Comparative Analysis of Detection of Depression using Transformer Models

S.Packiyalakshmi[1], G.Ramkumar[2]

[1] M.sc Student, Sri Kaliswari College(Autonomous), Sivakasi
[2] Assistant professor, Sri Kaliswari College(Autonomous), Sivakasi

Abstract- *Social media platforms play a major role in our day-to-day life and are considered as a virtual friend by many users, who use the social media to share their feelings all day. Many a time, the content which is shared by users on social media replicate their internal life. Nowadays people love to share their daily life incidents like happy or unhappy moments and their feelings in social media and it makes them feel complete and it has become a habit for many users. Social media provides a new chance to identify the feelings of a person through their posts. The aim of the shared task is to develop a model in which the system is capable of analyzing the grammatical markers related to onset and permanent symptoms of depression. We as a team participated in the shared task Detecting Signs of Depression from Social Media Text at LT- EDI 2022- ACL 2022 and we have proposed a model which predicts depression from English social media posts using the data set shared for the task. The prediction is done based on the labels Moderate, Severe and Not Depressed. We have implemented this using different transformer models like DistilBERT, RoBERTa and ALBERT by which we were able to achieve a Macro F1 score of 0.337, 0.457 and 0.387 respectively.*

I. INTRODUCTION

In the digital world, the usage of social media has become most common among the people. Social media is used without any limits to share happiness, joy, sadness, loneliness and all other personal emotions. The contents shared by the people reflect the mental state of the person and can act as an indicator of their depression level (Kamite and Kamble, 2020). Depression is a serious mental illness that negatively affects how you feel, the way you think and how you act. Fortunately, it is also treatable.

Depression causes feelings of sadness and/or a loss of interest in activities you once enjoyed. It can lead to a variety of emotional and physical problems and can decrease your ability to function at work and at home. Sometimes, social media could be the reason for the depression. It is necessary to measure the level of depression from the social media text to treat them and to avoid the negative consequences. Detecting levels of depression is a challenging task since it involves the mindset of the people which can change periodically. Our aim is to detect levels of depression with the use of deep learning Transformer Models to achieve the best results (Malviya et al., 2021).

The shared task Detecting Signs of Depression from Social Media Text was a part of LT-EDI 2022- ACL 2022 (Sampath et al., 2022). The task is based on English comments. The task was a classification problem based on the labels "Not Depressed", "Moderate" and "Severe". For example, "My life gets worse every year : That's what it feels like any- way " fall under the category Moderate, "Words can't describe how bad I feel right now : I just want to fall asleep forever." fall under the category Severe and "Is anybody else hoping the Coronavirus shuts everybody down?" fall under the category Not Depressed.

II. RELATED WORKS

Identifying depression from social media posts involves detecting whether the user associated with the posts could be identified for depression and this could be represented as a text classification prob lem. Various methods from rule based techniques to deep learning methods could be used for this purpose. Identification of depression markers and pre-processing the actual posts also play an important role in the performance of the model. The performance of the model used for classification mainly depends on the data set used for the purpose like the size of the data set and the distribution of the data in the data set. Hence the analysis of the data set is important for selecting the appropriate model for implementing the classification task. The contribution of different pre-processing techniques for improving the prediction efficiency of depression identification task (Figueredo and Calumby, 2020) had been presented. Depression-related markers in Facebook users had been identified by Socially Mediated Patient Portal (SMPP) (Hussain et al., 2020), which had used a data-driven approach with machine learning classification techniques for extracting such information. The syntactical markers related to onset and perpetual symptoms of depression (Kamite and Kamble, 2020) have been identified which when used together with statistical models had helped in effective and early

Learning Technique	Approaches used	Limitation
Traditional Approach	Statistical, Data driven, Rule based and lexicon based approaches	Requires specific features like syntactic markers, psycho-linguisticfeatures and temporal dependencies
Machine Learning approach	SVM, RF, DT, NB, KNN, LR	Requires proper fine tuning of parameters and does not show significant impact on the precision
Deep Learning approach	NN, RNN, LSTM and Transformer Models	Handling of heterogeneous and feature vector representation associated with the performance

Table 1: Summary of related work

identification of depression from social media posts. The impact of psycho-linguistic patterns on standard machine learning approaches had been illustrated for the classification of social media texts that are associated with depression (Trifan et al., 2020). Multi modal framework and statistical techniques had been used to discern depressive behaviours from a heterogeneous set of features including visual, textual, and user interaction data (Yazdavar et al., 2020) from social media posts. Multiple In- stance Learning methods (Mann et al., 2021) had been used for the task of identifying depression from social media posts which had implemented the classification by exploiting temporal dependencies between posts. Detection of mental health dis- orders, especially depression, had been predicted from Arabic posts using a lexicon based approach and machine learning approach (Alghamdi et al., 2020).

