Transformer-based multilabel classification for identifying hidden psychological conditions in online posts

Iurii Krak[1,2], Olexander Mazurets[3], Oleksandr Ovcharuk[3], Maryna Molchanova[3], Olexander Barmak[3] and Larysa Azarova[4]

[1] *Taras Shevchenko National University of Kyiv, Ukraine*
[2] *Glushkov Institute of Cybernetics of NAS of Ukraine, Kyiv, Ukraine*
[3] *Khmelnytskyi National University, Khmelnytskyi, Ukraine*
[4] *Vinnytsia National Technical University, Vinnytsia, Ukraine*

Abstract

The paper proposes the method of multilabel classification for identifying hidden psychological conditions in online posts was proposed. The method consists of the stages of tokenization, neural network analysis of texts and the formation of conclusions about the presence of hidden psychological conditions. The features of the tokenization stage are the addition of special tokens to fix the boundaries of text fragments, supplement or trim the text to the length of a given dimension. At the stage of text analysis, the presence of each type of hidden psychological conditions is determined by a separate neural network model. The output of the method is the conclusion about the presence of each type of conditions with their numerical measures of manifestations. The created method allows to obtain in the models an improved ability to distinguish specific features for each type of psychological condition, due to training on modified sets of text data, which reduces the probability of confusion between conditions, since the model learns to distinguish their characteristic features. The developed method provides an average value 92.3% of the F_1 metric for multilabel classification of hidden psychological conditions, while existing analogues provide an average value 64.5% of the F_1 metric for multiclass classification.

Keywords

transformer neural network, multilabel classification, hidden psychological conditions

CEUR-WS.org/Vol-4004/paper8.pdf

1. Introduction

With the development of digital technologies and social networks, the amount of user-generated content is growing rapidly [1, 2]. Social networks, blogs, forums and other online platforms have become an important source of information about peoples psychological state [3, 4]. Analysis of this content opens up new opportunities for detecting hidden psychological conditions at early stages, which allows for timely intervention and provision of assistance [5]. Psychological health has become a serious problem after the COVID-19 pandemic, and many researchers have applied various ML and DL algorithms to social network data for prediction and analysis of mental health [6].

MoMLeT-2025: 7th International Workshop on Modern Machine Learning Technologies, June, 14, 2025, Lviv-Shatsk, Ukraine

* Corresponding author.
yuri.krak@gmail.com (I. Krak); exe.chong@gmail.com (O. Mazurets); off4aruk@gmail.com (O. Ovcharuk); m.o.molchanova@gmail.com (M. Molchanova); alexander.barmak@gmail.com (O. Barmak); azarova.larusa@gmail.com (L. Azarova)
000-0002-8043-0785 (I. Krak); 0000-0002-8900-0650 (O. Mazurets); 0009-0008-3815-0035 (O. Ovcharuk); 0000-0001-9810-936X (M. Molchanova); 0000-0003-0739-9678 (O. Barmak); 0000-0002-2631-8151 (L. Azarova)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Current methods for diagnosing hidden psychological conditions are mainly based on clinical interviews, standardized psychometric questionnaires, and observation by specialists [7]. Although these approaches are generally recognized and widely used in practical psychology, they have a number of limitations that reduce their effectiveness in the context of timely detection of hidden psychological conditions, especially in the digital environment [8]. First of all, they require direct participation of a person who is not always ready or able to seek help, which leads to a delay in diagnosis [9]. In addition, such methods may not take into account contextual changes in the behavior of the individual outside the clinical environment, which limits the depth of understanding of his real psychological state. The subjectivity of self-assessment, cultural characteristics of perception, and social pressure can also contribute to the distortion of the results [10]. In the digital age, when much of emotional expression is transferred to the virtual space, traditional diagnostic approaches are not always able to effectively process voluminous, unstructured data that reflect the real psychological difficulties of users [11, 12]. That is why there is a need for new methods that can integrate automated analysis of digital communications for more accurate and sensitive diagnosis of hidden psychological conditions [13].

The main goal of the paper is to create the multilabel classification method for identifying hidden psychological conditions in online posts, which differs from existing ones in the ability to identify several psychological conditions at once, and not just the dominant one, without losing accuracy compared to multiclass classifications.

The main contributions of the paper can be summarized as follows:

- the multilabel classification method for identifying hidden psychological conditions in online posts has been developed;
- the effectiveness of the developed multilabel classification method for identifying hidden psychological conditions has been experimentally proven, which allows, unlike existing ones, to distinguish specific features for each condition, which is implemented by training a set of neural network transformers on specifically formed datasets.

2. Related works

For many years, the scientific community has been investigating monitoring approaches to detect certain hidden psychological conditions and risky behaviors, such as depression, eating disorders, gambling, and suicidal ideation, among others, in order to activate prevention or mitigation strategies and, in severe cases, clinical treatments [14].

Social networks are increasingly used as a source of data for psychological research, in particular for the early detection of depressive symptoms [15]. In this context, machine learning methods play a key role, as they allow processing large amounts of data [16, 17], predicting the probability of hidden psychological conditions, and modeling the effectiveness of potential diagnostic approaches [18].

