https://doi.org/10.31891/csit-2026-1-9

UDC 004.9

CROSS-LINGUAL TRANSFORMER-BASED SCREENING OF POST-TRAUMATIC STRESS DISORDER BASED ON COMPARATIVE ANALYSIS OF BERT AND XLM-ROBERTA WITH MACHINE TRANSLATION ADAPTATION FOR UKRAINIAN LANGUAGE

Andrii FEDORYCHKO
Student of Information Systems and Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
e-mail: andrii.fedorychko.sa.2022@lpnu.ua
https://orcid.org/0009-0005-1593-3481

Victoria VYSOTSKA
Doctor of Technical Sciences, Associate Professor of Information Systems and Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
e-mail: victoria.a.vysotska@lpnu.ua
https://orcid.org/0000-0001-6417-3689

Lyubomyr CHYRUN
Candidate of Technical Sciences, Associate Professor of Information Systems and Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
e-mail: lyubomyr.v.chyrun@lpnu.ua
https://orcid.org/0000-0002-9448-1751

The paper presents a comprehensive study on automated screening of post-traumatic stress disorder (PTSD) using transformer-based natural language processing models in a cross-lingual setting. The research aims to evaluate the feasibility of deploying an intelligent PTSD detection system in Ukraine under conditions of limited, localised training data. A balanced corpus of 4,822 text records was constructed by aggregating publicly available PTSD-related datasets, including 2,042 PTSD-positive texts and 2,780 control texts representing neutral content and other psychological conditions. The study compares the performance of the English-language BERT (bert-base-uncased) model and the multilingual XLM-RoBERTa (xlm-roberta-base) model applied to a Ukrainian-language corpus generated via machine translation using the Google Translate API and large language models for complex structures. Word cloud visualisation and semantic analysis confirmed preservation of core psychological markers during translation. Experimental results demonstrate high predictive performance for both architectures. The English-language model achieved an Accuracy of 0.90 and an ROC-AUC of 0.962. In contrast, the Ukrainian-language model achieved an Accuracy of 0.85 and an ROC-AUC of 0.940, significantly outperforming existing Ukrainian multi-class stress detection models (Accuracy ~0.45) and exceeding standard multilingual mental health benchmarks (0.78–0.82), establishing a robust state-of-the-art baseline for Ukrainian clinical NLP. Importantly, Recall remained identical (0.88) across both language settings, indicating strong sensitivity to PTSD markers despite translation-induced lexical noise. The minimal AUC degradation (2.3%) confirms the robustness of transformer architectures to cross-lingual adaptation. The findings validate the viability of combining machine translation with multilingual transformers for the rapid deployment of mental health screening systems in low-resource language environments. The proposed pipeline enables scalable and cost-effective digital PTSD monitoring while maintaining clinically relevant diagnostic sensitivity.

Keywords: post-traumatic stress disorder (PTSD); natural language processing (NLP); transformer models; BERT; XLM-RoBERTa; machine translation; cross-lingual transfer; text classification; mental health screening; ROC-AUC; deep learning; Ukrainian language corpus.

Received: 04/02/2026
Accepted:11/03/2026
Published:26/03/2026

Introduction

Post-traumatic stress disorder (PTSD) is a severe mental condition resulting from experiencing or observing traumatic events, such as a threat to life or health [2]. In recent years, there has been a rapid increase in studies applying machine learning (ML) techniques to predict and diagnose this disorder [3]. This trend is mainly due to the impact of the COVID-19 pandemic, which has become a global trauma and has led to a surge in stress worldwide [2]. Despite technological progress, there are still concerns in the scientific community about the clinical reliability and generalisability of the ML models obtained, which inhibit their implementation by practitioners [3].

Traditional diagnosis of PTSD through clinical interviews is often accompanied by subjectivity and the risk of misidentifying the disorder as depression, which significantly worsens the prognosis of treatment [2]. The stigmatisation of mental problems and the reluctance of patients to openly discuss symptoms also hinder timely care [5]. In this context, Natural Language Processing (NLP) offers powerful tools for analysing digital narratives: social media posts, therapy transcripts, and online forums in order to identify linguistic patterns specific to PTSD [5]. Text data, as digital biomarkers, offer significant advantages because they are safe for patients, require minimal collection costs, and directly reflect a person's emotional state through speech content [2].

Experimental studies using sentiment analysis and deep learning demonstrate high efficiency in screening, achieving an accuracy of more than 80% on interview data [2]. Patient testimonials collected using questionnaires that simulate DSM-5 clinical protocols have proven highly effective at rapidly identifying symptoms, especially among vulnerable populations [4]. However, systematic reviews emphasise that, for the successful implementation of such systems, the identified ML predictions must be clearly consistent with the existing theoretical understanding of the disorder [3].

Models built on the Transformer framework, in particular the BERT model, demonstrate a significant advantage in classifying symptoms and behavioural indicators compared to classical machine learning algorithms such as SVM or KNN [4]. The combination of advanced NLP techniques, including syntactic parsing, thematic modelling, and deep learning, enables the creation of forward-looking screening systems that ensure the objectivity of the diagnostic process in the digital environment [5]. Thus, the development of integrated models combining empirical ML results with fundamental theoretical knowledge is a key step towards improving the accessibility and accuracy of psychiatric care [3].

To address the gap in localised mental health screening tools, the main scientific contributions of this study are as follows:

- Creation of a localised synthetic corpus: We developed a specialised, balanced corpus for PTSD screening in the Ukrainian language by adapting validated English-language datasets using advanced machine translation techniques.
- Comparative framework: We designed a comprehensive comparative analysis to evaluate the screening efficacy of a native English Transformer (BERT) against a cross-lingual model (XLM-RoBERTa) under conditions of limited localised data.
- Identification of cross-lingual semantic shifts: Using Explainable AI (SHAP), we provided an in-depth analysis of specific lexical translation artefacts and morphological changes that impact the model's ability to accurately detect trauma markers in Slavic languages.

The rest of the paper is organised as follows: The subsequent sections review related works, define the problem statement, and provide a theoretical overview of text vectorisation, transformer models, and modern text analysis methods in mental health. Following this, the research methodology is detailed, encompassing data description, processing techniques, and the formalised model architecture. Next, the research results are presented, comparing the performance metrics of the evaluated models. Finally, the discussion interprets these findings in the context of cross-lingual translation artefacts, and the conclusions summarise the study and outline directions for future research.

Related Works

Modern neural network models based on the Transformer architecture have become a key stage in the evolution of machine translation (ML) systems, almost entirely replacing outdated statistical and rule-based approaches [6]. Analysis of the work of leading systems such as DeepL, Google Translate, and GPT demonstrates their high efficiency in technical and journalistic materials. However, it reveals significant difficulties in translating highly sensitive clinical and psychological texts, where emotional nuances and specific diagnostic markers can be distorted [6]. Experimental studies emphasise that, despite considerable time savings, the quality of automated translation still requires human monitoring and mandatory post-editing [6], [7]. The poor quality of adaptation of specialised texts is a serious barrier, since MP systems often achieve better results in their usual domain than in narrow professional fields [6].

The technical implementation of PTSD diagnostic systems is complicated by the problem of data imbalance, which arises due to the difficulty of collecting real clinical interviews [8]. To overcome this deficit, it is promising to use large language models (LLMs) to increase the volume of training data by synthesising text transcriptions and paraphrasing existing training samples [8]. At the same time, a comparative analysis of the architectures shows that specialised models adapted to the mental health domain are significantly superior to general transformer models in detecting symptoms, especially in cases with concomitant depression [9]. Despite the advancements in NLP, a critical research gap remains specifically in PTSD screening for the Ukrainian language. While a recent study explored automated psychoemotional stress detection among Ukrainian military personnel using mBERT, the complexity of their multi-class approach resulted in a relatively low accuracy of ~0.45 [1]. Currently, there is a severe lack of highly accurate, binary PTSD classification models for Ukrainian, primarily due to the absence of verified, clinically annotated datasets. Therefore, the urgent task of this study is to bridge this gap by evaluating whether cross-lingual adaptation (via machine translation) can compensate for the lack of native data while maintaining the high diagnostic accuracy required for clinical screening.

Problem statement

Post-traumatic stress disorder (PTSD) remains one of the most complex and socially significant mental health conditions, particularly in contexts of war, armed conflict, pandemics, and large-scale societal crises. Traditional diagnostic procedures based on structured clinical interviews require significant time, professional expertise, and direct patient engagement. These factors create substantial barriers to early detection, especially in regions where access to qualified psychiatric care is limited. Furthermore, subjective bias, social stigmatisation, and comorbid conditions such as depression complicate differential diagnosis and reduce screening reliability.