Early detection of different emotions of people including depression from their social media posts had been done using a hybrid model which is a combination of two machine learning algorithms namely Support Vector Machine and Naive Bayes algorithm (Smys and Raj, 2021). The performance measures of the model had been analyzed by fine tuning the parameters associated with the algorithms. Detection of depression from Bengali posts and commentaries had been implemented and evaluated using different machine learning algorithms like Support Vector Machine, Random Forest, Decision Tree, K-Nearest Neighbors, Naive Bayes (Multinomial Naive Bayes) and Logistic Regression. The results had shown that the same precision had been achieved by all the algorithms (Victor et al., 2020). Social media posts of high school students, college students and working professionals had also been considered in specific for identifying mental health using the above mentioned machine learning algorithms (Narayanrao and Lalitha Surya Kumari, 2020).

Use of deep learning models had been depicted for the prediction of mental disorders such as depression. A multi-task hierarchical neural network with topic attention had been used for identifying health issues from social media posts. Bidirectional gated recurrent units had been used to analyze the hierarchical relationship (Zhou et al., 2021) among documents, sentences and words based on which at- tention weights are enhanced for words. The posts with unstructured text data that display depression had been identified more effectively by deep learning models than by using supervised learning methods (Ahmad et al., 2020). The role of sentiment analysis in identifying depression had been shown which had improved the performance of the model by using different deep learning techniques for the process of classification (Banerjee and Shaikh, 2021). Better performance had been achieved when the heterogeneous and feature vector representation associated with social media posts had been handled and transformer based models (Garg, 2021) had been utilized for classification of depression and suicidal posts. Depression and associated negative emotions had been identified from Sina Weibo, using deep learning methods (Yao et al., 2019).

Table 1 summarises the approaches and the limitations associated with the models that exist for detecting depression from social media posts. It could be summarized from the related works that proper pre-processing, selection of markers and dominant feature extraction directly have an im- pact on the performance of the model. Different approaches like rule-based approach, statistical approach, machine learning approaches and deep learning approaches could be used for this purpose and the deep learning techniques tend to show bet- ter performance when compared with traditional and machine learning approaches. When appropriate text pre-processing and textual based featuring techniques (Zhou et al., 2021) had been used with machine learning classifiers, it had

been shown that depression associated social media texts could be effectively identified even when depression specific keywords were not present in the social media posts. The performance of the model based on an approach may not be the same for all data sets which is a major factor to be considered and this makes the problem of identifying depression from social media text as an important research field in the domain of Natural Language Processing.

III. DATA SET

The data set used for our model is a collection of Social Media Text provided by the organizers of the shared task (Sampath et al., 2022). The data set (Kayalvizhi and Thenmozhi, 2022) comprises training, development and test data set. The data files were in Tab Separated Values (tsv) format with

Data Set	Category	Instances
Train	Not Depression Moderate Severe	1971 6019 901
Validation	Not Depression Moderate Severe	1,830 2306 360

Table 2: Data set statistics

three columns namely posting id (pid), text data and label. The sample instances are as follows:
• Not depressed - This indicates the social media text is not depressed in nature Example: "Is anybody else hoping the Coronavirus shuts everybody down?"
• Moderate- This indicates the social media text is moderately depressed in nature Example: "My life gets worse every year : That's what it feels like anyway "
• Severe- This indicates the social media text is severely depressed in nature Example: "Words can't describe how bad I feel right now : I just want to fall asleep forever."
The distribution of the data in the data set is shown in Table 2. The training data set had 8,891 in- stances of which 1,971 instances were under the Not depressed category, 6019 instances were under Moderate category and 901 instances under Severe category. The validation data set provided for the evaluation of the model had 4496 instances with 1830, 2306 and 360 instances under the category Not depressed, Moderate and Severe respectively. The test data set provided for the purpose of prediction had 3245 instances.

IV. METHODOLOGY

The proposed methodology uses deep learning techniques for implementing the process of detecting depression from social media texts. From the existing systems it could be found that transformer based models exhibit better performance when

compared to Neural network based models and LSTM based models. Hence the proposed system uses three different Transformer models namely DistilBERT, ALBERT and RoBERTa for the task of detecting the depression level from social media text.

Model	DistilBert	Robert	AlBert
Accuracy	0.342	0.520	0.408
MacroPrecision	0.456	0.337	0.432
Macro F1Score	0.337	0.457	0.387
Mscro Recall	0.467	0.519	0.497

Table 3: Task Score

DistilBERT

DistilBERT (Sanh et al., 2019) is a general-purpose pre-trained version of BERT which had been pre- trained on the same corpus as BERT in a self- supervised fashion. This means it was pre-trained on the raw texts only, with no human labeling to generate inputs and labels from those texts using the BERT base model2. Distil-BERT has 97% of BERT's performance while being trained on half of the parameters of BERT. BERT-base has 110 parameters and BERT-large has 340 parameters, which are hard to deal with. For this problem's solution, distillation techniques are used to reduce the size of these large models3.