Work [19] aims to deepen the understanding of the linguistic manifestations of hidden psychological conditions and improve the explainability of deep learning models in this area. The problem of detecting psychological states is considered as an important public health task and the use of computational methods for detecting risky behavior in the online environment based on the analysis of data from social networks is proposed. Special attention is paid to the complexity of interpreting modern deep learning models based on neural networks in the context of automated diagnosis of hidden psychological conditions. The work proposes a multi-level interpretation of models that goes beyond classical techniques, and includes the analysis of hidden layer activations and errors related to emotional characteristics and thematic content of texts. The study was conducted on the basis of data from the social platforms Reddit and Twitter, the annotation of which covers four psychological conditions: depression, anorexia, post-traumatic stress disorder, and a tendency to self-harm.

The authors [20] emphasize the role of Natural Language Processing (NLP) and transformative models, such as BERT and GPT-4, in detecting emotional disorders through speech analysis. An approach is proposed that allows classifying human emotional states into six categories using a transformative architecture trained on a large English-language corpus. The model achieved accuracy of over 94% across all categories, demonstrating generalizability and stability of results on validation data. The paper emphasizes the potential of NLP models as tools for self-analysis and psychological health support, in particular through scalable support, language adaptation, and integration into decision-making processes.

The authors [21] investigated various linguistic indicators of psychological conditions based on the use of BERT architectures [22]. In the task of classifying 8 hidden psychological conditions, the authors achieved an accuracy of 0.645 according to the F_1 metric.

The article [23] considers the problem of detecting suicidal behavior through the prism of emotional analysis of suicide-related texts. The authors note that traditional studies mainly focus on identifying suicidal statements themselves, but leave out of consideration an in-depth analysis of the emotional state that precedes such tendencies. The proposed approach is based on identifying key negative emotions – such as anger, anxiety, guilt, fear, stress and sadness – in texts from social networks and suicide notes. For this purpose, a new method of assessing suicidal risk was introduced, which covers different levels of risk: from ideation to suicide. The authors proposed the CoDyn-BMHSA-CNN model, which combines bilateral LSTMs, multi-head attention mechanics and a convolutional network, which allows capturing the context and maintaining semantic flexibility in sequences of different lengths.

The results of the analysis of related works in the field of identifying mental disorders based on user content analysis revealed a rather low accuracy in multi-class classification, as well as the inability of existing models in research to identify several hidden psychological conditions simultaneously.

3. Problem statement

The existing multi-class classification when detecting hidden psychological conditions allows you to detect the most pronounced condition, while other types of psychological conditions are considered absent. This leads to a loss of time and missed opportunities to provide psychological assistance at the initial manifestations of psychological conditions [24]. Also, when detecting a more pronounced, but less important condition, less pronounced, but more severe psychological conditions may be missed. This situation contradicts the goals of the United Nations Development Program (UNDP), in particular, goal No. 3 «Good Health» [25].

When using multi-label classification, the problem of losing less pronounced hidden conditions disappears, but the problem of low classification accuracy arises, which is also present in multi-class. The problem of low accuracy is associated with the lack of correctly formed training data that would be labeled by specialists and would contain a sufficient number of records for training [26].

Another relevant problem in the task of detecting hidden psychological conditions is the presence of cross-features between different types of psychological conditions in text content, which complicates accurate classification. It is hypothesized that in order to increase the reliability of identifying hidden psychological conditions, it is necessary to carefully form training samples, where the target class will represent a separate type of psychological condition, and the control class will combine other types of hidden psychological conditions and texts without psychological condition features. In addition, it is assumed that the approach using a set of binary classifiers can provide higher efficiency compared to a single multi-class model. Therefore, it is necessary to develop a method that would take into account all the above-described aspects, as well as create software to study its efficiency.

4. Method design

The proposed method of multilabel classification for identifying hidden psychological conditions in online posts performs automatic classification of text user content for identifying hidden psychological conditions by their features.

The scheme of the method is shown in Figure 1. The input data of the method are: text for analysis and a set of transformer models *MTset* (1) and a set of corresponding tokenizers *MTTset* (2):

$$MTset = \{mt_1, mt_2, \ldots, mt_n\} \tag{1}$$

where mt_i is the i-th transformer model, $i=\overline{1,n}$, n – the number of transformer models, which corresponds to the number of hidden psychological conditions for identification.

$$MTTset = \{mtt_1, mtt_2, \ldots, mtt_n\} \tag{2}$$

where mtt_i – i-th tokenizer of the corresponding transformer model from *MTset*, $i=\overline{1,n}$, n – the number of tokenizers corresponding to the of hidden psychological conditions number for identification.

At stage 1, the text is tokenized for analysis by each tokenizer with *MTTset* for conversion into a vector representation. Tokenization also includes adding special tokens, such as [CLS] at the beginning and [SEP] at the end. Also, the text is supplemented or trimmed to the length of a given dimension [27]. In this study, the maximum length of the text is 128 tokens.

