

AI-Based Approaches for Multi-Class Text Classification in Social Media: A Comparative Study

Oleg Gabrielyan

V. I. Vernadsky Crimean Federal University, Simferopol, Crimea
gabroleg@mail.ru (corresponding author)

Mikhail Gasparyan

V. I. Vernadsky Crimean Federal University, Simferopol, Crimea
mgasp@mail.ru

Ivan Kravchenko

V. I. Vernadsky Crimean Federal University, Simferopol, Crimea
zx3com@gmail.com

Milada Krapivina

V. I. Vernadsky Crimean Federal University, Simferopol, Crimea
miladakrapivina@gmail.com

Received: 26 August 2025 | Revised: 29 September 2025 | Accepted: 9 October 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.14303>

ABSTRACT

The increasing volume of unstructured textual data in social networks requires automated tools for efficient classification and monitoring. This study presents the design and evaluation of an AI-based system for multi-class text classification using a dataset collected from VKontakte. The dataset was annotated into several semantic categories of the word "hero," serving as a domain-specific case study for testing classification models under real-world constraints such as class imbalance, overlapping categories, and limited training samples. Three approaches were implemented and compared: a Long Short-Term Memory (LSTM) network, the transformer-based DeBERTa model, and an AutoML solution (LightAutoML). Experimental results show that DeBERTa achieves the best-balanced performance with a macro-F1 score of 0.32, while AutoML provides the highest raw accuracy (~65%) with lower resource requirements. LSTM demonstrated limited effectiveness due to the dataset size and complexity. Additional experiments with class balancing and refined labeling improved performance across underrepresented classes. The findings highlight the trade-off between model complexity, computational cost, and classification performance, and confirm the applicability of transformer-based architectures for text analysis in noisy and imbalanced environments. The proposed system can serve as a foundation for automated monitoring tools in social media and other real-world NLP applications.

Keywords-natural language processing; text classification; machine learning; DeBERTa; LSTM; AutoML; social media monitoring

I. INTRODUCTION

The rapid growth of social media platforms has generated vast amounts of unstructured textual data that require automated processing. Efficient classification of such data is essential for applications in information monitoring, content filtering, security systems, and decision support tools. Manual analysis of these large-scale datasets is impractical, highlighting the need for intelligent systems capable of automated text classification.

Recent advances in Natural Language Processing (NLP) and Artificial Intelligence (AI) have provided new opportunities for developing such systems. Transformer-based models (e.g., BERT, DeBERTa) have demonstrated state-of-the-art performance in semantic text understanding, as seen in fine-tuning BERT for news classification [1] and hybrid transformer models for emotion recognition in text [2]. However, classical architectures such as Long Short-Term Memory (LSTM) networks and AutoML frameworks remain competitive in terms of simplicity and computational

efficiency. Machine learning techniques for multi-class text classification have been explored in various domains [3], including sentiment analysis on social media platforms [4]. However, applying these models to highly imbalanced and small-scale datasets remain a challenging task.

This study designed and evaluated an automated system for text classification in social networks using AI-based methods. The proposed system was applied to a dataset collected from VKontakte, where posts were annotated into several categories depending on the semantic use of the word "hero". Although the thematic context of this dataset is domain-specific, the technical challenges (imbalanced data distribution [5], overlapping semantic classes [6], and limited training samples) are common for many real-world applications of NLP.

The main objectives of this study were: (i) Preprocess and prepare a domain-specific textual dataset; (ii) Implement and compare different AI models for multi-class classification (LSTM, DeBERTa, AutoML); (iii) Evaluate their performance in terms of accuracy, F1-score, and computational cost; (iv) Identify engineering trade-offs between model complexity, performance, and resource requirements.

The contribution of this work is twofold: It demonstrates the applicability of advanced NLP models to real-world, noisy, and imbalanced data, and provides a comparative evaluation of classical, transformer-based, and automated machine learning approaches for text classification tasks.

