DOI: https://doi.org/10.15276/hait.08.2025.12
UDC 004.93

A model for constructing neural network systems for recognizing emotions of text fragments

Igor A. Tereikovskyi[1)]
ORCID: https://orcid.org/0000-0003-4621-9668; terejkowski@ukr.net. Scopus Author ID: 57195940293
Oleksandr S. Korovii[1)]
ORCID: https://orcid.org/0009-0002-2835-9173; zeusfsxtmp@gmail.com Scopus Author ID: 57644238900
[1)] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, Beresteiskyi Ave. Kyiv, 03056, Ukraine

ABSTRACT

Emotion Recognition in text is a crucial task in Natural Language Processing, particularly relevant given the exponential growth of textual data from social media and voice interfaces. However, developing effective emotion recognition systems for low-resource languages, such as Ukrainian, faces significant challenges, including linguistic informality, dialectal variations, and cultural specificities. This paper introduces a modular model (framework) for developing neural network-based tools for recognizing emotions in Ukrainian text fragments. The model encompasses a comprehensive data preprocessing pipeline, flexible architectural choices (including approaches based on Word to Vector, Long Short-Term Memory, and Transformers), and rigorous validation using standard metrics and interpretability methods. As part of an experimental study, two prototypes were implemented and compared: a lightweight classifier based on FastText and a more powerful classifier based on pretrained RoBERTa-base, both trained to recognize seven basic emotions. The results demonstrate that RoBERTa-base achieves high accuracy, significantly outperforming FastText and a baseline translation-based approach, yet it demands substantially more computational resources for inference. The study underscores the importance of creating Ukrainian-language corpora to enhance recognition capabilities and highlights the critical trade-off between accuracy and efficiency. It provides practical recommendations for model selection based on resource constraints and performance requirements for emotion analysis tasks in the Ukrainian language.

Keywords: Emotion recognition; natural language processing; neural networks based tools; modular model; low-resource languages

For citation: **Tereikovskyi I. A., Korovii O. S. "A model for constructing neural network systems for recognizing emotions of text fragments". *Herald of Advanced Information Technology*. 2025; Vol. 8 No. 2: 197–208. DOI: https://doi.org/10.15276/hait.08.2025.12**

INTRODUCTION

Admits the exponential growth in the volume of unstructured textual data, the scientific and practical task of developing effective models for its automated processing in universal computing systems becomes increasingly relevant. The aspect of recognizing the emotional tone of text fragments gains particular significance, as social media platforms have transformed into a key environment for interpersonal communication, generating vast amounts of textual content that reflect a wide spectrum of users' emotional states and reactions. Emotion analysis in social media texts presents unique challenges, especially for languages with limited natural language processing resources, such as Ukrainian.

Traditional approaches to emotion detection often face difficulties due to the informal nature of social media content, dialectal variations, and the complex interplay of emotions expressed in these texts.

Furthermore, the task of analyzing text obtained from speech recognition, which often occurs on low-resource computing devices, becomes particularly pressing. This is driven by the rapid development of voice interaction systems and the need to promptly determine the emotional context in real-time, even under conditions of limited computational capabilities.

The rapid growth in social media usage in Ukraine has created an unprecedented opportunity for understanding public sentiment, emotional reactions to events, and social dynamics, especially during periods of crisis. However, existing emotion detection systems often struggle to accurately process Ukrainian-language content for several reasons: the linguistic peculiarities of Ukrainian communication on social media, the presence of mixed-language content, and the unique cultural context that influences emotional expression.

Emotion detection systems enable valuable insights in finance, marketing, political analysis, and social-mood tracking, yet this study purposely uses a domain-neutral corpus to provide a broad benchmark that later domain-specific work can build upon.

Therefore, enhancing the efficiency of systems for recognizing emotional tone in texts is a pertinent

© Tereikovskyi I., Korovii O., 2025

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0 /deed.uk)

scientific and practical task. This directly contributes to the improvement of computer system interfaces and aligns with the strategic directions for the innovative development of modern information and communication technologies.

1. ANALYSIS OF LITERARY DATA

Emotion recognition in textual data is a priority research task in computational linguistics and Natural Language Processing (NLP), whose methodological apparatus has undergone significant transformation. Recent studies demonstrate a clear shift from traditional lexicon-based methods to methods based on deep learning, which provide enhanced accuracy and the ability to detect subtle emotional nuances in texts [1], [2].

Early approaches predominantly relied on lexicon-based methods, using predefined dictionaries such as SentiWordNet and VADER to identify emotionally charged words [1], [3]. These approaches formed the basis for the first emotion detection systems, employing resources like the NRC Word-Emotion Lexicon for basic emotion and sentiment classification [2]. However, these methods often struggled to account for context-dependent emotions and subtle linguistic variations. Machine learning methods, particularly Support Vector Machines (SVM), Random Forests, and Naïve Bayes classifiers, offered improvements by learning from labeled data [1], [2], [3], but remained limited in their ability to capture complex emotional expressions.

The advent of deep learning architectures marked a significant breakthrough in this field [2], [4]. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks with long-term memory, such as Long Short-Term Memory (LSTM), demonstrated superior performance in capturing sequential dependencies and contextual information in texts [1], [3]. Bidirectional LSTMs (Bi-LSTMs) further enhanced this capability by processing text in both forward and backward directions, providing a richer contextual analysis [4], [6]. Research indicates that these neural architectures significantly outperform traditional machine learning approaches, especially in handling complex emotional expressions and context-dependent situations [3], [4].

The most significant breakthrough in the field occurred with the introduction of architectures based on the self-attention mechanism, namely Transformers [4], [5], [7]. Transformer-based models, such as Bidirectional encoder representations from transformers (BERT), XLNet, and RoBERTa, consistently outperform traditional approaches in various emotion detection tasks [4]. Studies [4], [5], [6], [7] show that these models can process complex emotional expressions, achieving state-of-the-art results on standard benchmarks [4], [5]. The success of these models is attributed to their pre-training on massive text corpora and their ability to capture bidirectional context using self-attention mechanisms [4], [7]. Specifically, they have provided a significant increase in emotion detection accuracy, by more than 10 % compared to previous approaches [4], [6].

