ISSN 2007-9737

Automatic Depression Detection in Social Networks Using Multiple User Characterizations

Karla María Valencia-Segura, Hugo Jair Escalante,
Luis Villaseñor-Pineda

Instituto Nacional de Astrofísica, Óptica y Electrónica,
Department of Computer Science,
Mexico

{valencia.karla, hugojair, villasen}@inaoep.mx

Abstract. Depression is rapidly becoming one of the most common illnesses worldwide, currently affecting a significant number of people. These people may show different signs of depression depending on a number of characteristics (e.g., age, sex, personality, mood, etc.). Experts often use these signs to diagnose and monitor depression. However, due to the difficulty of obtaining this information through traditional methods, the use of social networks to characterize users has proven to be a valuable resource. In this paper, we study the potential of various user characterizations for the task of automatic depression detection. We consider two social networks and a variety of models for fusing the characterizations. In particular, we propose the use of a range of networks that learn to weight the contribution of each characterization from the data. We show that, using this model, the depression detection performance outperforms the state-of-the-art results on the two data sets considered. In addition, we present interesting findings on the correlation of the characterizations considered in the information fusion network.

Keywords. Automatic depression detection, user characterizations, social network analysis, information fusion.

1 Introduction

The World Health Organization (WHO) reports that over 280 million people worldwide experience depression, making it one of the leading causes of suicide. But only a tiny fraction of those with depression receive proper treatment [23], mostly due to the inherent difficulty of diagnosis.

This stigmatized illness prevents people with this condition to look for medical help [21]. For this reason, we continue to search for strategies that allow their appropriate detection.

Currently, there is a trend in research to develop improved methods of diagnosing traditional depression, such as interview-based methods.

These methods are based on the use of machine learning techniques to analyze user data. Among them, there are methods based on observing indicators of depression that people might show through written [26, 6].

This approach has become increasingly popular among the fields of language processing and machine learning, as it has shown that it is possible to detect depression to some extent with this type of techniques [7, 19, 1].

The adoption of methods based on textual information analysis to detect depression may have a great impact now more than ever. This is mainly due to the establishment of social networks as the main communication channel for people in the world [4].

However, the difficulty in finding useful indicators of this condition from text is not trivial. Mainly because there are many factors to take into account when dealing with users. Among them, age and gender [21], personality traits [10] and even users' interests and social context [9].

In this paper, we analyze different approaches to characterizing users in social media in order to identify indicators of depression that might be useful for detecting depression.

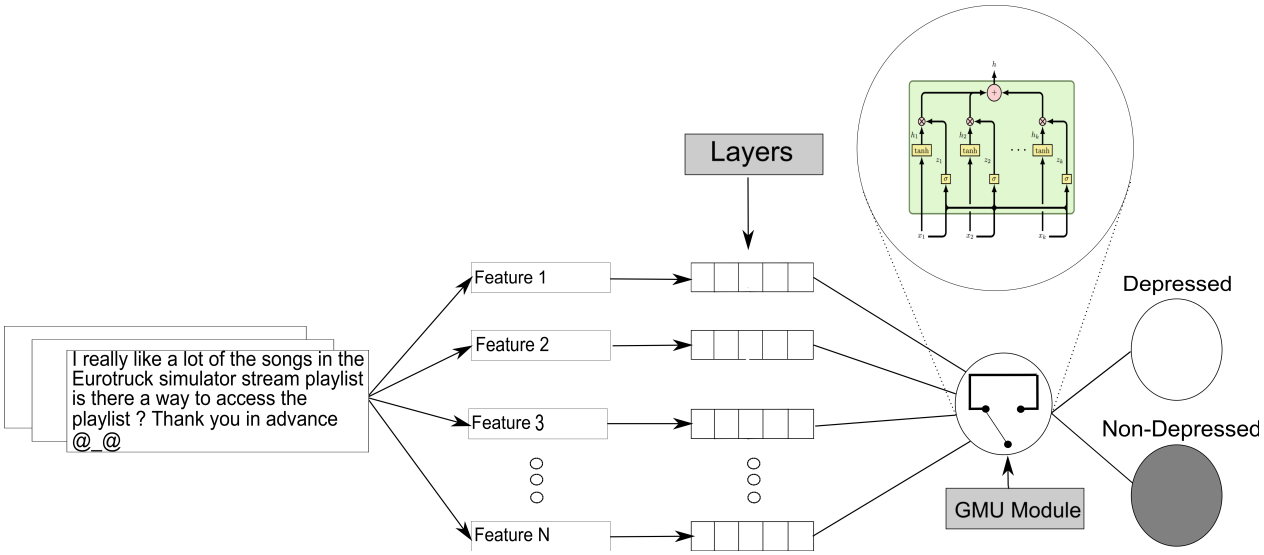


Fig. 1. General Model Architecture

Table 1. Statics of Reddit and Twitter datsets

DataSet	Training		Test	
	Depressive	**Non-Depressive**	**Depressive**	**Non-Depressive**
eRisk2018	135	752	79	741
Twiiter	2,626	5,373	-	-