Recent advances in Natural Language Processing (NLP) and transformer-based architectures have demonstrated high effectiveness in detecting psychological markers from textual self-narratives. However, most high-performing models are developed and validated primarily on English-language corpora. It creates a methodological and practical gap for low-resource language environments, including Ukrainian, where verified, clinically annotated datasets remain scarce.

An additional challenge arises from the necessity of cross-lingual adaptation. While multilingual transformer models such as XLM-RoBERTa enable knowledge transfer between languages, the impact of machine translation on diagnostic accuracy remains insufficiently studied. Machine translation may introduce lexical noise, distort emotional nuance, and affect the specificity of classification models. At the same time, the rapid deployment of automated screening systems in the Ukrainian digital space requires scalable solutions that do not rely on costly, time-consuming manual corpus annotation.

Thus, a critical scientific and practical problem emerges: how to ensure high diagnostic sensitivity and predictive reliability of transformer-based PTSD screening models in a low-resource language setting using machine-translated data, while preserving clinically meaningful linguistic markers. Addressing this problem requires a comparative evaluation of monolingual and multilingual architectures, analysis of translation-induced distortions, and validation of cross-lingual robustness using standard medical classification metrics.

The solution to this problem is essential for developing scalable, cost-effective, and clinically relevant digital mental health monitoring systems that can support both civilian populations and military personnel in real time. To address this scientific problem, this study formulates the following core research hypothesis: Semantic markers of PTSD, originally embedded in English clinical narratives, possess sufficient cross-lingual resilience to withstand the lexical and morphological noise introduced by machine translation into a Slavic language (Ukrainian). Consequently, it is hypothesised that a cross-lingual transformer model (XLM-RoBERTa) trained on this synthetic translation data can achieve baseline screening accuracy and predictive reliability comparable to those of a native monolingual model (BERT).

Text vectorisation and transformer models

The current stage of development of natural language processing technologies is characterised by a transition from classical methods of word nesting, such as Word2Vec, GloVe, and FastText, to more complex context-sensitive architectures [11]. A revolutionary breakthrough in this area was the emergence of the BERT (Bidirectional Encoder Representations from Transformers) model, which, through bidirectional text analysis, enabled a deeper understanding of the semantics and context of messages [12]. With its ability to cross-lingually transfer knowledge and efficiently perform a wide range of tasks, from tokenisation to sentiment analysis, the BERT architecture has become the standard for building intelligent language understanding systems [12].

The effectiveness of using transformer models, in particular RoBERTa and GPT, in emotional and semantic analysis tasks is confirmed today through a comparative analysis of standard classification metrics: accuracy, completeness, and F1 measures [14]. This approach allows us to identify the most appropriate scenarios for applying neural network architectures to the automated processing of complex linguistic patterns [14]. At the same time, optimisation of computing resources is critical for practical implementation in high-load systems. The DistilBERT model offers an effective solution to this issue, maintaining the high accuracy of the original BERT while significantly increasing text processing speed [13]. Its ability to understand context, combined with its fine-tuning, makes it an ideal tool for classifying and extracting key information [13].

A separate priority area is the development of models for Ukrainian. The use of community-collected corpora of texts allows not only to apply ready-made solutions, but also to train your own models adapted to specific subject areas [11]. The synergy of embedding models, the capabilities of fast encoder architectures, and high-quality language data provides the technological foundation for analysing specific content, particularly in areas where emotional accuracy and a deep understanding of the author's intentions are essential [13].

Overview of modern methods of text analysis in the field of mental health

The use of linguistic features in describing patients' own experiences opens up vast opportunities for automated screening of PTSD, since the choice of words directly reflects the state of mental health of a person [14]. Traditional quantitative analysis methods often struggle to handle unstructured text. Still, the use of keyword extraction algorithms (e.g., the Chi-square test) enables the construction of models with high consistency between computer diagnostics and psychiatric assessments [14]. Studies show that combining n-gram models with various machine learning algorithms, such as Naive Bayes or the Support Vector Machine (SVM), helps identify specific patterns of verbal behaviour [15]. At the same time, unigrams often yield the highest prediction accuracy, while multigrams help balance the model's sensitivity and specificity [15].

At the moment, NLP workflows already allow the analysis not only of short texts but also of extensive collections of electronic medical records (EMRs) numbering in the millions of clinical notes [16]. The use of pre-trained transformer models to extract research domain criteria (RDoC) enables visualisation of disease trajectories and detection of gender differences in the manifestations of disorders. For example, the analysis of clinical data indicates a higher level of sensorimotor impairment and abnormalities in the positive and negative valence systems in women compared to men, as well as in veterans compared to civilians. Such context-dependent real-time analysis is critical for monitoring the effectiveness of psychotherapy and identifying the risks of suicidal behaviour [16].

Despite significant successes, the challenge remains in differentiating PTSD from comorbidities, including depression [17]. Comparative analysis of different classifiers on specialised datasets such as DAIC-WOZ demonstrates that ensemble methods, namely Random Forest and XGBoost, achieve an accuracy of about 84%, which is significantly higher than that of classical SVM models [17]. It confirms that improving parameter settings and selecting appropriate architectures are key to advancing early detection of mental disorders [14]. Thus, integrating automated methods for analysing patients' self-expression into clinical practice not only reduces costs but also provides timely assistance in the early stages of the disorder [15].

Data description and processing

The study's basis was a representative corpus of text data, formed by aggregating several thematic datasets from the Kaggle platform. To ensure robust evaluation of the models across different text modalities, a corpus of 4,822 records was compiled from several distinct sources. The origin of the data significantly impacts the lexical composition, and utilising diverse platforms allows the model to generalise better. Specifically, the trauma-related and PTSD-positive narratives (totalling 2,048 records) were sourced from three distinct modalities:

- Clinical Narratives (120 records consist of formal, structured clinical text and transcripts, providing dense, professional diagnostic markers).
- Reddit Threads (1,264 records were aggregated from three separate psychological and mental health subreddits, contributing 82, 531, and 651 long-form posts, respectively). This modality represents detailed, self-reported psychological discussions and personal trauma disclosures.
- Twitter Posts: 664 records were collected from Twitter. These represent short-form, informal expressions of stress, characterised by colloquialisms, high emotional density, and abbreviated syntax.
- Non-PTSD texts (the remaining 2,774 records constitute the control group necessary to achieve the final dataset size and maintain appropriate class balance for training). This multi-source approach ensures that the transformer models learn a wide spectrum of lexical compositions, ranging from structured clinical terminology to informal social media phrasing.

To strengthen the validity of the chosen approach to sample formation and the visualisation of qualitative data composition, the word cloud method was used. This tool plays a critical role in the Intelligence Data Analysis (EDA) phase, as it allows you to instantly identify the most frequent tokens dominating each of the classes. In the context of the differential diagnosis of PTSD, such imaging makes it possible to determine the presence in texts of specific clinical terminology and descriptions of symptoms, such as "trauma", "flashback" or "nightmare", which make up the semantic core of the disorder. It confirms that the model will learn from meaningful markers of traumatic experiences rather than from commonly used vocabulary that reflects general negative sentiment.

Fig. 1 The given word cloud allows you to conduct a deep semantic analysis of the lexical composition of the target class containing signs of PTSD. The central place and the largest size of the word "feel" indicate that the corpus is composed of texts saturated with subjective emotional experiences and descriptions of internal states. It shows a high concentration of self-narratives, where authors focus on their own feelings. The word "ptsd", as one of the key visual centres, confirms the sample's thematic integrity and the respondents' direct self-identification with the condition during questioning or record-keeping. The high frequency of the verbs "know", "want", "think" and "make" indicates active cognitive processes: attempts to make sense of one's experiences, express needs, or describe the complexity of decision-making in a post-traumatic context.

An important aspect is the presence in the cloud of specific clinical terms that directly correlate with the diagnostic criteria for PTSD. Words such as "trauma", "nightmare", "flashback", "trigger" and "anxiety" form the core of traits on which deep learning models will rely for classification[5]. Their pronounced presence indicates that the dataset contains relevant descriptions of invasion and hyperexcitability symptoms. The presence of the words "therapy", "therapist", "medication", and "diagnosed" indicates that a significant portion of the texts describe patients' experiences interacting with the healthcare system and the treatment process, which adds clinical validity to the data.