An important aspect of recent research is the use of standardized datasets [1], [2], [6]. Frequently used datasets include SemEval, ISEAR, and EmoBank, which provide labeled data for various emotional categories [1], [3]. The International Survey on Emotion Antecedents and Reactions (ISEAR) has been particularly valuable for cross-cultural emotion detection research [1], [5]. However, studies also highlight the limitations of existing datasets, particularly concerning multilingual coverage and cultural representation [5], [6]. Recent research emphasizes the importance of dataset quality, as factors such as representativeness, readability, and structural characteristics significantly impact model performance [6].

Recent studies show that attention-based models provide higher accuracy compared to traditional approaches [4], [5]. For instance, models built on BERT consistently demonstrate improved performance in detecting subtle emotional nuances and handling context-dependent expressions [4]. This advantage is particularly evident in addressing complex tasks such as sarcasm detection and multi-emotion recognition, which have traditionally been challenging for lexicon-based and conventional machine learning methods [1], [4]. The effectiveness of Transformer models is especially striking in multilingual contexts, where they demonstrate a superior ability to recognize emotional nuances across different languages [5].

Current research also emphasizes the importance of pre-training and fine-tuning strategies for Transformer models [4], [5]. Studies suggest that domain-specific fine-tuning can significantly improve performance, especially when working in specialized contexts or with specific types of emotional expressions [4]. This has led to the

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

development of various specialized variants of Transformer models, each optimized for specific aspects of emotion detection [4], [5]. Research indicates that meticulous pre-training strategies can lead to performance improvements of up to 10% in specific emotion detection tasks [4], [6].

Despite significant progress in using neural networks for emotion recognition, drawbacks exist related to interpretability, data quality, contextual understanding, and computational costs. The issue of the lack of representative corpora for the Ukrainian language is particularly acute, while existing datasets are mostly concentrated on English-language content [3], [5], [6].

Emotion recognition is a complex task, the solution to which, according to [8], []9], [10, [11], requires the application of several combined approaches that integrate traditional language processing methods with modern neural networks, as well as regular model updates considering changes in the linguistic environment and cultural specifics. This, in turn, necessitates the creation of a model for constructing the process of neural network-based emotion recognition in Ukrainian-language texts, which considers the peculiarities of the emotion recognition process and allows for the use of multiple types of neural networks.

2. THE PURPOSE AND OBJECTIVES OF THE RESEARCH

The main purpose of the research is to develop a model for constructing artificial neural network-based tools for recognizing emotions of Ukrainian-language domain-neutral text fragments, which will enhance the efficiency of relevant computer systems.

3. RESEARCH METHODS

Based on studies [8], [9], [10], [11], we propose embedding a modular approach into the architecture of the emotion-recognition process, enabling the recognition tools to be adapted to the specific requirements of the task. The model consists of several key components that can be tuned to implementation needs and includes three main blocks, depicted in Fig. 1.

The primary objective of the proposed model is to quantitatively assess the emotional coloration of text fragments.

The model is designed to predict the intensity of a predefined set of seven fundamental emotions: Joy, Love, Sadness, Anger, Fear, Disgust, and Neutrality.]

Formally, for an input text x, the model produces an output vector

$$y = (y_e)_{e \in \mathcal{E}},$$

where $\mathcal{E} = \{Joy, \text{Love}, \ldots, \text{Neutrality}\}$ is the target emotion set. Each component of vector y_e corresponds to an emotion $e \in \mathcal{E}$ and takes values in the continuous range $[0,1]$. This value is interpreted as the model-predicted intensity or degree of presence of the corresponding emotion in the text x. It is important to note that although the current configuration targets these seven emotions, the architecture is modular, allowing the target set to be flexibly expanded for task-specific needs or more fine-grained analyses in the future.

Data preprocessing constitutes the first stage in detecting the emotions in textual fragments. As illustrated in Fig. 2, this stage is implemented through six steps.

Step 1 – Data collection. Accumulating textual content from heterogeneous sources – including social-media posts and online news articles – while deliberately diversifying inputs to ensure a representative sample. Such diversity in context, writing style, and source domain is critical for achieving high accuracy and effective generalization of the neural-network model, fostering robustness to contextual variability.

Step 2 – Filtering. A multi-level selection procedure that eliminates non-Ukrainian texts, removes fragments below a predefined length threshold, and discards duplicates. This process yields a high-quality corpus for downstream analysis.

Step 3 – Cleaning. Focused on text normalization: removing special characters, standardizing punctuation, unifying formatting, correcting spelling mistakes, and ensuring consistent character encoding. These operations are essential for input-data consistency.



Fig. 1. **Main stages of constructing a neural-network emotion-recognition model process**
Source: **compiled by the authors**

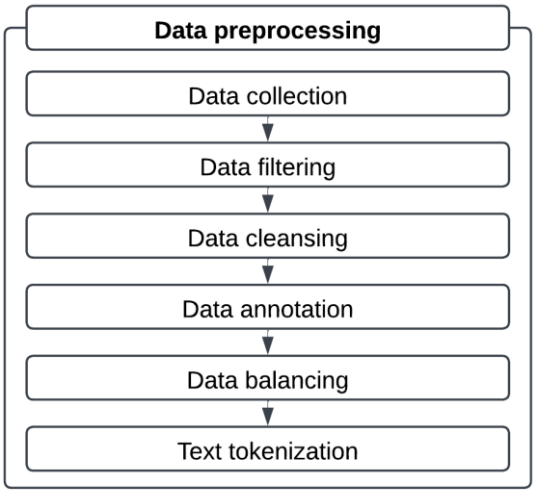


Fig. 2. **Data-preprocessing stage**
Source: **compiled by the authors**

Step 4 – Annotation. Performing through human labelling of textual fragments. Modern large language models (LLMs) may act as auxiliary tools, accelerating the workflow; studies [12], [13] report a 40–60 % increase in annotation efficiency when LLM suggestions are subsequently verified by a domain expert.

Step 5 – Balancing. Ensuring an optimal class distribution in the training set – crucial for preventing the model from over-fitting to over-represented emotion labels.