At stage 2, the text tokenized by each tokenizer with *MTTset* is analyzed by the corresponding neural network model with *MTset*. The result of this stage is an assessment from 0 to 1 of the strength of manifestation of each of the studied mental disorders.

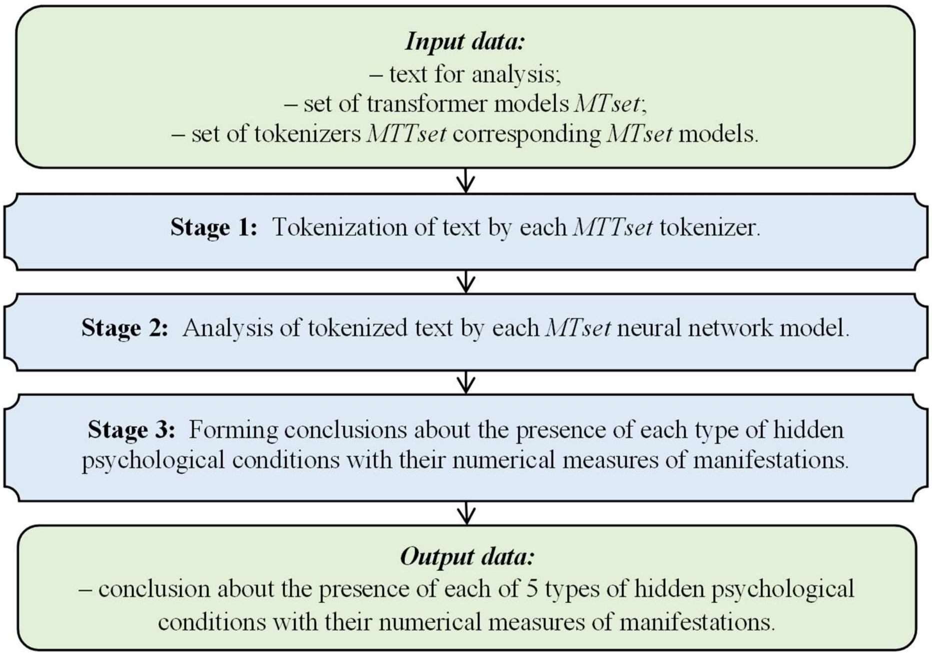

Figure 1: Stages of multilabel classification method for identifying hidden psychological conditions in online posts.

Stage 3 is responsible for forming conclusions about the presence of each of the 5 studied types of hidden psychological conditions with their numerical measures of manifestations. A psychological condition is considered to be present if the strength of its manifestation is higher than the threshold [28], which is established experimentally. In the study, the threshold is 0.5, but it can be fine-tuned for the person being studied for each type of psychological conditions individually.

The output of the method is the conclusion about the presence of each of the 5 types of psychological conditions with their numerical measures of manifestations.

The key aspect of the method is the formation of input data, namely - trained transformer models (*MTset*) and their tokenizers (*MTTset*). For training neural networks, training samples are formed in a specific way: target and control classes. The target class consists of exclusively text data with manifestations of the *i*-th psychological condition. To prevent confusion of psychological conditions with each other and taking into account that in 1 text there may be manifestations of other disorders, records in the control category are formed according to certain rules:

- the number of posts in the control category corresponds to or approaches the target (error no more than 10 posts);
- the control category consists of equal proportions of the remaining texts with other types of psychological conditions and texts that do not contain such manifestations, or contain them to a very small extent (up to 0.3 on a scale from 0 to 1).

An example of dataset forming with class «Narcissistic condition» is shown in Figure 2.

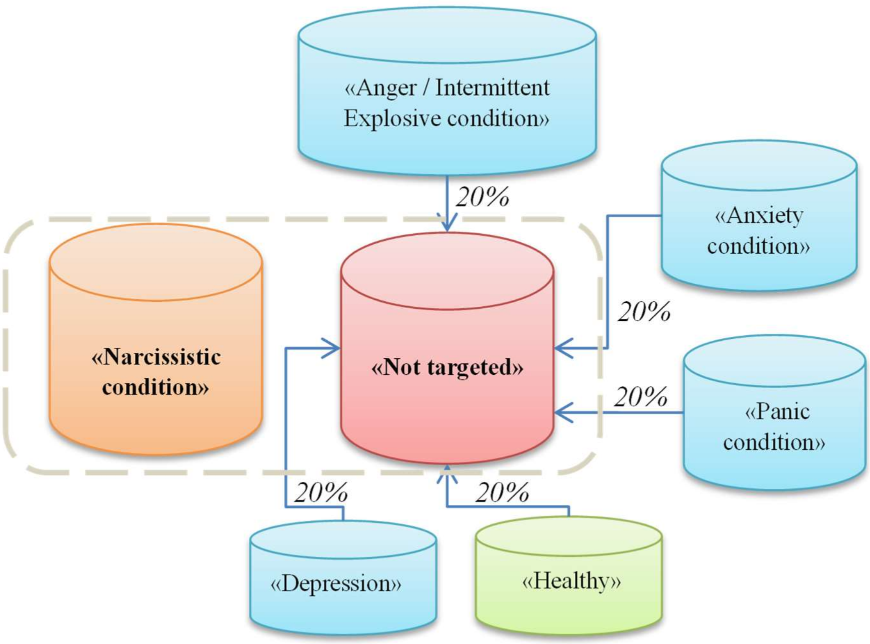

Figure 2: Dataset formation scheme using the example of «Narcissistic condition».

