Context-enriched approach to students depression monitoring in education using BERT-GPT hybrid model

Olexander Mazurets[1], Roman Vit[1], Maryna Molchanova[1], Illia Tymofiiev[1] and Olena Sobko[1]

[1]Khmelnytskyi National University, 11, Instytuts'ka str., Khmelnytskyi, 29016, Ukraine

Abstract

The paper proposes a new context-enriched approach to students' depression monitoring in education using a BERT-GPT hybrid model. The neural network, which is the basis of the approach, is trained taking into account the loss function, which is optimised simultaneously for both submodels, and the pre-trained weights of BERT and GPT-2 are updated during fine-tuning, taking into account the specifics of psycho-emotional text patterns. Thus, the model not only detects the presence of a depressive state, but is also enriched with educational and contextual features of the language. The developed approach increases the accuracy of detecting depressive states by at least 0.0525 compared to the closest analogues, and on average by 0.14, which indicates the superiority of the combined architecture in the binary classification task. The achieved accuracy at the level of 0.99 indicates the effectiveness of combining contextual features of two transformer-type models, which provides deeper modelling of semantic and syntactic information in texts. The proposed approach promotes early identification of psycho-emotional risks among students, which allows for the timely implementation of preventive measures and psychosocial support. For the education sector, this means increasing the level of academic success, reducing the number of cases of emotional burnout and improving the overall mental well-being of the student community, which in the long term contributes to the formation of a more sustainable, healthy and productive educational environment.

Keywords

neural networks, GPT2, BERT, student depression, education

1. Introduction

The mental health of students is gaining increasing attention in the context of ensuring quality education [1] and supporting academic success [2]. Depressive states and emotional burnout associated with academic workload and social challenges significantly affect cognitive functions [3], motivation and general well-being of students [4]. Traditional methods of diagnosing depression, based on surveys and clinical interviews [5], often turn out to be insufficiently operational or unsuitable for mass screening [6]. In this regard, there is a need to develop automated methods for detecting and classifying signs of depression based on the analysis of text data that students leave in social networks [7], forums or specialised platforms [8]. Modern approaches to natural language processing, particularly neural architectures based on transformers, demonstrate effectiveness in classifying mood, emotions [9], and mental states [10]. However, to increase the accuracy of detecting depressive manifestations in an educational context, it is important to consider both the text's lexical features and contextual information that forms a deeper understanding of the user's psychological state. Combining models that integrate context-enriched analysis with generative capabilities opens up new horizons in automatic recognition of depression.

The paper aims to develop and test a method for context-enriched student depression monitoring in education using a BERT-GPT hybrid model. The paper's main contributions are the combination of a two-stream neural network architecture, which simultaneously uses the BERT model for deep syntactic and contextual understanding of the text and the GPT-2 model, which provides semantic enrichment through the transformer's generative properties.

DigiTransfEd 2025: 4th Workshop on Digital Transformation of Education, (September 24, 2025, Kyiv, Ukraine)

exe.chong@gmail.com (O. Mazurets); vit.roman.vit@gmail.com (R. Vit); m.o.molchanova@gmail.com (M. Molchanova); ilia.tumofiev@gmail.com (I. Tymofiiev); olenasobko.ua@gmail.com (O. Sobko)

0000-0002-8900-0650 (O. Mazurets); 0009-0009-6958-4730 (R. Vit); 0000-0001-9810-936X (M. Molchanova); 0009-0006-4610-5889 (I. Tymofiiev); 0000-0001-5371-5788 (O. Sobko)

2. Related works

Recently, machine learning and natural language processing methods have been widely used to detect signs of depression.

Scientists [11] study the problem of depression among students as a factor that negatively affects their academic performance, social functioning, and future professional prospects. The authors apply machine learning methods to predict the likelihood of depressive states, considering personal, academic, and behavioural variables. A comparison of several models showed that logistic regression provides the best accuracy (85%) with balanced metrics of precision, completeness, and F_1-measure. The SMOTE method was used to solve the class imbalance problem, which increased the model's sensitivity to underrepresented cases. The results demonstrate the potential of using predictive analytics to develop targeted psychosocial interventions within educational institutions, and also confirm the effectiveness of machine learning in identifying psycho-emotional risks in the student environment.