Step 6 – Text tokenization. Implemented via segmentation algorithms. Principal approaches include word-based, character-based, and subword tokenization, the latter commonly implemented with Byte Pair Encoding (BPE) [14], WordPiece [15], or SentencePiece [16]. The chosen tokenization method strongly influences the quality of subsequent text vectorization.

Adhering to this preprocessing structure yields a high-quality training corpus that underpins reliable model training and evaluation.

The model-construction stage, in turn, consists of three core steps (see Fig. 3).

1. Selection of the neural-network architecture type.

2. Parametrization of the neural-network architecture.

3. Model training.

The recognition model is built using one of three neural-network architectures – or an ensemble thereof [17].

1. Word-Embedding – based architecture.

This approach, typically implemented with Word to Vector (Word2Vec) (Continuous Bag of Words (CBOW) or Skip-Gram variants) [18], represents every word w from the vocabulary V by a dense embedding vector $\mathrm{v_w} \in \mathrm{R^d}$, where R - is set of real numbers, d is the embedding-space dimensionality.

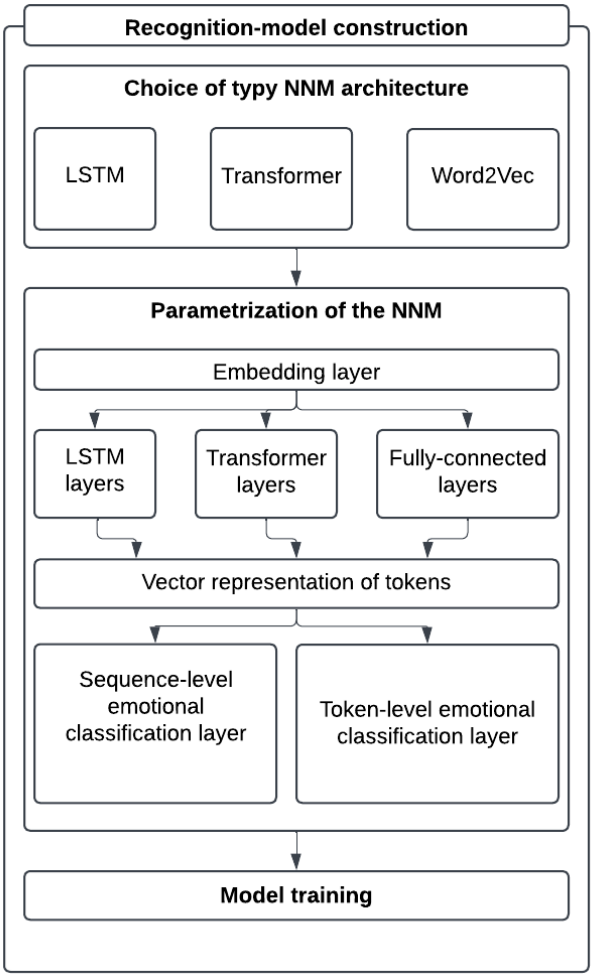


Fig. 3. **Structural diagram of the neural-network emotion-recognition model-construction stage**
Source: **compiled by the authors**

For a text fragment$T = \{w_1, w_2, \ldots, w_n\}$, of n words, an aggregated representation – e.g., the mean of word vectors (Bag-of-Embeddings) – is often used:

$$v_T = \frac{1}{n}\sum_{i=1}^{n} v_{w_i} \ . \qquad (1)$$

The resulting vector v_T is then fed to a classification layer (e.g., logistic regression or a shallow multi-layer perception (MLP) $f_{\text{classifier}}$) to predict the emotion vector y.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

These models train quickly and are computationally efficient, making them suitable for real-time, resource-constrained applications, although they capture word order and complex context only crudely. Word2Vec is also useful as an initial baseline for hypothesis validation.

2. Recurrent architecture with Long Short-Term Memory cells.

Long Short-Term Memory (LSTM) networks [19] are designed to capture sequential dependencies and long-range context.

The input text is represented as a sequence of token vectors $(x_1, x_2, \dots, x_n)$, де $\boldsymbol{x_t} \in R^d$.

The LSTM processes the sequence step-by-step, updating the hidden state h_t and cell state c_t at each time step t:

$$h_t, c_t = LSTM(x_t, h_{t-1}, c_{t-1}), \quad (2)$$

where the $LSTM$ function employs input, forget, and output gates to control information flow.

Long Short-Term Memory offers a balance between contextual expressiveness and computational demand relative to simpler models, making them suitable for many practical tasks.

3. Attention-based architecture – Transformer

Transformer models [7] have become the de-facto standard for Nature Language Processing (NLP) thanks to self-attention; this allows the model to determine the importance of different tokens in the input sequence.

The token sequence $(x_1, \dots, x_n)$ augmented with positional encodings P, forms input representations $X_{\mathbf{input}} \in R^{n \times d}$.

The core computation is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (3)$$

where Q (Query), K (Key), and V (Value) are linear projections of X_{input} and d_k is the key dimensionality. Multi-Head Attention (MHA) lets the model attend to multiple information facets in parallel. Each Transformer block combines MHA, position-wise Feed-Forward Networks (FFN), layer normalization, and residual connections. Transformers excel at contextual understanding and capture subtle emotional nuances, achieving state-of-the-art results, but they demand substantially more compute for training and inference.

Each of these architectures presents its own advantages and trade-offs in the context of emotion recognition tasks. This approach implements a comprehensive text-processing pipeline that is adapted to the specific requirements of Ukrainian-language processing and social-media content analysis, while accounting for potential resource constraints.

The parameterization step of the neural architecture for emotion tone recognition is based on a multi-level structure comprising differentiated text-processing components. The fundamental element of the architecture is the embedding layer, which provides the initial vector representation of textual tokens across all architecture types. The configuration of this layer is adapted according to the chosen base module: Word2Vec-based approaches employ static lexical vector representations; LSTM-based architectures utilize pre-trained embeddings that respect the sequential nature of text; Transformer architectures implement both positional and contextual embeddings, ensuring a comprehensive representation of the semantic and syntactic characteristics of the text.