This distribution allows models to develop improved ability to distinguish between specific features for each condition. This reduces the likelihood of confusion between conditions as the model learns to distinguish between their characteristic features.

5. Experiment

5.1. Dataset for experiment

This study used two datasets to train psychological condition classification models. The first dataset contains cleaned texts from Reddit, marked according to the presence or absence of depressive states, «Depression: Reddit Dataset (Cleaned)» [29]. The total number of posts is 7650, of which 3900 do not contain signs of depression or other psychological conditions. It was from this resource that the records of users who do not show signs of psychological conditions were selected.

The second dataset, «COMSYS-T1» [30], includes posts from Twitter, which are characterized by the presence of linguistic markers inherent in different types of psychological conditions. The total number of tweets is 740, of which 208 belong to the «Depression» class, 158 to the «Narcissistic Condition» class, 154 to the «Anger/Intermittent Explosive Condition» class, 153 to the «Anxiety Condition» class, and 112 to the «Panic Condition» class.

Both resources contain marked-up data covering both texts without signs of pathologies and messages with potential hidden psychological conditions. Combining these datasets allows to form training samples with a clear division into target classes of disorders and a control group [31]. This helps to improve the ability of models to differentiate between different types of psychological conditions.

5.2. Experiment description

To investigate the effectiveness of the multilabel classification method for identifying hidden psychological conditions in online posts, the existing approach to multiclass classification was compared with the existing multilabel classification approach and with the proposed approach based on the use of a set of binary classifiers for each type of mental disorder.

The DistilBert neural network model was used and trained as a multiclass classifier for 5 classes: «Narcissistic condition», «Anxiety condition», «Panic condition», «Anger / Intermittent Explosive condition» and «Depression». The DistilBert model was trained for multiclass classification of the above disorders, and the *MTTset* and *MTset* sets were also trained for each of the 5 types of hidden psychological conditions.

For all experiments, software was created in the form of an IPython Notebook with CPU runtime environments (for training 5 *MTset* models) and TPU v2-8 for training multiclass and multilabel classifiers.

6. Results and discussion

Regarding the results obtained, with multi-class classification, the following results were obtained by macro-metrics: Accuracy: 0.854, Precision: 0.867, Recall: 0.854, F_1-score: 0.854. The confusion matrix is shown in Figure 3.

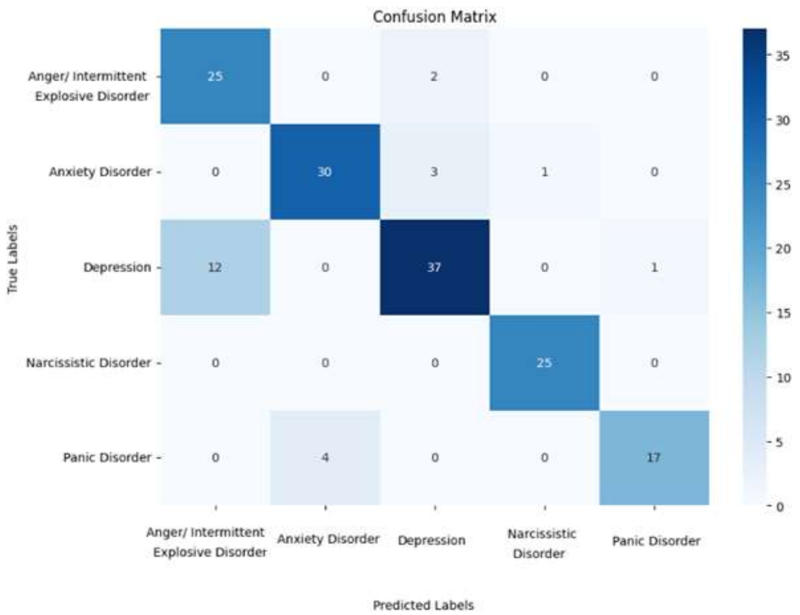

Figure 3: Confusion matrix for multiclassification.

As for the micrometrics for each class, which characterizes the presence of hidden psychological conditions, their values are given in Table 1. The number of samples for analysis is 157.

Table 1
Micro-metrics of multiclass classification

Classes/Metrics	Precision	Recall	F_1-score
«Anger/Intermittent Explosive condition»	0.68	0.93	0.78
«Anxiety condition»	0.88	0.88	0.88
«Depression»	0.88	0.74	0.80
«Narcissistic condition»	0.96	1.00	0.98
«Panic condition»	0.94	0.81	0.87

According to the table, the model demonstrates the highest performance for the classification of «Narcissistic condition» (Precision – 0.96, Recall – 1.00, F_1-score – 0.98), while the classes «Anger/Intermittent Explosive condition» and «Depression» have lower F_1-measures (0.78 and 0.80, respectively). From the confusion matrix (Figure 3) and the metric values, it can be seen that the classification of conditions is often accompanied by a significant number of errors between classes. This may be due to the fact that different conditions have common symptoms and manifestations, and several psychological conditions can be present in one text at the same time, which contradicts the multiclass classification approach. The experiment with multi-label classification also did not show high results, the following metric values were obtained: Precision: 0.878, Recall: 0.737, F_1-score: 0.801. The Precision value is higher compared to multiclassification, but Recall and F_1-score are lower. As in the previous experiment, there is significant confusion of psychological conditions.

