UNED_MRES Team at MentalRiskES2024: Exploring Hybrid Approaches to Detect Mental Disorder Risks in Social Media

Modesto Sierra-Callau[1,*], Miguel Ángel Rodríguez-García[1], Soto Montalvo-Herranz[2] and Raquel Martínez-Unanue[1]

[1]*Universidad Nacional de Educación a Distancia, Spain*
[2]*Universidad Rey Juan Carlos, Spain*

Abstract

Depression is a widespread mental disorder that significantly contributes to suicide worldwide. Research has identified a correlation between this psychological disorder and people's emotions, feelings, thoughts, and communication methods. This has made it very challenging to develop techniques to identify and analyze the communication methods of people suffering from this disorder. In this work, we describe the system built for the MentalRisk 2024 shared task, specifically for two of the three proposed activities: mental disorder detection (anxiety/depression/none) and context detection (for those subjects identified in the previous task). Two techniques were primarily designed to tackle the challenge: a Deep Learning model based on transformer architecture, and a Machine Learning model using traditional classifiers. We studied various transformer architectures and pre-processing techniques for the Deep Learning part. In the Machine Learning part, we employed several embedding combination methods to analyze the performance of traditional classifiers. The results indicate a slight difference between the two selected strategies, with a significant improvement compared to the established baseline. The best results were achieved by the transformer architecture, with a precision score of 0.64 and a recall score of 0.62.

Keywords

Depression detection, Transformers-based language models, Machine learning models, Early risk prediction of depression

1. Introduction

Depression is currently one of the most common psychiatric disorders affecting people worldwide [1]. Its prevalence is increasing every year, posing a serious medical, economic, and social problem [2]. This disorder directly impacts a person's growth, influencing their thoughts, feelings, and behaviors [3]. Therefore, early detection plays a vital role in reducing the number of affected individuals [4].

One symptom that affects people suffering from this disorder is the linguistic footprint, which reflects subtle changes in speech production [5]. This has enabled the community to

IberLEF 2024, September 2024, Valladolid (Spain)
*Corresponding author.
msierra@barbastro.uned.es (M. Sierra-Callau); miguelangel.rodriguez@lsi.uned.es (M. Á. Rodríguez-García); soto.montalvo@urjc.es (S. Montalvo-Herranz); raquel@lsi.uned.es (R. Martínez-Unanue)
0009-0003-1256-1163 (M. Sierra-Callau); 0000-0001-6244-6532 (M. Á. Rodríguez-García); 0000-0001-8158-7939 (S. Montalvo-Herranz); 0000-0003-1838-632X (R. Martínez-Unanue)

develop Natural Language Processing techniques to detect depression from text. Consequently, the detection and identification of mental disorders have become an attractive task, posing interesting challenges that have attracted the research community's attention.

Given its impact, there have been several proposed challenges to develop tools for early detection of mental disorders. In fact, in its second edition, the MentalRiskES2024 challenge [6] (which is part of the IberLEF2024 [7] shared evaluation campaign) presents three tasks related to the early detection of mental disorder risks in the Spanish language (Disorder detection, context detection, and suicidal ideation detection). In particular, Task 1 is a classification problem involving the detection of mental disorders such as depression, anxiety, or none. Task 2 involves the detection of contexts associated with the disorder through a multilabel classification problem and the identification of early risk. The third subtask involves the detection of suicidal ideation. The team participated in Task 1 with three runs using two different approaches (run 0 for DL and run 1 and run 2 for ML) and also participated in Task 2 with DL. In this work, we describe the approach presented for the challenge. We have experimented with two types of approaches: deep learning and machine learning. From a deep learning perspective, we have explored various language models, including ROBERTA, BETO, and BERT. On the other hand, from a machine learning perspective, we utilized Logistic Regression, Multilayer Perceptron, Naive Bayes, and Random Forest. The evaluation results indicate that language models are better able to generalize knowledge for classifying unseen data.