Time markers such as "time," "year," "day," "since," and "always" emphasise the chronic nature of the conditions described. The frequency of the word time may be related to the description of symptom duration or to specific moments of trauma, which is critical for differential diagnosis. The social dimension of the disorder is clearly traced through relational vocabulary: "people", "friend", "family", "mother", "mom" and "kid". It demonstrates the impact of PTSD on interpersonal relationships and the role of the social environment in the process of living traumatic experiences. The presence of the words "help", "never", "nothing" and "hard" creates a specific linguistic profile of helplessness or a call for support, which often accompanies severe forms of mental disorders.

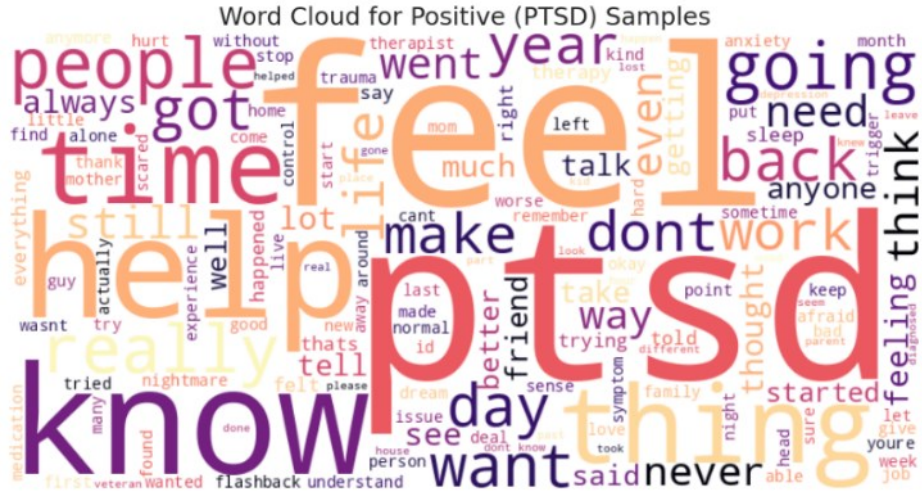

Fig. 1. Word cloud of positive meanings of the original English dataset

The process of preparing the Ukrainian-language dataset involved creating a Ukrainian-language version of the corpus using the Google Translate API. For each English-language unit, the corresponding machine translation was generated, preserving the original dataset's class distribution and structure. The visualisation of the Ukrainian-language corpus, generated via the Google Translate API, demonstrates the impressive stability of the semantic core and the preservation of key psychological markers from the original English-language dataset. The dominant lexemes in the word cloud are first-person verbs such as "know", "want" and "feel", which fully correspond to the English-language centres "know", "want" and "feel". It confirms that even with automated adaptation of texts, the subjective nature of self-narratives and the focus on inner experiences remain unchanged, which is critical for learning a multilingual model. Particular attention is drawn to the word "stressful", which, together with the less pronounced, but present abbreviation marker "PTSD", forms the terminological focus of the disorder. It indicates that the translator correctly identified the context of "Post-Traumatic Stress Disorder", expanding it into a full-fledged adjectival description, which is often found in Ukrainian professional and journalistic vocabulary.

Fig. 2. Word cloud of positive meanings for the Ukrainian-language dataset

A comparative analysis of the two visualisations reveals a deep conceptual symmetry: the English terms "trauma", "nightmare" and "flashback" are accurately reflected in the words "trauma", "nightmares" and "memories". The presence of the lexeme "hard" (equivalent to the English "hard") emphasises the emotional weight of the texts, and the words "happened", "life" and "think" indicate the process of reflection on the traumatic experience experienced. It is essential to note the presence of relational markers such as "people", "help" and "family", indicating the preservation of the social context of PTSD and its impact on interpersonal relationships, which was a characteristic feature of the original data. Despite caveats about the quality of machine translation in specialised domains, this visual analysis confirms that, for screening purposes, the content accuracy of key concepts was sufficient to form semantic embeddings.

In the XLM-RoBERTa multilingual architecture used in this study, lexical saturation of the Ukrainian-language word cloud is key to successful cross-lingual knowledge transfer. Since the cloud clearly reflects the symptoms (the words "symptoms", "pain", "fear"), the model gets the opportunity to build stable associations between speech pairs at the level of deep investments. It allows you to neutralise the imbalance and shortage of local data by leveraging a robust linguistic foundation laid during pretraining on large-scale datasets. Thus, the Ukrainian-language word cloud serves not just as an illustration of the translation but also as a means of validating the quality of data preparation, ensuring that the model will operate with relevant concepts rather than noise artefacts of machine translation.

Comparative analysis of English- and Ukrainian-language word clouds reveals a high degree of isomorphism in semantic structures, confirming the validity of using machine translation to prepare the data corpus. The most pronounced feature is the complete preservation of the dominant subjective experience: the English semantic core "feel", "know" and "want" has been transformed into the exact Ukrainian equivalents "feel", "know" and "want", which retains the focus on the first person, characteristic of self-narratives in the field of mental health. This lexical symmetry is critical for multilingual architectures, as it allows the model to build identical vector embedding for emotional states regardless of the language of expression.

In the terminological aspect, there is an interesting transformation. If in the original dataset the abbreviation "ptsd" occupies a central place, then in the Ukrainian version it is supplemented and partially replaced by the adjective "stressed". It suggests that the Google Translate API has successfully recognised the disorder's context by adapting it to the syntactic constructions of the Ukrainian language, while retaining specific markers such as "PTSD" and "symptoms". The clinical cluster, which includes the concepts of "trauma", "nightmare"/"nightmares" and "flashback"/"memories", demonstrates almost identical frequency in both buildings, which guarantees the preservation of diagnostic signs of the disorder after translation.

Relational and social contexts also demonstrate persistence: the frequent use of the words "people", "friend", "family" in the English segment is synchronised with the appearance of the lexemes "people", "friends", "family" in Ukrainian, which indicates the persistence of descriptions of interpersonal conflicts and social isolation inherent in PTSD. The appearance of the word "hard" in the Ukrainian cloud, as a direct semantic equivalent of the English "hard," further enhances the emotional tonality of the sample. Such a deep semantic correspondence allows us to conclude that the formed Ukrainian-language corpus has not lost its diagnostic value, and the automated adaptation of professional vocabulary has passed without significant semantic distortions, creating a reliable basis for further classification.

Model architecture

The architectural basis of this research is a fundamental shift from static methods of inserting words to deep, context-dependent representations enabled by the Transformer architecture. According to the latest reviews of neural network technologies, it is this architecture that has enabled qualitatively reproducing complex syntactic structures and accounting for contextual relationships previously inaccessible to rule-oriented or statistical approaches [10]. The selected models are based on self-attention, which allows the system to dynamically assign weights to each word in a sentence relative to others, forming an integrated semantic field. To build the English-language corpus, the bert-base-uncased model was used, which, with 12 encoder layers and 110 million parameters, provides a deep bidirectional understanding of the text [12]. The use of the "uncased" version allows you to neutralise the influence of the case, which is critical for the analysis of informal texts from social networks and forums, where users often ignore the rules of punctuation and capitalisation. The process of preparing data for this architecture uses the WordPiece tokeniser, which breaks words into sub-units, effectively addressing the limited vocabulary problem and allowing the model to correctly interpret complex medical terms even when they are spelt fragmentarily. Moving on to the analysis of the Ukrainian-language segment, the choice of the multilingual model xlm-roberta-base was due to its exceptional ability to cross-lingual knowledge transfer, enabling the leveraging of experience from large English-language corpora to improve classification quality in Ukrainian. Unlike standard BERT, this architecture has been trained using the RoBERTa methodology, with dynamic token masking and without predicting the following sentence, which, as confirmed by recent research, leads to more stable and accurate investments. Of particular note is the use of SentencePiece's byte-level tokenisation, which is ideal for Ukrainian, with its rich morphology and inflexion, as it allows the model to capture the semantics of word roots regardless of their prefixes or suffixes. It directly correlates with the task of detecting PTSD, where the emotional state of the author is often encoded in specific verb forms and adjectival constructions.

To ensure the reproducibility of our experiment and formalise the transformation of input text into a diagnostic decision, we define the training and classification pipeline mathematically. Both BERT and XLM-RoBERTa share a core Transformer encoder architecture, differing primarily in their tokenisation algorithms

(WordPiece vs SentencePiece) and the designation of the initial sequence token ([CLS] for BERT, <s> for XLM-RoBERTa).

