Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

A semi-Automatic Annotation with a Novel Deep Learning Model for Terrorism Tweets

Seror Imad Abbas [1*], Enas Fadhil Abdullah [1]

[1]University of Kufa, College of Education, Department of Computer Science, Iraq.
E-mail: sururi.abbas@student.uokufa.edu.iq
E-mail: inasf.altueky@uokufa.edu.iq
*Corresponding author E-mail: almwswyasdallh@gmail.com

Article's Information

Received: 09.01.2026
Accepted: 20.03.2026
Published: 31.03.2026

Keywords:
classification, deep learning , Annotation , Labeling, XlM-R Large

Abstract

Terrorist groups are resorting more to social networks to promote their operations through Twitter. To attract new members, these groups employ radical propaganda by posting content. Sentiment-based natural language processing methods are widely used in practice to pre-process such posts or tweets .A Semi-automatic annotation approach was used. the first set of labels were created automatically with the help of a Zero-Shot classifier (ZSc) , which text to category classification, after which random tweets were manually verified. The labelled dataset was then finally used to train the model. This paper gives a developed system of terrorism-content classification utilizing the Cross-linguistic Language Model RoBERTa Large (XLM-R Large) to categorize information in the Arabic and English language in Twitter using structured and labelled information. The accuracy of our system, which involved the combination of automated labeling and deep learning, was 85% ,and it was successful in identifying extremist and terrorist materials.

https://doi.org/10.46649/fjiece.v5.1.7a.31.3.2026
*Corresponding author: almwswyasdallh@gmail.com

1. INTRODUCTION

Social media sources are now one of the most powerful channels of information distribution and shaping the opinion of the audience of the digital age. No longer serving as a means of social interaction, these platforms have turned into virtual arenas for building political and ideological discourses and influencing audiences psychologically. With the initial growth of the internet and the development of the social media space, the digital space has become a main arena of intellectual and ideological conflict, as an increasing number of extremist actors and groups use the digital space to spread ideological narratives, polarize and exert influence networks, and manipulate the opinion of the population on the national and global scales [9]. The revolution has worked well in making the social media sites open for intellectual

rivalry and ideological debate. Twitter has become one of the most powerful social media platforms in this category as it quickly disseminates content and can use short texting as its fundamental means of communication. These features have been utilized by extremist and terrorist organizations to deliver short, but very effective tweets and posts through which they share extremist ideologies, create propaganda scripts, and gain adherents or followers in the online environment. Recent research shows that these groups are turning to these platforms more and more to manipulate individuals and disseminate their ideas, which all points to the fact that smart systems that are capable of picking up subtle indicators of linguistic messages hidden in such types of material are highly needed [1]. Twitter increased the maximum size of the tweet by twice and reached 280 characters in November 2017, which is an additional consideration in the analysis of Twitter data [8]. On the one hand, such an enlargement allows more expressiveness; on the other hand, it adds to the risks of platform abuse by offering the opportunity to make condensed , emotionally loaded messages that can be spread very quickly and produce a strong psychological effect in just a few seconds. This has made Twitter a rich ground where terrorism and extremist discourse can proliferate. The analysis of multilingual text has become even more precise than the classical one with the fast development of Natural Language Processing (NLP) and deep learning models, specifically, Transformer-based ones, including BERT, RoBERTa, and XLM-RoBERTa [2],[3]. Some of the studies have concentrated on early radicalization recognition based on deep contextual representation , hence enhancing the capability of models to categorize extremist material [3]. The effectiveness of these models has been proven by other studies showing the presence of these models in identifying extremist discourse on different social media [4], and how the use of Twitter-specific features ,including linguistic symbols and lexical patterns, has been important in detecting violent or extremist discourse [5].Recent experiments have also noted the need to focus on multilingual or code-mixed settings, like Arabic-English ones, in which more complex multilingual models have been found to be the method that works better in addressing such difficulties [6],[7]. However, major gaps exist in the previous studies ,such as a lack of longitudinal study, failure to emphasize bilingual material content, and enough attention of implicit or symbolic manifestations that are usually expressed in extremist writings. In order to fill these gaps, the current research paper applies the XLM-RoBERTa Large model to perform an elaborate linguistic and lexical analysis of the extremist content on Twitter. By extensive pre-processing and annotation-labeling approach, the model is optimized to categorize terrorist and extremist-related tweets, and it offers a highly robust, scalable, and multilingual framework, which can identify even the slightest and implicit extremist utterances. Therefore, this study will contribute to the advancement of the research on the dynamics of extremist communication and open the way to results with empirical scientific and practical significance.