The paper is organized as follows: Section 2 analyzes proposed works related to the detection of mental disorders. Section 3 describes the architecture of the system. Section 4 collects and discusses the performance of the techniques employed in the tasks that we participated in. Finally, Section 5 reflects on the outcomes obtained during our participation and provides future directions to consider.

2. Related work

Due to evidence that the most common mental disorders impact a person's daily activities [8], there has been considerable interest in the community to research the application of Artificial Intelligence techniques to early infer these disorders from non-invasive data, such as speech or writing [9, 10]. In the literature, there are studies that employ different methods to infer depression disorder from natural language, particularly writing. Thus, Figuerêdo, Maia, and Calumby [11] proposed a method for detecting early depression in social media users. The method combines neural networks, word embeddings, and data fusion techniques. It uses a Convolutional Neural Network for classification, context-independent word embeddings for knowledge representation, and two data fusion approaches: Early Fusion, which combines raw data before feature extraction, and Late Fusion, which combines classification results [12]. In their evaluation, they utilized the eRisk 2017 dataset [13], which compiles over 2000 posts from Reddit users in the English language. Their study involved comparing the performance of their approaches with the best ones achieved in the eRisk 2017 challenge, as well as a main baseline configured from FastText Wiki and Meta LR (Learning Rates) models. They proposed eleven approaches by combining word embeddings obtained from four different models: FastText Crawl, FastText WN, Glove Crawl, and Glove WN, and employing two data fusion techniques.

As a result, they arrived at two main conclusions. Firstly, for this task and the selected dataset, the Late Fusion approach outperformed the Early Fusion, exceeding a 10% improvement in F-measure. Secondly, the model achieved better performance, as combining four types of embeddings resulted in a slightly better performance than using only one.

In a similar research study, Fatima et al. [14] examined the performance of various Machine Learning and Deep Learning models. These models included: i) Naive Bayes, ii) Convolutional Neural Network (CNN), iii) Long Short Term Memory networks (LSTM), and iv) CNN + BiLSTM with attention model. The proposed architecture consisted of five layers. The first layer managed the dataset, the second layer preprocessed the dataset by performing cleaning and tokenization tasks, the third layer served as the embedding layer to transform the text representation into a format understandable by neural network architectures, the fourth layer functioned as the learning layer where the models were located to carry out the classification tasks, and finally, the softmax layer provided the probability outcomes of the classification. The analysis of the discussion behind the results of the eRisk 2018 challenge concludes that, in this case, more complex models yield better results than less complex ones.

Finally, in their study, Islam et al. [15] approach depression detection from a different perspective by using traditional Machine Learning techniques. They focus on four classifiers and their variants: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision trees (DT), and Ensemble, which involves a weighted combination of multiple classification models. The study analyzes users' depressive behaviors from three perspectives: emotional process, temporal process, and linguistic style. The evaluation is conducted on a raw dataset collected from Facebook, which was processed and formatted into a set of columns organized based on the mentioned perspectives. After analyzing the study results, it is not possible to identify the best technique across all dimensions of the analysis. This is because the selected variants produce significantly different results. For example, techniques like Decision Trees and Ensemble Classifiers achieve high precision values for the emotional process dimension, but their performance in the linguistic style dimension is comparatively lower.

Considering the various approaches analyzed, we selected certain features to design our architecture. We decided to tackle the challenge by creating a hybrid approach that combines two main classifier methods: transformer models and traditional models. The following sections will outline the proposed architecture.

3. Material and methods

This section describes two crucial elements in this contribution: the dataset distribution provided by the challenge for the assessment phases and the solution devised to address it.

3.1. Dataset analysis

The MentalRiskES2024 dataset [16] is composed by a set of messages sent to groups on the Telegram platform. Messages correspond to 885 users in total (20 users for trial, 465 for training and 400 for testing) (see Table 1). There is a variable number of messages for each user (with an average number of 50 messages for each of them). These messages are grouped by subject and

each of these subjects is labelled in a gold standard file for each task (task 1 and task 2). Test dataset was not available for the research time until the test phase.