After the chosen base module, should choose number of hidden layers. Recommendations based on works [7], [19], [20], is next:

- Long Short-Term Memory layers effectively capture long-range dependencies in sequential data [19]. In text classification tasks – such as emotion recognition – configurations of one to three LSTM layers are common, see formula (2) often implemented in a bidirectional variant to process context in both forward and backward directions of the sequence, frequently yielding performance improvements. Adding more layers beyond this range may not produce significant accuracy gains and increases computational complexity.
- Transformer encoder blocks realize a self-attention mechanism, shown in formula (3) for parallel processing of global contextual relationships within text. The number of layers varies substantially across models: from six layers in lightweight architectures (e.g., DistilBERT) to twelve layers in base versions of popular models (e.g., BERT-base, RoBERTa-base), and twenty-four or more layers in large models (e.g., BERT-large, RoBERTa-large) [7], [20]. Emotion-analysis tasks often employ pre-trained models with their original layer counts (for example, twelve layers for RoBERTa-base) to leverage their powerful language-representation capabilities.
- Fully-connected layers are typically positioned after the Word2Vec mean of vectors (1), LSTM, or Transformer blocks and perform non-linear transformations of the resulting vector representations. They serve to aggregate information and map it into the target emotion-class space. A

typical configuration includes one or two fully-connected layers preceding the final output layer (e.g., a Softmax or Sigmoid function for classification) [18].

Vector representations of tokens aggregate the outputs of preceding layers to form a unified depiction of textual elements. Let $T = t_1, t_2, \dots, t_n$ denote the sequence of input text tokens, where $n$the total number of tokens is. After processing by the previous layers, we obtain a matrix of vector representations$V = v_1, v_2, \dots, v_n$, where $v_i \in R^d$ is the d-dimensional vector corresponding to token t_i.

Based on this representation, two parallel classification mechanisms can be implemented:

1. Sequence-Level Emotion Classifier (E_{seq}), analyzes the overall emotional dynamics of the text fragment:

$$E_{seq} = f_{seq}\big(g(V)\big), \quad (4)$$

where $g: R^{n \times d} \to R^m$ aggregates all token vectors into a single context vector, and $f_{seq}: R^m \to R^k$ classifies it into k emotion classes.

2. Token-Level Emotion Classifier (E_{token}): Performs a fine-grained analysis of emotional characteristics at the individual token level:

$$E_{token} = \{ f_{token}(v_1), f_{token}(v_2), \dots, f_{token}(v_n) \}, \quad (5)$$

where $f_{token}: R^d \to R^k$ classifies each token's emotional tone.

Thus, the model provides both a global evaluation of the text's emotional tone (4) and a local evaluation for each token (5), thereby enhancing the accuracy and flexibility of the emotion recognition system.

Last step – **the training process** – includes flexible optimization strategies and adaptive learning schedules. The system supports various batch-size configurations depending on the available computational resources.

The model development stage also includes a comprehensive validation and testing framework that ensures the reliability of the neural network model through rigorous evaluation procedures (Fig. 4).

Validation procedures assess the model's generalization capability, and continuous monitoring of performance metrics allows for a detailed analysis of model behavior and resource utilization. Model performance is evaluated using a comprehensive set of metrics that provide insights into different aspects of classification quality. **Accuracy** serves as the primary metric, indicating the overall correctness of emotion classification across the entire dataset. However, considering potential class imbalance in emotional content, **Precision** and **Recall** offer a more detailed perspective on model performance. Precision defines the proportion of correct positive predictions, whereas recall measures the model's ability to identify all relevant instances of each emotion category. The **F1 - score** (the harmonic mean of precision and recall) provides a balanced assessment of model performance, particularly valuable in scenarios with uneven class distributions, which is typical for emotional content on social networks.

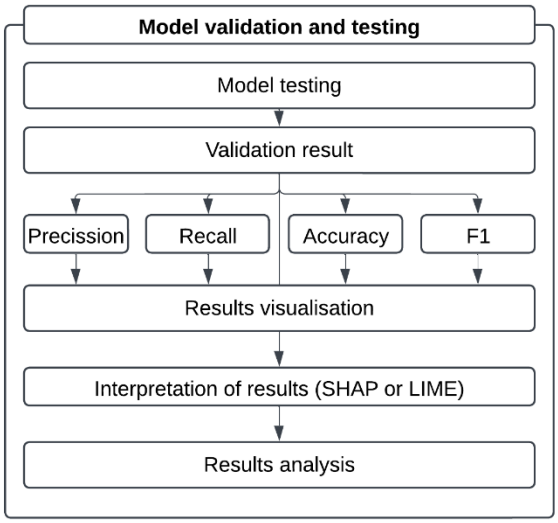


Fig. 4. **Schematic representation of the model validation and testing stage**
Source: **compiled by the authors**

Thus, the overall classification quality metric can be described by the following expression:

$$Quality = \alpha \cdot Accuracy + \beta \cdot Precision + \gamma \cdot Recall + \delta \cdot F1 - score, \quad (6)$$

where $\alpha, \beta, \gamma, \delta$ are the importance coefficients for the four metrics, such that $\alpha + \beta + \gamma + \delta = 1$; the larger a coefficient, the greater that metric's contribution to the composite quality score (6).

To enhance transparency and trust in the emotion-recognition system, the architecture incorporates two state-of-the-art model-interpretation approaches: SHAP (SHapley Additive exPlanations) [21] and LIME (Local Interpretable Model-agnostic Explanations) [22]. SHapley Additive exPlanations provides a unified framework for understanding feature importance and model decisions. By computing Shapley values for each feature, this method reveals how specific words and phrases influence particular emotion classifications. SHapley Additive exPlanations is especially useful for analyzing model behavior

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

across the entire dataset and uncovering patterns in emotional-content recognition. Local Interpretable Model-agnostic Explanations complements SHAP by providing local explanations for individual predictions. By generating locally faithful approximations, LIME helps clarify why certain texts are classified with specific emotions. This approach is particularly beneficial for examining complex cases where multiple emotional elements interact within the text.

