

Modules for mental and physical health states analysis based on user text input

Artem Bashtovyi^a, Andriy Fechan^a, Vitaliy Yakovyna^{a,b}

^a Lviv Polytechnic National University, 12 Bandera str., Lviv, 79013, Ukraine

^b University of Warmia and Mazury in Olsztyn, 2 Michała Oczapowskiego str., Olsztyn, 10-719, Poland

Abstract

Given the pandemic situation around the world, mental and physical wellbeing become a crucial part of our lives, having said that health support is one of the most popular topics nowadays. In order to keep mental stability, we have to feel when the actual help is needed. Sometimes it's extremely hard to analyze our mental and physical health and the time when we need to ask for help. In this work, we built the module for mental illness classification, which can identify mental states based on human text input. The module is part of a platform for physical and mental state identification based on user journaling. The platform will allow daily journaling and automatically identify user mental and physical states based on the texts.

Keywords

Natural language processing, classification, multiclass analysis, mental health, journaling, BERT, physical state.

1. Introduction

Human health on of the most important factors for longevity and happy life. Nowadays most a lot of people neglect the importance of mental health as far as mental issues are not always related to physical symptoms. Usually, people reach out to doctors for additional help and treatment, albeit problems with the mental issues are related to the physical state, sometimes it's not that straightforward to identify them. Furthermore, the question of supporting mental health is crucial during the pandemic, when the lockdown affects our social life. The COVID-19 pandemic has a negative impact on people's mental well-being, this even provoked the term "covid depression". With an appropriate treatment and care method, many individuals are able to quickly identify the issues and work on them as long as they elaborate with specialists. Some people may resist going to the specialists out of fear, embarrassment, lack of support, and resources. At the same time usage of IT (information technologies) can help with basic identification of mental state, thus creating a first step to the treatment. One of the examples could be the application that would help to analyze mental state without a specialist's help, just sitting at home.

Information technologies help us identify health problems more easily, for instance, usage of image and motion recognition, IoT sensors for physical human health, and potential diagnostic. In addition, one of the approaches which do not require additional equipment and sensors - text analysis of human-written texts, including the texts published on the internet. In this work, we worked on the model for multi-class text classification, which will be one of the components of a web application for identifying mental and physical states based on user text input. The main purpose is to provide an effective self-analysis platform for people who want to support their mental and physical well-being.

IDDM-2021: 4th International Conference on Informatics & Data-Driven Medicine, November 19–21, 2021 Valencia, Spain

EMAIL: artem.bashtovyi.mpz.2020@lpnu.ua (A. 1); andrii.v.fechan@lpnu.ua (A. 2); yakovyna@matman.uwm.edu.pl (A. 3)

ORCID: 0000-0003-4304-8605 (A. 1); 0000-0001-9970-5497 (A. 2); 0000-0003-0133-8591 (A. 3)



© 2021 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)

The platform will allow classifying user's mental state(anxiety, depression, happiness, etc) and physical state(fatigue, pain, energy).

2. Machine learning and medicine

Numerous innovations and new tech solutions are already on the market of healthcare, it's hard to imagine modern medicine without computer-supported information systems. Machine learning(ML) algorithms are specifically useful for analyzing human health[1,2,3] and pandemic affect prediction[4]. Artificial intelligence(AI) is used in healthcare specifically for solving a wide range of tasks[5], like the following:

1. Predicting suicide risk at hospitalization via ML algorithms based on social media data.
2. Identifying the correlation between antidepressant usage and deprivation based on behavior.
3. Data analysis for identifying drugs abuse.

On the whole, AI is becoming more popular for usage in combination with precision medicine, usage of ML algorithms could potentially help to solve problems that will retrofit the analysis of patients[6]. As the example in Figure 1, the components potentially could be improved and optimized. According to the research from Clinical and Translational Science(CTS) AI methods and spread out the approaches for patient treatment. AI is making medicine easier and more efficient nonetheless, complete AI integration into healthcare requires solving the issues like data misclassification, external factors which are not suitable for models training, user privacy.