The last experiment was to train the sets *MTTset* and *MTset* for each of the 5 types of hidden psychological conditions. The parameters of the neural network models were as follows: batch size = 16, epoch = 5, learning_rate = 2e-5. With this approach, the results shown in Table 2 were obtained.

Table 2
Metrics of binary models from *MTset*

Classes/Metrics	Accuracy	Precision	Recall	F_1-score
«Anger/Intermittent Explosive condition»	0.934	0.934	0.934	0.934
«Anxiety condition»	0.869	0.863	0.87	0.866
«Depression»	0.843	0.864	0.845	0.841
«Narcissistic condition»	0.984	0.986	0.982	0.984
«Panic condition»	1	1	1	1

The confusion matrices of binary classifiers with *MTset* on the validation data are shown in Figure 4 (a–e).

For the classes «Anger/Intermittent Explosive condition» and «Narcissistic condition», the model showed high metric values, which indicates the possibility of clear identification of these types of psychological conditions.

For the classes «Anxiety condition» and «Depression», slightly lower metric values are observed compared to other classes. Although the model showed satisfactory results, some mixing with other classes is possible, which may be due to the similarity of symptoms. The class «Panic condition» shows 100% for all metrics, which indicates that the model copes well with the identification of this condition, and this is confirmed by the confusion matrix shown in Figure 4(b).