RoBERTa

RoBERTa (Liu et al., 2019) is a transformer model pre-trained on a large corpus of English data and is based on BERT model and modifies key hyper- parameters and training is implemented with larger mini-batches and learning rates4. RoBERTa is a Robust BERT method which has been trained on a far extra large data set and for a whole lot of large quantities of iterations with a bigger batch length of 8k.

ALBERT

The ALBERT (Lan et al., 2020) model is a Lite BERT which improves the training and results of BERT architecture by using different techniques like parameter sharing, factorization of embedding matrix and Inter sentence Coherence loss.
We have used "ALBERT–base-v1" model for the task which is also a pre-trained model on English language. This model is uncased5 and it consists of 12 repeating layers, 128 embedding, 768-hidden, 12-heads and 11M parameters6. The first step associated with the task is to prepare the data set which involves pre-processing the text from the data set for effective modeling. As the text from social media posts does

not have a standard structure and use of symbols, tags and URLs are common, the texts need to be pre-processed by converting the complete text into lower case words and removing the stop-words, URLs, numbers and tags which do not contribute much for the classification task. Then the three different transformer models namely DistilBERT, ALBERT and RoBERTa had been used to implement the classification of the texts into Moderate, Severe and Not Depressed texts. The labels were converted to equivalent integer categorical values so that it can be given as input to the trans- former models. The models are trained using the training set provided as a part of the shared task. The evaluation of the model was carried out using the evaluation data set provided by the shared task. Finally the required predictions were done using the test data set provided by the shared task. The number of epochs that were considered for training were 5 for DistilBERT and ALBERT and 1 epoch was used for RoBERTa.

V. EXPERIMENTAL SETUP

We have used the virtual GPU (Tesla k80) provided by Google Colab for implementing different transformer models.The processing time was found to be 5.43 min, 15.46 min, 5.48 min for DistilBERT, RoBERTa and AlBERT models respectively. The memory usage of our model was calculated to be 3583MiB.

Label	Precision	Recall	F1-Score	Support
Not Depression	0.60	0.45	0.52	1830
Moderate	0.39	0.72	0.65	2306
Severe	0.31	0.29	0.30	360
Accuracy			0.37	4496
Macro Avg	0.50	0.49	0.49	4496
Weighted Avg	0.576	0.57	0.57	1830

Table4:ClassificationReport

VI. RESULTS

The evaluation of the shared task was done using the metric namely Macro F1-score. The other metrics that were used to represent the performance of the model were Accuracy, Macro Recall and Macro Precision. The ratio of the number of correct pre- dictions to the total number of input samples is represented by the metric Accuracy. The ratio of the number of correct positive results to the number of positive results predicted by the classifier is represented by Precision. The model's ability to detect positive samples is represented by recall. F1 score is an overall measure of a model's accuracy that combines precision and recall. A high F1 score means that the classification has resulted with low number of false positives, and low false negatives. The metric values that were scored by the three different models on the test data set provided for the shared task are given in Table 3. When using the DistilBERT the values that were obtained for the different metrics were 0.342 for accuracy, 0.337 for Macro F1-score, 0.467 for Macro recall and 0.456 for Macro precision. By using the trans- former model ALBERT for classification the metrics were improved to 0.408 for accuracy, 0.387 for Macro F1-score, 0.497 for Macro Recall and 0.432 for Macro Precision. The metrics were further improved when the RoBERTa model was used for implementing the classification task which resulted in an accuracy of 0.510, Macro F1-score of 0.457, Macro recall of 0.519 and Macro precision of 0.461 which brought us to the rank of 23 in the leader board.

VII. ERROR ANALYSIS

The model RoBERTa had resulted in an F1 score of 0.457 which is low compared to 0.583 which is the F1 score obtained by the topper of the leader board. This shows that more false positive and false negative classification has occurred in our proposed model. The data set provided is highly imbalanced in nature which could also be considered as a reason for the poor performance of the model. The data set could be converted to a balanced data set by using different up-sampling and down-sampling techniques.

The classification report generated during the evaluation of the model is shown in Table 4. It could be found that the instances that fall under the category 'Severe' have a low F1 score of 0.30, which means more false positives and false negatives have occurred under this category. Most of the posts associated with the category Severe do not use the depression related markers directly which can also be considered as a reason for poor performance of the model.

VII. CONCLUSION

As social media platforms play a crucial role in to-day's world and the posts shared replicate the internal mental state of the user, the task of identifying depression from social media posts have become an important research area. The methodology associated with our submission used three different transformer models to implement the above said task namely DistilBERT, ALBERT and RoBERTa of which the RoBERTa model had shown a better performance with a F1 score of 0.457.