For a given input text T, the classification process is formalised in Algorithm 1.

Algorithm 1. Transformer-based Text-to-Diagnosis Classification

Input: Raw text sequence T, True label $y \in \{0, 1\}$, Pre-trained Transformer Model $\mathcal{M}$

Output: Predicted class probability $\hat{y}$ and optimised model weights

Hyperparameters: Learning rate $\eta = 5 \times 10^{\{-5\}}$, Batch size $B = 16$, Max length $L = 128$, Dropout $p = 0.3$

1. Text Preprocessing $T_{clean} \leftarrow Remove\_Noise(T)$. It removes URLs, emojis, and normalises Unicode.
2. Tokenisation $E \leftarrow Tokenizer(T_{clean}, L)$. It uses WordPiece for BERT or SentencePiece for XLM-RoBERTa.

$$E = [t_{CLS}, t_1, t_2, \dots, t_k, t_{SEP}, PAD, \dots, PAD]$$

3. Transformer Forward Pass. Let E_{emb} be the sum of token, position, and segment embeddings.

$H \leftarrow TransformerEncoder\ (E_{emb})$,

where $H \in \mathcal{R}^{L \times d_{model}}$ is the sequence of hidden states.

4. CLS Vector Aggregation. Extract the hidden state of the first token (the classification token), which aggregates the global semantic representation of the sequence $h_{CLS} \leftarrow H[0]$, where $h_{CLS} \in \mathcal{R}^{d\_model}$.
5. Classification Head (Dropout & Linear Layer). Apply dropout to mitigate overfitting:

$$h'_{CLS} \leftarrow Dropout(h_{CLS}, p)$$

Compute logits Z using the classification weight matrix W and bias b:

$$Z \leftarrow W \cdot h'_{CLS} + b$$

6. Probability Distribution. Apply the Softmax function to obtain the probability for the PTSD class:

$$\hat{y} \leftarrow \frac{e^{Z_1}}{e^{Z_0} + e^{Z_1}}$$

7. Loss Calculation (Training phase). Compute the Cross-Entropy Loss for the batch:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{B} (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

8. Optimisation Step and compute gradients $\nabla \mathcal{L}_{CE}$, apply gradient clipping:

$$Clip\_Norm(\nabla \mathcal{L}_{CE}, 1.0)$$

Update weights using AdamW optimiser $W \leftarrow W - \eta \cdot AdamW\_Step\ (\nabla \mathcal{L}_{CE})$.

An essential element of the architecture is the final classification layer, which is implemented directly above the initial state of the special token [CLS]. This token accumulates an aggregate representation of the entire sequence fed to the input, and it is this vector that the linear layer uses to decide whether the text belongs to the "PTSD" or "Non-PTSD" class. This approach allows you to leverage the powerful capabilities of pre-trained transformers by fine-tuning only on a specific dataset, significantly increasing efficiency compared to training models from scratch. Each of the 12 layers of the model contains 12 attention heads that analyse various aspects of the linguistic structure in parallel, from basic grammar at lower levels to complex semantic concepts and emotional states at higher levels. It creates a solid technological foundation that allows you to identify implicit markers of traumatic stress, which are often overlooked by classical machine learning algorithms. The whole process is implemented using the Hugging Face ecosystem and the Transformers library, which provides high reproducibility of experiments and the flexibility to configure training hyperparameters to achieve an accuracy of 84% or higher, as reported in comparative studies on similar datasets.

Research methodology

Fig. 3 The activity diagram shows the architectural logic and sequence of technological operations that form the basis of the developed system for automated screening of signs of PTSD. The workflow is initiated by the input reception stage, after which the architecture involves branching into parallel threads to ensure cross-linguistic validity of the study. One branch focuses on working with an authentic English-language corpus, while the parallel segment prepares a Ukrainian-language sample. To ensure high-quality cross-lingual adaptation of the dataset, a hybrid translation approach was implemented. The Google Translate API served as the primary translation tool for the majority of the corpus, efficiently processing standard-length self-reported symptoms and short statements. However, the criterion of "structural complexity", which necessitated the use of a Large Language Model (LLM), was explicitly defined by text length and contextual continuity. Specifically, texts that exceeded the translation API's character limit or long, multi-paragraph personal narratives, where standard APIs often truncate output or lose contextual tracking, were routed to an LLM. For these extended narratives, we utilised the ChatGPT-4o model (OpenAI). This specific version was selected for its advanced context window capabilities, ensuring that lengthy psychological texts were translated as a single cohesive unit without losing the emotional nuance and diagnostic integrity of the original English narrative.

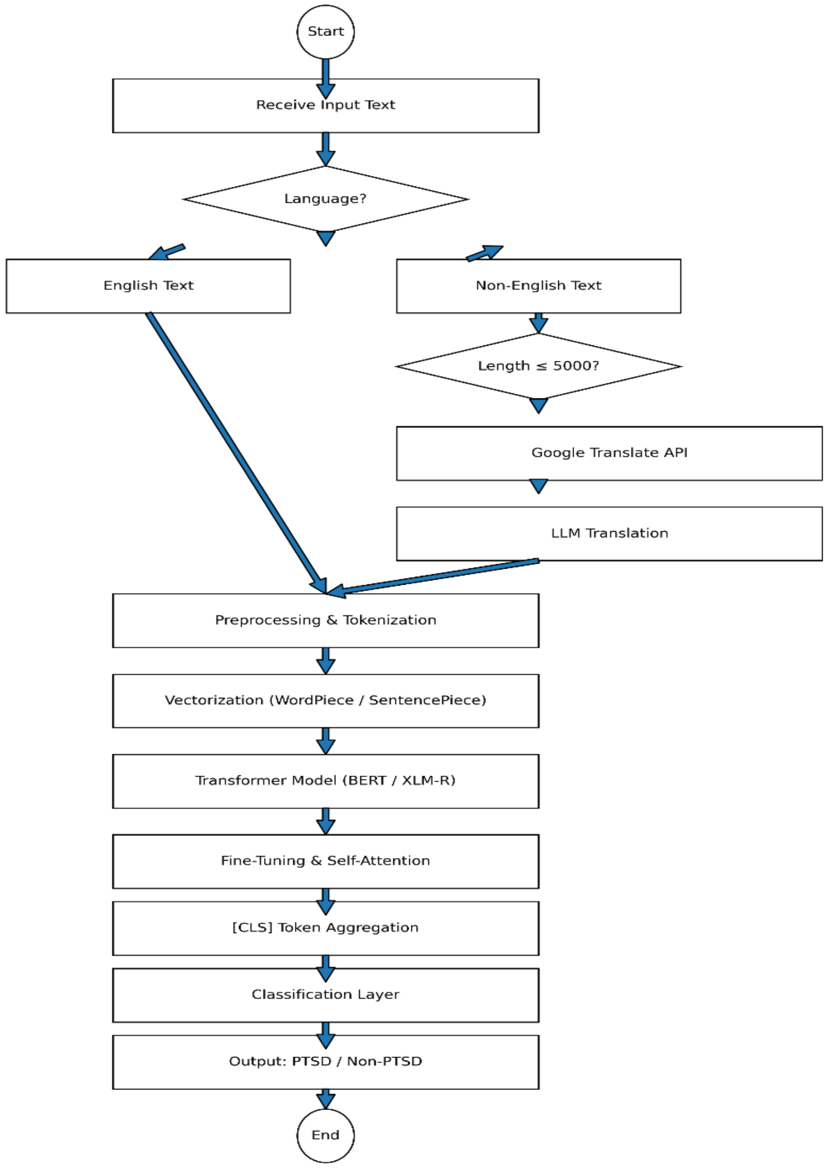

Fig. 3. UML-Activity Diagram

After synchronising the original and adapted data flows, the pipeline proceeds to the in-depth linguistic processing stage, where tokenisation and embedding are integrated. At this level, the system transforms unstructured text sequences into multidimensional vector representations, applying embedding algorithms that correctly vectorise even specific morphology and terminology. The central stage of the diagram is the model's direct training, based on the transformer architecture, using a self-attention mechanism to identify hidden markers of traumatic stress. The final phase of the activity focuses on obtaining and interpreting the results, in which the aggregated contextual representations pass through the classification layer to yield an objective conclusion on the presence of signs of the disorder. Such a structured sequence of actions ensures high reproducibility of the experiment and enables you to effectively combine machine translation capabilities with advanced deep learning methods.