We then combine such indicators with information fusion methodologies to improve detection performance.

The underlying hypothesis is that the indicators considered are complementary to each other and information fusion methodologies can leverage such a benefit.

Our goal is to gain an in-depth understanding of the contribution of each type of feature and the impact of using a highly effective data fusion methodology.

Specifically, we explore different fusion methods, achieving better results using a multimodal fusion method (Gated Multimodal Unit [2]).

We experimentally evaluated the proposed methodology on two widely used datasets for depression recognition. The experimental results show the effectiveness of the adopted approach, outperforming the state-of-the-art results for both datasets. In addition, we find that the relevance of features varies according to the social network associated with the dataset, while emotion-related features have a relevant impact on the recognition of depression in social media users for both datasets.

We anticipate that our study will motivate further research on the use and combination of user characterizations for depression recognition.

The remainder of the article is organized as follows: Section 2 presents related work. Section 3 describes the proposed methodology.

Section 4 reports the main experiments and results. Section 5 shows additional analyses of the proposed model. Finally, Section 6 presents our conclusions and future work.

Table 2. Individual performance of the 7 characterizations on Reddit and Twitter

Feature	Precision	Recall	F1-score
Reddit			
Polarity	0.10	0.89	0.18
Emo-Lex	0.60	0.58	0.59
CNN-Emotions	0.26	0.56	0.35
Personality	0.18	0.86	0.20
Thematic	0.76	0.43	0.55
Gender	0.32	0.62	0.42
Age	0.22	0.96	0.36
Twitter			
Polarity	0.41± 0.03	0.85±0.03	0.61±0.02
Emo-Lex	0.51±0.04	0.37±0.03	0.70±0.02
CNN-Emotions	0.40± 0.02	0.71± 0.02	0.60±0.02
Personality	0.48±0.02	0.70±0.02	0.59±0.02
Thematic	0.79± 0.01	0.76±0.01	**0.72±0.01**
Gender	0.34±0.01	**0.95 ±0.01**	0.52±0.01
Age	0.41±0.02	0.90±0.03	0.56±0.02

Table 3. Coincident Failure Diversity on Reddit and Twitter features

Feature fusion	CFD
Reddit	
Emo-Lex, CNN-Emotions, Thematic, Gender	0.87
All characterizations	0.62
Twitter	
Emo-Lex, CNN-Emotions, Thematic	0.59
All characterizations	0.47

2 Related Work

Different works have already been performed on seeking signs of depression in social networks through the analysis of history posts of users.

For example, Chen et al. [7] used an emotion-based characterization approach to detect depression users on Twitter, concluding that emotions have an important impact on detecting depression.

On the other hand, Preoţiuc-Pietro et al. [19] estimated demographic information such as age and gender through the analysis of posts on Twitter. Obtaining high performance to identify users with depression.

However, most works extract several characterizations to represent the users ignoring the existing complementary among these characterizations.

Some works focused on adopting a data fusion method to solve this problem. Peng et al. [18] used a model based on SVM Multi-Kernel to select optimal kernels and combat the heterogeneity of three characterizations (text from microblogs, information on the user profile, and the behavior of the user); to identify depression users on Sina Weibo.

On the other hand, Meng et al. [15] used facial expressions and audio characterizations to predict depression, applying the linear opinion pool method as a fusion technique.

Other works have been proposed to analyze textual information using the same datasets used in this work. Some of these implement fusion techniques.

In the case of the Twitter dataset in [22], the authors propose a multimodal approach that combines characteristics such as emotions, personal information, topics, etc.

Where they apply a learning dictionary to fuse the features between the modalities and learn the sparse joint representation to obtain the latent features.

On the other hand, in [24], the authors propose the creation of specific classifiers (gender and age) and consider the feelings expressed in the messages through a new text representation that captures their polarity to improve the detection of depressive users.