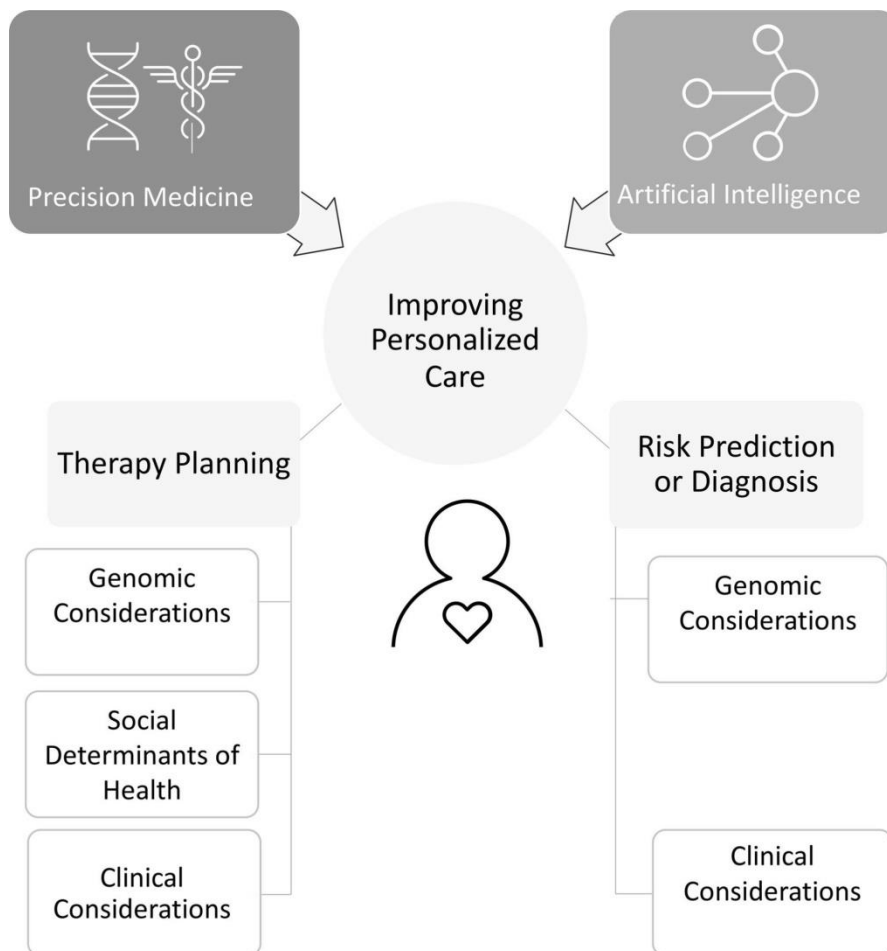


Figure 1: Precision medicine and AI combination

3. NLP for physical and mental states identification

Natural language processing (NLP) gives us the ability to process the text and extract meaningful information based on user input. It allows doing the classification (a process of feature selection), tagging, and parsing data in order to understand the human intention. In fact, NLP has been widely used in the medical field, specifically for mental and physical health analysis. Clinical data and even regular data from social media is a great source of useful information for mental illness detection [7], in addition, specific data sources like suicide notes can be used for algorithms development [8]. Classification approaches based on clinical data showed success in predicting mental problems with a precision of about 71% [9]. Recently, the research from the MIT group created a Long short-term memory (LSTM) based model for depression predisposition by processing user text input and voice input sequences.

Physical health issues classification is a pretty hard task in contrast to mental health issues due to the specific data processing and huge physical states diversification. Specific models are required for such issues classification, for example, this research includes specific models for low back pain detection [10]. It includes a comprehensive analysis of data from user medical records, which are formed by doctors. Such tasks include a lot of effort in terms of data selection. Another research describes a concrete approach for migraine classification[11]. Physical issues are hard to describe and detect from the solid user input text, consequently, physical state detection is a complex task that requires specific data selection and processing.