In a study analysing data from Reddit, the platform is seen as an important complement to the traditional healthcare system due to the speed of exchange of ideas, the versatility of emotional expression, and the use of medical terminology. Analysis of comments and posts with suicidal intent using NLP confirmed that subreddits are reliable online resources for obtaining help and providing reliable data on the mental state of users. The use of machine learning algorithms such as Naive Bayes, SVM, logistic regression, and random forest showed the effectiveness of identifying individuals at risk, where logistic regression achieved 77.29% accuracy and an F_1 score of 0.77 [12].

Another study focused on sentiment analysis based on Twitter data using classical approaches such as TF-IDF, Bag of Words, and Multinomial Naive Bayes. These methods, applied to real-time tweets, demonstrated accuracy that allows them to be considered as additional tools for diagnosing depression, with the possibility of adapting to different languages [13].

The authors [14] investigated the problem of detecting depression among students based on data from the Chinese social network Weibo. The authors proposed a hybrid approach that combines text features obtained through pre-trained BERT with manually calculated user meta-features. Based on these two groups of features, several "multimodal fusion" strategies were implemented, among which the best results were demonstrated by the late fusion method at the logistic regression level. The maximum accuracy of the model reached 93.75%, which indicates the effectiveness of involving the structured user context as a complement to text analysis.

In [15], which focuses on improving the diagnosis of depression using machine learning and NLP, the difficulties of detecting depressive states in the presence of comorbid disorders, such as post-traumatic stress disorder, are emphasised. Using data cleaning procedures, feature selection and classifier comparison based on the DAIC-WOZ set, it was shown that the Random Forest and XGBoost models achieve about 84% accuracy, significantly exceeding the SVM performance (72%).

The study [16] aimed to identify the risk of anxiety and depressive disorders among US college students by building predictive machine learning models. The authors use a large array of empirical data for modelling based on the XGBoost, Random Forest, Decision Tree and logistic regression algorithms. According to the results, the highest classification quality was achieved with AUC = 0.77, indicating the models' stable discriminatory power. The proposed approach has the potential to be implemented in educational counselling as a tool for early detection of mental disorders, allowing for preventive measures before the development of clinical symptoms.

Another study is devoted to the early detection of depression among cancer patients through the analysis of messages from a secure portal. Classifiers built based on logistic regression, support vector machines and BERT transformer models, trained on a large corpus of messages, showed high efficiency. The BERT and RedditBERT models outperformed other algorithms in terms of AUROC (88% and 86%, respectively). In addition, the obtained predictions correlated with diagnoses and treatment of depression, which confirms the practical value of such models in clinical applications [17].

Despite significant progress in applying machine learning and natural language processing methods to detect depressive states, several scientific and applied problems remain open. Most existing approaches often ignore the importance of complex contextual information, which is particularly important in

the student environment due to the specifics of the educational process, linguistic and behavioural features of communication, and the dynamics of the psycho-emotional state during the academic year. Considering these factors, particularly through a combination of semantic and sequential aspects of the text, allows modelling deeper behaviour and mood patterns. Based on the combined neural architecture BERT-GPT, the proposed approach will provide such a context-enriched representation of text that will increase the accuracy and reliability of the binary classification of depressive states in the student environment.

3. Method design

The approach implements cognitive-semantic integration, which is effective in emotional monitoring tasks in the educational environment, allowing for the identification of psychoemotional risks of depression at early stages with high accuracy. The scheme and stages of the approach are shown in figure 1.