In the case of Reddit in [25] an ensemble was implemented as a fusion technique to fuse multiple features such as linguistic metadata at the user level, bag of words, neural word embeddings, and Convolutional Neural Networks (CNN).

In [16] used various classification techniques (Ada Boost, Random Forest, and Recurrent Neural Network (RNN)), using a bag of words and metamaps features; where the best F1-score was obtained using Random Forest. Finally, in [20]

Table 4. Comparison of the various fusion methods vs. the model proposed for this work on Reddit and Twitter dataset

Fusion Method	Precision	Recall	F1-score
Reddit			
Concatenation	0.93	0.49	0.63
Features-Union	0.61	0.58	0.59
MKL-Average	1.0	0.46	0.63
MKL-Gram	1.0	0.42	0.59
EmbraceNet	0.59	0.72	0.66
Proposed model	0.58	0.87	0.70
Twitter			
Concatenation	0.76 ±0.01	0.77±0.01	0.76±0.02
Features-Union	0.88±0.02	0.75±0.02	0.80±0.02
MKL-Averag	0.89± **0.01**	0.77±0.02	0.81±0.02
MKL-Gram	0.89± **0.03**	0.75±0.00	0.83±0.02
EmbraceNet	0.87±0.02	0.86±0.03	0.89±0.02
Proposed Model	0.88±0.01	0.97± **0.02**	0.92± **0.01**

Table 5. Results of the proposed model in the depressive class vs. the top three places in eRisk 2018

Model	Precision	Recall	F1-score
Trotzek et al.	**0.64**	0.65	0.64
Paul et al.	0.63	0.64	0.63
Ramiandrisoa et al.	0.38	0.67	0.48
Model proposed	0.58	0.87	0.70

two models were built, one for the extraction of 18 characterizations because the combination of these provided the best results.

The second model corresponds to vectorization using doc2vec, and a voting ensemble determines if the user is depressed or not. The previous review shows many methods to detect depression from social media posts.

However, most of these works use complex models with various features and sophisticated approaches to text representations.

Instead, this work proposes a simple model that analyses and evaluates different representations (emotions, demographics, sentiment, personality, and thematic information) extracted from the user.

Please note that even when fusion methods have been used to approach this problem, they usually use different modalities to represent the user. However, in this work, all characterizations associated with the user were extracted from the same modality.

With this in mind, we aim to show that selecting characterizations with high diversity and using an adequate fusion method results in a model that improves the prediction of depressed users on social networks.

3 Combination of Users' Characterizations for Depression Recognition

This study aimed to evaluate the utility of a depression detection model based on the characteristics of social network users. One working hypothesis of this work is that combining different features will lead to better recognition performance.

In doing so, we suggest using a Gated Multimodal Units (GMU) based network. The remainder of this section elaborates on the characterizations and the fusion method that was adopted.

3.1 Feature Extraction

3.1.1 Pre-Processing and Feature Selection:

From data collections, preprocessing involves removing unnecessary information, such as special characters, numbers, URLs, and punctuation. After the user text was preprocessed, different features were selected to represent the user, described below:

1. **Sentiment Analysis:** This characterization was performed using two approaches. The first one employs NRC Emotion Lexicon (EmoLex)[1].

   The EmoLex lexicon contained eight basic emotions (Anger, Sadness, Fear, Anticipation, Trust, Surprise, Joy, and Disgust).

   The second approach employs CNN[2] to identify four emotions (Sadness, Fear, Anger, and Joy).

[1] https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[2] https://github.com/lukasgarbas/nlp-text-emotion

2. **Demographic Information:** For this characterization, we focused on inquiring about two attributes (Gender and Age); using a lexical resource available at World Well-Being Project[3].

   The prediction was performed at the publication level. So, the weights of each one are added up, and the final sum represents the result of the prediction in the user's history.

   The gender is indicated by the sign obtained from the result, a positive result indicates a female user, and a negative result indicates a male user. While age is represented by the sum of the weights.

3. **Polarity:** For polarity, each publication in the user's history, was analyzed to calculate polarity with the help of the lexical resource SentiWordNet[4], assigning the polarity (positive or negative) to the words.

   The positive and negative scores found for each term were added separately to obtain two different scores for each publication: the positive (S_+) and the negative (S_-).

   This was carried out for each of the publications of depressive and non-depressive users.

4. **Personality:** For personality prediction, we used the model developed in [14].