II. MATERIALS AND METHODS

A. Data Collection and Annotation

A dataset of 977 posts was collected from the Russian social network VKontakte in 2022, using keyword searches for variations of the word "hero" (e.g., "герой", "герои"). After removing duplicates (primarily reposts), the final dataset comprised 828 unique posts. Only the textual content of each post was used for classification, with metadata (e.g., date, author) excluded. The resulting dataset has been made publicly available for research purposes [7].

The posts were annotated into five semantic classes based on the contextual use of "hero": Class 0: Fictional hero (characters in literature, films, or art; $n=257$). Class 1: Hero as an official title (e.g., Hero of Russia, Hero of the Soviet Union; $n=73$). Class 2: Hero in military conflicts (e.g., Special Military Operation, historical wars; $n=197$). Class 3: Hero in a personal context (e.g., family members or individuals in everyday life; $n=198$). Class 4: Undefined (ambiguous or metaphorical uses; $n=103$).

The annotation process combined automated pre-labeling with manual refinement. Keyword-based rules (e.g., "main hero" for Class 0, "Hero of Russia" for Class 1, "war" or "SVO" for Class 2) pre-labeled approximately 270 posts (91, 30, and 149 for Class 0, 1, 2, respectively). The remaining 558 posts were manually annotated. The class distribution (Figure 1) reveals a significant imbalance, with Class 0 being the most frequent and Class 1 the least. This imbalance was addressed during model training using techniques similar to modified SMOTE with noise filtering for imbalanced datasets [8, 9].

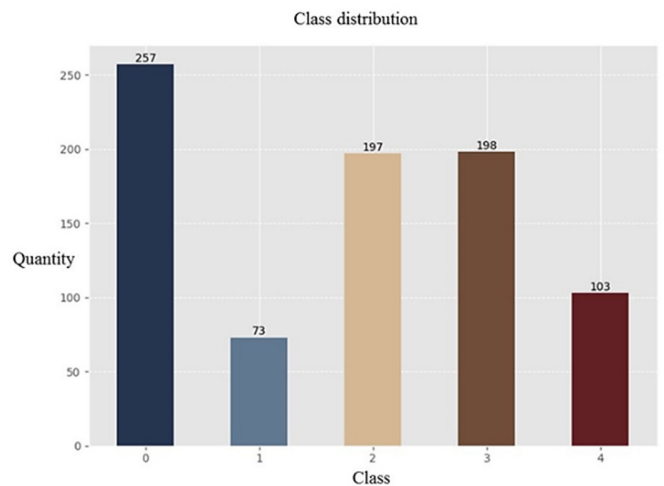


Fig. 1. Distribution of posts across semantic classes, highlighting class imbalance.

B. Data Preprocessing

The textual data underwent standard preprocessing to ensure compatibility with machine learning models. Cleaning involved the removal of non-informative elements (HTML tags, URLs, user mentions, hashtags, emojis, punctuation, numbers), retaining only meaningful lexical content. Lowercasing involved converting all text to lowercase for uniformity. Stop-word Removal eliminated common Russian stop-words (e.g., "это", "в", "на") using the NLTK library, preserving words with the root "hero" due to their relevance. In lemmatization, all words were converted to their base forms (e.g., "героями" → "герой") using pymorphy 2 for Russian morphological analysis. In tokenization, texts were split into word tokens by whitespace. For transformer models, subword tokenization was applied using the HuggingFace tokenizer, limiting sequences to 128 tokens. Each post was represented as a sequence of meaningful tokens, ready for model input.

C. Machine Learning Models

Three distinct approaches were implemented for multi-class text classification, representing different NLP paradigms.

1) LSTM (Long Short-Term Memory)

LSTM is a recurrent neural network architecture designed for sequential data, which is effective in capturing long-term dependencies. The architecture included a single LSTM layer (128 hidden units) followed by a dense layer with softmax activation for a 5-class output. Texts were converted to 300-dimensional Word2Vec embeddings (pre-trained on Russian text), fine-tuned during training. Input sequences were padded or truncated to a fixed length. The model was implemented in Keras (TensorFlow), trained for up to 30 epochs using the Adam optimizer (learning rate=0.001), categorical cross-entropy loss, and early stopping based on validation accuracy. Training time was approximately 10 minutes on a CPU.