4. Platform architecture

The purpose of this work is to create a model for NLP multi-class text analysis. The model is the main component of the self-analysis platform for mental and physical states detection based on user input. Figure 2 shows the main component of the platform.

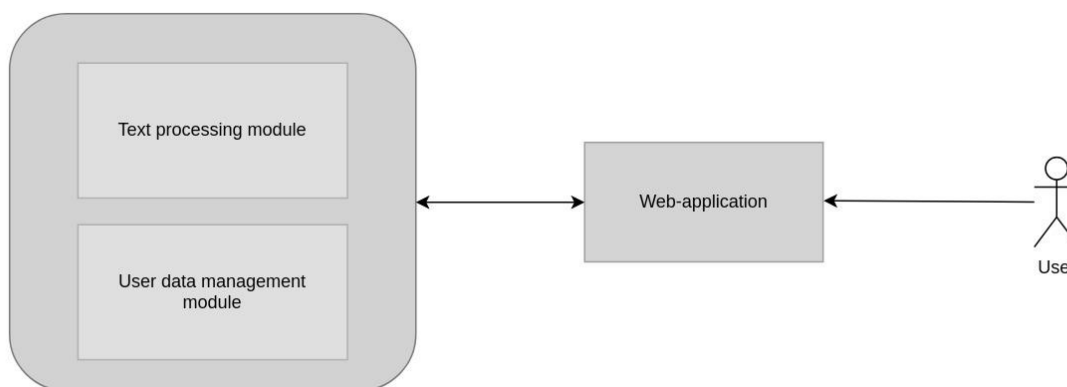


Figure 2: High-level platform architecture

Text processing module - main module for text processing and tagging, based on defined mental and physical states classes. This module is related to the following steps and operations:

- Data handling for classification
- Test classification for the mental and physical state

User data management module - basic component for storing, editing, and processing user records in the journals.

Web application - main client for user interaction and journaling.

The text processing module requires the main components described in Figure 3.