2. RELATED WORK

The prior studies on the detection of extremist and terrorist content have involved a plethora of analytical and computational methods, which are the consequence of the nature of extremist discussions on social media resources. The same Kaggle dataset was used in the current study to identify the contextual semantic differences between pro and non-ISIS users; however, the lexicon-based methodology used by Fernandez and Alani [10] lacked the capability to identify implicit or nuanced extremist meanings. Fuhrman et al. [11] suggested a similar hybrid, in an opposing direction, where automated processing is used to extract geographical indicators using manual coding on extremist texts. However, they were limited by the lack of detailed multilingual gazetteers , especially in Arabic-English formation.

As the field of deep learning has developed, a methodological shift to neural architectures has occurred in this field. Alshaabi et al. [12] implemented Transformer-based models (BERT) and LSTM networks on large-scale datasets of Twitter, and a significant positive change in performance was observed. However, their models still experienced the issues of data noise, multilinguality, and contextual overlap that are typically present in the social media data in real-life. The MIWS multi-ideology dataset by Crampton et al. [13] was a new resource on extremist-content analysis, but it had a significant shortcoming, namely with a class imbalance and a small seed, which limited the ability to generalize a model. Equally, Al-Hassan and Al-Dossari [14] reported the usefulness of deep models like AraBERT in the classification of Arabic extremist-related content, and noted that the challenges associated with the classification of extremist content were still present, including the problem of dialectal variation and the problem of implicit hate speech.

On the data-annotation (labeling) level, several works have highlighted the extreme level of cost, subjectivity, and lack of scalability of manual annotation in terms of extremist content. Recent efforts to overcome these drawbacks have considered weakly supervised and automatic labeling (pseudo-labeling) techniques, based on pre-trained language models, label-name classification, and either zero-shot or self-training systems to build large-scale annotated datasets [18],[19]. Despite the success of these methods, especially in a multilingual and low-resource context, they come with the issue of label noise and confusion between related categories like extremism and terrorism. In spite of them, pseudo-labeling is being seen as a reasonable trade-off between annotation quality and scalability. On the dataset level, Karimi et al. [15] published a massive longitudinal corpus of over 9.9 million tweets about ISIS, which provides valuable time-related information about the patterns of extremist communications. Nevertheless, deleting accounts and limiting APIs influenced the completeness of datasets. Arabic-targeted experiments by Al-Qurishi et al. [16] have obtained good classification performance with machine-learning pipelines, but their models have been prone to dialect, as well as satirical expression. Later deep-learning research, including [17], has shown that RoBERTa is able to perform highly in multi-class classification of extremism, though not on English-language tweets. Al-Zahrani et al. [20] have also recently presented the RADAR ensemble, which has state-of-the-art performance; however, surrounding ambiguity, particularly of sensitive extremist terms, is still a puzzle.

In general, even though significant advancements were achieved in the construction of datasets and the detection of extremists and their content with the help of deep learning, serious gaps are still present in the field of multilingual robustness and scalable annotation. Specifically, the combination of automatic annotation and labeling mechanisms and multilingual Transformer-based models is not well studied. This literature void is filled with the current study proposing a bilingual XLM-RoBERTa-based model.

3. NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) is one of the central fields of artificial intelligence that deals with text processing and semantic information retrieval [21]. NLP methods were used in this research, where a structured pre-processing phase was carried out which involved the ability to remove linguistic noise, cut symbols and URLs , and also the normalization of Arabic and English text in the ISIS twitter census data. The pseudo-labeling method was then used to label the hitherto unlabelled data using multilingual transformer-based language models deployed on the Hugging Face platform [22]. This method minimized the use of manual annotation and contributed to the growth of the labelled data. All in

all, NLP techniques allowed converting raw textual data into structured forms that could be used to classify it using deep learning in the proposed framework.