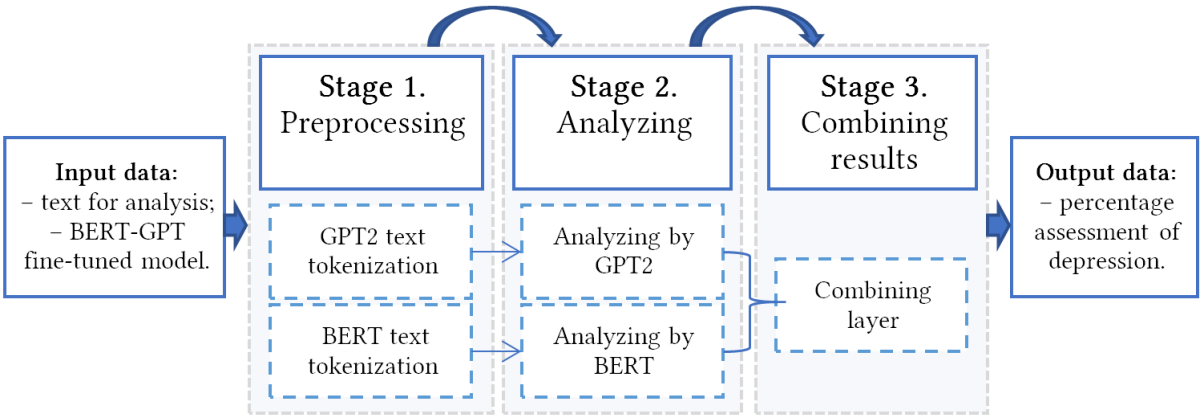

Figure 1: Stages of context-enriched approach to students depression monitoring in education.

The approach to context-based detection of depressive states in students in an educational environment consists of transforming input text data using a trained two-stream neural network model BERT-GPT [18], which provides integrated syntactic-semantic processing to obtain a numerical score reflecting the probability of depressive manifestations in a student.

At the first stage, responsible for pre-processing, the input text of student self-expression undergoes parallel tokenisation using the BERT and GPT-2 models. This provides a two-vector representation of the input data: taking into account the syntactic structure (via BERT) and the semantic context (via GPT-2) [19], which is critically important for capturing non-obvious manifestations of depressive states.

At the second stage, a separate data pass is performed through each model. The BERT model extracts syntactic features, especially focusing on grammatical patterns and formal language constructions [20], which may indicate a cognitive change. While GPT-2 analyses the text's deep semantic space, [21] allows it to consider hidden emotional content, allusions, and the style of speech characteristic of affective disorders.

At the third stage, the vectors of the output features from the two streams are combined into a single representative vector through the concatenation layer. This combined vector is passed to the dense layer, which performs logistic regression using sigmoid activation. The obtained probability value is interpreted as an assessment of the degree of depression of the student in the educational process, expressed in percentage format. Accordingly, the output data is a percentage assessment of depression.

Since the key part of the methodology is the trained neural network model, it is advisable to provide an algorithm 1 for its training.

The model was trained on a curated dataset of student-authored texts, where each sample was

Algorithm 1 Fine-tuning BERT-GPT2 Architecture for Depression Prediction

Let:

- $D = \{(x_i, y_i)\}_{\{i=1\}}^{n}$ be the dataset of student texts x_i and binary labels $y_i \in \{0, 1\}$;
- $f_B(x)$ — output of BERT encoder (syntactic features);
- $f_G(x)$ — output of GPT-2 encoder (semantic context);
- v_i — combined feature representation for sample x_i;
- $W \in \mathbb{R}^d$ — weight vector of the linear output layer;
- $b \in \mathbb{R}$ — bias term;
- $\hat{y}_i$ — predicted probability that x_i indicates a depressive state;
- $\sigma(\cdot)$ — sigmoid function;
- $\mathcal{L}$ — binary cross-entropy loss;
- θ_B — trainable parameters of the BERT encoder;
- θ_G — trainable parameters of the GPT-2 encoder.

Training procedure:

1. Initialize pretrained BERT and GPT-2 models with frozen vocabulary.
2. For each training sample x_i:
   a) Tokenize x_i using both BERT and GPT-2 tokenizers.
   b) Encode x_i via both encoders:
   $$h_B \leftarrow f_B(x_i), \quad h_G \leftarrow f_G(x_i)$$
   c) Extract [CLS] representation from BERT and first-token from GPT-2:
   $$v_i \leftarrow \text{concat}(h_B^{[CLS]}, h_G^{[0]})$$
   d) Compute prediction:
   $$\hat{y}_i \leftarrow \sigma(W^\top v_i + b)$$
   e) Compute loss:
   $$\ell_i \leftarrow \mathcal{L}(\hat{y}_i, y_i)$$
3. Backpropagate gradients and update parameters θ_B, θ_G, W, b using Adam optimizer.
4. Iterate over mini-batches until convergence criteria or maximum epochs reached.