The English-language bert-base-uncased model was trained using the AdamW optimiser with an initial learning rate of $5\times 10^{-5}$and a weight decay coefficient of 0.01. The process consisted of 3 epochs, using a linear planner and a dropout rate of 0.3 to prevent overtraining. For the Ukrainian-language segment, the xlm-roberta-base model lasted 5 epochs with a learning rate of 3×10^{-5}, a packet size of 16, and a sequence length limit of 128 tokens.

The choice of these specific training hyperparameters was not arbitrary; it was strictly grounded in established best practices for fine-tuning Transformer models on classification tasks and tailored to the specifics of our dataset to ensure reproducibility. A maximum sequence length of 128 tokens was determined to be optimal for capturing the core semantic markers of short self-reported PTSD narratives while maintaining computational efficiency The batch size of 16 and the selected learning rates fall within the optimal stability range recommended for BERT architectures. Specifically, the lower learning rate ($3 \times 10^{-5}$and a higher number of epochs (5) for XLM-

RoBERTa were intentionally selected to allow the cross-lingual model to adequately converge on the structurally noisier, machine-translated Ukrainian text. In contrast, the native English BERT converges faster (3 epochs). Finally, a dropout rate of 0.3 and a weight decay of 0.01 were applied as strict regularisation measures to mitigate the risk of overfitting, which is a common vulnerability when training highly parameterised models on relatively small, domain-specific psychological corpora.

Research results

This section presents the comparative performance metrics of the baseline English-language model (BERT) and the cross-lingual Ukrainian-language model (XLM-RoBERTa). The evaluation is based on confusion matrices, classification reports, the area under the Receiver Operating Characteristic (ROC) curve (ROC-AUC), Precision-Recall curves, and statistical significance testing. The Confusion Matrix is used to evaluate the distribution of correct and incorrect predictions, which is critical in medical screening, where false negatives carry significant clinical consequences.

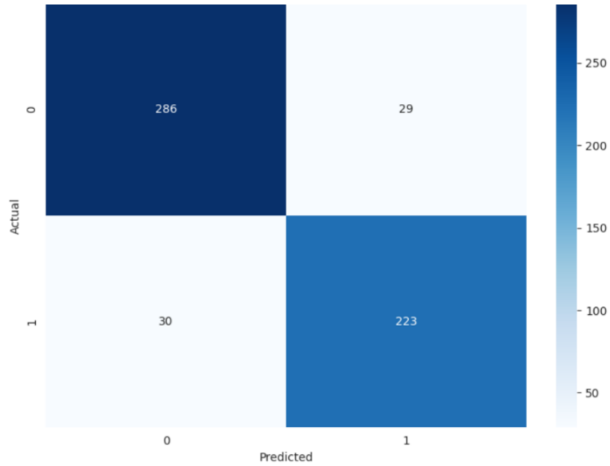

Fig. 4. Confusion matrix of the model for English

As shown in Fig. 4, the original BERT-based model demonstrates a balanced set of predictions. The algorithm successfully identified 223 cases of PTSD (True Positives) and 286 control texts (True Negatives). The type I and type II errors are symmetrically distributed: 29 false positives and 30 false negatives.

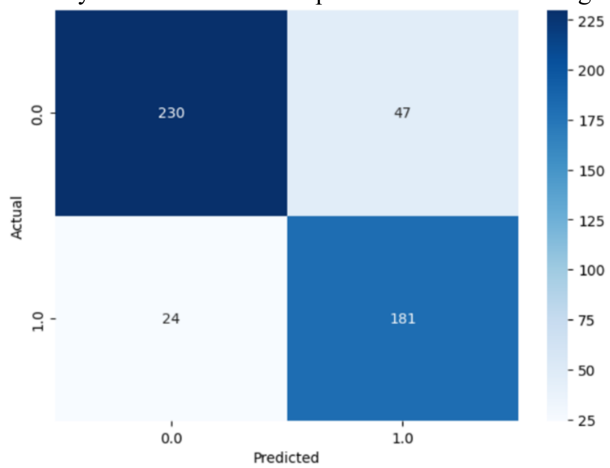

Fig. 5. Confusion matrix of the model for the Ukrainian language

As shown in Fig. 4, the original BERT-based model demonstrates a balanced set of predictions. The algorithm successfully identified 223 cases of PTSD (True Positives) and 286 control texts (True Negatives). The type I and type II errors are symmetrically distributed: 29 false positives and 30 false negatives. To assess the classifiers' ability to distinguish between the PTSD and control groups across various probability thresholds, ROC and Precision-Recall (PR) curves were generated.

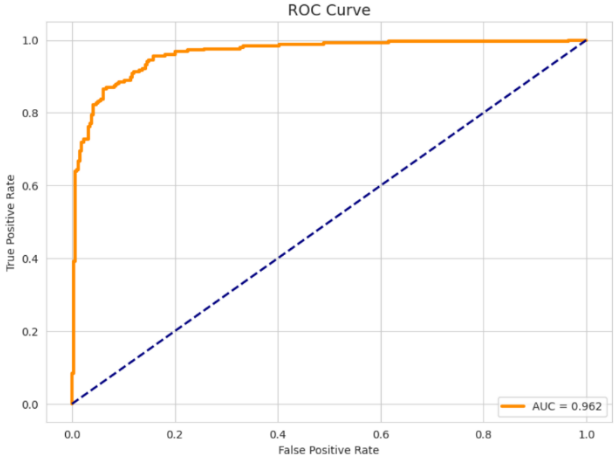

Fig. 6. ROC Curve Model Graph for English

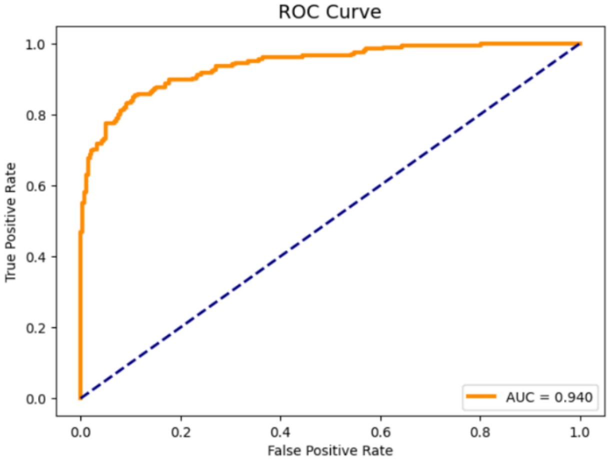

Fig. 7. Graph of the ROC-curve model for the Ukrainian language

The ROC curve for the English-language model (Fig. 6) yields an AUC of 0.962. For the Ukrainian-language model (Fig. 7), the AUC is 0.940, indicating a minimal degradation of 2.3% in the classifier's overall resolution.

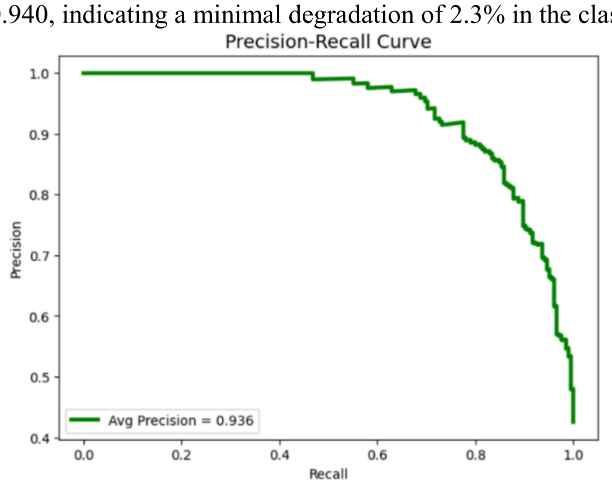

Fig. 8. Precision-Recall curve model for English

The Precision-Recall curves evaluate the classification quality specifically for the target PTSD class. The English model (Fig. 8) achieved an Average Precision of 0.953. The Ukrainian-language model (Fig. 9) maintained a highly competitive Average Precision of 0.936, displaying a slight decrease in precision primarily at the highest recall thresholds.

The English BERT baseline achieved an Accuracy of 0.90, with Precision, Recall, and F1-Score all stabilising at 0.88. Following cross-lingual adaptation, the Ukrainian XLM-RoBERTa model achieved an Accuracy of 0.85. While Recall remained perfectly stable at 0.88 across both language segments, precision experienced a drop from 0.88 to 0.79, subsequently reducing the F1-Score to 0.84.