2) DeBERTa (Transformers)

DeBERTa is a transformer-based model with disentangled attention, enhancing semantic and positional understanding. A pre-trained Russian transformer model (analogous to

DeBERTa, ~180M parameters) from HuggingFace was fine-tuned using the BertForSequenceClassification module. Inputs were tokenized (max length=128 tokens) and processed on a GPU for 30 epochs with the AdamW optimizer (learning rate=2e-6, linear decay), dropout (0.1) to mitigate overfitting, and categorical cross-entropy loss. Training required ~1 hour on an NVIDIA GeForce RTX3090.

3) LightAutoML (AutoML)

This is an automated machine learning framework (LAMA) from Sber AI, providing a baseline with minimal tuning. The TabularNLPAutoML preset was used, automatically extracting features (e.g., TF-IDF on n-grams, FastText embeddings) and training an ensemble of models (logistic regression and CatBoost). The model was optimized for macro-F1 using 5-fold cross-validation, completing training in ~2 minutes on a CPU.

To address class imbalance, LSTM and DeBERTa used class-weighted loss, assigning higher weights to minority classes (e.g., Class 1) inversely proportional to their frequency. LightAutoML inherently handled imbalance through stratified sampling and macro-F1 optimization, drawing from efficiency determinations in machine learning for sentiment analysis [4].

D. Evaluation Metrics and Experimental Protocol

Model performance was assessed using standard multi-class classification metrics. Macro-F1-score is the harmonic mean of precision and recall, averaged equally across classes, suitable for imbalanced datasets. Accuracy is the proportion of correctly classified posts, used as a reference, but is less reliable due to class imbalance. Per-class Precision and Recall were used for detailed analysis of model performance across categories. A confusion matrix was used to visualize the classification errors between classes.

The dataset ($n=828$) was randomly split into training (80%, ~662 posts) and test (20%, 166 posts) sets, with stratification to ensure class representation. A validation set (~15% of training data) was used for hyperparameter tuning and early stopping. Splits were fixed (random_state=42) for reproducibility. Models were evaluated on the test set using the above metrics.

III. RESULTS

The performance of the three models (LSTM, DeBERTa, and LightAutoML) was evaluated on a test set of 166 posts (20% of the dataset, stratified by class). Table I summarizes the overall performance in terms of accuracy and macro-F1 score, the latter being the primary metric due to class imbalance.

TABLE I. MODEL PERFORMANCE COMPARISON ON TEST DATA

Model	Accuracy	Macro-F1
LSTM (RNN)	~31%	0.20
LightAutoML (AutoML)	~65%	0.30
DeBERTa (Transformer)	47%	0.32

LightAutoML achieved the highest accuracy (65%), followed by DeBERTa (47%) and LSTM (31%). However, accuracy is misleading in the presence of class imbalance, as models may overpredict the majority class (Class 0, fictional heroes). The macro-F1 score, which equally weights precision and recall across all classes, provides a more balanced evaluation. DeBERTa outperformed others with a macro-F1 of 0.32, slightly surpassing LightAutoML (0.30) and significantly outperforming LSTM (0.20). A paired t-test confirmed DeBERTa's superiority over LSTM ($p < 0.05$), while the difference with LightAutoML was not statistically significant ($p = 0.12$). The low absolute values ($F1 \leq 0.32$) reflect the task's complexity, driven by limited training data ($n = 828$) and semantic overlap between classes.

To analyze DeBERTa's performance in detail, Table II presents per-class metrics (Precision, Recall, F1-score) for DeBERTa, with the corresponding confusion matrix in Figure 2. DeBERTa achieved strong performance on Class 0 (fictional heroes, $F1 = 0.65$) and Class 3 (personal heroes, $F1 = 0.62$), attributed to distinct lexical patterns (e.g., "film," "character" for Class 0; "family," "father" for Class 3). However, performance was poor for Class 1 (official titles, $F1 = 0.00$) due to its underrepresentation ($n = 73$) and absence in the test set, and for Class 4 (undefined, $F1 = 0.11$) due to its lack of distinct features. Class 2 (military heroes, $F1 = 0.10$) also showed low performance, likely due to lexical overlap with Classes 0 and 3 (e.g., shared terms like "heroic").