Validation: Evaluate trained model on held-out validation set using: accuracy, precision, recall, F_1-score, confusion matrix.

labelled with a binary indicator of depressive tendencies. The dataset was partitioned into training and validation subsets in a 90/10 ratio, ensuring class distributional balance. During training, the model optimised the parameters of the BERT and GPT-2 encoders and the final linear classification layer using the Adam optimiser [22]. The objective was to minimise the binary cross-entropy loss between the predicted probabilities and ground-truth labels. Fine-tuning was performed iteratively in mini-batches, with convergence monitored via validation performance metrics such as accuracy, precision, recall, and F_1-score [23].

Thus, the methodology and algorithm for obtaining a trained context-based two-stream architecture for neural network analysis, which combines syntactic and semantic text processing for monitoring depressive manifestations in students, are presented.

4. Experiment

4.1. Dataset

This paper utilises the "Student-Depression-Text dataset" [24], an annotated corpus comprising 7.489 English-language textual samples from social media platforms, including user comments from Facebook

and related forums. The dataset is structured in Excel format and includes the following attributes: raw text content, binary class label (0 = non-depressed, 1 = depressed/anxious), participant age, age group, and gender. The dataset was balanced to avoid bias [25]. All contributors to the dataset are students aged between 15 and 17, fluent in English, and representative of an adolescent demographic particularly vulnerable to affective disorders.

Each text entry is manually labelled as either indicative of a normal psychological state or suggestive of depressive or anxiety-related expression. These annotations were designed to reflect linguistic cues associated with underlying emotional distress. Including demographic variables enables the exploration of correlations between mental health indicators and age or gender, thereby offering a richer context for modelling.

The dataset is the foundation for training a dual-stream neural architecture to identify signs of depression in educational contexts based on student-generated textual data.

4.2. Experimental software

The experimental software prototype for automated detection of depressive tendencies in educational contexts was implemented using Python [26] and specialised natural language processing and machine learning libraries. The system integrates pretrained transformer-based models to extract syntactic and semantic features from student-generated textual data.

The core functionality was deployed as a web-based application using the Flask microframework, enabling an accessible RESTful interface for model inference and interaction. Model development and fine-tuning were conducted using PyTorch as the primary deep learning framework [27], with transformer models sourced via the Hugging Face Transformers library [28]. Feature extraction leveraged the “BertTokenizer”, “BertModel”, “GPT2Model” components for obtaining contextualized embeddings [29].

Data preprocessing and management were facilitated through Pandas [30] and NumPy [31], which supported structured manipulation of input samples and efficient tensor operations during training and evaluation. The final application allows users to submit input text and receive probabilistic feedback on the presence of depressive markers based on learned linguistic patterns. The interface of the developed software prototype is shown in figure 2.

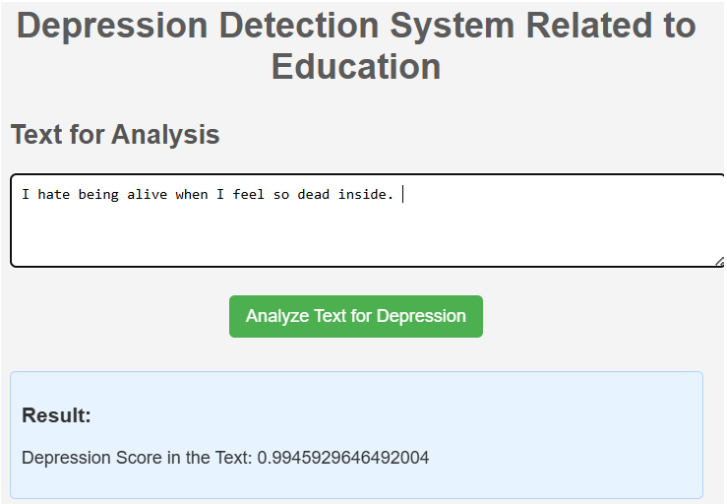

Figure 2: Interface of the developed software prototype.