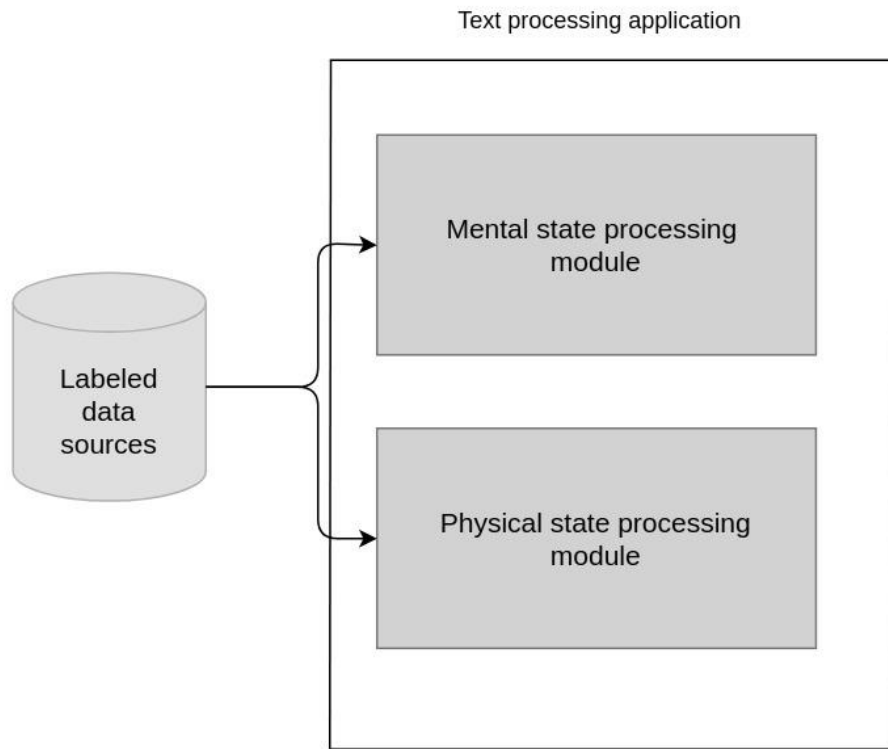


Figure 3: Text processing module structure

4.1. Dataset

Choosing the right dataset is crucial for NLP tasks, especially the processing of health-related text that contains disease or state information. Protecting personal data is very important for medical companies, the data must be fully confidential, although some of the sources provide open anonymous user information, finding relevant data is relatively hard. For the sake of simplicity and privacy policy regulation, the open-source data set from Reddit network was chosen[12]. The dataset contains the classified mental diseases data that is extracted from respective Reddit topics. It includes 13727 posts(records), the main 5 mental disorders classes are defined: Attention deficit hyperactivity disorder(adhd), Post-traumatic stress disorder(ptsd), anxiety, depression, bipolar disorder(bipolar). The additional class undefined(none) is associated with topics that are not connected to the mental state discussion(music, travel, science, politics). Table 1 shows the distribution of classes within the dataset.

Table 1

Data set labels distribution

Class(subreddit)	Posts count	Avg. No of words per post
adhd	2465	152.74
anxiety	2422	170.38
pstd	2001	233.55
bipolar	2407	203.28
depression	2450	152.74
undefined(none)	1982	238.52

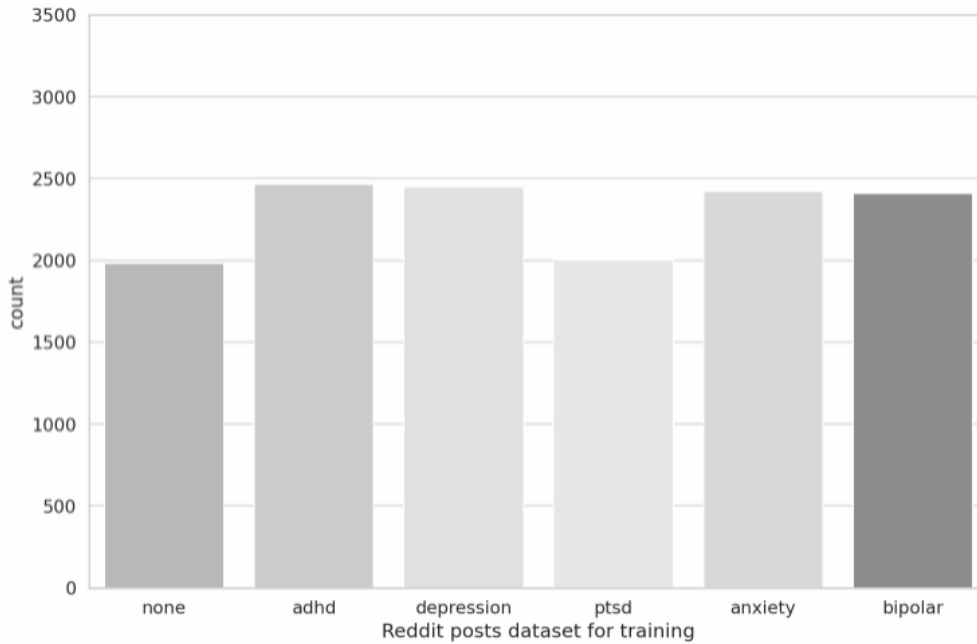


Figure 4: Dataset classes distribution

4.2. Initial model

The classification module is based on BERT BASE[13] pre-trained model, the neural network with 12-layers, 768-hidden, 12-heads, 110M parameters, trained on lower-cased English text. BERT approach is specifically useful for the current task because it has a “fine-tuning” ability for domain-specific data(in this case mental and physical illness). Having the proper datasets for physical and mental states gives us the ability to fine-tune the model by applying the softmax function. Python, Pytorch, Transformers are core technologies used for the classification module.