The objectives of these interpretation methods are:

- to verify the model's learned patterns by checking their consistency with linguistic and psychological principles of emotion expression;
- to identify potential biases or artifacts in the model's decision‑making process;
- to provide actionable insights for model improvement and optimization;
- to increase system transparency for end users and stakeholders.

The combination (6) of robust performance metrics and modern interpretation techniques ensures comprehensive validation of the model's capabilities, providing valuable information for its continual refinement. It also facilitates the model's adaptation to evolving requirements for emotion analysis in Ukrainian social-media texts.

4. EXPERIMENT AND RESULTS

For practical verification of the proposed approach using the developed model, two prototype emotion-recognition modules were implemented, based on the FastText [18] and RoBERTa-base [20] architectures.

The implementation was carried out using the Python programming language, due to its extensive ecosystem and suitability for data-driven applications. Key libraries used include Pandas for data preprocessing and manipulation, and PyTorch for constructing and training a deep-learning model capable of classifying emotional content. For user interaction, the Gradio library was employed to build a web-based graphical user interface, allowing users to input free-form text and receive immediate feedback on the predicted emotional category (e.g., joy, anger, sadness).

The choice of these specific architectures as the foundation was motivated by [23], [24] the need to explore the fundamental trade-off between computational efficiency and classification accuracy in the task of emotion recognition in Ukrainian-language texts:

• FastText represents a lightweight solution based on averaging word vectors and character n-grams, with minimal system requirements. This makes it an attractive candidate for resource-constrained systems, such as mobile devices or embedded systems, where inference speed is crusial;

• RoBERTa-base, by contrast, embodies a state-of-the-art Transformer-based approach with self-attention mechanisms. This model can capture complex contextual dependencies in text, typically yielding significantly higher classification accuracy, albeit at the cost of substantially greater computational cost during both training and inference.

To adapt these base architectures to the specifics of our task, we modified their standard configurations. In particular, on top of each model's output representations we added a specialized classification block consisting of two sequential fully connected layers with ReLU activation, followed by a final Softmax layer to produce a probability distribution over the seven target emotions. This extension preserves the powerful knowledge encoded in the pre-trained layers (especially important for RoBERTa-base) while providing the model with additional flexibility for fine-tuning and learning the specific patterns characteristic of emotion expression in Ukrainian-language content.

The data preprocessing stage was unified for both models and included.

1. Constructing a balanced dataset of ~18,000 Ukrainian-language posts randomly sampled from X (formerly Twitter) between December and February 2025, with no topical or domain filters applied, thereby providing a broad cross-section of general-purpose user discourse rather than content tied to any specific subject area, it was annotated for seven basic emotions – sadness (3 144), anger (2 434), love (1 463), surprise (1 638), fear (2 161), joy (3 057) and neutral (4 103) – providing a balanced yet sufficiently large foundation for robust modelling.
2. Filtering content by language (Ukrainian only), minimum length (at least 5 words), and uniqueness (removal of duplicates).
3. Cleaning texts of special characters, normalizing punctuation, and standardizing encoding.
4. Corpus was pre-annotated with using LLMs like GPT-4o [25] with verification of three independent volunteers for annotation consistency.
5. Balancing the sample to ensure uniform representation of emotion classes.

6. Tokenization processes differing by architecture: for FastText, word-based tokenization using a Ukrainian-language dictionary; for RoBERTa-base, subtokenization via SentencePiece with a custom Ukrainian dictionary.

The model-construction stage was implemented according to the following principles.

1. **FastText classifier**

Architecture built on word-embedding representations (Word2Vec) using the SkipGram approach. Parameterization included: embedding dimension of 100, context window size of 5, minimum word frequency of 3. Setting the minimum word frequency to 3 filters out noise while preserving almost all meaningful vocabulary, giving more reliable vectors, faster convergence, and a smaller model footprint without sacrificing coverage [18]. Classification layer implemented as a softmax classifier on top of additional fully connected layers over the averaged word vectors. Total parameters count of approximately 3 million. Chosen as a lightweight solution for resource-constrained, real-time scenarios.

2. **RoBERTa-base classifier**

Architecture based on the Transformer encoder module with self–attention mechanisms. The base RoBERTa–base model was further pre–trained on a Ukrainian-language corpus to adapt to the linguistic features of Ukrainian, justifying its use in the experiment. Parameterization included: 12 Transformer encoder layers, hidden-state dimension of 768. Classification layer built on top of two sequential fully connected layers over the averaged token representations, mapping into the seven emotion-class space. Total parameters count of approximately 125 million. Selected as a representative of modern high-performance architectures for scenarios prioritizing recognition accuracy.

Training parameters:

- for FasText classifier: optimizer – stochastic gradient descent; mini-batch size – 32; learning rate – 0.1 with linear decay; number of epochs – 100;
- for RoBERTa-base classifier: optimizer – AdamW; mini-batch size – 32; initial learning rate – 1e-5 with a cosine decay schedule; number of epochs – 5.

Training loss showed on Fig. 5 illustrates the training dynamics of the FastText baseline, whose loss declines gradually from ≈ 2.3 to ≈ 0.35 over 100 epochs, with minor oscillations typical of stochastic optimization on shallow architectures. In stark contrast, the RoBERTa-base fine-tuning curve in Fig. 6 shows a precipitous drop from 1.103 to < 0.05 within only five epochs, after which the loss effectively plateaus. The 20x faster convergence rate of RoBERTa underscores the benefit of leveraging large-scale pre-training for low-resource downstream tasks. Moreover, the order-of-magnitude lower terminal loss achieved by RoBERTa suggests a substantially higher capacity to capture nuanced semantic features that the bag-of-subwords approach in FastText cannot model. These findings align with prior work reporting that transformer-based encoders require fewer gradient updates to reach equivalent – or superior – performance compared to shallow embedding models [26].