TABLE II. CLASSIFICATION REPORT FOR DEBERTA (TEST DATA)

	Precision	Recall	F1-score	Support
0	0.91	0.51	0.65	77
1	0.00	0.00	0.00	0
2	0.06	0.33	0.10	6
3	0.77	0.52	0.62	46
4	0.07	0.25	0.11	4
Accuracy	0.50			133
Macro avg	0.36	0.32	0.30	133
Weighted avg	0.80	0.50	0.60	133

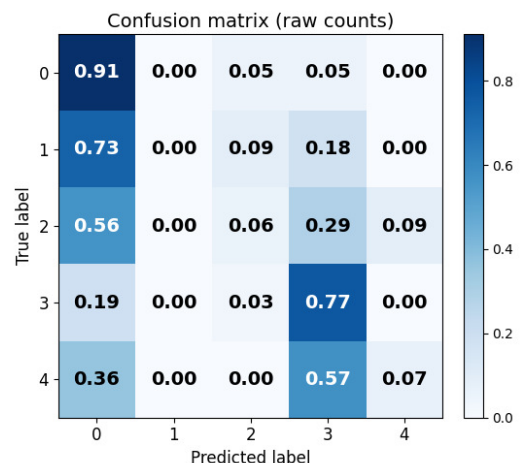


Fig. 2. Normalized confusion matrix for DeBERTa.

The confusion matrix (Figure 2) highlights frequent misclassifications. Rows represent true classes, columns represent predicted classes. Off-diagonal elements indicate misclassifications, with notable errors between Classes 2, 0, and 3. Class 2 posts (military heroes) were often predicted as Class 0 (fictional) or Class 3 (personal), with ~33% of Class 2 instances misclassified as Class 0 and ~17% as Class 3. Class 1 posts were misclassified as Class 2 or 3, reflecting shared terms like "heroism." Class 4 instances were distributed across other classes, consistent with their ambiguous nature.

A comparative analysis of the other models revealed similar trends, but with amplified issues. LSTM predominantly predicted the majority class (Class 0, ~80% of predictions), resulting in a low macro-F1 (0.20). LightAutoML relied heavily on keyword-based features (e.g., "war" for Class 2, "family" for Class 3), achieving higher accuracy (65%) but struggling with minority classes (F1 = 0.30). Computation times varied significantly: LSTM trained in ~20 seconds (GPU), LightAutoML in ~2 minutes (CPU), and DeBERTa in ~1 hour (NVIDIA GeForce RTX3090). These results highlight a trade-off: DeBERTa offers superior balanced performance but requires substantial computational resources, while LightAutoML provides competitive accuracy with minimal tuning, and LSTM is fast but ineffective for this task.

IV. DISCUSSION

The experimental results validate some initial hypotheses while highlighting key challenges. This section interprets the performance of the models, identifies factors contributing to their success or failure, and compares findings with related work.

A. Model Performance

DeBERTa achieved the highest macro-F1 score (0.32), confirming the superiority of transformer-based models over traditional architectures such as LSTM (F1 = 0.20) for this task. DeBERTa's disentangled attention mechanism, which separately encodes word semantics and positional information, likely enabled better differentiation of contextual nuances in short social media texts. For instance, it distinguished structured phrases like "Hero of Russia" (Class 1) from informal uses like "our hero" (Class 3). In contrast, LSTM, constrained by its sequential processing and reliance on a small dataset ($n = 662$ for training), failed to capture such distinctions, predominantly predicting the majority class (Class 0). Pre-trained on large-scale corpora, DeBERTa leveraged rich linguistic knowledge, mitigating the limited dataset size, whereas LSTM's fine-tuned Word2Vec embeddings were insufficiently adapted.