4. ZERO-SHOT TEXT CLASSIFICATION(ZSC)

The concept of zero-shot classification is regarded as one of the recent trends in natural language processing , as it allows defining texts by pre-existing categories, and no task-specific and labelled training data are required to do so. In textual scenarios, models that have been trained on Natural Language Inference (NLI) tasks typically provide zero-shot classification ,with the semantic relationship between the input text and the candidate labels being determined based on contextual knowledge instead of a simple, surface-level lexical match **[**18]. The label that has the greatest semantic compatibility is chosen. Previous studies have shown that this model has proven effective when using large, multilingual data sets, where it can achieve good levels of semantic accuracy[18][25].

5.DEEP LEARNING (XLM-ROBERTA LARGE)

XLM-RoBERTa Large model is an efficient multilingual transformer-based architecture and an improved development of the evolutionary trend, which BERT started and evolved by RoBERTa. It was made especially to address the shortcomings of monolingual models by being trained on very large-scale multilingual corpora of over 100 languages. The model is designed on the basis of the Transformer architecture and the self-attention mechanism, allowing it to model relationships between all the tokens in a sentence and extract fully contextualized representations of words using a deep stack of processing layers [23]. The Large variant has 24 encoder layers and has a high amount of parameters (~550M), which offers a better representational capacity of deep contextual knowledge, implicit semantic knowledge, ambiguity resolutions, and complicated semantic relationship modelling. These qualities render the model especially appropriate in sensitive and difficult tasks, including the classification of extremist content in a multilingual environment. To achieve its results in this study, the XLM-RoBERTa Large was used to produce the high-quality semantic representations of both Arabic and English texts, which led to the significant enhancement of differentiating between terrorism and extremism and the further improvement of the overall accuracy and strength of the proposed binary classification system. The latter was also manifested by the quality of higher accuracy and F1-score in the last experimental analysis.

The XLM-RoBERTa Base model is the base version of the same architecture and shares the same architecture principles and mechanisms of self-attention , but only 12 encoder layers, which significantly decreases computational complexity. Therefore, it allows training and inference in less time and using fewer resources , and at competitive speeds in text classification and feature extraction activities. Hence, the Large variant should be used in cases where the major goal is to maximize classification accuracy and the Base variant is a more convenient and efficient option in cases where resources are limited. The building of the XLM-RoBERTa model is presented in Figure 1, and its main structural parameters are summarized in Table 1, and facilitate multilingual semantic representation [23].

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

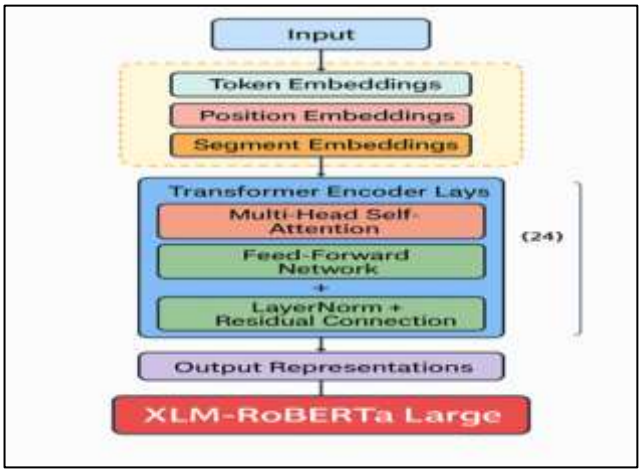

Fig. 1. Architecture of XLM-RoBERTa Large.

Table 1. Structure of XLM-RoBERTa Large (Parameter Initialization).

Specification	XLM-RoBERTa Large
Number of Encoder Layers	24
Attention Heads per Layer	16
Hidden Size	1024
Feed-Forward Network Size	4096
Maximum Sequence Length	514
Vocabulary Size	~250,000
Total Parameters	~550 Million
Architecture Type	Transformer Encoder Only
Languages Supported	>100 Languages

6. MATERIALS AND METHODS

6.1.PROPOSED METHODOLOGY

Figure 2 shows the general architecture of the suggested system. The framework is structured into a chain of states having six main steps: dataset gathering, exploratory data analysis preceding labeling, preprocessing, annotation\labeling, model training and evaluation.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