Tokenization plays a crucial role in NLP text processing, BERT model uses a special kind of tokenizer (BertTokenizer) based on reserved classes. BERT is a sequence-to-sequence model, having said that it requires a fixed-length input sequence. Based on the conducted experiments it was defined the number of tokens per sequence - 512. The padding and truncation were used as methods for proper sequence length generation. After careful data examination and test runs, 10 epochs was the most relevant value with a learning rate $2e-5$. AdamW(one of the fastest algorithms[14]) served as the main optimization approach, it’s the combination of AdaGrad and RMSProp algorithms, recently became popular in the computer vision branch. In order to avoid overfitting, we used dropout probability as 0.3. Table 2 shows epoch details:

Table 2

Train epoch details

Epoch	Loss	Accuracy
1	0.8162	0.7653
2	0.5332	0.8420
3	0.4421	0.9355
4	0.2671	0.9184
5	0.1288	0.9577
6	0.0857	0.9612
7	0.0749	0.9692
8	0.0506	0.9702
9	0.0411	0.9734
10	0.0391	0.9790

Figure 5 shows the training and validation accuracy difference. The last step shows almost no difference in training accuracy, thus we decided to stay with base parameters.

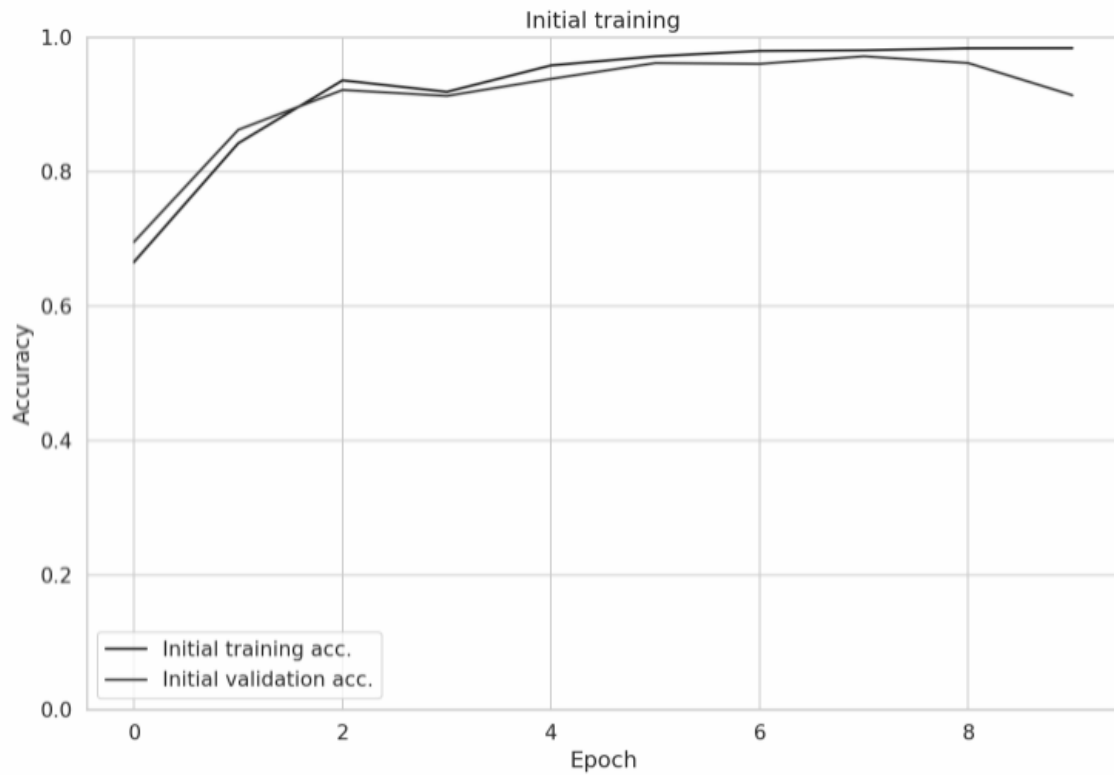


Figure 5: Training and validation accuracy results

Table 3

General results

Precision	Recall	F1	Accuracy
0.88	0.88	0.89	0.88

Table 4

BERT multiclass performance results. Data set labels distribution

Class name	Precision	Recall	F1
adhd	0.90	0.91	0.90
anxiety	0.86	0.84	0.85
bipolar	0.88	0.83	0.86
depression	0.81	0.88	0.84
ptsd	0.87	0.86	0.89
undefined(none)	0.98	0.97	0.98