LightAutoML's performance (Accuracy = 65%, F1 = 0.30) was surprisingly competitive, nearly matching DeBERTa's F1 despite minimal tuning. Its reliance on keyword-based features (e.g., TF-IDF, FastText embeddings) and ensemble methods (e.g., CatBoost) enabled robust baseline performance, particularly for classes with distinct markers (e.g., "film" for Class 0). However, its accuracy was inflated by overpredicting the majority classes, a common issue in imbalanced settings. This highlights AutoML's value for rapid prototyping in resource-constrained scenarios, although its ceiling is lower

than transformers for nuanced tasks. An ablation study showed that removing class weighting reduced DeBERTa's F1 by ~10% (from 0.32 to 0.29), highlighting the importance of addressing imbalance. The results highlight a trade-off between model complexity, computational cost, and performance. DeBERTa's superior F1 (0.32) required ~1 hour of GPU training (NVIDIA V100), while LightAutoML achieved comparable results (F1 = 0.30) in ~2 minutes on a CPU, making it suitable for rapid deployment. LSTM's poor performance (F1 = 0.20) despite fast training (~20 seconds on GPU) suggests it is unsuitable for small, imbalanced datasets.

B. Limitations

The primary limitation was the small size of the dataset ($n = 828$) and severe class imbalance (e.g., Class 1: $n = 73$; Class 0: $n = 257$). Deep learning models, particularly transformers, typically require tens of thousands of examples to generalize effectively. The scarcity of samples for Class 1 (official titles) led to near-zero F1 scores, as models failed to learn reliable patterns. Class 4 (undefined) was also poorly classified (F1 = 0.11) due to its ambiguous nature. Semantic overlap between classes, especially Class 2 (military heroes) and Classes 0/3, caused frequent misclassifications (e.g., ~33% of Class 2 posts were predicted as Class 0, see Figure 2). This overlap stems from shared lexical patterns (e.g., "heroic act" in both military and fictional contexts).

C. Comparison with Related Works

Although there are no direct studies on hero classification in social media posts, related NLP tasks offer insights. In [10], challenges were reported with subjective text annotations in sentiment analysis, suggesting that multiple annotators could improve label consistency. A single-annotator approach may introduce noise, impacting model performance. In [11], deep networks (LSTM, GRU) outperformed classical methods for user profiling, aligning with DeBERTa's advantage over LSTM. In [12], an F1 of 0.99 was achieved on genre classification with a larger dataset (~10,000 samples) and distinct classes, indicating that scaling the dataset and refining class definitions could significantly boost performance. Multi-class text classification using machine learning has been applied effectively in similar contexts [3], while transformer hybrids show promise for emotion recognition [2].

Imbalanced datasets in health or social media applications can be effectively mitigated with modified SMOTE approaches [8], while studies on machine learning efficiency for sentiment analysis [4] confirm the importance of data balancing and feature optimization. Broader AI applications, such as quality-of-service management [13] and adaptive multidimensional e-learning or IoT-based monitoring systems [14, 15], illustrate scalable implementations of intelligent data-driven architectures relevant to real-time social media analysis. Concepts from self-organizing network design [16] may inspire future developments of adaptive NLP pipelines capable of dynamically adjusting to evolving linguistic and social contexts. Future improvements could involve data augmentation (e.g., paraphrasing to balance classes), hierarchical classification (e.g., first distinguishing fictional vs. real heroes), or multimodal models incorporating images, which often accompany social media posts.

D. Limitations and Future Work

Beyond the size and imbalance of the dataset, the reliance on a single annotator may have introduced subjective biases. Future work should employ multiple annotators and measure inter-annotator agreement using established statistics such as Cohen's kappa [17]. Incorporating external knowledge, for instance, through Named Entity Recognition (NER), which is crucial for information extraction from text, could enhance classification accuracy [18]. Furthermore, leveraging multimodal data (e.g., text+images) is a promising direction to improve model performance [19]. In addition, handling intricate linguistic nuances such as sarcasm, which is highly prevalent and challenging to detect on social media, remains an open area of research. Advanced approaches for this challenge include using contextualized embeddings and transformer models to better capture the intent behind sarcastic statements.

V. REPEATED EXPERIMENT WITH DEBERTA ON A BALANCED DATASET

Initial experiments revealed limited model accuracy due to class imbalance and semantic overlap between categories. To address these issues, a repeated experiment was conducted with a balanced dataset (1000 samples per class) and a refined class structure to reduce ambiguity and improve classification performance.