Table 1
Dataset characteristics

Dataset	n records
Trial	20
Train	465
Test	400
Total	885

The preliminary analysis of the trial and training datasets, showed that the distribution of the labels in the datasets was not uniform and the sets of subjects belonging to each of the categories were not balanced. Table 2 shows the distribution of labels in the dataset for the two tasks in which the team took part.

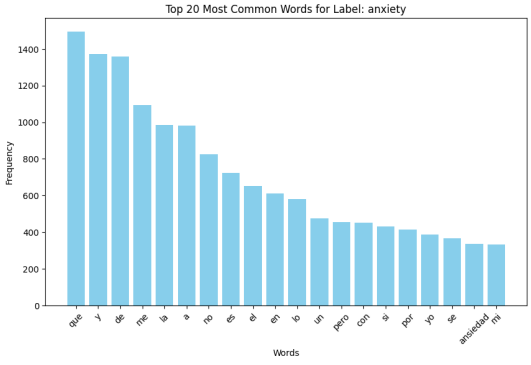
Table 2
Distribution of labels in the provided dataset for two tasks.

Set	N. subjects	Task 1: Mental Disorder Detection						
		none			**depression**		**anxiety**	
Train	20	10			5		5	
Trial	465	213			164		88	
Train+Trial	485	223			169		93	
Set	**N. subjects**	Task 2: Context Detection						
		addiction	**emergency**	**family**	**work**	**social**	**other**	**none**
Train	10	0	0	0	3	0	3	4
Trial	252	12	17	61	17	88	56	74
Train+Trial	262	12	17	61	17	91	59	78

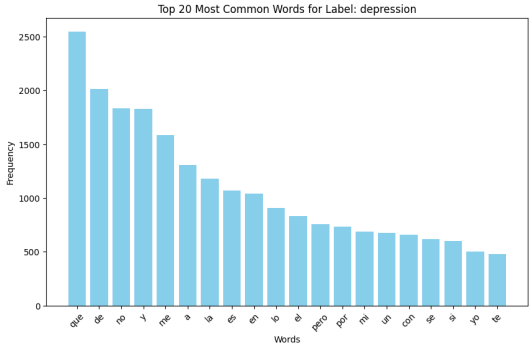
In order to gain a better understanding of the dataset distribution, we conducted a thorough analysis to examine how the dataset was compiled in terms of word distribution for Task1. We believed that this analysis could provide valuable insights for the feature extraction process. First, we performed a frequency analysis to identify the most commonly used words in the messages labeled as “Anxiety” and “Depression”. Then, we analyzed the most common words related to them. Figure 1 illustrates the most frequently used words associated with Anxiety and Depression.

The initial analysis of these frequency diagrams reveals that, apart from ”anxiety”, the most frequent words consist of meaningless stopwords.

After removing the stopwords using the NLTK library [17], the frequency histogram displayed in Figure 2 shows that the word ”cara” (face) is one of the most common words in both sets. This may be related to the use of emoticons expressing emotions in social media. As a result, we decided to take into account replacement of the most frequent emoticons with sentences that describe their meaning.

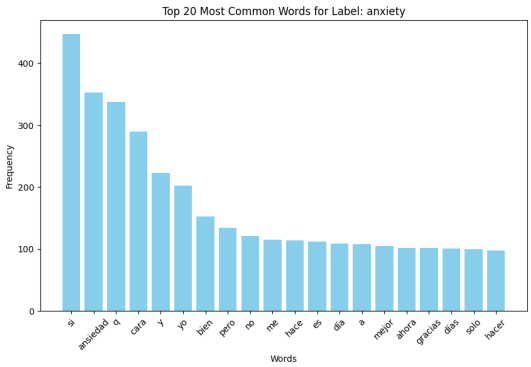


(a) Most common words “Anxiety”