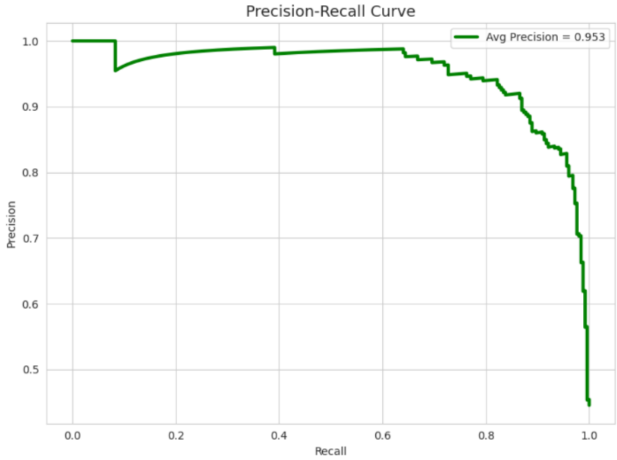

Fig. 9. Precision-Recall model curve for the Ukrainian language

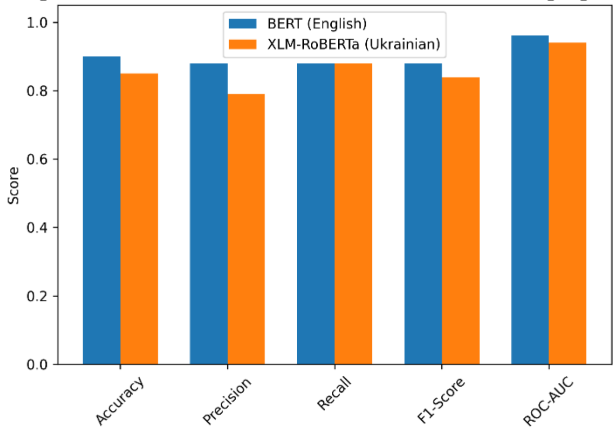

Fig. 10. Comparison of Model Performance Metrics

To ensure that the observed difference in accuracy (0.90 for English BERT vs 0.85 for Ukrainian XLM-RoBERTa) was not merely due to data-splitting variance, we conducted a statistical significance test. Given that both models were evaluated on the same paired test instances (yielding nominal correct/incorrect outcomes), McNemar's test was the most appropriate statistical method for comparing the classifiers. The contingency matrix for the predictions is presented in Table 1.

Table 1

Results of McNemar's test

	Correct Ukrainian prediction	Wrong Ukrainian prediction
Correct English prediction	463	46
Wrong English prediction	21	38
Statistic (Chi-squared): 8.597		
p-value: 3.36714e-03		

The analysis revealed 46 instances in which the English model was correct while the Ukrainian model failed, and 21 instances in which the Ukrainian model succeeded despite the English model failing. McNemar's test produced a test statistic of χ^2 = 8.597 with a corresponding p-value of 0.00336. Since the p-value is strictly less than the conventional alpha level ($p < 0.05$), we reject the null hypothesis. It confirms that the performance degradation following machine translation is statistically significant. However, while the translation noise introduces a measurable drop in performance, the retained accuracy of 0.85 remains highly acceptable and establishes a robust baseline for low-resource clinical NLP.

Table 2

Comparison of model classification reports

Metrics	BERT (English)	XLM-RoBERTa (Ukrainian)
Accuracy	0.90	0.85
Precision	0.88	0.79
Recall	0.88	0.88
F1-Score	0.88	0.84
ROC-AUC	0.962	0.940

Precision drops from 0.88 (English) to 0.79 (Ukrainian). It reflects an increase in false positives in the Ukrainian model, meaning the multilingual model tends to slightly over-diagnose PTSD after translation. Both models maintain identical Recall = 0.88. It is a critical finding: the Ukrainian model preserves sensitivity to PTSD markers despite machine translation. For medical screening systems, maintaining high Recall is more vital than maximising precision. The F1-score decreases slightly from 0.88 to 0.84, reflecting the trade-off caused by reduced precision. However, the score remains high and clinically acceptable. The ROC-AUC decreases minimally from 0.962 to 0.940 (only 2.3% degradation). Both models fall into the "excellent classifier" category according to standard medical evaluation scales.

Discussion

The quantitative performance indicators obtained from the BERT and XLM-RoBERTa architectures provide crucial insights into the viability of cross-lingual mental health screening. The English-language model, serving as a baseline, achieved an Accuracy of 0.90, which is highly consistent with recent NLP research on identifying psychological markers in native-language environments. The stability of its Precision, Recall, and F1-Score (all at 0.88) confirms the robust differential diagnostic capability of the BERT architecture when processing structurally coherent, untranslated clinical narratives.

The most significant finding of our comparative analysis is the impressive stability of the Recall metric for the Ukrainian language model, which remained at 0.88 and is completely identical to the English baseline. This result directly confirms that key semantic and emotional markers of PTSD, such as descriptions of trauma triggers, obsessive memories, and specific phobias, possess strong cross-lingual resilience and are successfully preserved during machine translation. Maintaining a high level of Recall is an absolute priority for primary psychological screening systems, as the fundamental clinical goal is to minimise false-negative results and prevent missing patients who urgently require professional help.

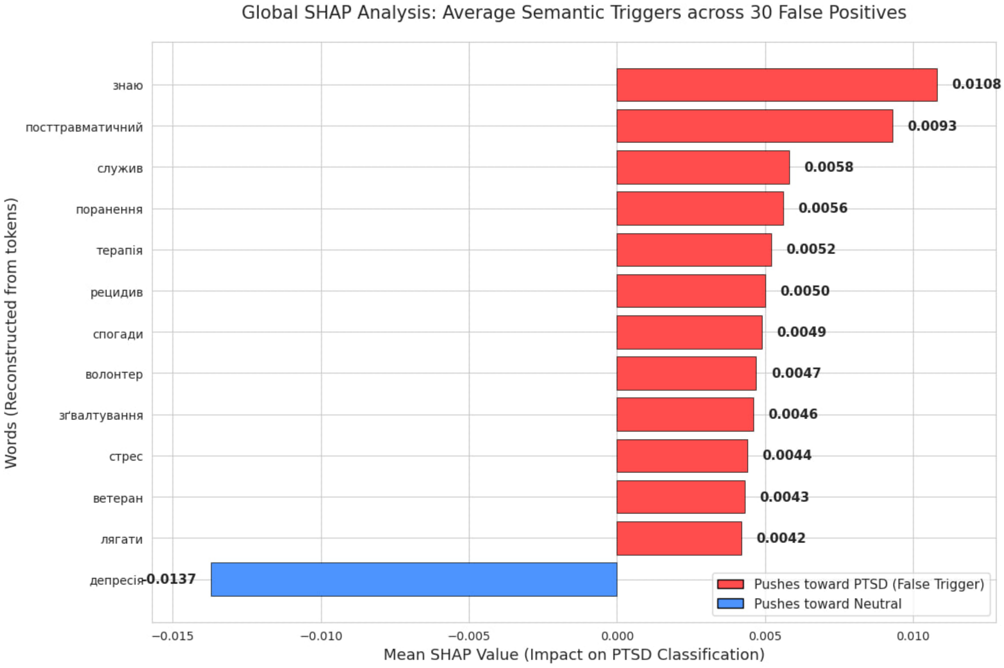

Fig. 11. Global SHAP analysis illustrating average semantic triggers reconstructed from sub-word tokens across false-positive predictions

It is important to note that this SHAP visualisation represents a heuristic approximation. Because the XLM-RoBERTa model processes text using sub-word tokenisation (SentencePiece), the SHAP values were originally calculated at the token level and subsequently aggregated to reconstruct whole words. Therefore, while this analysis highly accurately reflects semantic trends, the displayed terms represent clusters of impactful tokens rather than absolute word-level certainties.

As illustrated in Fig. 11, the global SHAP analysis of false positives reveals that the Ukrainian model assigns disproportionately high positive weights (pushing the prediction toward a PTSD diagnosis) to specific lexical triggers. The highest impact is observed for the unexpectedly common verb "знаю" (I know), followed by highly contextual trauma-related terms such as "посттравматичний" (post-traumatic), "служив" (served), "поранення" (injury), and "волонтер" (volunteer).