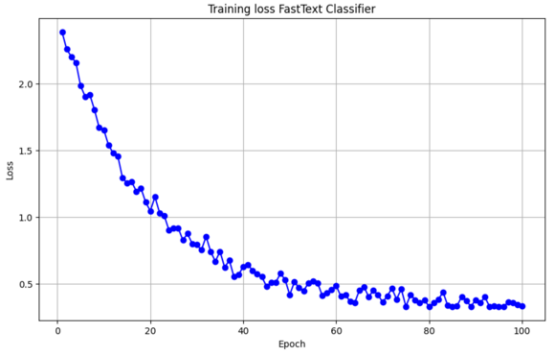


Fig. 5. **Train loss for FastText classifier**
Source: **compiled by the authors**

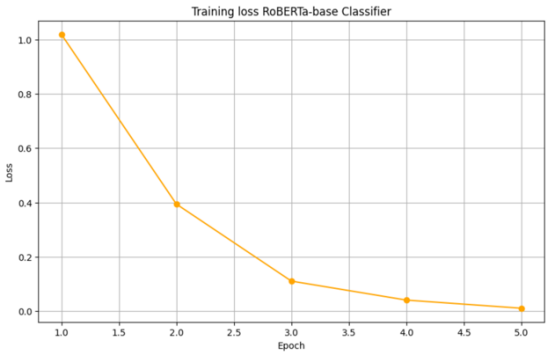


Fig. 6. **Train loss for RoBERTa-base classifier**
Source: **compiled by the authors**

Training of the FastText model was performed on the CPU, whereas the RoBERTa-base model was trained on GPU NVIDIA Tesla L10 GPU with 24 GB of video random accessed memory (VRAM).

For a comprehensive assessment of the effectiveness of the developed neural-network tools, we compared them against existing alternatives. Given the scarcity of specialized classifiers for emotion recognition in Ukrainian-language texts, two alternative approaches were selected. The first employed the proprietary multilingual GPT-4o model, which demonstrates high accuracy in text

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

analysis across multiple languages, including Ukrainian. The second approach combined two models: translation of Ukrainian text into English (Ukr2Eng Translation) followed by a specialized English-language emotion classifier (Eng-Emotion Classifier).

The validation, testing, and comparison phase of the models was conducted according to the following criteria:

• Computational complexity (number of model parameters);

• Inference speed (time to process a 20-word text);

• Classification quality (Precision, Recall, Accuracy, F1-score metrics).

The experimental study presented in Table 1 demonstrated significant differences in the effectiveness of four approaches to emotion recognition in Ukrainian texts. For calculation (6) was used default values of importance coefficients, where $\alpha = 0.25, \beta = 0.25, \gamma = 0.25, \delta = 0.25$.

The RoBERTa-base model, built on a Transformer encoder architecture, achieved the highest accuracy among models suitable for local deployment, attaining an F1-score of 0.91. This represents an 18-percentage-point improvement over the FastText results (F1-score 0.73), confirming the hypothesis that the attention mechanism more effectively captures emotional nuances in Ukrainian-language texts. However, this gain in accuracy is accompanied by a substantial increase in computational complexity: RoBERTa-base contains 125 million parameters, which is 41.7 times more than the 3 million parameters of FastText. This increase directly impacts inference speed: the inference time for a single text fragment on RoBERTa-base is 9 ms, compared to just 5 µs for FastText (a 1,800-fold speed difference). Therefore, the FastText model remains the optimal choice for resource-constrained systems requiring low latency, despite its lower accuracy.

Fig. 7 illustrates SHAP explanations for the *Joy* prediction on the sentence *"Я дуже радий, що ти прийшов на мій день народження"*. Starting from the corpus-level base value (0.448), SHAP adds the contribution of each token: affect-laden words such as *радий*, *дуже*, and the personal pronoun *я* (shown in red) cumulatively raise the log-odds to 0.993, whereas function words like *ти*, *мій*, and prepositions (in blue) exert only slight negative offsets. The colour-coded bars therefore reveal that the classifier's decision is dominated by valence-bearing content words, providing an intuitive, additive breakdown of how individual tokens shape the final emotion label.

The approach based on the sequential application of two models (Ukrainian-to-English translation followed by a specialized English-language emotion classifier) yielded the lowest performance among the tested methods (F1-score 0.69). These results support the hypothesis that significant emotional nuance is lost during translation. Moreover, employing two models in sequence greatly increases overall computational complexity (157 million parameters) and raises the inference time to 593 ms, limiting its practicality in real-time systems.

Table 1. **Comparative analysis of emotion recognition models**

Model name	Number of parameters	Inference speed	Precision	Recall	Accuracy	F1	Quality
FastText Classifier	3 M	5 µs	0.72	0.72	0.72	0.72	0.72
RoBERTa-base Classifier	125 M	9 ms	0.91	0.91	0.91	0.91	0.91
Ukr2Eng Translation + Eng-Emotion Classifier	74M+ 83M	593 ms	0.59	0.59	0.59	0.59	0.59
GPT-4o	–	2343 ms	0.95	0.96	0.96	0.95	0.95

Source: **compiled by the authors**

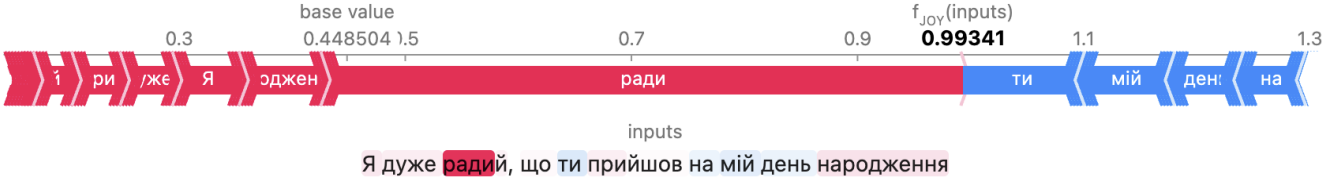


Fig. 7. **Visualization of how SHAP decomposes a RoBERTa-base classifier's "Joy" prediction into token-level contributions**

Source: **compiled by the authors**

Despite achieving the highest classification accuracy (F1-score 0.95), the practical use of the GPT-4o model is constrained by its inability to be deployed locally, high API costs, lack of transparency in its internal architecture, and potential bias that is difficult to assess due to the model's closed nature. In addition, its inference time of 2,343 ms per text makes it unsuitable for real-time applications.