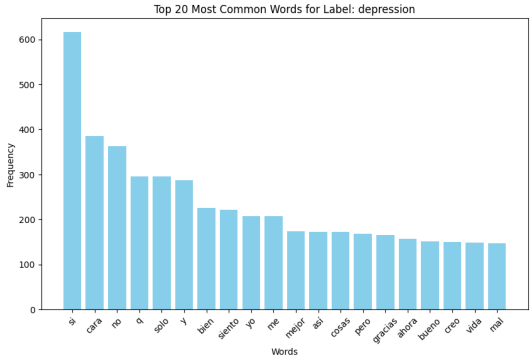


(b) Most common words “Depression”

Figure 1: Comparison of word usage in texts labeled with “Anxiety” and “Depression”



(a) Most common words “Anxiety”



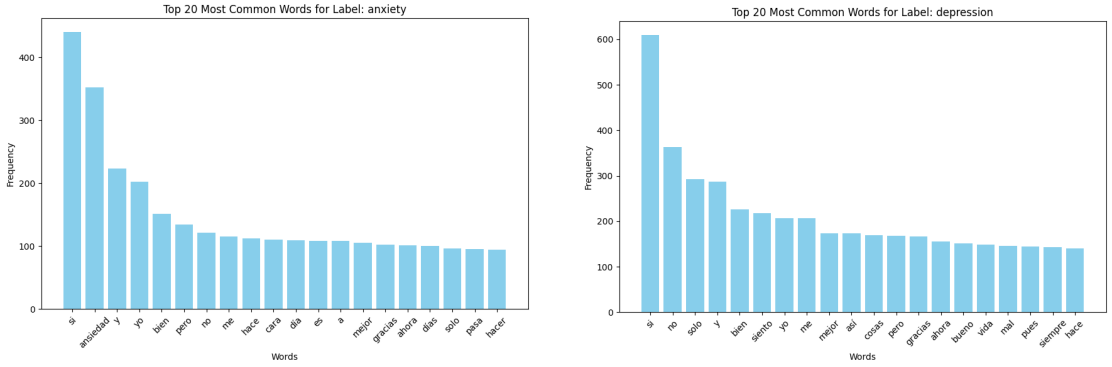
(b) Most common words “Depression”

Figure 2: Comparison of word usage in texts labeled with “Anxiety” and “Depression” removing stopwords

Replacing emoticons with text and removing stopwords configure the bar charts of word frequencies in Figure 3. The analysis indicates that the word “cara” occurs less frequently in both datasets. Some words, such as “pero”, “bien”, “gracias” and “mejor” are common in both datasets with similar frequencies. However, other words, such as “siento”, “bueno”, “vida”, “mal” and “siempre” are more related to “Depression” while “ansiedad”, “dia” and “gracias” are more associated with “Anxiety”.

3.2. System architecture

The system proposed for the challenge consists of two main modules: 1) A Deep Learning module that incorporates a transformers model with two different configurations, one for each task. Implementing this required studying the most common Spanish language transformers that could be applied to the proposed tasks. 2) A Machine Learning module that includes two different approaches for the first task. It aims to combine innovative ideas about word



(a) Most common words "Anxiety"

(b) Most common words "Depression"

Figure 3: Comparison of word usage in texts labeled with "Anxiety" and "Depression" with emoticon substitution by description and removing stopwords

embeddings and dataset oversampling with traditional approaches. The architecture of the proposed system, which includes both approaches, is shown in Figure 4.

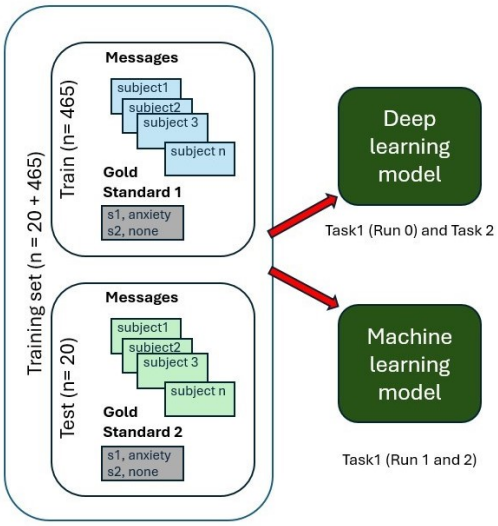


Figure 4: The architecture of the system proposed.