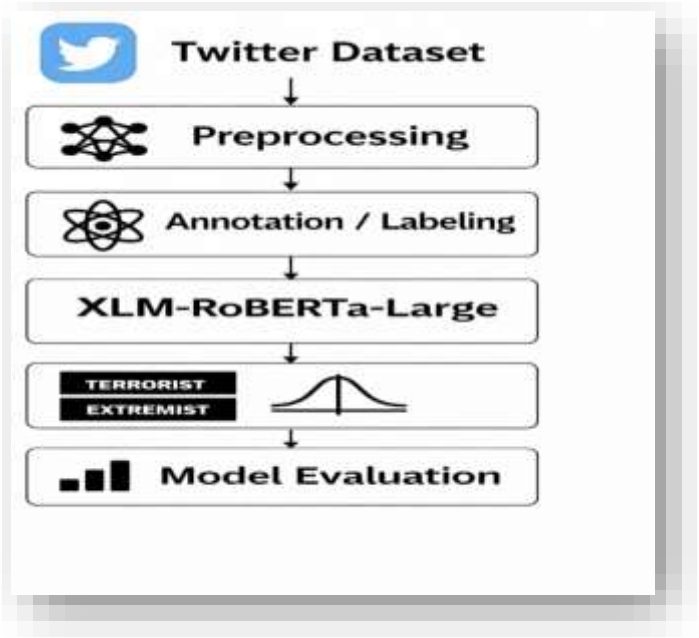

Fig. 2. Proposed Methodology Diagram.

6.2.DATASET COLLECTION

A Twitter dataset that was relevant to the subject of terrorism was obtained using the Kaggle platform [24] and filtered to only Arabic and English tweets in order to provide linguistic uniformity. The same dataset has been used in the previous research, including the analysis conducted by Fernandez and Alani [10], who took the ISIS Twitter Census as the basis to study the extremist language and observed the semantic contextual variations between pro-ISIS and non-ISIS users. That further represents the usefulness and significance of this dataset in the study of extremist content. An exploratory analysis was performed on the dataset to determine the most repeated terms and the prominent themes before any sort of labeling or categorization was carried out. Visualization was also created in the form of a word cloud, in order to give an initial visualization of the most common linguistic patterns in the data.

6.3.TEXT PREPROCESSING (TWEETER)

The whole cleaning and linguistic normalization of the texts were done according to the familiar Natural Language Processing (NLP) techniques to standardize the text representations, reduce the noise, and deliver the consistency of the information. This initial preparation step trained the data in a fully-fledged structured and normalized form of data to the extent that it could easily move to the subsequent phase of the entire process, which is the process of classification using the deep learning technique.

Considering the noisy and unstructured character of the ISIS-related Twitter data, the pre-processing pipeline aimed at eliminating symbolic irrelevantness, repetitive material, and linguistic inconsistency that might have a detrimental impact on the model performance. These procedures converted the unprocessed archival data into a clean and analyzable resource to be used in modeling using deep learning methods.

The pre-processing enabled the maintenance of meaningful contextual information in both Arabic and English tweets due to the use of solid multilingual normalization techniques. It increased the quality and consistency of data, which increased the capacity of large prior trained language models to acquire discriminative linguistic and semantic patterns throughout the training process, which eventually lead to more reliable and consistent classification results.

6.4.ANNOTATION/LABELING

Since the data used in the given study was not initially labelled, an annotation (labeling) phase was added as a basic preparatory measure before supervised learning. This step was performed with the help of model-assisted automatic annotation with reference to large-scale multilingual pre-trained language models, which have high-energy functions of capturing deep semantics and hidden meanings in the content of tweets. Notably, no task-specific model training or parameter updates were used in this process.

Using the contextual representation strength of current transformer-based language models, the annotation procedure facilitated the discovery of latent semantic conventions and indirect clues that are linked to the discourse of extremists on more than just lexical patterns. All the tweets were automatically grouped into two ideologically categorized sets, namely, terrorism and extremism, based on the score of semantic compatibility of the pre-trained models.

In order to increase the reliability of the labels and minimize the effects of the potential annotation noise, (≥ 0.7) confidence threshold was used on the model outputs. Tweets with confidence scores that are at least confidence scores were seen as high-confidence cases and were not edited, but twitter tweets with low confidence scores were termed as low-confidence cases. The researcher then manually revised and edited a randomly selected sample of low-confidence tweets to enhance the quality of annotated tweets. The checked labels were subsequently recombined into the full set of annotations, which were done automatically, into the annotations of the reviewed samples.