The experimental validation was conducted on an ARM-architecture processor (Apple Silicon M3 Pro with 11 CPU cores), which is representative of modern mobile and desktop computing systems.

Thus, the obtained results validate the effectiveness of the proposed neural-network–based emotion-recognition modules and clearly illustrate the trade-off between classification quality and computational requirements. The modular approach enables selection of the optimal model configuration according to specific practical requirements for accuracy, speed, and available computational resources.

CONCLUSIONS AND PROSPECTS OF FURTHER RESEARCH

In this work, proposed a flexible modular framework for constructing neural-network–based tools to recognize the emotional tone of Ukrainian texts, which represents a significant advancement given the limited availability of neural-network resources for the Ukrainian language. The model is founded on a modular approach encompassing stages from preprocessing through validation, enabling the architecture to be adapted to specific tasks. Thus, the proposed solutions allow the neural-network architecture for text emotion detection to be tailored to requirements of accuracy and computational resources.

Experimental comparisons of the implemented prototypes clearly demonstrated the principal trade-off: RoBERTa-base achieved high accuracy (F1 = 0.91) required for in-depth analysis, whereas FastText delivered extremely low-latency inference (F1 = 0.72, ~5 μs), which is critical for real-time systems. Moreover, the results of the comparative experiments indicate the advantages of the proposed model over known approaches that rely on translating Ukrainian texts into English before emotion recognition.

Future research should be directed toward expanding the set of recognizable emotions for more granular analysis, exploring hybrid and ensemble architectures, and improving the robustness of the models to linguistic features specific to Ukrainian — such as sarcasm and regional dialects.

REFERENCES

1. Nandwani, P. & Verma, R. "A review on sentiment analysis and emotion detection from text". *Social Network Analysis and Mining*. 2021; 11 (81), https://www.scopus.com/sourceid/19700177337. DOI: https://doi.org/10.1007/s13278-021-00776-6.

2. Acheampong, F. A., Wenyu, C. & Nunoo-Mensah, H. "Text-based emotion detection: Advances, challenges, and opportunities". *Engineering Reports*. 2020. DOI: https://doi.org/10.1002/eng2.12189.

3. Maruf, A. A., Khanam, F., Haque, M. M., Jiyad, Z. M., Mridha, M. F. & Aung, Z. "Challenges and opportunities of text-based emotion detection: A survey". *IEEE Access*. 2024; 12: 18416–18450. DOI: https://doi.org/10.1109/ACCESS.2024.3356357.

4. Acheampong, F. A., Nunoo-Mensah, H. & Chen, W. "Transformer models for text-based emotion detection: a review of BERT-based approaches". *Artificial Intelligence Review*. 2021; 54: 5789–5829. DOI: https://doi.org/10.1007/s10462-021-09958-2.

5. Zhang, X., Mao, R. & Cambria, E. "Multilingual Emotion Recognition: Discovering the Variations of Lexical Semantics between Languages". *International Joint Conference on Neural Networks (IJCNN)*. Yokohama, Japan. 2024. p. 1–9. DOI: https://doi.org/10.1109/IJCNN60899.2024.10651409.

6. De León Languré, A. & Zareei, M. "Improving text emotion detection through comprehensive dataset quality analysis". *IEEE Access*. 2024; 12: 70489–70500. DOI: https://doi.org/10.1109/ACCESS.2024.3491856.

7. Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. "Attention is all you need". *Advances in Neural Information Processing Systems. 31st Conference on Neural Information Processing Systems (NIPS 2017*). 2017. p. 5999–6009, https://www.scopus.com/sourceid/23669.

8. Korchenko, O., Tereikovskyi, I., Ziubina, R., Tereikovska, L., Korystin, O., Tereikovskyi, O. & Karpinskyi, V. "Modular Neural Network model for biometric authentication of personnel in critical

infrastructure facilities based on facial images". *Applied Sciences.* 2025; 15 (5): 2553, https://www.scopus.com/sourceid/21100829268. DOI: https://doi.org/10.3390/app15052553.

9. Tereikovska, L. & Tereikovskyi, I. "Conceptual principles of neuronal recognition of phonemes in the voice signal of distance learning system members". *Science, Technology and Innovation in the Modern World: Scientific Monograph. Baltija Publishing*. 2023. DOI: https://doi.org/10.30525/978-9934-26-364-4-5.

10. Abdullah, T. & Ahmet, A. "Deep Learning in Sentiment Analysis: Recent Architectures". *ACM Comput. Surv*. 2022; 55 (8): 1–37. DOI: https://doi.org/10.1145/3548772.

11. Tan, K. L., Lee, C. P. Anbananthen, K. S. M. & Lim, K. M. "RoBERTa-LSTM: A hybrid model for sentiment analysis with transformer and recurrent Neural Network". In; *IEEE Access*. 2022; 10: 21517–21525. DOI: https://doi.org/10.1109/ACCESS.2022.3152828.

12. Pang, J., Wei, J., Shah, A. P., Zhu, Z., Wang, Y., Qian, C., & Wei, W. "Improving Data Efficiency via Curating LLM-Driven Rating Systems". *arXiv*, 2024. DOI: https://doi.org/10.48550/arXiv.2410.10877.

13. Wang, X., Kim, H., Rahman, S., Mitra, K. & Miao, Z. "Human-LLM collaborative annotation through effective verification of LLM labels". In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024. p. 1–21.

14. Bostrom, K. & Durrett, G. "Byte Pair encoding is Suboptimal for Language Model Pretraining". In *Findings of the Association for Computational Linguistics: EMNLP*. 2020. p. 4617–4624.

15. Song, X., Salcianu, A., Song, Y., Dopson, D. & Zhou, D. "Fast WordPiece Tokenization". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2021. p. 2089–2103. DOI: https://doi.org/10.18653/v1/2021.emnlp-main.160.

16. Kudo, T. & Richardson, J. "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing". *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.* 2018. p. 66–71. DOI: https://doi.org/10.18653/v1/D18-2012.