The system functions as follows: it takes in a user message as input, and, depending on the task being executed, one module or both modules are involved in the classification process. For example, messages from task 1 will be classified by both modules, while those received for task 2 will only be processed by the deep learning module. In the following subsections, each module is detailed.

3.2.1. Deep Learning module

To design this module, an intensive analysis of Spanish language models was carried out. The list below details each model analyzed to tackle the first two tasks proposed in the challenge.

- dccuchile/bert-base-spanish-wwm-cased (BETO cased) [18]. BETO is a BERT model trained on a big Spanish corpus. BETO has a size similar to the one of BERT-Base and was trained with the Whole Word Masking technique [19].
- dccuchile/bert-base-spanish-wwm-uncase (BETO UNCASED) [18]. In the uncased version of the BETO model, the text is lowercased before WordPiece tokenization.
- bertin-project/bertin-roberta-base-spanish [20]. BERTIN is a series of BERT-based models for Spanish based on the RoBERTa base model.
- PlanTL-GOB-ES/roberta-base-bne [21]. (RoBERTa BNE) The roberta-base-bne is a transformer-based masked language model for the Spanish language. It is also based on the RoBERTa [22] base model and has been pre-trained using the largest Spanish corpus, compiled from the National Library of Spain ("Biblioteca Nacional de España") from 2009 to 2019.

Once the models were selected, the next step was to prepare the text for training. The design of the message pre-processing method was based on the following two ideas:

- Removing stopwords. We employed the Natural Language Toolkit (NLTK) [17] [1] to filter out those words that are considered insignificant.
- Translating emoticons. We developed a text processing function that implements a list of emoticons and their corresponding meanings. It is designed to translate emoticons within a piece of text. Listing 1 shows some examples about how this translation is made.

```
emoticons = {
    "cara llorando de risa": "me estoy riendo mucho",
    "corazón rojo": "amor",
    "corazón roto": "es doloroso",
    "cara con ojos de corazón": "estoy enamorado",
    "cara feliz con ojos sonrientes": "estoy feliz",
    "pulgar hacia arriba": "de acuerdo",...}
```

Listing 1: Emoticon substitution with expression

The transformer architecture is designed to support text lengths of up to 512 tokens. However, the concatenation of messages for one user provided in the trial dataset either exceed this threshold. To address this issue, rather than training the model on individual messages, we concatenated all messages from each user and split them into strings that could meet the length requirement of the model. As a result, the initial trial and training datasets, which originally consisted of messages from 20 and 465 subjects respectively, were transformed into datasets of different sizes (e.g. 130 and 2390 dictionary entries for one of the combinations).

[1] https://www.nltk.org/

3.2.2. Machine Learning module

This module offers two solutions for task 1, both based on combining two strategies: contextual embeddings to condense textual representation into numerical values that models can process, and the use of synthetic oversampling techniques to augment the dataset distribution. These values are extracted from a transformer architecture and a traditional machine learning technique. Creating this numerical representation required time, as the transformer architecture provides 12 layers from which the embeddings can be extracted. After a literature review and a series of tests, we decided to use the last layer, as this configuration achieved better performance than others. To construct this representation, we attempted to use the same transformer models mentioned earlier. For the machine learning part, we implemented various classifier techniques: logistic regression, multilayer perceptron, naive bayes, random forest, decision tree, and support vector machine. The algorithm selection was conducted during the training phase. All algorithms were trained, and the best performing one was employed in the test.