   Which predicts the Big5 (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) personality scores, the model integrates psycholinguistic features, including five lexical and LIWC features, along with features of BERT language model, using a Multilayer Perceptron (MLP).

5. **Thematic:** Finally, a vectorization of the users' histories was made using a weighted TF-IDF, which assigns weights to the words to discriminate between classes (depressive and non-depressive)

Emotion traits are included based on Chen et al. showing that emotion-based features have a positive impact on depression detection.

[3]https://wwbp.org/lexica.html

[4]http://sentiwordnet.isti.cnr.it/

Table 6. Results of the proposed model in the depressive class vs. the top three places in eRisk 2018

Model	F1-score
Shen et al. [22]	0.85
Titla-Tlatelpa et al. [24]	0.88
Model proposed	**0.92**

In addition to emotions, polarity analysis, associated with the vocabulary shared by users on social networks, may contain relevant information.

For example, when words associated with different everyday situations, are mentioned in positive or negative contexts.

In addition, profile information, such as demographic attributes, has been associated with depressive indicators. In addition, several studies conducted by psychologists have associated neurotic personality with emotional stability.

In addition to these, other results on the detection of mental disorders have concluded that personality is a relevant trait for discriminating those suffering from a mental disorder [10].

Finally, a thematic analysis provides a structured and systematic approach to understanding sentiment and allows the detection of patterns.

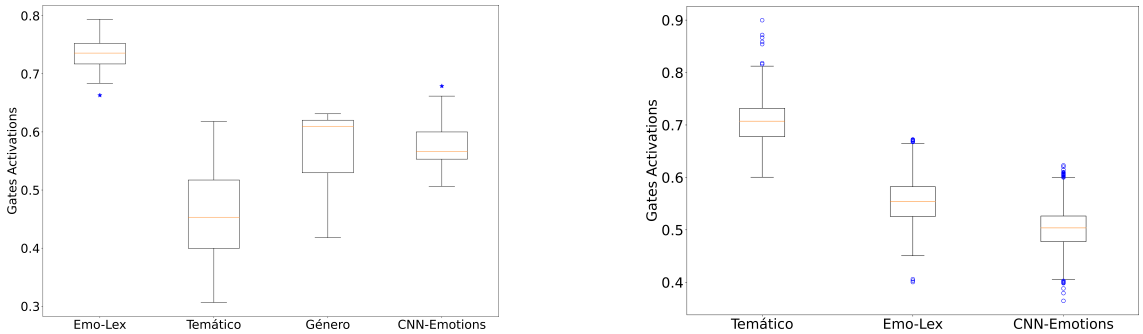
3.2 Information Fusion with Gated Multimodal Units

The information obtained through various characterizations should be fed into a predictive model that learns to distinguish between depressed and non-depressed users.

Since the characterizations considered capture different aspects of the user, models that exploit feature complementarity and redundancy are expected to yield better performance.

Although there are many methods to fuse information from multiple modalities for classification purposes, there are no established methods with proven performance.

After an extensive literature review, we found a promising model that could be used for this purpose: GMU-based networks.



(a) Reddit (b) Twitter

Fig. 2. Relevance of characterization in the depressive class for Reddit and Twitter according to GMU

This model has the advantage of finding the best representation between late and early fusion by combining attributes of different modalities and determining the relevance of each modality in the model to the predictive phase.

Figure 1 illustrates the proposed model, starting from the features extracted from the text. The features were normalized to pass through the GMU module as Arévalo et al., who further describes this module.

Once obtained, the GMU module is passed through a fully connected hidden layer, a dropout layer, and finally, an output layer with sigmoid activation, for final classification.

4 Experiments and Results

This section presents an extensive experimental evaluation of the methodology described in Section 3. We describe the considered datasets, baselines, and experimental results.

4.1 Corpora and Evaluation Metrics

To evaluate the proposed model, we considered two datasets related to depression on two social networks (Reddit[5] and Twitter [22]).

These two collections were built for the English language. In the case of Reddit, the dataset was constructed from a collection of depression-related posts.

This dataset was named e-Risk2018 and was employed throughout the CLEF evaluation forum.

On the other hand, the Twitter dataset was collected via API, and for this collection, users were considered depressed if any of their posts contained a self-reported clinical diagnosis of depression, such as: ”I am/am/was/was/have been diagnosed with depression”.

In Table 1 we can see some statistics for these datasets. Since the objective is to identify depressed users, and we do not have much positive sample data to work with, we proposed the use of a penalized model.