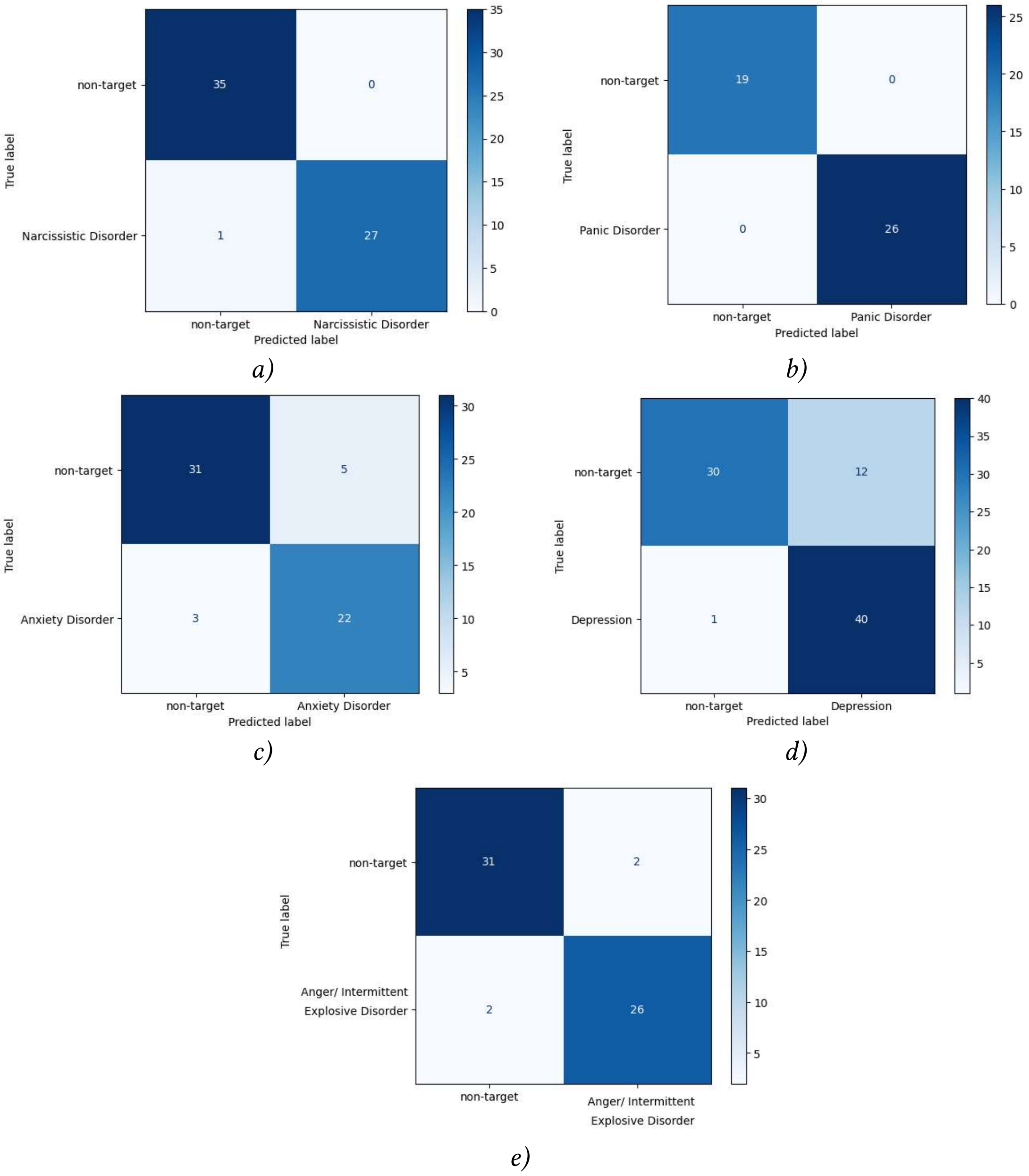

Figure 4: The confusion matrices of binary classifiers:
a) «Narcissistic condition» class; b) «Panic condition» class; c) «Anxiety condition» class;
d) «Depression» class; e) «Anger/Intermittent Explosive condition» class.

Regarding the errors obtained in «Anxiety condition» and «Depression» (Figures 4(c) and 4(d)), they are not critical, since non-target classes are incorrectly identified.

Figure 5 shows a comparison of the multiclass classification approach and the binary one for the Precision, Recall, and F_1-score metrics.

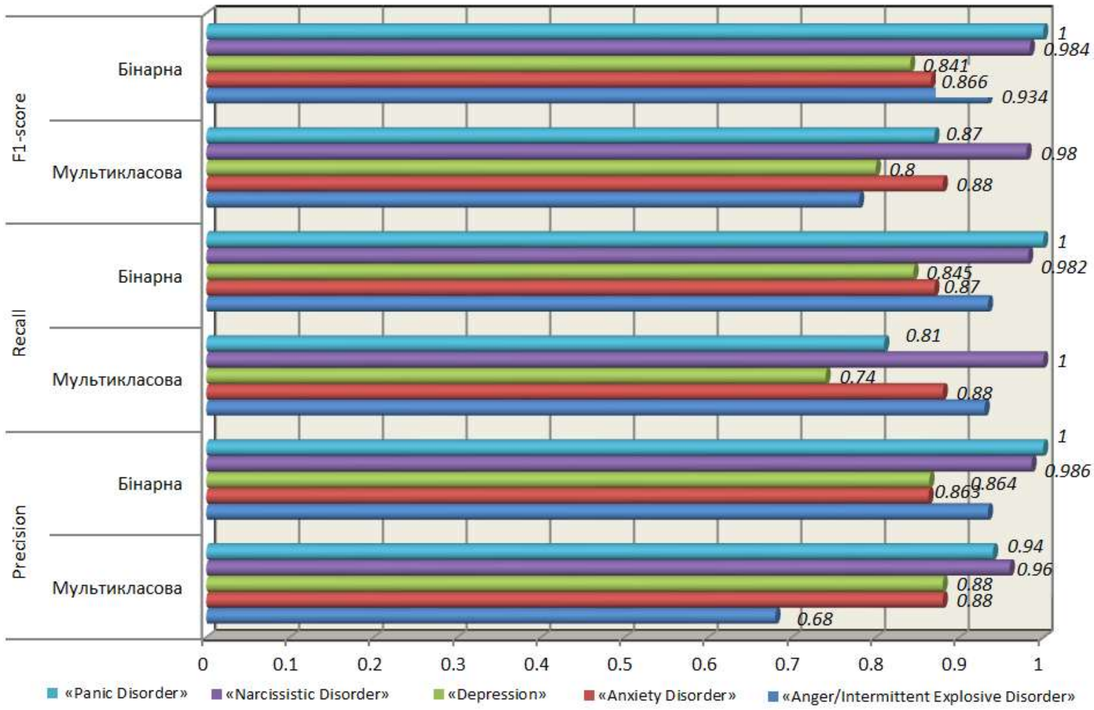

Figure 5: Comparison of binary and multiclass classification approaches.

From Figure 5 and Tables 1, 2 it is clear that for the most part the binary approach has an advantage over the multiclass one. By using the binary approach, it was possible to increase the identification of «Anger/Intermittent Explosive condition» by the Precision metric by 25.4%, by the Recall metric by 0.4%, and by the F_1-score metric by 15.4%. The identification of «Depression» by the Precision metric decreased by 1.6%, but by the Recall metric increased by 10.5%, and by the F_1-score metric increased by 4.1%. For the class «Narcissistic condition», Precision increased by 2.6%, F_1-score increased by 0.4%. However, Recall decreased by 1.8%. For the class «Panic Disorder», it was possible to achieve an increase in the values of the metrics as follows: Precision increased by 6%, Recall by 19%, and F_1-score increased by 13%. For the class «Anxiety condition» improvement was not achieved.

Regarding the comparison with similar studies, in [21] F_1 for multiclass classification was about 0.645, with the proposed BERT-trained model for multiclass classification 0.854 was obtained, and with binary classification the average metric value was 0.923.