The two solutions proposed in this module differ in the way embeddings are used to train traditional classifiers. In the first approach, embeddings are directly used to train the classifiers, while in the second approach, a technique called Synthetic Minority Oversampling Technique (SMOTE) is employed to balance the dataset by creating 300 samples for each label. Augmentation techniques can be applied to text data or embeddings. In the first case, it can be applied at different levels, such as characters, words, sentences, and documents. Techniques like translation and synonym replacement are used in this case. In the second case, techniques like combination, swapping, and mathematical operations such as mean and sum are employed to oversample datasets. In our work, we utilized the SMOTE technique, which creates synthetic examples by randomly replacing samples considering the nearest neighbor for augmentation.

4. Results and discussion

This section outlines the details of the experimentation conducted for the challenge. For the Machine Learning approach, no pre-processing tasks were conducted over the dataset, and as mentioned in the preceding section, we utilized a transformer model as an extractor tool. On the other hand, the transformer architectures used for the challenge were configured with the following hyperparameters: 3.00E-05, 8, 0, 3, learning rate, batch size, number of warm up steps, and number of epochs, respectively. For the first task, we evaluated four different transformer architectures using the training set provided by MentalRiskES2024. Additionally, three different pre-processing techniques were employed: with stopwords, without stopwords, and emoticon substitution. Also different tests were made with only the first messages of a user (for early detection) and also some others by generating entries in the data dictionary splitting messages in entries with lengths suitable for the model (as described in section 3.2.1). Below, Table 3 collects the preliminary results obtained on the training and test dataset provided.

When analyzing the Machine Learning approach, the best results are achieved by combining RoBERTa with LR and RoBERTa with LR and SMOTE. The former reaches a Precision, Recall, and F1 score of 0.80, while the latter obtains 0.75, 0.70, and 0.71, respectively. On the other hand, the worst results were obtained by RoBERTA combined with NB and RoBERTA combined with DS and SMOTE. The first approach achieved scores of 0.24 for Precision, 0.30 for Recall, and

Table 3
Results of all combinations of machine learning and deep learning models with preprocessing techniques. LR stands for Logistic Regression, NN for Neural Network, NB for Naive Bayes, RF for Random Forest, DS for Decision Tree, and SVM for Support Vector Machine.

Model	Stop	Emoticon	TrainDS	TestDS (trial)	P	R	F1
RoBERTa + LR	NO	NO	465	20	0.80	0.80	0.80
RoBERTa + NN					0.75	0.75	0.75
RoBERTa + NB					0.24	0.30	0.23
RoBERTa + RF					0.67	0.65	0.65
RoBERTa + DS					0.67	0.50	0.49
RoBERTa + SVM					0.75	0.75	0.75
RoBERTa + LR + SMOTE	NO	NO	465	20	0.75	0.70	0.71
RoBERTa + NN + SMOTE					0.70	0.65	0.66
RoBERTa + NB + SMOTE					0.55	0.40	0.39
RoBERTa + RF + SMOTE					0.71	0.70	0.70
RoBERTa + DS + SMOTE					0.25	0.25	0.25
RoBERTa + SVM + SMOTE					0.72	0.70	0.70
BETO c	NO	NO	2390	130	0.56	0.50	0.50
	NO	YES	2382	128	0.52	0.50	0.50
	YES	NO	1716	89	0.58	0.55	0.54
	YES	YES	1713	89	0.56	0.53	0.53
BETO u	NO	NO	2390	130	0.57	0.51	0.51
	NO	YES	2382	128	0.46	0.42	0.43
	YES	NO	1716	89	0.50	0.44	0.43
	YES	YES	1713	89	0.55	0.51	0.5
BERTIN	NO	NO	2390	130	0.17	0.33	0.23
	NO	YES	2382	128	0.56	0.54	0.53
	YES	NO	1716	89	0.17	0.33	0.23
	YES	YES	1713	89	0.56	0.57	0.56
RoBERTa	NO	NO	2390	130	0.55	0.51	0.52
	NO	YES	2382	128	0.56	0.56	0.55
	YES	NO	1716	89	0.51	0.50	0.50
	YES	YES	1713	89	0.56	0.55	0.54

0.23 for F1, while the second approach scored 0.25 for Precision, 0.25 for Recall, and 0.25 for F1. Based on these results, it can be inferred that the oversampling method does not improve the model's ability to identify important features for better performance. This is evident as the results obtained from this method are significantly worse compared to other machine learning approaches. This suggests that the features extracted by the chosen transformer architectures lack the necessary distinguishing power to accurately perform the task, even when instances are augmented.