Table 3 contains the final results after all training stages, the general model precision is 0.87, which is a pretty high result dataset based on Reddit topics. The hardest part in multi-class health issues classification is the high correlation between different classes. The following approaches are used for precision, recall, F1, and accuracy calculation:

- Precision - true positives divided by sum of true positives and false positives
- Recall - true positives divided by sum of true positives and false negatives
- F1 - weighted average of precision and recall
- Accuracy is defined based on the validation dataset with provided targets

The results from Table 4 gives us precise information about text analysis and specific classes prediction. We can definitely say that the precision of “none” class classification is high, it gives us hope that the model is relatively resistant to the false-positive results. Also, we can say that the top classes from the results are “ptsd” and “adhd” diseases. The worst results are “anxiety” and “depression”, there are a few reasons for this consequence. Those two classes are highly correlated with other classes. The “bipolar” class contains around 30% “depression” posts, about 11% “anxiety” posts, and about 12% posts related to the “depression” in the same manner. Consequently, the correlation between the classes impairs the precision results, which requires careful dataset classification. We identified that we can observe text misclassification on the user input due to the connection of the mental states. The proper base dataset classification, including the interaction with experts in medical records and diagnosis, will help to improve the situation.

4.3. Adjusting elements of the model

The pre-trained model that was used for the experiment performed relatively well, since BERT is based on transformer architecture it has achieved state-of-the-art performance for NLP tasks on the dataset. In the task described above, we performed fine-tuning of the model for mental disorders classification that gives us the model for specific tasks. In spite of good model precision, the testing steps took a lot of resources, specifically the time for processing the test dataset. The module for mental health state prediction requires constant improvement in the future for experiments with different states and higher classification precision, having said that we decided to modify the pre-trained model in hope to reduce testing time in the future and change the model size. Fortunately, there are a lot of approaches[15] to tune BERT configuration(bertology) and pre-trained model in order to gain performance, precision boost, reduce the time and size of the transformer-based models.

Following studies explored methods and approaches improving the original BERT model for various tasks. XLNet model research[16] suggested adding autoaggressive capabilities to BERT, improving the quality of the original model, though at the cost of extra compute and time requirements. Well-known RoBERTa model provides higher performance[17]. The recent studies explain the “head-pruning” approach to save testing time and model size without sacrificing the performance too much. This can be useful for tasks that require constant model testing and fine-tuning.

The approach described in the study[18] helps prune unimportant heads. Basically, two main approaches are used during head pruning. First to prune the whole head from the model, the second prune only least useful heads weights for the given head. The head pruning requires defining the importance of each head, the following formula suggested in the study and was used:

$$I_h = E_{x \sim X} \left| Att_h(X)^T \frac{\partial L(X)}{\partial Att_h(X)} \right|, \quad (1)$$

where X - the data distribution, $L(x)$ - loss for sample x

4.4. Results from the experiment

After conducting the experiments with head pruning, we highlighted the best parameters for testing the configured model. Basically, the least head importance was defined for encoder multi-head attention(EMHA). Hence the final pruning was performed mostly on EMHA. Figure 6 shows particular steps in 10% head-pruning and the effect on the precision. Basically, the results we discovered - the model precision drops drastically only after 38% pruned heads, which gives us a positive attitude. The precision was reduced by almost 8% for the testing stage at that point, however, test data validation took about 25% less time. Consequently, the adjusted model configuration for the given dataset is suitable in terms of testing the model quickly. The final result showed that 8 heads per

layer with total precision of almost 80% and time lesser by almost 24% is the best option for the selected dataset and model architecture.