To do this, we pass the weights of each class in order to try to balance out the depressive class and thereby get the model to ”pay more attention” to samples from the underrepresented class.

In the case of Twitter, as there is no test set, a 5-fold cross-validation was performed for this dataset.

Here, we show different evaluation metrics to evaluate the performance of our model: Precision, Recall and F1-score of the depressive class. Note that in this work, we focused on enhancing the F1-score of the depressive class.

4.2 Data Fusion Baselines

We considered four data fusion methods as a baseline:

— Concatenation: Different works have concluded that a simple concatenation of representation could be good [2].

[5] https://early.irlab.org/2018/index.html

Table 7. Dataset analysis measures

Dataset	Jaccard	Imbalance Degree
Reddit	0.706261	0.33099
Twitter	0.255112	0.13735

— Features-Union [17]: This method is quiet similar to concatenation method; the principal difference between both is that this method assign the same weight to each modality and it is useful to combine several feature extraction mechanisms into a single.

— Multi-kernel: This method shows good performance among heterogeneous data. We implemented two Multi-Kernel learning styles with SVM: 1) MKL(Average) [11].

This is a simple wrapper that defines the combination as average base kernels; and 2) MKL (GRAM) [12].

Gradient-based RAdius-Margin optimization, this method focuses on finding the combination of the kernel that simultaneously maximizes the margin between classes while minimizing the resulting kernel's radius.

— EmbraceNet: This neural network-based method ensures compatibility with any learning model and correctly handles different modalities [8].

Furthermore, this model has been compared with several neural network fusion techniques, achieving better performance.

4.3 Experimental Results

As we mentioned before, we explored the individual performance of each of the previously described features using a Support Vector Machine with a linear kernel as a classification method.

Table 2 shows the evaluation metrics on the depressive class of the seven characterizations extracted from the users' posts on Reddit and Twitter. The Table above shows that the different characterizations behave differently for each dataset.

This behaviour might be related to the data and the way the models were built to extract the features, as a few of these methods were built for different types of datasets.

Once obtaining these results, we calculated the Coincident Failure Diversity (CFD) measure, to assess the diversity and complementarity between each fusion of the seven characterizations in the tow datasets (Reddit and Twitter).

This measure[6] is employed to determine the probability that members of the same system commit mistakes coincidentally.

Reddit. In Table 3, at the top, we can see the diversity between the fusion of each characterization for Reddit, as we can see the highest CFD of 0.87 was obtained with the fusion of the characterizations Emo-Lex, CNN-Emotions, thematic information, and gender.

This suggests that for the Reddit dataset, these four characterizations provide the highest diversity and complementarity. Since the table is so large, we only show the highest CFD obtained with the four characterizations mentioned previously and the CFD of the fusion of the seven characterizations.

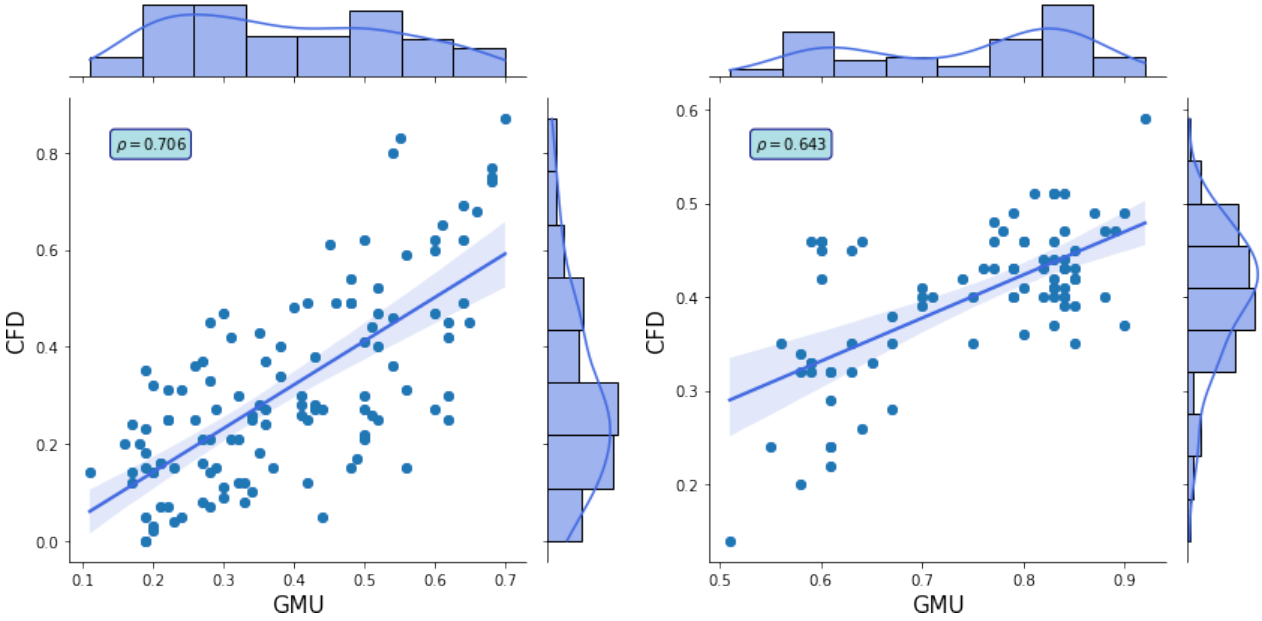
Twitter. In Twitter's case, the characterizations with the highest Coincident Failure Diversity are Emo-Lex, CNN-Emotions, and Thematic Information. As shown in Table 3 at the bottom, these characterizations achieve a CFD of 0.59.