For the Deep Learning Approach, and after analyzing the preliminary results, we decided to use the model dccuchile/bert-base-spanish-wwm-uncase (BETO UNCASED) [18] without

any additional preprocessing such as removing stop words or substituting emoticons. Despite the slightly worse results than other selected transformer architectures, we choose the uncased version of BETO to address the tasks. We wanted to assess if its performance was superior to the other chosen architectures, as has been the case in other medical domain-related challenges [23] Once we have chosen the best configuration, we will implement the selected approaches on the test data using the evaluation service provided by the committee. The performance achieved in the test dataset is summarized in Table 4. The results indicate that the RoBERTa baseline models achieve better results compared to the BETO uncased option without preprocessing. This suggests that while BETO may have been initially more suitable for non-English texts, the current implementations of multilingual RoBERTa, which have been trained with large corpora, exhibit better performance in certain tasks.a

Table 4
Results of NLP UNED MRES Team from MentalRiskES2024 for Multi-class classification for task 1 and task 2

Model	Task	Rank	Run	Acc	Macro_P	Macro_R	Macro_F1
BETO u (NO/NO)		22	0	0.557	0.644	0.62	0.561
RoBERTa + LR	1	29	1	0.352	0.564	0.402	0.264
RoBERTa + LR + SMOTE		30	2	0.318	0.664	0.383	0.237
BETO u (NO/NO)	2	16	0	0.55	0.63	0.608	0.55

For the first approach, the results obtained with the Test server for task 1 are very similar to those obtained during the training process for the model selection (Accuracies of 0.557 and 0.561 respectively). Other results as the Macro Precision, Macro Recall and Macro F1 scores were slightly better at the Test server, indicating a more balanced performance across different classes. This might due to the fact that the number of records used for training the final model was bigger than the original one as it included also trial data. Results in terms of efficiency also show that there is still room for improvement. The average values obtained for CO2 emissions are 2.06E-03 for task 1 and 2.02E-03 for task 2. These values rank low when compared to those from other teams participating in the tasks.

The analysis of the results shows that the performance obtained by the two presented approaches can still be improved. We believe that the regular performance might be due to the following points: i) selecting a non-optimal transformer model; ii) selecting non-optimal hyperparameters; iii) inefficient use of augmentation techniques in the constructed pipeline.

5. Conclusions

In this work, we present our contribution to the MentalRiskES2024 challenge, which aims to detect specific mental disorders through the behavior of social network users. The challenge includes three different tasks: detecting mental disorders, detecting context for these mental disorders and also early risk for suicide prevention. These tasks involve binary classification, a combination of binary and multiclass classification, and regression problems. In this work, we focus on describing our contribution to the first two tasks.

The main objectives of this work were to pave the way for the application of NLP techniques in the mental health field, and explore different alternatives based on both ML and DL models. Both objectives have been accomplished. Results show that there are approaches that have obtained better performance in the application of these techniques, so further work must be done in order to examine what are the shortcomings of our approach.

In future work, we plan to explore additional augmentation techniques at various levels as we believe this is one of the main drawbacks that has affected our system's performance. Additionally, we intend to employ other feature extraction methods to consider not only the embeddings generated by the transformation architectures.

Acknowledgments

This work has been partially supported by the projects DOTT-HEALTH (PID2019-106942RB-C32, MCI/AEI/FEDER, UE); GELP (TED2021-130398B-C21, MCI/AEI/10.13039/501100011033 and NextGenerationEU/PRTR), and EDHER-MED (PID2022-136522OB-C21 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE).