The final decision was to keep parameters for the initial model described in section 4.1, and perform head pruning based on the final results above. Furthermore, we adjusted the number of hidden layers respectively with pruned heads.

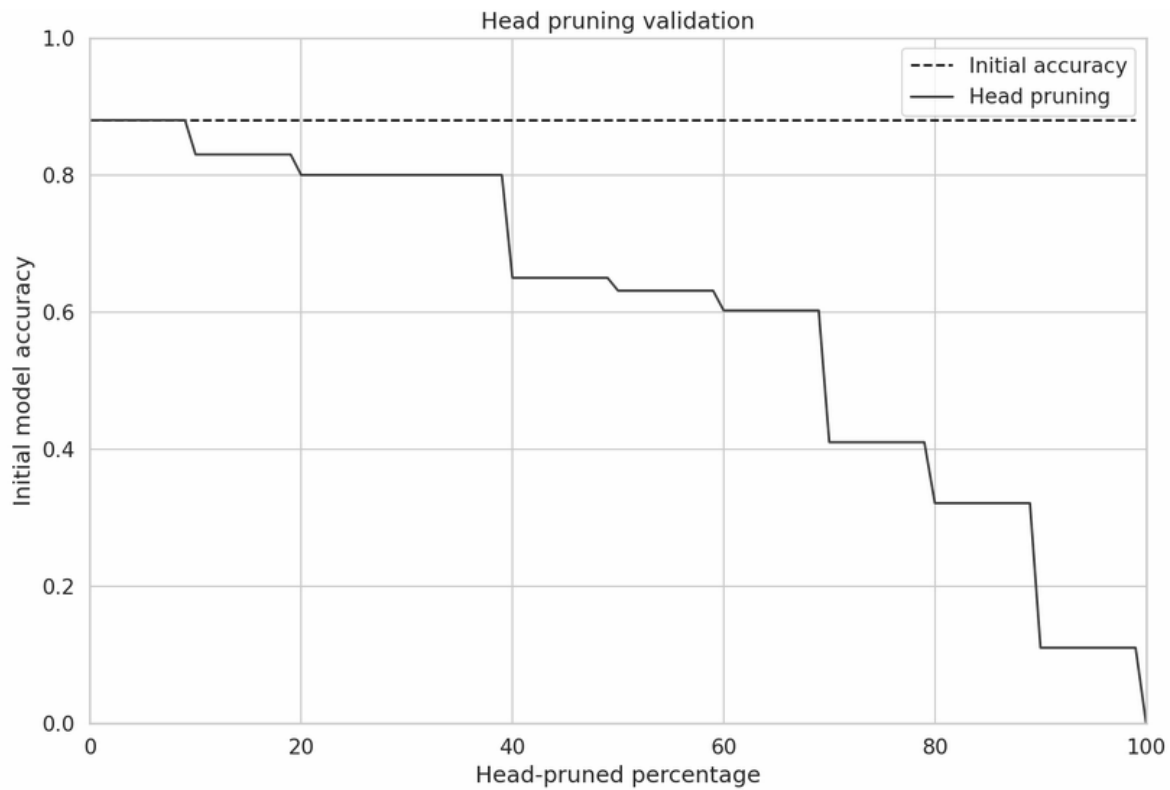


Figure 5: Head pruning stages with 10% step

Table 5

The original and pruned model with final parameters test performance results

Model	Precision	Recall	F1	Accuracy
Original pre-trained	0.88	0.88	0.89	0.88
Modified model with 38% heads pruned	0.81	0.82	0.79	0.81

The positive results give us the ability to perform the testing stage for different mental issues records faster. Additionally, we plan to use the approach described above for physical state classification model tuning and configuration, thereby reducing testing time on physical states dataset.

The future work is related to the second module development for physical state classification. Physical diseases dataset requires careful processing, furthermore, the model development includes constant feedback and collaboration with doctors and medical companies. The final step is to combine to model into a single text classification model and create the system with a web application for the full journaling process.