Comprehensively, this annotation process, as a set of model-assisted automatic labeling and human verification, generated a consistent and trustworthy labelled dataset to be further used in downstream supervised training and evaluation of deep learning models. Importantly, the manually corrected samples were only utilized to refine the final dataset and were not used to retrain or update the annotation models and thus provide a clear methodological distinction between the annotation phase and the following classification phase[18][19].

To assess annotation noise, 300 low-confidence tweets (score < 0.7) were manually reviewed. Of these, 114 samples were revised, resulting in an estimated noise rate of approximately 38%. This relatively high correction rate is expected, as the evaluation focuses on low-confidence predictions that are inherently more uncertain. The findings demonstrate that the selected confidence threshold effectively isolates potentially noisy labels, while high-confidence samples are expected to exhibit significantly lower noise levels. Furthermore, the combination of manual validation and the use of Focal Loss during training helps mitigate the negative impact of label noise on overall model performance.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

The specifics of the 2 steps of the tweet preprocessing and labeling were presented in Algorithm (1), the following way:

Algorithm (1): Multilingual Social Media and Annotation/Labeling (MSMaAL)

Input: Multilingual Twitter tweets in their raw form (unstructured)

Output: List of tweets, which are cleaned, structured, and labelled, and which are grouped based on confidence.

Begin

Step 1: Pre-processing of texts (t)

1. Change English words into lowercase.
2. Remove punctuation marks (e.g., ! $ * ? / & . ; ") and special symbols.
3. Remove numbers (e.g., 3, 4, 567).
4. Remove URLs (http, https, www).
5. Remove user mentions and hash tags (e.g., @user, #topic).
6. Get rid of emojis and non-textual characters.
7. Normalize Arabic text:
    - Unify Alef forms (ا → آ ،إ ،أ).
    - Remove diacritics (tashkeel).
8. Deal with Arabic and English stopwords.
9. Eliminate unnecessary gaps and cut down the text.

Step 2: Dataset Cleaning

10. Eliminate the duplicated tweets in the dataset.
11. Cleaning After cleaning, remove empty or very short tweets.

Step 3: Automatic Annotation/Labeling (Zero-Shot)

12. Pre- trained multilingual zero-shot classifier (Hugging Face).
13. Target labels: Extremism, Terrorism.
14. For each cleaned tweet t:
    - Obtain prediction scores for the two labels.
    - Assign **label(t)** as the label with the highest score.
    - Let **score(t)** be the highest prediction score.

Step 4: Confidence Grouping (Thresholding)

15. Apply confidence threshold θ=0.7\theta = 0.7θ=0.7:
    - If **score(t) ≥ 0.7**, assign **confidence_group = High**.
    - Else, assign **confidence_group = Low** (candidate for manual review).

Step 5: Return

16. Response to each tweet ,sending: tweet-id ,cleaned text, predicted label, prediction score , and confidence group.

End.

6.5.MODEL TRAINING

The labelled data were stratified with a stratified sampling technique into 70% training, 10% validation, and 20% testing sets such that the original class proportions were maintained in all the splits. To deal with the problem of the imbalance in classes, class weighting with Focal Loss was used

during the training process to improve the capacity of the model to pick patterns on the minority class and minimize bias in predictions.

Fine-tuning on the model XLM-RoBERTa Large was applied to binary tweet classification depending on the following categories :Terrorism (1) and Extremism (0). The validation set was used to track the model performance throughout the training process, to encourage early termination ,as well as to decide on the proper level of decision. The test set, on the contrary, was to be used only in the final evaluation stage in order to provide an objective and unbiased evaluation of the model's performance.

Also, the training process was based on the rich textual representations in the context of contextual texts, which were offered by pretrained multilingual models via the Hugging Face library. Such representations allowed good fine-tuning and allowed the model to enjoy the advantages of good multilingual semantic features that eventually resulted in the robustness, stability, and general accuracy of the classification system.