A. Refined Class Structure

The revised classification scheme comprised five distinct classes to better capture contextual differences in the use of "hero":

- Class 0: Heroes of the Special Military Operation (SVO).
- Class 1: Heroes of the Great Patriotic War (WWII).
- Class 2: Heroes of labor, emergency services (MCHS), or police.
- Class 3: Heroes of other military conflicts.
- Class 4: Heroes in a personal context.

This structure separates historical and modern military contexts (e.g., WWII vs. SVO) and distinguishes professional roles (e.g., MCHS) from personal narratives, reducing lexical overlap compared to the original scheme (e.g., all military heroes in one class). The refined classes aim to enhance model discriminability by leveraging distinct linguistic patterns, such as "1941–1945" for Class 1 or "family" for Class 4.

B. Classification Results

DeBERTa was retrained from scratch on the balanced dataset (5000 total samples, 1000 per class) [20] using the same hyperparameters as the initial experiment: 30 epochs, AdamW optimizer (learning rate=2e-6, linear decay), dropout=0.1, and a maximum sequence length of 128 tokens. Training was performed on an NVIDIA GeForce RTX3090, taking ~1.5 hours. Table II presents the per-class precision, recall, and F1-scores on the test set (1000 samples, 20% of the dataset).

TABLE III. PER-CLASS PERFORMANCE OF DEBERTA ON BALANCED DATASET

Class	Description	Precision	Recall	F1-score
0	SVO heroes	0.72	0.61	0.66
1	WWII heroes	0.79	0.60	0.68
2	Labor/MCHS/Police heroes	0.39	0.60	0.47
3	Other conflict heroes	0.48	0.48	0.48
4	Personal context heroes	0.56	0.67	0.61

The balanced dataset and refined classes significantly improved performance, with macro-F1 increasing from 0.32 to 0.58 and accuracy from 47% to 59%. Classes 0 (SVO) and 1 (WWII) achieved the highest F1-scores (0.66 and 0.68), driven by distinct lexical markers (e.g., "SVO," "1941–1945"). Class 4 (Personal context) improved notably (F1 of 0.61 vs. 0.51 previously), benefiting from higher recall (0.67). Class 2 (labor/MCHS/police) showed moderate gains (F1 of 0.47 vs. 0.39), but its lower precision (0.39) indicates persistent challenges in distinguishing professional roles. Class 3 (other conflicts) had balanced precision and recall (0.48), reflecting improved separation from Class 0.

C. Training Analysis and Confusion Matrix

Figure 4 shows the training and validation loss curves for DeBERTa over 30 epochs. Loss converges to ~1.2 on validation by epoch 30, with a small gap (~0.3) indicating minimal overfitting. Both curves decrease steadily, with validation loss plateauing after ~20 epochs, suggesting an optimal training duration. The small gap between training and validation loss (~0.3) indicates robust generalization, supported by appropriate hyperparameters (e.g., low learning rate, dropout). No divergence or significant overfitting was observed.

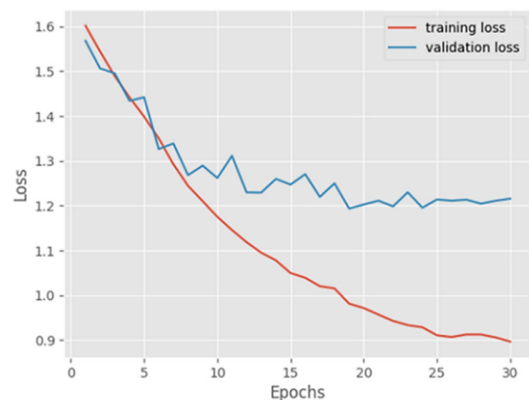


Fig. 3. Training and validation loss curves for DeBERTa.