This modular software system serves as a functional prototype and an experimental platform for evaluating the effectiveness of transformer-based feature fusion in mental health detection tasks.

5. Result and discussion

Initially, an experiment was conducted to investigate the influence of training hyperparameters on the performance of the developed neural model. Table 1 presents the comparative results of four configurations based on a combined BERT-GPT2 architecture, each trained with different hyperparameter settings. The evaluation employed standard classification metrics, including accuracy, precision, recall, and F_1-score. This analysis aimed to identify optimal combinations of learning rate, batch size, number of epochs, and input sequence length for efficient model convergence and robust generalisation.

Table 1
Performance of BERT-GPT2-based models under different training configurations.

Model ID	Epochs	Learning Rate	Batch Size	Seq. Length	Accuracy	Precision	Recall	F_1-score
“Gpt_Bert1”	1	7.00E-05	4	512	0.95	0.94	0.96	0.95
“Gpt_Bert2”	3	3.00E-04	8	256	0.96	0.95	0.96	0.96
“Gpt_Bert3”	2	2.00E-05	16	128	0.99	0.98	0.98	0.98
“Gpt_Bert4”	2	1.00E-03	32	64	0.98	0.97	0.98	0.98

“Gpt_Bert3” model demonstrated the highest performance across all metrics, suggesting that a lower learning rate and moderate batch size may contribute positively to generalisation. The results underline the sensitivity of transformer-based models to sequence length and learning rate, which should be carefully tuned in downstream mental health prediction tasks.

However, the rest of the results are also quite high, indicating all trained models’ ability to identify depressive states associated with learning in educational institutions correctly.

The best model was also tested on a labeled sample of over 7000 text samples, the result is shown in figure 3 in the form of confusion matrix.

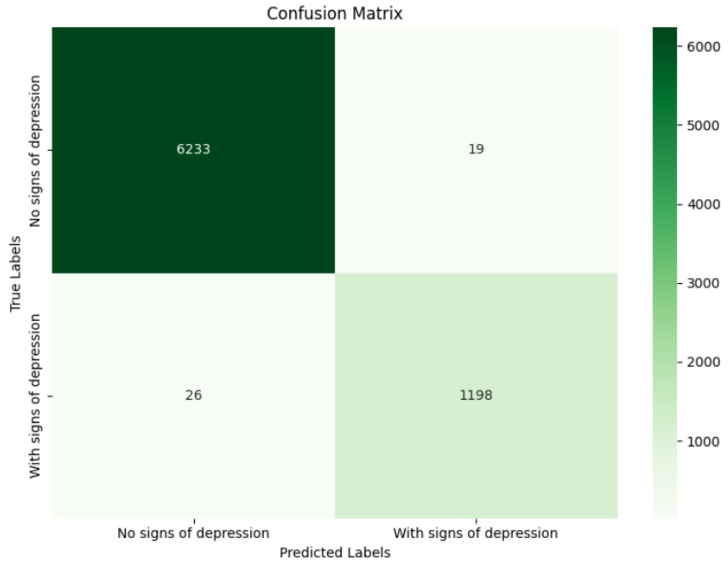

Figure 3: Confusion matrix.

As the confusion matrix shows, the neural network has a minimum rate of false positives.

The developed approach was also compared with known analogues, the results are given in table 2. The proposed model showed the highest accuracy values (0.99) and F_1-measure (0.98) among the compared approaches. In the study [17], the BERT and RedditBERT models achieved accuracies of 0.88 and 0.86, respectively, which are 0.11 and 0.13 lower. Logistic regression in the student environment [11] showed an accuracy of 0.86, which is 0.13 lower. In the Reddit study with the analysis of suicidal intentions [12], these indicators were 0.77, which is 0.22 lower. In the case of the multimodal approach

with Weibo [14], the model showed 0.9375 accuracy, which is 0.0525 lower. The classical algorithms Random Forest and XGBoost [15] achieved an accuracy of 0.84, which is 0.15 lower. The results indicate the superiority of the combined BERT-GPT2 architecture in binary classification of depressive states.

