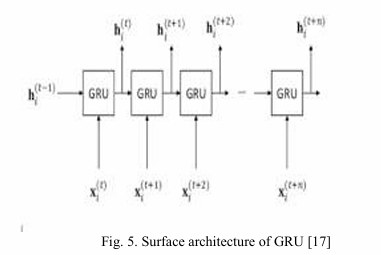
During the training phase both forward and backward models are combined to revise model parameters .



1. GRU :

Gated recurrent unit model consist of interconnected GRU cells that process the input one step at a time sequentially .  
Cell - memory cell , forget gate and update gate

Cell calculates it output by combing current input and memory cell through non-linear activation function .Then the ouput is sent as input to next cell .



1. Bi-directional GRU :

Extended version of GRU which can process data in forward and backward direction .

Uses two sepreate GRU networks one for processing input in forward direction and one in backward direction .

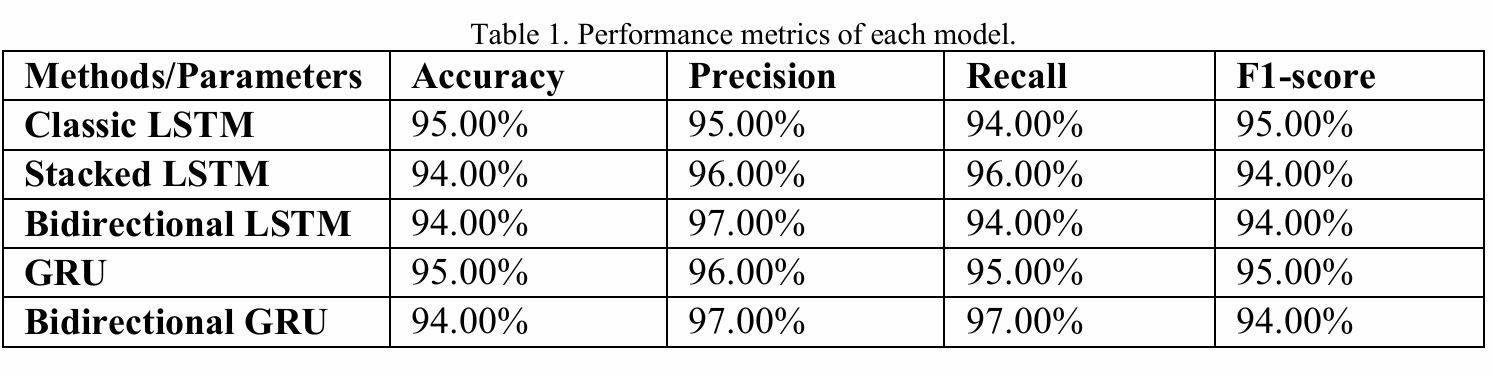
1. **Model Training** : Every LSTM model was trained on a training set of 10 epoches , batch size of 32 using Adam optimizer. .

Error Analysis the binary cross-enotropy loss function was used .

1. **Model Evaluation** :Performance of the model was evaluated using precision, recall, accuracy, and F1-score . Conclusion matrix was formed to visualize .

**Results :**

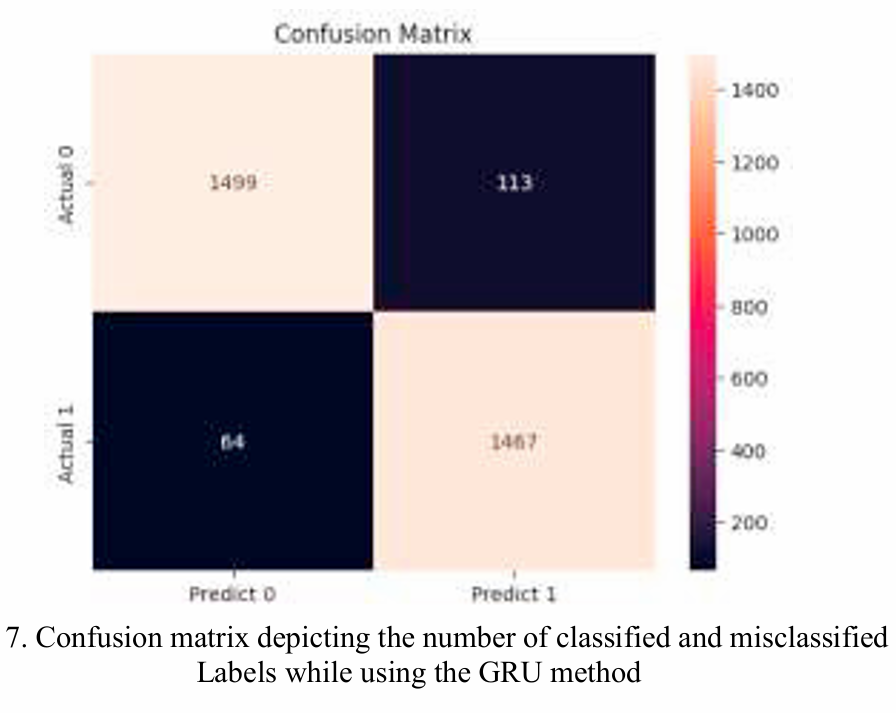
Classical LSTM and GRU shows best precision but when recall , F1, accuracy was tested GRU shows best results .



GRU shows that it more efficient when working with lasrge set of data and its is more producitve and accurate when data is lengthy .

Because GRU has Long and Short term memory combined together .

Fig. 7, comprising the confusion matrix, depicts the number of correctly and incorrectly classified data for all the methods used for the classification of data.



**Conclusion :**

That it can provide classification on social media data (twitter + reddit ) by using NLP techqniques like LSTM variants . And the study porves that Classical LSTM and GRU shows better precision than their bi-directional variants .

**Research Gap :**

Even though it considered using Reddit and Twitter dataset but still it cant be robust to any other environment as it is data centric to reddit and twitter .

This study also just collaborated with text data not including images and audio files .