Again, we only show the highest CFD achieved with the three above-mentioned characterizations and the CFD of the fusion of the seven characterizations.

Table 3 shows that the characterizations with the highest CFD, change throughout the datasets. Although, in both cases, emotions and thematic information have a relevant impact on obtaining the highest diversity.

With this in mind, if we choose an appropriate fusion method, we can achieve a model with better performance. To evaluate the GMU module

[6] When $CFD = 0$, it indicates that the faults are the same for all the characteristics; therefore, there is no diversity. On the contrary, a $CFD = 1$, indicates that all the faults are unique.



(a) Reddit (b) Twitter

Fig. 3. Pearson correlation between GMU performance and CFD

properly, we considered the use of four different fusion methods, previously described.

All the methods presented here were both trained and tested with the same erisk2018 and Twitter datasets. For this experiment, we expect to confirm the hypothesis presented in this paper, and we also seek to determine the effectiveness of using GMU over traditional fusion methods.

Reddit. In the case of Reddit, we trained and tested the fusion methods with the four characterizations that obtain the highest CFD. In this dataset, the results can be observer at the first part of Table 4.

Twitter. In the second part of the Table 4, we can see the performance of each fusion method on the Twitter dataset, using the three characterizations with the highest CFD for this dataset.

For both the Reddit and Twitter datasets 4, the neural network-based fusion methods are the highest F1 scores, of which the proposed model outstood by obtaining the best performance.

This is because the GMU module achieves better learning of the diversity of characterizations compared to the other fusion methods for both cases.

4.4 Comparison with the State of Art

Reddit. In the case of Reddit, we compare ourselves with various works presented in the CLEF 2018 evaluation forum, which make use of fusion techniques, where they use the same dataset as in this work (eRisk2018). The results can be observed in Table 5.

Twitter. On Twitter, we compare ourselves with other works that use the same dataset described above. These works only report the F1-score. The results can be observed in Table 6.

For both cases, the proposed method outperformed the state-of-the-art works. Thus, of the results obtained, the following stand out:

1. The implementation of a GMU in a simple neural architecture, together with an adequate selection of characterizations, outperformed various traditional fusion techniques implemented as a baseline, indicating that fusing the different characterizations at a deeper level is indeed relevant for depression detection.

2. Our approach outperformed the eRisk2018 winner approaches and the state-of-the-art works for Twitter. It is key to note that some of the works presented in the state of the art tried different complex models with a wide range of features using traditional fusion methods such as late and early fusion.

   In contrast, the one presented here was only based on the use of 4 characterizations for Reddit and 3 characterizations for Twitter, and a GMU module as a fusion method.

5 Analysis of the Proposed Model

With this analysis, we expect to observe how GMU determines the relevance of each characterization according to the dataset. In Fig 2, we can observe the relevance of each feature of the depressive class for Reddit and Twitter.

From Figure 2, we can observe how the activation of each characterization changes depending on the dataset. For example, for Reddit, Fig.2(a), the most relevant characterization is Emo-Lex, while for Twitter, Fig.2(b), thematic information has the greatest relevance.

This could be related to the construction of the datasets and the similarity of the classes (depressive and non-depressive).