Therefore, the proposed approach contributes to increasing the correct detection of hidden psychological conditions, and allows detecting within one user text not only the most pronounced disorder, but also other existing hidden psychological conditions, provided that the disorder is considered to be present.

7. Conclusions

The multilabel classification method for identifying hidden psychological conditions in online posts was proposed, which performs automatic classification of text user content for identifying hidden psychological conditions by their features. The method consists of the stages of tokenization, neural network analysis of texts and the formation of conclusions about the presence of hidden psychological conditions. The features of the tokenization stage are the addition of special tokens to fix the boundaries of text fragments, supplement or trim the text to the length of a given dimension. At the stage of text analysis, the presence of each type of hidden psychological conditions is determined by a separate neural network model. The output of the method is the conclusion about the presence of each type of psychological conditions with their numerical measures of manifestations. The created method allows to obtain in the models an improved ability to distinguish specific features for each type of psychological condition, due to training on modified sets of text data, which reduces the probability of confusion between conditions, since the model learns to distinguish their characteristic features. The developed method provides an average value 92.3% of the F_1 metric for multilabel classification of hidden psychological conditions. While existing analogues provide an average value 64.5% of the F_1 metric for multiclass classification.

The minimum F_1 metric value for the developed method is 84.1% (for the "Depression" hidden psychological condition). It has been found that there are certain difficulties with the classification of some conditions, in particular «Anxiety condition» and «Depression», which may be related to their clinical similarity. This requires further research to improve the model and, possibly, to involve additional data or other text characteristics to increase the accuracy of classification. Further research will be aimed at expanding the training data sets and studying other types of psychological conditions, in addition to the 5 considered, and at improving the metrics for the considered hidden psychological conditions.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] J. J. Van Bavel, C. E. Robertson, K. del Rosario, J. Rasmussen, S. Rathje, Social Media and Morality, Annual Review of Psychology 75.1 (2023). doi:10.1146/annurev-psych-022123-110258.

[2] G. Pietrabissa, M. Semonella, G. Marchesi, S. Mannarini, G. Castelnuovo, G. Andersson, A. A. Rossi, Validation of the Italian Version of the Web Screening Questionnaire for Common Mental Disorders, Journal of Clinical Medicine 13.4 (2024) 1170. doi:10.3390/jcm13041170.

[3] J. Ham, S. Li, J. Looi, M. S. Eastin, Virtual humans as social actors: Investigating user perceptions of virtual humans emotional expression on social media, Computers in Human Behavior (2024) 108161. doi:10.1016/j.chb.2024.108161.

[4] A. P. Wibowo, M. Taruk, E. Tarigan, M. Habibi, Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders, Indonesian Journal of Data and Science 5.1 (2024) 8–14. doi:10.56705/ijodas.v5i1.122.

[5] J. Philip Thekkekara, S. Yongchareon, V. Liesaputra, An attention-based CNN-BiLSTM model for depression detection on social media text, Expert Systems with Applications 249 (2024) 123834. doi:10.1016/j.eswa.2024.123834.

[6] M. Saleem, H. Afzal, A Review of Mental Health Analysis Through Social Media Using Machine Learning and Deep Learning Approaches, 2024 International Conference on Engineering & Computing Technologies (ICECT), IEEE, 2024. doi:10.1109/icect61618.2024.10581373.

[7] Y. Bao, J. Liu, C. DiStefano, R. Ding, Identifying Children at Risk for Emotional and Behavioral Problems: A Diagnostic Classification Model Approach, Psychology in the Schools (2025). doi:10.1002/pits.23394.

[8] D. Zhang, L. Zhou, J. Tao, T. Zhue, G. Gao, A Knowledge-Enhanced Transformer-Based Approach to Suicidal Ideation Detection from Social Media Content, Information Systems Research (2024). doi:10.1287/isre.2021.0619.

[9] K. Dellarmo, M. J. Tasse, Diagnostic Overshadowing of Psychological Disorders in People With Intellectual Disability: A Systematic Review, American Journal on Intellectual and Developmental Disabilities 129.2 (2024) 116–134. doi:10.1352/1944-7558-129.2.116.

[10] G. Young, L. Erdodi, L. Giromini, R. Rogers, Detection Systems Related to Malingering and Invalid Response Set in Psychological Injury Assessments, Psychological Injury and Law (2024). doi:10.1007/s12207-024-09526-3.