In the original English narratives, the strict syntax usually clarifies whether such terms relate to the patient's direct traumatic experience or are merely observational. However, in the translated Ukrainian texts, the accumulation of these high-impact tokens often overwhelms the broader context. For instance, if a user writes a general statement like, "Я знаю одного волонтера, який служив і має посттравматичний стрес" (I know a volunteer who served and has post-traumatic stress), the model hyper-focuses on the localised semantic weight of the tokens "знаю", "волонтер", "служив", and "стрес". Consequently, it misleads the classifier into a positive PTSD prediction, despite the subject not describing their own trauma.

Conversely, the chart demonstrates that the model correctly identifies the word "депресія" (depression) as a strong indicator of the control class (producing a significant negative SHAP value of -0.0137). It proves the model actively attempts to differentiate comorbid depression from PTSD. Yet, when depression-related vocabulary is mixed with morphologically ambiguous trauma-associated tokens in a translated sentence, the classifier tends to lean towards a cautious over-diagnosis. This SHAP analysis definitively confirms that the observed drop in precision is heavily driven by the cross-lingual transfer of lexical noise and morphological ambiguity, rather than a fundamental flaw in the transformer's architecture.

To further illustrate the impact of machine translation on classification accuracy, a qualitative analysis of specific misclassified narratives was conducted. Two particularly striking examples demonstrate how translation artefacts can completely distort clinical semantics and mislead the cross-lingual model. In the first instance, a simple typographical error in the source text fundamentally shifted the translation's domain. The original English narrative contained, "My fiancé suffers from suspected PTSD," where the user clearly intended to write the word "fiancé". The native English BERT model easily inferred the human context from the surrounding narrative and correctly predicted PTSD with a probability of 0.99. However, the machine translation API processed the typo literally and produced the Ukrainian phrase "Мої фінанси страждають від підозри на посттравматичний стресовий розлад" (My finances suffer from suspected PTSD). As a result, the XLM-RoBERTa model interpreted the text as an economic metaphor or spam, and the predicted probability of PTSD drastically dropped to 0.01.

A second critical failure mode involves the destruction of psychological semantics through literal translation (calque). In one narrative, a user described a well-known PTSD symptom, olfactory flashbacks, by asking, "Does anyone 'smell' things that are in pictures?" The translation engine rendered this literally into Ukrainian as "Хтось «пахне» речами, які зображені на фотографіях?" (Does anyone smell like the things depicted in the pictures?) This translation transformed a profound psychological trauma trigger into a nonsensical physical description, which completely erased the diagnostic marker. The XLM-RoBERTa model, therefore, treated this severe translation failure as random semantic noise and failed to flag the text. These examples clearly highlight how the literal translation of source-language typos and idioms in automated pipelines can critically mask trauma markers and demonstrate the necessity of robust, context-aware adaptation in cross-lingual clinical NLP.

While the proposed cross-lingual approach demonstrates high screening efficacy, several limitations must be acknowledged. Primarily, the reliance on machine-translated English datasets means the XLM-RoBERTa model is fundamentally trained on Western cultural expressions of trauma. Consequently, the automated translation process may strip away or fail to capture the unique cultural contexts, idioms, and specific psychological vernacular of trauma inherent to Ukrainian society. For instance, localised coping mechanisms or specific stress markers arising from the current wartime environment may not have direct equivalents in the original English source data. As a result, while the model successfully identifies universal PTSD symptoms, it may lack sensitivity to highly nuanced, native Ukrainian narratives. Future research must prioritise the collection and clinical annotation of native, untranslated Ukrainian text corpora to close this cultural gap and further refine diagnostic precision.

Conclusions

Within the framework of the conducted scientific and practical research, a comprehensive methodology for automated classification of text data for identifying signs of post-traumatic stress disorder (PTSD) has been developed and comprehensively verified. The primary focus of the work is the comparative analysis of the effectiveness of modern transformer architectures, in particular the distilbert-base-uncased and xlm-roberta-base models, in the context of working with original English-language texts and an adapted Ukrainian-language segment. The results of the experiments convincingly demonstrate that deep learning technologies can successfully recognise complex linguistic and emotional markers of mental disorders, maintaining high predictive value even when automatic machine translation is used to prepare educational materials. It opens significant opportunities for the rapid deployment of psychological monitoring systems in the Ukrainian digital environment, where there is an acute shortage of verified local datasets. The representativeness of the results is based on a balanced data corpus of 4822 records, including 2042 texts from the target class that cover a wide range of verbal expression of traumatic experiences. A crucial methodological advantage was the formation of a control group of 2780 entries based on a combined principle that included not only neutral content but also descriptions of other psychological states. This approach enabled the models to go beyond superficial sentiment analysis and to isolate specific clinical markers of PTSD, a capability rigorously validated using Explainable AI (SHAP) rather than simple frequency-based word clouds. The semantic analysis demonstrated a substantial, yet imperfect, degree of cross-lingual isomorphism. While the core English trauma indicators were successfully transferred to maintain high diagnostic Recall, their Ukrainian equivalents (such as «знаю», «посттравматичний», «волонтер») often carried morphological ambiguity. Consequently, while the fundamental diagnostic value of the narratives is strongly preserved after linguistic adaptation, the process is not entirely lossless; translation-induced lexical noise necessitates a careful trade-off between sensitivity and precision in the target language. Quantitative performance indicators demonstrated the high stability of the selected architectures.

The English-language model achieved a reference accuracy of 0.90 with a balanced F1-score of 0.88. The most significant result is the performance of the Ukrainian-language version based on xlm-roberta-base, which achieved an overall accuracy of 0.85 while maintaining the same level of completeness (Recall) as the original (0.88). It suggests that, despite some degradation in specificity and an increase in false positives due to machine translation, the model retains its ability to detect genuine cases of the disorder. In the context of primary medical screening, this priority of completeness over accuracy is critically important, since the cost of a missed case (false negative) is significantly higher than the cost of false overdiagnosis. The analysis of ROC curves confirms the excellent predictive quality of both models: the AUC was 0.962 for English and 0.940 for Ukrainian. The minimal 2.3% reduction in AUC when transitioning between languages is a key proof of the resilience of transformer architectures to interlanguage conversion. It supports the hypothesis that neural networks successfully recognise the invariant semantic structures of traumatic experiences, regardless of the morphological complexity of the language used to express them. The revealed shift in the Ukrainian-language model towards "false positive" errors is a natural consequence of automated adaptation of professional vocabulary. Still, the overall result remains sufficiently practical for implementation. The practical significance of the work lies in the possibility of creating intelligent mental health support systems that can operate in real time. The developed pipeline, which includes a multi-level translation strategy using the Google Translate API for short entries and an LLM for complex structures, provides high-speed, stable data processing. Such tools can serve as the basis for digital platforms to support military personnel and civilians affected by armed aggression. Further development of this study involves validating the current model on purely native Ukrainian test sets, training future models on non-translated corpora, and expanding upon the Explainable Artificial Intelligence (XAI) foundations established in our current analysis. Since PTSD is a severe medical diagnosis, the implementation of "black-box" machine learning models remains insufficient for real-world clinical practice, where ethical standards and the cost of diagnostic errors are extremely high. While our initial SHAP analysis successfully identified global translation artefacts and morphological noise, future research will focus on the deep, instance-level integration of dynamic XAI methods, such as Local Interpretable Model-agnostic Explanations (LIME) and token-level attention-weight visualisation.

The specific goal of this future research is to develop real-time, interpretable clinical dashboards that generate visual heatmaps of text segments. It will allow mental health professionals to transparently verify exactly which specific word-markers, emotional triggers, and syntactic structures drive individual positive predictions, explicitly differentiating genuine clinical markers from hyper-medicalised everyday slang. Ultimately, advancing from static XAI evaluation to interactive interpretability is crucial for building deep clinical trust, bridging the gap between automated NLP screening and expert psychiatric evaluation, and ensuring that digital mental health tools serve as reliable, explainable assistants to medical practitioners, aligned with established guidelines such as the DSM-5 criteria.

ADDITIONAL INFORMATION

AUTHOR CONTRIBUTIONS

Conceptualisation, V.V.; methodology, A.F. and V.V.; software, A.F.; validation, A.F., V.V. and L.C.; formal analysis, A.F.; data curation, L.C. and A.F.; writing—original draft preparation, A.F.; writing—review and editing, V.V. and L.C.; visualisation, A.F.; supervision, V.V.; project administration, V.V. All authors have read and agreed to the published version of the manuscript.