REFERENCES

[1] Hussain Ahmad, Dr. Muhammad Asghar, Fahad Alotaibi, and Ibrahim Hameed. 2020. Applying deep learning technique for depression classification in social media text. Journal of Medical Imaging and Health Informatics, 10:pp. 2446–2451(6).

[2] Norah Saleh Alghamdi, Hanan A. Hosni Mahmoud, Ajith Abraham, Samar Awadh Alanazi, and Laura García-Hernández. 2020. Predicting depression symptoms in an arabic psychological forum. IEEE Access, 8:57317–57334.

[3] Satyaki Banerjee and Nuzhat F. Shaikh. 2021. A survey on mental health monitoring system via social me- dia data using deep learning framework. In Techno- Societal 2020, pages 879–887, Cham. Springer Inter- national Publishing.

[4] José Figueredo and Rodrigo Calumby. 2020. On text preprocessing for early detection of depression on social media. In Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde, pages 84–95, Porto Alegre, RS, Brasil. SBC.

[5] Muskan Garg. 2021. Quantifying the suicidal ten- dency on social media: A survey. arXiv preprint arXiv:2110.03663.

[6] Jamil Hussain, Fahad Ahmed Satti, Muhammad Afzal, Wajahat Ali Khan, Hafiz Syed Muhammad Bilal, Muhammad Zaki Ansaar, Hafiz Farooq Ahmad, Taeho Hur, Jaehun Bang, Jee-In Kim, Gwang Hoon Park, Hyonwoo Seung, and Sungyoung Lee. 2020. Exploring the dominant features of social media for depression detection. Journal of Information Science, 46(6):739–759.

[7] Sangeeta R. Kamite and V. B. Kamble. 2020. Detec- tion of depression in social media via twitter using machine learning approach. In 2020 International Conference on Smart Innovations in Design, Environ- ment, Management, Planning and Computing (IC- SIDEMPC), pages 122–125.

[8] S Kayalvizhi and D Thenmozhi. 2022. Data set cre- ation and empirical analysis for detecting signs of de- pression from social media postings. arXiv preprint arXiv:2202.03047.

[9] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. Albert: A lite bert for self-supervised learning of language representations.

[10] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Man- dar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining ap- proach. arXiv preprint arXiv:1907.11692.

[11] Keshu Malviya, Bholanath Roy, and SK Saritha. 2021. A transformers approach to detect depression in so- cial media. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pages 718–723.

[12] Paulo Mann, Aline Paes, and Elton H. Matsushima. 2021. Screening for depressed individuals by us- ing multimodal social media data. Proceedings of the AAAI Conference on Artificial Intelligence, 35(18):15722–15723.

[13] Purude Vaishali Narayanro and P. Lalitha Surya Ku- mari. 2020. Analysis of machine learning algorithms for predicting depression. In 2020 International Con- ference on Computer Science, Engineering and Ap- plications (ICCSEA), pages 1–4.

[14] Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022. Findings of the shared task on Detecting Signs of Depression from Social Media. In "Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion". Association for Computational Linguistics.

[15] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

[16] S Smys and Jennifer S Raj. 2021. Analysis of deep learning techniques for early detection of depression on social media network-a comparative study. Jour- nal of trends in Computer Science and Smart technol- ogy (TCSST), 3(01):24–39.

[17] Alina Trifan, Rui Antunes, Sérgio Matos, and Jose Luís Oliveira. 2020. Understanding depression from psy- cholinguistic patterns in social media texts. In Ad- vances in Information Retrieval, pages 402–409, Cham. Springer International Publishing.

[18] Debasish Bhattacharjee Victor, Jamil Kawsher, Md Shad Labib, and Subhenur Latif. 2020. Machine learning techniques for depression analysis on social media- case study on bengali community. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pages 1118–1126.

[19] Xiaoxu Yao, Guang Yu, Xianyun Tian, and Jingyun Tang. 2019. Patterns and longitudinal changes in negative emotions of people with depression on sina weibo. Telemedicine journal and e-health : the offi- cial journal of the American Telemedicine Associa- tion.

[20] Amir Hossein Yazdavar, Mohammad Saeid Mah- davinejad, Goonmeet Bajaj, William Romine, Amit Sheth, Amir Hassan Monadjemi, Krishnaprasad Thirunarayan, John M. Meddar, Annie Myers, Jy- otishman Pathak, and Pascal Hitzler. 2020. Multi- modal

mental health analysis in social media. PLOS ONE, 15(4):1–27.

[21] Deyu Zhou, Jiale Yuan, and Jiasheng Si. 2021. Health issue identification in social media based on multi- task hierarchical neural networks with topic attention. Artificial Intelligence in Medicine, 118:102119.