[11] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, R. Bahrii, O. Sobko, O. Barmak, Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network, CEUR Workshop Proceedings, 3688 (2024) 16–28. URL: https://ceur-ws.org/Vol-3688/paper2.pdf.
[12] A. Malhotra, R. Jindal, XAI transformer based approach for interpreting depressed and suicidal user behavior on online social networks, Cognitive Systems Research (2023) 101186. doi:10.1016/j.cogsys.2023.101186.
[13] P. Cuijpers, C. Miguel, M. Ciharova, M. Harrer, D. Basic, I. A. Cristea, N. de Ponti, E. Driessen, J. Hamblen, S. E. Larsen, Absolute and relative outcomes of psychotherapies for eight mental disorders: a systematic review and meta–analysis, World Psychiatry 23 2 (2024) 267–275. doi:10.1002/wps.21203.
[14] A. Montejo-Ráez, M. D. Molina-González, S. M. Jiménez-Zafra, M. Á. García-Cumbreras, L. J. García-López, A survey on detecting mental disorders with natural language processing: Literature review, trends and challenges, Computer Science Review 53 (2024) 100654. doi:10.1016/j.cosrev.2024.100654.
[15] A. Aldkheel, L. Zhou, Depression Detection on Social Media: A Classification Framework and Research Challenges and Opportunities, Journal of Healthcare Informatics Research (2023). doi:10.1007/s41666-023-00152-3.
[16] I. Krak, V. Didur, M. Molchanova, O. Mazurets, O. Sobko, O. Zalutska, O. Barmak, Method for political propaganda detection in internet content using recurrent neural network models ensemble, CEUR Workshop Proceedings, 3806 (2024) 312–324. URL: https://ceur-ws.org/Vol-3806/S_36_Krak.pdf.
[17] J. Aina, O. Akinniyi, M. M. Rahman, V. Odero-Marah, F. Khalifa, A Hybrid Learning-Architecture for Mental Disorder Detection using Emotion Recognition, IEEE Access (2024) 1. doi:10.1109/access.2024.3421376.
[18] A. Khan, R. Ali, Unraveling minds in the digital era: a review on mapping mental health disorders through machine learning techniques using online social media, Social Network Analysis and Mining 14.1 (2024). doi:10.1007/s13278-024-01205-0.
[19] A. S. Uban, B. Chulvi, P. Rosso, On the Explainability of Automatic Predictions of Mental Disorders from Social Media Data, Natural Language Processing and Information Systems, Springer International Publishing, Cham, 2021, pp. 301–314. doi:10.1007/978-3-030-80599-9_27.
[20] A. R. Mishra, A. Rai, D. Nandan, U. Kshirsagar, M. K. Singh, Unveiling Emotions: NLP-Based Mood Classification and Well-Being Tracking for Enhanced Mental Health Awareness, Mathematical Modelling of Engineering Problems 12.2 (2025). doi:10.18280/mmep.120228.
[21] Z. Jiang, S. I. Levitan, J. Zomick, J. Hirschberg, Detection of Mental Health from Reddit via Deep Contextualized Representations, Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis, Association for Computational Linguistics, Stroudsburg, PA, USA, 2020. doi:10.18653/v1/2020.louhi-1.16.
[22] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, E. Manziuk, O. Barmak, Method for neural network detecting propaganda techniques by markers with visual analytic, CEUR Workshop Proceedings, 3790 (2024) 158-170. URL: https://ceur-ws.org/Vol-3790/paper14.pdf.
[23] D. Kodati, R. Tene, Emotion mining for early suicidal threat detection on both social media and suicide notes using context dynamic masking-based transformer with deep learning, Multimedia Tools and Applications (2024). doi:10.1007/s11042-024-19411-5.
[24] E. O. Ogunseye, C. A. Adenusi, A. C. Nwanakwaugwu, S. A. Ajagbe, S. O. Akinola, Predictive Analysis of Mental Health Conditions Using AdaBoost Algorithm, ParadigmPlus 3.2 (2022) 11–26. doi:10.55969/paradigmplus.v3n2a2.
[25] T. K. Oswald, M. T. Nguyen, L. Mirza, C. Lund, H. G. Jones, G. Crowley, D. Aslanyan, K. Dean, P. Schofield, M. Hotopf, Interventions targeting social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews, Psychological Medicine (2024) 1–25. doi:10.1017/s0033291724000333.
[26] M. Razavi, S. Ziyadidegan, R. Jahromi, S. Kazeminasab, E. Baharlouei, V. Janfaza, A. Mahmoudzadeh, F. Sasangohar, Machine Learning, Deep Learning and Data Preprocessing

Techniques for Detection, Prediction, and Monitoring of Stress and Stress-related Mental Disorders: A Scoping Review (Preprint), JMIR Mental Health (2023). doi:10.2196/53714.
[27] Y. V. Krak, O. V. Barmak, O. V. Mazurets, The practice investigation of the information technology efficiency for automated definition of terms in the semantic content of educational materials, Problems in Programming, 2–3 (2016) 237–245. doi:10.15407/pp2016.02-03.237.
[28] I. Krak, O. Sobko, M. Molchanova, I. Tymofiiev, O. Mazurets, O. Barmak, Method for neural network cyberbullying detection in text content with visual analytic, CEUR Workshop Proceedings, 3917 (2025) 298-309. URL: https://ceur-ws.org/Vol-3917/paper57.pdf.
[29] Infamouscoder, Depression Reddit Cleaned, 2021. URL: https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned.
[30] Kajimi, COMSYS-T1, 2023. URL: https://www.kaggle.com/datasets/kajimi/comsys2023.
[31] O. Sobko, O. Mazurets, M. Molchanova, I. Krak, O. Barmak, Method for analysis and formation of representative text datasets, CEUR Workshop Proceedings, 3899 (2024) 84-98. URL: https://ceur-ws.org/Vol-3899/paper9.pdf.