Nevertheless, both datasets agree that the characteristics attached to emotions and thematic are the most important to identify depressive users from non-depressive users.

5.1 GMU Error Analysis

Based on the results obtained, we decided to evaluate and determine why some users were misclassified and if there is any common reason between the data sets. For both datasets the errors were related to the amount of information.

Very short publication histories do not provide sufficient information for the model. Whereas very long histories the features may vary over a long period of time. Therefore, in both cases, the signs of depression are not so clear to the model, so it cannot adequately discriminate between classes.

5.2 Diversity and Model Performance

Looking at Figure 3, we can see that in the case of Reddit the Pearson [5][7] correlation between the diversity obtained with the CFD measure and the F1-score of the depressive class in the GMU architecture is 0.71, which indicates that there is a strong correlation between both variables, so the more diversity, the higher the performance for the detection of depressive users.

The same for the case of Twitter where we obtain a correlation of 0.64 between diversity and the F1-score of the depressive class of the GMU model.

5.3 Differences between Reddit and Twitter

In this analysis, we decided to compare some elements of both datasets to provide information on the differences between the relevance of the characterizations in the datasets.

To address this, we decided to assess the similarity between the classes, depressives and non-depressives, in both datasets (Reddit and Twitter).

As we can see in Table 9, the first column indicates the vocabulary overlap between the classes with Jaccard coefficient.

This coefficient is a statistical measure used in natural language processing to compare the similarity between documents [13].

[7] Pearson's correlation coefficient assigns values between -1 and 1, where 0 indicates that there is no correlation, 1 is a total positive correlation and -1 indicates a total negative correlation

In our case, we can see that the Jaccard coefficient is higher on Reddit than on Twitter. This tells us that in the Reddit dataset the vocabulary used by both classes is close enough.

While the one used by Twitter differs significantly between both classes, which could explain the main reason why the thematic characterization is more relevant for Twitter than for Reddit.

On the other hand, in column 2 we show the degree of imbalance between classes where, again, Redditt is higher than Twitter.

The degree of imbalance between classes is an important feature to take into account when working with corpora, since depending on the degree of imbalance there may be different levels of difficulty [3]. Therefore, we may find it more difficult to classify depressive users on Reddit than on Twitter.

6 Conclusion and Future Work

The task of detecting depression in social networks is not trivial. However, different works give us insight into this disorder, showing that including a wide range of features is not necessarily helpful for the performance of the model.

Also, the use of deep fusion models has been compared in recent years with traditional fusion methods, showing improvements in several situations. However, these models have been poorly explored for the task of depression detection.

In this paper, we present a study of the importance of selecting characterizations based on their diversity along with the integration of a neural network-based deep fusion method.

We compared the performance of GMU with other fusion methods such as late, and early, to identify depressive users, where the neural network-based fusion techniques showed better performance, highlighting the GMU module.

Furthermore, we compare the results obtained in this work with the works presented in the state of the art for both datasets, showing that our model outperforms these works.

In future work, we expect to explore the approach proposed here, in other languages to observe the behavior of the characterizations related to the language used by users in social networks and see if they are affected by cultural differences and thereby determine whether the model presented here can be adapted in other languages.

Acknowledgments

This work was supported by CONACyT/México (scholarship 782579 and grant CB-2015-01-257383).

Authors thank CONACYT for the computational resources provided by the Deep Learning Platform for Language Technologies.

References

1. **Aragón, M. E., López-Monroy, A. P., González-Gurrola, L. C., Montes, M. (2019).** Detecting depression in social media using fine-grained emotions. Proceedings of the NAACL-HLT, Vol. 1, pp. 1481–1486. DOI: 10.18653/v1/n19-1151.
2. **Arevalo, J., Solorio, T., Montes-y-Gomez, M., González, F. A. (2020).** Gated multimodal networks. Neural Computing and Applications, pp. 1–20. DOI: 10.48550/ARXIV.1702.01992.
3. **Barrón-Cedeño, A., Rosso, P., Pinto, D., Juan, A. (2008).** On cross-lingual plagiarism analysis using a statistical model. Proceedings of the International Conference on Uncovering Plagiarism, Authorship and Social Software Misuse, Vol. 212, pp. 9–14.
4. **Baruah, T. D., Kanta, K., State, H. (2012).** Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study. International journal of scientific and research publications, Vol. 2, No. 5.
5. **Boslaugh, S. (2012).** Statistics in a nutshell: A desktop quick reference. 2 edition.
6. **Bucci, W., Freedman, N. (1981).** The language of depression. Bulletin of the Menninger Clinic, Vol. 45, No. 4, pp. 334–358.