Table 2
Comparison of developed and existing approaches.

#	Model / Architecture	Accuracy	F_1-score
1	Proposed BERT-GPT2	0.99	0.98
2	BERT [17]	0.88	–
3	RedditBERT [17]	0.86	–
4	Logistic Regression (student data) [11]	0.85	0.85
5	Logistic Regression (Reddit, suicide) [12]	0.77	0.77
6	BERT + user features (Weibo, multimodal) [14]	0.9375	–
7	Random Forest / XGBoost (DAIC-WOZ) [15]	0.84	–

However, the given approach has limitations. The main limitation is the language – currently, the neural network works only with English-language texts. Another limitation concerns the length of the input sequence. Within the framework of the study, the maximum number of input tokens is 128.

Further research will aim to continue training the neural network of the dual architecture with different parameters, such as the number of epochs, batch size, and learning rate, to reduce error and work with languages other than English.

6. Conclusion

The paper proposes a new context-enriched approach to students' depression monitoring in education using a BERT-GPT hybrid model. The neural network, which is the basis of the approach, is trained taking into account the loss function, which is optimised simultaneously for both submodels, and the pre-trained weights of BERT and GPT-2 are updated during fine-tuning, taking into account the specifics of psycho-emotional text patterns. Thus, the model not only detects the presence of a depressive state, but is also enriched with educational and contextual features of the language.

The developed approach increases the accuracy of detecting depressive states by at least 0.0525 compared to the closest analogues and on average by 0.14, which indicates the superiority of the combined architecture in the binary classification task. The achieved accuracy at the level of 0.99 indicates the effectiveness of combining contextual features of two transformer-type models, which provides deeper modelling of semantic and syntactic information in texts.

The proposed approach promotes early identification of psycho-emotional risks among students, which allows for the timely implementation of preventive measures and psychosocial support. For the education sector, this means an increase in students' academic performance. It should be noted that the constructed model for monitoring students' depression has not been tested in the real educational process to assess its effectiveness using mathematical statistics methods.

In addition, the conclusion that implementing the proposed system will improve academic performance is too optimistic since diagnosis does not yet mean that students will receive adequate assistance to overcome existing problems within the study period. Additional experimental studies are needed to confirm this conclusion.

Author contributions

Conceptualization, Olexander Mazurets; methodology, Roman Vit and Maryna Molchanova; formulation of tasks analysis, Ilia Tymofiiev and Olena Sobko; software, Roman Vit and Maryna Molchanova; writing–original draft, Roman Vit and Maryna Molchanova; analysis of results, Olexander Mazurets

and Olena Sobko; visualization, llia Tymofiiev; reviewing and editing, Olexander Mazurets. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability statement

No new data were created or analysed during this study. Data sharing is not applicable.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors express their gratitude to Tetiana Vakaliuk, Viacheslav Osadchyi and Olha Pinchuk to scientific and organizational support of 4th Workshop on Digital Transformation of Education 2025.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Y. Wang, J. Zhang, L. Huang, X. Li, Y. Zhong, Association between healthy lifestyle choices and mental health among students: a cross sectional study, BMC Public Health 25 (2025) 247. doi:10.1186/s12889-025-21482-4.

[2] A. Madrid-Cagigal, C. Kealy, C. Potts, M. D. Mulvenna, M. Byrne, M. M. Barry, G. Donohoe, Digital Mental Health Interventions for University Students With Mental Health Difficulties: A Systematic Review and Meta-Analysis, Early intervention in psychiatry 19 (2025) e70017. doi:10.1111/eip.70017.

[3] S. Hussain, Y. Ali, S. U. Khan, N. Khan, A. Ullah, S. Abas, F. ur Rehman, S. Khan, Mental health issues: Stress, anxiety, and depression in diploma and degree health care students, Journal of Medical & Health Sciences Review 2 (2025) 3453–3471. doi:10.62019/5vm15e38.