17. Petrescu, A., Truica, C. -O., Apostol, E. -S. & Paschke, A. "EDSA-Ensemble: an Event Detection Sentiment Analysis Ensemble Architecture". In *IEEE Transactions on Affective Computing*. 2025; 16 (2): 555–572. DOI: https://doi.org/10.1109/TAFFC.2024.3434355.

18. Joulin, A., Grave, E., Bojanowski, P. & Mikolov, T. "Bag of tricks for efficient text classification". *15th Conference of the European Chapter of the Association for Computational Linguistics Eacl.* 2017; 2: 427–431. DOI: https://doi.org/10.18653/v1/e17-2068.

19. Hochreiter, S. & Schmidhuber, J. "Long Short-Term Memory." *Neural Computation.* 1997; 9 (8): 1735–1780, https://www.scopus.com/sourceid/24782. DOI: https://doi.org/10.1162/neco.1997.9.8.1735.

20. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D. & Stoyanov, V. "Roberta: A robustly optimized bert pretraining approach". *arXiv*. 2019. DOI: https://doi.org/10.48550/arXiv.1907.11692.

21. Lundberg, S. & Lee, S.-I. "A unified approach to interpreting model predictions". *Advances in Neural Information Processing Systems.* 2017. p. 4766–4775, https://www.scopus.com/sourceid/23669. DOI: https://doi.org/10.48550/arXiv.1705.07874.

22. Ribeiro, M. T., Singh, S. & Guestrin, C. "Why should i trust you?" Explaining the predictions of any classifier". In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. p. 1135–1144. DOI: https://doi.org/10.48550/arXiv.1602.04938.

23. Korovii, O. & Petrashenko, A. "Adaptation of Distilling Knowledge Method in Natural Language Processing for Sentiment Analysis". In *Advances in Computer Science for Engineering and Manufacturing. ISEM*. 2022; 463, https://www.scopus.com/sourceid/21100901469. DOI: https://doi.org/10.1007/978-3-031-03877-8_15.

24. Korovii, O., & Tereikovskyi, I. "Conceptual model of the process of determining the emotional tonality of the text". *Computer-Integrated Technologies: Education, Science, Production*. 2024; (55): 115–123. DOI: https://doi.org/10.36910/6775-2524-0560-2024-55-14.

25. "OpenAI. "GPT-4o system card". – Available from: https://openai.com/index/gpt-4o-system-card/

26. Liu, N. F., Gardner, M., Belinkov, Y., Peters, M. E. & Smith N. A. "Linguistic Knowledge and Transferability of Contextual Representations". In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. *Association for Computational Linguistics*. 2019. p 1073–1094. DOI: https://doi.org/10.18653/v1/N19-1112.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article
Author Igor A. Tereikovskyi is a member of the Editorial Board of this journal. This role had no influence on the peer review process or editorial decision regarding this manuscript

Received 03.04.2025
Received after revision 04.06.2025
Accepted 17.06.2025

DOI: https://doi.org/10.15276/hait.08.2025.12
УДК 004.93

Модель побудови нейромережевих засобів розпізнавання емоційного забарвлення фрагментів текстів

Терейковський Ігор Анатолійович[1)]
ORCID: https://orcid.org/0000-0003-4621-9668; terejkowski@ukr.net. Scopus Author ID: 57195940293
Коровій Олександр Сергійович[1)]
ORCID: https://orcid.org/0000-0002-2019-2527; zeusfsxtmp@gmail.com
[1)] Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Берестейський, 37. Київ, 03056, Україна

АНОТАЦІЯ

Розпізнавання емоційного забарвлення тексту є ключовою задачею в обробці природної мови, особливо актуальною в умовах експоненційного зростання текстових даних із соціальних мереж та голосових інтерфейсів. Однак розробка ефективних систем розпізнавання емоцій для мов з обмеженими ресурсами, таких як українська, стикається зі значними викликами, включаючи неформальність мови, діалектні варіації та культурну специфіку. Ця робота представляє модульну модель (фреймворк) для побудови нейромережевих засобів розпізнавання емоцій у фрагментах україномовних текстів. Модель охоплює комплексний конвеєр препроцесингу даних, гнучкий вибір архітектури (включаючи підходи на основі Word2Vec, LSTM та Transformer), та ретельну валідацію з використанням стандартних метрик та методів інтерпретації. В рамках експериментального дослідження було реалізовано та порівняно два прототипи: легковаговий класифікатор на основі FastText та потужний класифікатор на основі RoBERTa-base, навчені розпізнавати сім базових емоцій. Результати демонструють, що RoBERTa-base досягає високої точності , значно перевершуючи FastText та базовий підхід з перекладом, але потребує суттєво більших обчислювальних ресурсів. Дослідження підкреслює критичний компроміс між точністю та ефективністю, надаючи практичні рекомендації щодо вибору моделі залежно від ресурсних обмежень та вимог до продуктивності для задач аналізу емоцій в українській мові.

Ключові слова: розпізнавання емоцій; нейронні мережі; обробка природної мови; мови з обмеженими ресурсами; модульна модель

ABOUT THE AUTHORS



Ihor A. Tereikovskyi – Doctor of Engineering Sciences, Professor, Professor of System Programming and Specialized Computer Systems Department. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37, Beresteiskyi Ave. Kyiv, 03056, Ukraine
ORCID: https://orcid.org/0000-0003-4621-9668; terejkowski@ukr.net. Scopus Author ID: 57195940293
***Research field*:** Neural Networks, Voice Recognition, Cybersecurity, Specialized Computer Systems

Терейковський Ігор Анатолійович – доктор технічних наук, професор. Професор кафедри Системного програмування і спеціалізованих комп’ютерних систем. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Берестейський, 37. Київ, 03056, Україна



Oleksandr S. Korovii – PhD student. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37, Beresteiskyi Ave. Kyiv, 03056, Ukraine
ORCID: https://orcid.org/0000-0002-2019-2527; zeusfsxtmp@gmail.com, Scopus Author ID: 57644238900
***Research field*:** Neural Networks, Natural Language Processing, System Programming, Specialized Computer Systems

Коровій Олександр Сергійович – аспірант кафедри Системного програмування і спеціалізованих комп’ютерних систем. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», пр. Берестейський, 37. Київ, 03056, Україна