7. **Chen, X., Sykora, M., Jackson, T., Elayan, S., Munir, F. (2018).** Tweeting your mental health: An exploration of different classifiers and features with emotional signals in identifying mental health conditions. Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 3320–3328.
8. **Choi, J. H., Lee, J. S. (2019).** EmbraceNet: A robust deep learning architecture for multimodal classification. Information Fusion, Vol. 51, pp. 259–270.
9. **Jackson, P. B., Williams, D. R. (2006).** Culture, race/ethnicity, and depression. Women and Depression: A Handbook for the Social, Behavioral, and Biomedical Sciences, pp. 328–359. DOI: 10.1017/cbo9780511841262.016.
10. **Kendler, K. S., Gatz, M., Gardner, C. O., Pedersen, N. L. (2006).** Personality and major depression: a swedish longitudinal, population-based twin study. Archives of general psychiatry, Vol. 63, No. 10.
11. **Lauriola, I., Aiolli, F. (2020).** MKLpy: a python-based framework for multiple kernel learning. DOI: 10.48550/ARXIV.2007.09982.
12. **Lauriola, I., Polato, M., Aiolli, F. (2017).** Radius-margin ratio optimization for dot-product boolean kernel learning. Proceedings of the Artificial Neural Networks and Machine Learning – ICANN, Springer, pp. 183–191. DOI: 10.1007/978-3-319-68612-7_21.
13. **Manning, C., Schutze, H. (1999).** Foundations of statistical natural language processing.
14. **Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E., Eetemadi, S. (2020).** Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. 2020 IEEE International Conference on Data Mining (ICDM), IEEE, pp. 1184–1189.
15. **Meng, H., Huang, D., Wang, H., Yang, H., Ai-Shuraifi, M., Wang, Y. (2013).** Depression recognition based on dynamic facial and vocal expression features using partial least square regression. Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, pp. 21–30. DOI: 10.1145/2512530.2512532.
16. **Paul, S., Jandhyala, S. K., Basu, T. (2018).** Early detection of signs of anorexia and depression over social media using effective machine learning frameworks. Proceedings of the 9th Conference and Labs of the Evaluation Forum Association Conference, pp. 2125.
17. **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011).** Scikit-learn: Machine learning in python. Journal of Machine Learning Research, Vol. 12, pp. 2825–2830.
18. **Peng, Z., Hu, Q., Dang, J. (2019).** Multi-kernel svm based depression recognition using social media data. International Journal of Machine Learning and Cybernetics, Vol. 10, No. 1, pp. 43–57. DOI: 10.1007/s13042-017-0697-1.
19. **Preoţiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., Schwartz, H. A., Ungar, L. (2015).** The role of personality, age, and gender in tweeting about mental illness. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 21–30. DOI: 10.3115/v1/w15-1203.
20. **Ramiandrisoa, F., Mothe, J., Benamara, F., Moriceau, V. (2018).** IRIT at e-Risk 2018. Proceedings of the 9th CLEF Association Conference, pp. 1–12.
21. **World Health Organization (2020).** Depression.
22. **Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T. S., Zhu, W. (2017).** Depression detection via harvesting social media: A multimodal dictionary learning solution. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 3838–3844.
23. **Skaik, R., Inkpen, D. (2020).** Using social media for mental health surveillance: A review. ACM Computing Surveys, Vol. 53, No. 6, pp. 1–31. DOI: 10.1145/3422824.
24. **Titla-Tlatelpa, J. d. J., Ortega-Mendoza, R. M., Montes-y-Gómez, M., Villaseñor-Pineda, L. (2021).** A profile-based sentiment-aware approach for depression detection in social media. EPJ Data Science, Vol. 10, No. 1. DOI: 10.1140/epjds/s13688-021-00309-3.
25. **Trotzek, M., Koitka, S., Friedrich, C. M. (2018).** Word embeddings and linguistic metadata at the clef 2018 tasks for early detection of depression and anorexia. Proceedings of the 9th Conference and Labs of the Evaluation Forum Association Conference, pp. 10–14.
26. **Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., Wolf, M. (2017).** First-person

pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. Clinical psychology and psychotherapy, Vol. 24, No. 2, pp. 384–391.

Article received on 21/08/2022; accepted on 01/12/2022.
Corresponding author is Luis Villaseñor-Pineda.