[4] J. Wang, Z. Wang, Y. Yang, T. Wang, H. Lin, W. Zhang, X. Chen, C. Fu, Academic burden and emotional problems among adolescents: A longitudinal mediation analysis, Journal of Adolescence 97 (2025) 989–1001. doi:10.1002/jad.12471.

[5] Y.-R. Lee, J.-S. Lee, Real-time monitoring to predict depressive symptoms: study protocol, Frontiers in Psychiatry 15 (2024) 1465933. doi:10.3389/fpsyt.2024.1465933.

[6] T. Vu, M. Yamamoto, J. T. Tay, N. Watanabe, Y. Kuriya, A. Oya, P. N. H. Tran, M. Araki, et al., Prediction of depressive disorder using machine learning approaches: findings from the NHANES, BMC Medical Informatics and Decision Making 25 (2025) 83. doi:10.1186/s12911-025-02903-1.

[7] W. A. Osman, Social media use and associated mental health indicators among University students: a cross-sectional study, Scientific Reports 15 (2025) 9534. doi:10.1038/s41598-025-94355-w.

[8] I. Levkovich, Is artificial intelligence the next co-pilot for primary care in diagnosing and recommending treatments for depression?, Medical Sciences 13 (2025) 8. doi:10.3390/medsci13010008.

[9] O. Zalutska, M. Molchanova, O. Sobko, O. Mazurets, O. Pasichnyk, O. Barmak, I. Krak, Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network, in: V. Lytvyn, A. Kowalska-Styczen, V. Vysotska (Eds.), Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Machine Learning Workshop, Kharkiv, Ukraine, April 20-21, 2023, volume 3387 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 344–356. URL: https://ceur-ws.org/Vol-3387/paper26.pdf.

[10] Ş. H. Gemicioğlu, M. Sönmez, G. Kara, B. Bozer, Ö. Güder, I. Oztürk, Examination of depression, anxiety, and stress in nursing students receiving distance education during the COVID-19 pandemic, Journal of Psychiatric Nursing 16 (2025) 38–47. doi:10.14744/phd.2025.26566.

[11] A. O. Hassan, I. M. Jamal, S. D. Ahmed, A. U. Abdullahi, Predicting Student Depression using Machine Learning: A Comparative Analysis of Machine Learning Algorithms for Early Depression Detection in Students, AITU SCIENTIFIC RESEARCH JOURNAL 4 (2025) 28–35. doi:10.63094/AITUSRJ.25.4.1.4.

[12] P. Jain, K. R. Srinivas, A. Vichare, Depression and suicide analysis using machine learning and NLP, in: Journal of Physics: Conference Series, volume 2161, IOP Publishing, 2022, p. 012034. doi:10.1088/1742-6596/2161/1/012034.

[13] A. Mali, R. R. Sedamkar, Prediction of Depression Using Machine Learning and NLP Approach, in: V. E. Balas, V. B. Semwal, A. Khandare (Eds.), Intelligent Computing and Networking, volume 301 of *Lecture Notes in Networks and Systems*, Springer Nature Singapore, Singapore, 2022, pp. 172–181. doi:10.1007/978-981-16-4863-2_15.

[14] Y. Luo, Z. Ye, R. Lyu, Detecting student depression on Weibo based on various multimodal fusion methods, in: S. U. Jan (Ed.), Fourth International Conference on Signal Processing and Machine Learning (CONF-SPML 2024), volume 13077, International Society for Optics and Photonics, SPIE, 2024, p. 130770M. doi:10.1117/12.3027177.

[15] G. Lorenzoni, C. Tavares, N. Nascimento, P. Alencar, D. Cowan, Assessing ML classification algorithms and NLP techniques for depression detection: An experimental case study, PloS one 20 (2025) e0322299. doi:10.1371/journal.pone.0322299.

[16] Y. Zhai, Y. Zhang, Z. Chu, B. Geng, M. Almaawali, R. Fulmer, Y.-W. D. Lin, Z. Xu, A. D. Daniels, Y. Liu, Q. Chen, X. Du, Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students, Journal of Counseling & Development 103 (2025) 110–125. doi:10.1002/jcad.12543.