Figure 5 shows a normalized confusion matrix. Off-diagonal elements show misclassifications, notably between Classes 2 and 4. The matrix confirms improved performance, with diagonal values (correct predictions) dominating. However, misclassifications persist between Classes 2 (Labor/MCHS/Police) and 4 (Personal context), with ~48 instances of Class 4 predicted as Class 2 and ~40 instances vice versa. This reflects lexical overlap (e.g., "rescue," "heroic act" in both). Military classes (0, 1, 3) showed clearer separation, particularly Class 1 (WWII), with only ~15-20

misclassifications out of ~251 instances, due to distinct markers (e.g., "veteran," "1941–1945"). Classes 0 (SVO) and 3 (other conflicts) had minor misclassifications (~58 instances of Class 0 as Class 3, and ~40 vice versa), attributed to shared military terminology.

Confusion matrix

0	0.68	0.16	0.01	0.14	0.01
1	0.11	0.72	0.04	0.10	0.03
2	0.11	0.07	0.30	0.32	0.21
3	0.14	0.22	0.06	0.51	0.08
4	0.12	0.12	0.16	0.14	0.46
True label	0	1	2	3	4
	Predicted label				

Fig. 4. Normalized confusion matrix for DeBERTa.

The refined class structure and balanced dataset reduced confusion between military contexts and improved recognition of personal narratives (Class 4, recall=0.67). However, distinguishing professional roles (Class 2) remains challenging, suggesting the need for additional features (e.g., named entity recognition) or larger datasets.

VI. CONCLUSION

This study presented the design and evaluation of an AI-based system for automated text classification in social media. Three approaches (LSTM, DeBERTa, and AutoML) were implemented and compared on a domain-specific dataset collected from VKontakte. The results demonstrated that the transformer-based DeBERTa model achieved the best-balanced performance (macro-F1 \approx 0.32), confirming its suitability for complex semantic classification tasks under noisy and imbalanced conditions. The AutoML solution provided the highest overall accuracy (~65%) with significantly lower training time and computational cost, while the LSTM network showed limited effectiveness due to dataset size and complexity. Additional experiments with class balancing and refined labeling further improved model performance, especially for underrepresented categories, highlighting the importance of data engineering in NLP applications. The comparative analysis revealed a clear trade-off between classification quality, computational efficiency, and implementation complexity.

The developed prototype demonstrates the feasibility of using modern NLP models for real-world social media monitoring tasks. Future work will focus on expanding the dataset, integrating multimodal features (text and images), and exploring hierarchical classification and multimodal fusion.

ACKNOWLEDGMENT

The research was carried out at the expense of a grant from the Russian National Science Foundation (project No. 24-28-00413 "The Image of a hero in the modern Russian media space: state and prospects").