5. References

- [1] Y. Bashtyk, J. Campos, A. Fechan, S. Konstantyniv, V. Yakovyna, Computer monitoring of physical and chemical parameters of the environment using computer vision systems: Problems and prospects. CEUR Workshop Proceedings, 2020, 2753, pp. 437–442.
- [2] Z. Mykytyuk, A. Fechan, V. Petryshak, G. Barylo, O. Boyko, Optoelectronic multi-sensor of SO₂ and NO₂ gases. Modern Problems of Radio Engineering, Telecommunications and Computer Science, Proceedings of the 13th International Conference on TCSET 2016, 2016, pp. 402–405.
- [3] A. Kucher, O. Boyko, K. Ilkanych, A. Fechan, N. Shakhovska Retrospective analysis by multifactor regression in the evaluation of the results of fine-needle aspiration biopsy of thyroid nodules CEUR Workshop Proceedings, 2020, 2753, pp. 443–447.
- [4] V. Yakovyna, N. Shakhovska, Modelling and predicting the spread of COVID-19 cases depending on restriction policy based on mined recommendation rules. Mathematical Biosciences and Engineering 18(3), 2021 2789–2812,.
- [5] Ayesha Kamran Ul haq, Amira Khattak, Noreen Jamil, M. Asif Naeem, Farhaan Mirza, "Data Analytics in Mental Healthcare", Scientific Programming, 2020, doi:10.1155/2020/2024160
- [6] K.B. Johnson, W.Q. Wei, D. Weeraratne, M.E. Frisse, K. Misulis, K. Rhee, J. Zhao, J.L. Snowden, Precision medicine, AI, and the future of personalized health care, Clin. Trans. Sci., 14, 2020, doi:10.1111/cts.12884
- [7] A. Le Glaz, Y. Haralambous, D. Kim-Dufor, P. Lenca, R. Billot, T. Ryan, J. Marsh , J. DeVylder, M. Walter, S. Berrouiguet, Lemey C Machine Learning and Natural Language Processing in Mental Health: Systematic Review J Med Internet, Res 2021, doi: 10.2196/15708
- [8] R. Cavlo, D. Milne, M. Hussain, H. Christensen, Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 2017 23(5). doi:10.1017/S1351324916000383
- [9] N. Viani, R. Botelle, Kerwin, J. et al. A natural language processing approach for identifying temporal disease onset information from mental healthcare text, 2021, Sci Rep 11, 757. <https://doi.org/10.1038/s41598-020-80457-0>
- [10] Judd, Michael & Zulkernine, Farhana & Wolfrom, Brent & Barber, David & Rajaram, Akshay Detecting Low Back Pain from Clinical Narratives Using Machine Learning, 2018, doi:10.1007/978-3-319-99133-7_10.
- [11] M. Katsuki, N. Narita, Y. Matsumori, N. Ishida, O. Watanabe, S. Cai, T. Tominaga. Preliminary development of a deep learning-based automated primary headache diagnosis model using Japanese natural language processing of medical questionnaire, 2020, doi:10.25259/SNI_827_2020.
- [12] Reddit topics classified dataset URL: <https://github.com/amurark/mental-health-classification>
- [13] <https://huggingface.co/bert-base-uncased>
- [14] Gentle Introduction to the Adam Optimization Algorithm for Deep Learning URL: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning>
- [15] Rogers, Anna et al. "A Primer in BERTology: What We Know About How BERT Works." Transactions of the Association for Computational Linguistics 8,2020,: 842-866.
- [16] Yang, Zhilin et al. "XLNet: Generalized Autoregressive Pretraining for Language Understanding." NeurIPS ,2019, arXiv:1906.08237.
- [17] Liu, Yinhan et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach.", 2019,: n. pag., ArXiv abs/1907.11692.
- [18] Michel, Paul et al. "Are Sixteen Heads Really Better than One?" NeurIPS, 2019, arXiv:1905.10650