[17] M. M. van Buchem, A. A. de Hond, C. Fanconi, V. Shah, M. Schuessler, I. M. Kant, E. W. Steyerberg, T. Hernandez-Boussard, Applying natural language processing to patient messages to identify depression concerns in cancer patients, Journal of the American Medical Informatics Association 31 (2024) 2255–2262. doi:10.1093/jamia/ocae188.

[18] M. Salıcı, Ü. E. Ölçer, Impact of Transformer-Based Models in NLP: An In-Depth Study on BERT and GPT, in: 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP), 2024, pp. 1–6. doi:10.1109/IDAP64064.2024.10710796.

[19] O. V. Barmak, O. V. Mazurets, I. V. Krak, A. I. Kulias, L. E. Azarova, K. Gromaszek, S. Smailova, Information technology for creation of semantic structure of educational materials, in: R. S. Romaniuk, M. Linczuk (Eds.), Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, volume 11176, 2019, p. 1117623. doi:10.1117/12.2537064.

[20] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, B. Alshemaimri, BERT applications in natural language processing: a review, Artificial Intelligence Review 58 (2025) 166. doi:10.1007/s10462-025-11162-5.

[21] G. Bharathi Mohan, R. Prasanna Kumar, S. Parathasarathy, S. Aravind, K. B. Hanish, G. Pavithria, Text Summarization for Big Data Analytics: A Comprehensive Review of GPT 2 and BERT Approaches, Springer Nature Switzerland, Cham, 2023, pp. 247–264. doi:10.1007/978-3-031-33808-3_14.

[22] Adam, 2025. URL: https://keras.io/api/optimizers/adam/.

[23] G. Naidu, T. Zuva, E. M. Sibanda, A Review of Evaluation Metrics in Machine Learning Algorithms,

in: R. Silhavy, P. Silhavy (Eds.), Artificial Intelligence Application in Networks and Systems, volume 724 of *Lecture Notes in Networks and Systems*, Springer International Publishing, Cham, 2023, pp. 15–25. doi:10.1007/978-3-031-35314-7_2.

[24] N. Yadav, Student-depression-text, 2023. URL: https://www.kaggle.com/datasets/nidhiy07/student-depression-text.

[25] O. Sobko, O. Mazurets, M. Molchanova, I. Krak, O. Barmak, Method for analysis and formation of representative text datasets, in: T. Hovorushchenko, E. Zaitseva, S. Lysenko, V. G. Levashenko (Eds.), Proceedings of the 1st International Workshop on Advanced Applied Information Technologies with CEUR-WS, Khmelnytskyi, Ukraine, Zilina, Slovakia, December 5, 2024, volume 3899 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 84–98. URL: https://ceur-ws.org/Vol-3899/paper9.pdf.

[26] Python, 2025. URL: https://www.python.org/.

[27] M. Molchanova, V. Didur, O. Mazurets, O. Sobko, O. Zakharkevich, Method for Construction and Demolition Waste Classification Using Two-Factor Neural Network Image Analysis, in: N. Shakhovska, A. T. Augousti, S. Liaskovska, O. Duran (Eds.), Proceedings of the 2nd International Conference on Smart Automation & Robotics for Future Industry (SMARTINDUSTRY 2025), Lviv, Ukraine, April 3-5, 2025, volume 3970 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 168–182. URL: https://ceur-ws.org/Vol-3970/PAPER14.pdf.

[28] Hugging face, 2025. URL: https://huggingface.co/docs/transformers/index.

[29] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, E. Manziuk, O. Barmak, Method for neural network detecting propaganda techniques by markers with visual analytic, in: A. Pakstas, Y. P. Kondratenko, V. Vychuzhanin, H. Yin, N. Rudnichenko (Eds.), Proceedings of the 12th International Conference Information Control Systems & Technologies (ICST 2024), Odesa, Ukraine, September 23-25, 2024, volume 3790 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 158–170. URL: https://ceur-ws.org/Vol-3790/paper14.pdf.

[30] pandas, 2025. URL: https://pandas.pydata.org/.

[31] NumPy, 2025. URL: https://numpy.org/.