6.6.MODEL EVALUTION

Standard measures were used to analyze the performance of the model on the test set, i.e. Accuracy, Precision, Recall, and F1-Score, as well as the Confusion Matrix. Besides, the Receiver Operating Characteristic (ROC) curve and associated Area Under the Curve (AUC) were also provided in order to further determine the discriminatory potential of the model. Combining these evaluation indicators, one can get a detailed picture of the performance of the classifier in terms of its ability to confidently discriminate between the category of terrorism and the category of extremism.

7. RESULTS AND DISCUSSION

After the language-filtering process, only Arabic and English tweets were left to make sure that there was linguistic consistency. The results of the distribution of the remaining tweets are shown in Table 2) after the cleaning process has been done. The dataset post-filtering comprises English tweets as the highest percentage, as illustrated.

Table 2. Distribution of Arabic and English Tweets After Language Filtering and Cleaning.

language	No.tweets	percentage
Arabic	631	4.26%
English	14.164	95.73
Total	14.795	100%

The findings showed good results throughout the evaluation process, and the proposed model recorded a general accuracy of 85% on the test set. To Further confirm the utility of the proposed model, a comparative analysis has been done using XLM-RoBERTa Base as a baseline model under the same experimental conditions. As shown in the table 3, the Large model is always better in all assessment

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

measures compared to the Base model. Specifically, the enhancement can be traced more in the minority category (Terrorist), in which the Large model has scored a higher F1-score than the Base model. This increase in performance is explained by the richer architecture and greater representational capacity of XLM-RoBERTa Large. Nevertheless, it continued to be a competitive Base model with reduced computational cost.

Table 3. Performance Evaluation of XLM-RoBERTa and SVM Models in class-wise.

Model	Class	Precision	Recall	F1-score	Support
XLM-RoBERTa Large	Extremist	0.91	0.89	0.90	2292
	Terrorist	0.65	0.72	0.68	667
XLM-RoBERTa Base	Extremist	0.90	0.86	0.88	2292
	Terrorist	0.58	0.69	0.63	667
SVM (TF-IDF)	Extremist	0.89	0.82	0.85	2292
	Terrorist	0.52	0.67	0.58	667

XLM-RoBERTa Large model performed the most with the highest accuracy of 0.85, whereas the Base model had an accuracy of 0.82, and the SVM model had an accuracy of 0.78. Since there is an imbalance in the datasets, accuracy on its own is not accurate; hence, class-wise testing offers a more valid assessment, especially on the minority classification. It was also reported that XLM-RoBERTa was more effective than traditional approaches as it was able to capture contextual representations.

Moreover, Figure 3 shows the confusion matrix on the basis of the test data , which allows seeing more deeply into the classification behavior of the model and its capability to recognize both of the semantic classes properly. Similarly, Figure 4 represents the ROC curve, which resulted in a large AUC value, attesting to the strong discriminating power of the model at various classification undercuts.

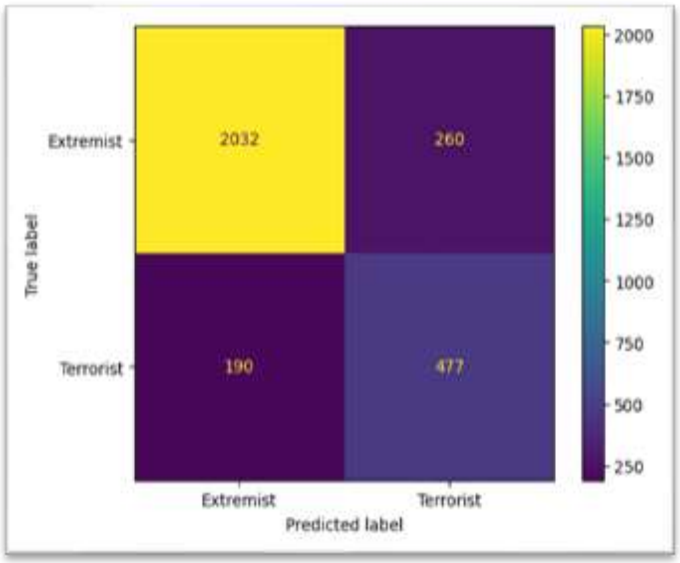

Fig. 3. Confusion Matrix of the Proposed Multilingual Classification Model.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