References

[1] M. Gałecka, K. Bliźniewska-Kowalska, M. Maes, K.-P. Su, P. Gałecki, Update on the neurodevelopmental theory of depression: is there any 'unconscious code'?, Pharmacological Reports 73 (2021) 346–356.

[2] M. Kowalczyk, J. Szemraj, K. Bliźniewska, M. Maes, M. Berk, K.-P. Su, P. Gałecki, An immune gate of depression–early neuroimmune development in the formation of the underlying depressive disorder, Pharmacological reports 71 (2019) 1299–1307.

[3] S. Mahato, N. Goyal, D. Ram, S. Paul, Detection of depression and scaling of severity using six channel eeg data, Journal of medical systems 44 (2020) 1–12.

[4] S. Smys, J. S. Raj, Analysis of deep learning techniques for early detection of depression on social media network-a comparative study, Journal of trends in Computer Science and Smart technology (TCSST) 3 (2021) 24–39.

[5] K. Milintsevich, K. Sirts, G. Dias, Towards automatic text-based estimation of depression through symptom prediction, Brain Informatics 10 (2023) 4.

[6] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del-Arco, M. D. Molina-González, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Overview of MentalriskES at IberLEF 2024: Early Detection of Mental Disorders Risk in Spanish, Procesamiento del Lenguaje Natural 73 (2024).

[7] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.

[8] U. Zetsche, P.-C. Bürkner, J. Bohländer, B. Renneberg, S. Roepke, L. Schulze, Daily emotion

regulation in major depression and borderline personality disorder, Clinical Psychological Science 12 (2024) 161–170.
[9] I. Calixto, V. Yaneva, R. M. Cardoso, Natural language processing for mental disorders: an overview, Natural Language Processing In Healthcare (2022) 37–59.
[10] T. Zhang, A. M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental illness detection: a narrative review, NPJ digital medicine 5 (2022) 1–13.
[11] J. S. L. Figuerêdo, A. L. L. Maia, R. T. Calumby, Early depression detection in social media based on deep learning and underlying emotions, Online Social Networks and Media 31 (2022) 100225.
[12] S. Y. Boulahia, A. Amamra, M. R. Madi, S. Daikh, Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition, Machine Vision and Applications 32 (2021) 121.
[13] D. E. Losada, F. Crestani, J. Parapar, erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings 8, Springer, 2017, pp. 346–360.
[14] B. Fatima, M. Amina, R. Nachida, H. Hamza, A mixed deep learning based model to early detection of depression, Journal of Web Engineering 19 (2020) 429–455.
[15] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, Health information science and systems 6 (2018) 1–12.
[16] A. M. Mármol Romero, A. Moreno Muñoz, F. M. Plaza-del Arco, M. D. Molina González, M. T. Martín Valdivia, L. A. Ureña-López, A. Montejo Ráez, MentalRiskES: A new corpus for early detection of mental disorders in Spanish, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 11204–11214. URL: https://aclanthology.org/2024.lrec-main.978.
[17] E. Loper, S. Bird, Nltk: the natural language toolkit, CoRR cs.CL/0205028 (2002). doi:10.3115/1118108.1118117.
[18] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020, pp. 1–9.
[19] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018). arXiv:1810.04805.
[20] J. D. la Rosa y Eduardo G. Ponferrada y Manu Romero y Paulo Villegas y Pablo González de Prado Salas y María Grandury, Bertin: Efficient pre-training of a spanish language model using perplexity sampling, Procesamiento del Lenguaje Natural 68 (2022) 13–23. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403.
[21] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, Maria: Spanish language models, Procesamiento del Lenguaje Natural 68 (2022). URL: https://upcommons.upc.edu/handle/2117/367156#.YyMTB4X9A-0.mendeley. doi:10.26342/2022-68-3.
[22] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer,

V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach, arXiv preprint arXiv:1907.11692 (2019).

[23] M. Polignano, M. de Gemmis, G. Semeraro, et al., Comparing transformer-based ner approaches for analysing textual medical diagnoses., in: CLEF (Working Notes), 2021, pp. 818–833.