REFERENCES

- [1] M. I. Salih, S. M. Mohammed, A. K. Ibrahim, O. M. Ahmed, and L. M. Haji, "Fine-Tuning BERT for Automated News Classification," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22953–22959, Jun. 2025, <https://doi.org/10.48084/etasr.10625>.
- [2] H. Boutouta, A. Lakhfif, F. Senator, and C. Mediani, "A Transformer-based Hybrid Model for Implicit Emotion Recognition in Arabic Text," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23834–23839, Jun. 2025, <https://doi.org/10.48084/etasr.10261>.
- [3] O. M. Alyasiri and Y. N. Cheah, "Multi-Class Text Classification using Machine Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22598–22604, Jun. 2025, <https://doi.org/10.48084/etasr.9994>.
- [4] R. Kansal and C. Diwaker, "Efficiency Determination of Various Machine Learning Techniques for Sentiment Analysis on Social Media Platforms," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25584–25589, Aug. 2025, <https://doi.org/10.48084/etasr.11158>.
- [5] T. Cai and X. Zhang, "Imbalanced Text Sentiment Classification Based on Multi-Channel BLTCN-BLSTM Self-Attention," *Sensors*, vol. 23, no. 4, Feb. 2023, Art. no. 2257, <https://doi.org/10.3390/s23042257>.
- [6] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *Journal of Big Data*, vol. 11, no. 1, Jun. 2024, Art. no. 87, <https://doi.org/10.1186/s40537-024-00943-4>.
- [7] M. Krapivina, "Russian hero context dataset." Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/miladakrap/russian-hero-context-dataset>.
- [8] T. Widiyaningtyas, H. Hairani, D. D. Prasetya, U. Pujiyanto, and W. Caesarendra, "A Modified SMOTE with Noise Filtering and Manhattan Distance Metric Approach to Address Imbalanced Health Datasets," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25452–25459, Aug. 2025, <https://doi.org/10.48084/etasr.11925>.
- [9] R. Suguna, J. Suriya Prakash, H. Aditya Pai, T. R. Mahesh, V. Vinoth Kumar, and T. E. Yimer, "Mitigating class imbalance in churn prediction with ensemble methods and SMOTE," *Scientific Reports*, vol. 15, no. 1, May 2025, Art. no. 16256, <https://doi.org/10.1038/s41598-025-01031-0>.
- [10] A. V. Kolmogorova, A. A. Kalinin, and A. V. Malikova, "Linguistic Principles and Computational Linguistics Methods for the Purposes of Sentiment Analysis of Russian Texts," *Current Issues in Philology and Pedagogical Linguistics*, no. 1, pp. 139–148, 2018, [https://doi.org/10.29025/2079-6021-2018-1\(29\)-139-148](https://doi.org/10.29025/2079-6021-2018-1(29)-139-148).
- [11] V. D. Oliseenko, M. V. Abramov, and A. L. Tulupyev, "Neural networks with lstm and gru in application to the task of multiclass classification of text posts of social network users," *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, no. 4, pp. 130–141, Dec. 2021, <https://doi.org/10.17308/sait.2021.4/3803>.
- [12] K. V. Lagutina, "Classification of Russian Texts by Genres Based on Modern Embeddings and Rhythm," *Modeling and Analysis of Information Systems*, vol. 29, no. 4, pp. 334–347, Dec. 2022, <https://doi.org/10.18255/1818-1015-2022-4-334-347>.
- [13] Z. Ayan, B. Alimjan, M. Olga, Z. Timur, and Z. Toktalyk, "Quality of service management in telecommunication network using machine learning technique," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, Nov. 2023, Art. no. 1022, <https://doi.org/10.11591/ijeecs.v32.i2.pp1022-1030>.
- [14] N. Kunicina, K. Rakhimova, J. Caiko, and M. Mansurova, "Adaptive Multidimensional E-Learning Systems," in *2025 IEEE 12th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, Vilnius, Lithuania, May 2025, pp. 1–6, <https://doi.org/10.1109/AIEEE66149.2025.11050765>.

-
- [15] B. Zhao *et al.*, "Design and Optimization of an Internet of Things-Based Cloud Platform for Autonomous Agricultural Machinery Using Narrowband Internet of Things and 5G Dual-Channel Communication," *Electronics*, vol. 14, no. 8, Jan. 2025, Art. no. 1672, <https://doi.org/10.3390/electronics14081672>.
- [16] A. Solochshenko, A. Baikenov, V. Tikhvinskiy, and J. Caiko, "Research of Self – Organizing Networks (SON) Algorithms Efficiency Applying on Fourth – Generation Mobile Networks," *Transport and Telecommunication Journal*, vol. 22, no. 4, pp. 444–452, Nov. 2021, <https://doi.org/10.2478/tjt-2021-0034>.
- [17] M. Li, Q. Gao, and T. Yu, "Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters," *BMC Cancer*, vol. 23, no. 1, Aug. 2023, Art. no. 799, <https://doi.org/10.1186/s12885-023-11325-z>.
- [18] W. L. Seow, I. Chaturvedi, A. Hogarth, R. Mao, and E. Cambria, "A review of named entity recognition: from learning methods to modelling paradigms and tasks," *Artificial Intelligence Review*, vol. 58, no. 10, Jul. 2025, Art. no. 315, <https://doi.org/10.1007/s10462-025-11321-8>.
- [19] N. A. Helal, A. Hassan, N. L. Badr, and Y. M. Afify, "A contextual-based approach for sarcasm detection," *Scientific Reports*, vol. 14, no. 1, Jul. 2024, Art. no. 15415, <https://doi.org/10.1038/s41598-024-65217-8>.
- [20] M. Krapivina, "Russian hero context dataset (balanced)." Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/miladakrap/russian-hero-context-dataset-balanced>.