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

In preparing this work, the authors used the Google Translate API and ChatGPT-4o specifically for the cross-lingual adaptation of the English-language clinical dataset into Ukrainian, as detailed in the Research Methodology section. Additionally, ChatGPT-4o and DeepL were used for partial assistance with the English translation of the manuscript, grammar and spelling checks, and stylistic rephrasing of the academic text. After using these tools and services, the authors thoroughly reviewed, critically evaluated, and manually edited all content to ensure scientific accuracy and proper academic tone. The authors take full responsibility for the final content, originality, and integrity of this publication.

REFERENCES

1. Korbylo, R. Ya., & Hrytsiuk, Yu. I. (2025). Systema otsiniuvannia psykholohichnoho stanu viiskovosluzhbovtsiv metodamy oblroblennia pryrodnoi movy system for assessing the psychological state of military personnel using natural language processing methods. Scientific Bulletin of the National Technical University of Ukraine, 35(6). https://doi.org/10.36930/40350617

2. Sawalha, J., Yousefnezhad, M., Shah, Z., & Greiner, R. (2022). Detecting the presence of PTSD using sentiment analysis from text data. *Frontiers in Psychiatry, 12*, Article 811392. https://doi.org/10.3389/fpsyt.2021.811392

3. Blekic, W., D'Hondt, F., Shalev, A. Y., & Schultebraucks, K. (2025). A systematic review of machine learning findings in PTSD and their relationships with theoretical models. *Nature Mental*

Health, 3, 139–158. https://doi.org/10.1038/s44220-024-00365-4

4. Yuan, Y., Liu, Z., Miao, W., & Tian, X. (2024). Automatic screening for post-traumatic stress disorder in early adolescents following the Ya'an earthquake using text mining techniques. *Frontiers in Psychiatry, 15*, Article 1439720. https://doi.org/10.3389/fpsyt.2024.1439720

5. Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. NPJ digital medicine, 5(1), 46. https://doi.org/10.1038/s41746-022-00589-7

6. Shevchuk, A. V., Yanovets, A. I., & Sternichuk, V. B. (2025). Neural network models in the translation of German-language and English-language texts. *Academic Studies. Series "Humanities"*, (3), 102–107. https://doi.org/10.52726/as.humanities/2025.3.12

7. Olkhovska, A., & Andrushchenko, Y. (2024). Experimental study of the influence of machine translation on the quality of translation. *Bulletin of V. N. Karazin Kharkiv National University. Series "Foreign Philology"*, (99), 76–84. https://doi.org/10.26565/2786-5312-2024-99-09

8. Wu, Y., Chen, J., Mao, K., & Zhang, Y. (2023). Automatic post-traumatic stress disorder diagnosis via clinical transcripts: A novel text augmentation with large language models. *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1957–1962. https://doi.org/10.1109/BioCAS58349.2023.10388714

9. Chen, F., Ben-Zeev, D., Sparks, G., & et al. (2025). *Detecting PTSD in clinical interviews: A comparative analysis of NLP methods and large language models*. arXiv. https://doi.org/10.48550/arXiv.2504.01216

10. Sharma, C., Vaid, A., & Sharma, K. (2024). Natural language processing in SAP: Enhancing user interactions and data analysis through NLP. *International Journal of Computer Science and Engineering Technology (IJCSET), 2*, 58–76. https://doi.org/10.5281/zenodo.14064999.

11. Zhurko, D., & Bilous, I. (2025). Use of models of word insertion in the processing of natural language. *Technical Sciences and Technologies*, (1), 151–160. https://doi.org/10.25140/2411-5363-2025-1(39)-151-160

12. Gardazi, N. M., Daud, A., Malik, M. K., & et al. (2025). BERT applications in natural language processing: A review. *Artificial Intelligence Review, 58*(166). https://doi.org/10.1007/s10462-025-11162-5

13. Chernyak, I. O. (2024). Study of the possibilities of the DistilBERT model for further use in the system of automation of electronic document management. *Scientific Notes of the Vernadsky TNU. Series: Technical Sciences, 35*(74), 47. https://doi.org/10.32782/2663-5941/2024.5.1/47

14. He, Q., Veldkamp, B. P., Glas, C. A. W., & de Vries, T. (2017). Automated Assessment of Patients' Self-Narratives for Post-traumatic Stress Disorder Screening Using Natural Language Processing and Text Mining. *Assessment*, *24*(2), 157-172. https://doi.org/10.1177/1073191115602551

15. He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for post-traumatic stress disorder using verbal features in self narratives: a text mining approach. *Psychiatry research*, *198*(3), 441–447. https://doi.org/10.1016/j.psychres.2012.01.032

16. Miranda, O., Kiehl, S.M., Qi, X., Brannock, M.D., Kosten, T., Ryan, N.D., Kirisci, L., Wang, Y., & Wang, L. (2024). Enhancing post-traumatic stress disorder patient assessment: leveraging natural language processing for research of domain criteria identification using electronic medical records. BMC Medical Informatics and Decision Making, 24. https://doi.org/10.1186/s12911-024-02554-8

17. Lorenzoni, G., Tavares, K., Nasimento, N., & et al. (2025). Depression and PTSD detection: An assessment of machine learning and NLP techniques. *Journal of Medical Systems, 49*, Article 102. https://doi.org/10.1371/journal.pone.0322299

Андрій ФЕДОРИЧКО, Вікторія ВИСОЦЬКА, Любомир ЧИРУН
Національний університет «Львівська політехніка»

МІЖЛІНГВАЛЬНИЙ ТРАНСФОРМАТОРНИЙ СКРИНІНГ ПОСТТРАВМАТИЧНОГО СТРЕСОВОГО РОЗЛАДУ НА ОСНОВІ ПОРІВНЯЛЬНОГО АНАЛІЗУ BERT ТА XLM-ROBERTA З АДАПТАЦІЄЮ МАШИННОГО ПЕРЕКЛАДУ ДЛЯ УКРАЇНСЬКОЇ МОВИ

У статті представлено комплексне дослідження автоматизованого скринінгу посттравматичного стресового розладу (ПТСР) за допомогою трансформаторних моделей обробки природної мови в міжмовному середовищі. Метою дослідження є оцінка можливості розгортання інтелектуальної системи виявлення ПТСР українською мовою за умов обмежених локалізованих навчальних даних. Збалансований корпус із 4822 текстових записів був створений шляхом агрегації загальнодоступних наборів даних, пов'язаних з ПТСР, включаючи 2042 тексти з позитивним ПТСР та 2780 контрольних текстів, що представляють нейтральний контент та інші психологічні стани. У дослідженні порівнюється продуктивність англомовної моделі BERT (bert-base-uncased) та багатомовної моделі XLM-RoBERTa (xlm-roberta-base), застосованої до україномовного корпусу, згенерованого за допомогою машинного перекладу з використанням API Google Translate та великих мовних моделей для *складних структур. Візуалізація хмари слів та семантичний аналіз підтвердили збереження основних психологічних маркерів під час перекладу. Експериментальні результати демонструють високу прогностичну ефективність для обох архітектур. Англомовна модель досягла точності 0,90 та ROC-AUC 0,962, тоді як україномовна модель досягла точності 0,85 та ROC-AUC 0,940, що суттєво перевершує існуючі українські моделі багатокласового виявлення стресу (Accuracy ~0.45) та перевищує стандартні багатомовні показники у сфері психічного здоров'я (0.78–0.82), встановлюючи надійний передовий базовий рівень для української обробки природної мови (NLP) у клінічній сфері[1]. Важливо, що показник Recall (повторне сприйняття) залишився ідентичним (0,88) в обох мовних налаштуваннях, що свідчить про високу чутливість до маркерів ПТСР, незважаючи на лексичний шум, викликаний перекладом. Мінімальна деградація AUC (2,3%) підтверджує стійкість архітектур трансформаторів до міжмовної адаптації. Результати підтверджують життєздатність поєднання машинного перекладу з багатомовними трансформаторами для швидкого розгортання систем скринінгу психічного здоров'я в середовищах з обмеженими мовними ресурсами. Запропонований конвеєр дозволяє масштабований та економічно ефективний цифровий моніторинг ПТСР, зберігаючи при цьому клінічно значущу діагностичну чутливість.*

Ключові слова: посттравматичний стресовий розлад (ПТСР); обробка природної мови (НЛП); трансформерні моделі; BERT; XLM-RoBERTa; машинний переклад; міжмовний переказ; класифікація текстів; скринінг психічного здоров'я; ROC-AUC; глибинне навчання; корпус української мови.