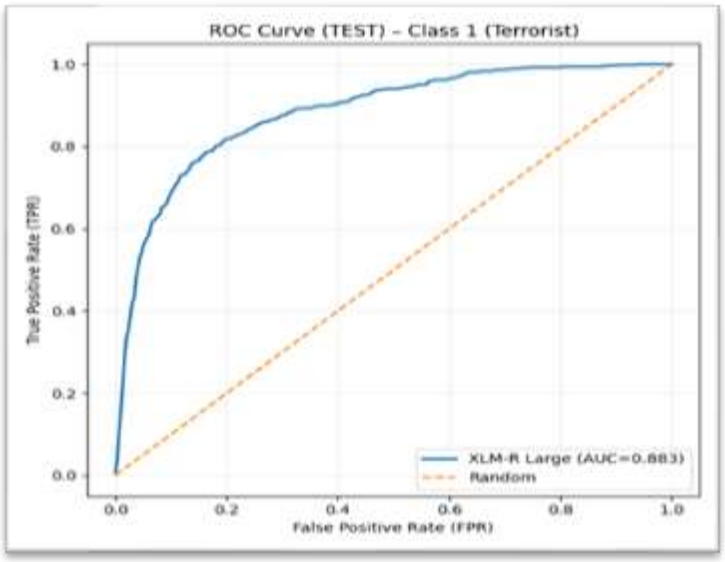

Fig. 4. ROC Curve for Terrorist Tweet Classification.

Table 4. Pre-processing and annotation-Labeling Results of Tweets.

Tweet Before Preprocessing	Tweet After Preprocessing	score	Tweet Before Preprocessing
RT Hamza_sy9 بحجة داعش قصفت إيران و روسيا و التحالف الدولي أطفالنا وأهلنا ومدننا ودمروا منازلنا فهل علمتم لماذا وجدت داعش	rt hamzasy0 بحجة داعش قصفت إيران و روسيا و التحالف الدولي أطفالنا وأهلنا ومدننا ودمروا منازلنا فهل علمتم لماذا وجدت داعش	0.965	Terrorist
Masive Attack khilafah soldiers safawi army position injarayshi nd tarahroad	masive attack khilafah soldiers safawi army position injarayshi tarahroad	0.820	Extremist
Killed 12iraqi soldiers by ISIS attack began wit 2car mines on army HQ in alurdani hospital East Fallujah amp ended with control on it	killed iraqi soldiers by isis attack began with car mines on army hq alurdani hospital east fallujah ended control	0.504	Terrorist

Table 4 shows some representative samples of tweets prior to and following the preprocessing step, and the results of the applied classification. The score of the confidence of each of the tweets and the class that is attributed to it (Terrorism / Extremism) is also reported in the table, which underlines the success of the classification process that is adopted.

4. CONCLUSIONS

The paper has shown that automated and manual labeling strategies can be used together to offer a strong platform on which to develop effective terrorism-content classification models. The labeling procedure was founded on the automatic labeling with the help of ZSC, which was supplemented with the manual labeling of randomly sampled data, after which the latter were combined with the primary dataset

to facilitate the possibility of training the model on this supplemented dataset. The findings demonstrate that a unified solution, namely, intensive pre-processing, hybrid labeling, and sophisticated models of multilingual deep learning, such as XLM-R Large, can be effective in reaching high accuracy and reliability in differentiating between extremist and terrorist material posted on social media in Arabic and English. XLM-RoBERTa Large model scored the highest accuracy of 0.85 as opposed to Base model (0.82) and SVM model (0.78), thus showing that it has a great potential to deal with multilingual data. Another significant finding of the results is that it is effective in combating the issue of class imbalance and enhancing the process of minority classes detection due to its sophisticated capacity to represent contexts.

REFERENCES

[1] Q. Hu, Y. Fang, and X. Ruan, "Detecting Extremist Messages in Social Media Using BERT," *arXiv preprint arXiv:2103.05314, 2021*.
[2] M. Mozafari, R. Farahbakhsh, and N. Crespi, "Multilingual Hate Speech Detection Using Transformer Models," *arXiv preprint arXiv:*2205.12438, 2022.
[3] S. Mansour et al., "Early Detection of Radicalization on Social Media Using Deep Learning," *arXiv preprint arXiv:*2202.03625, 2022.
[4] F. Alam et al., "A Transformer-Based System for Detecting Online Extremism," *arXiv preprint arXiv*:2305.19043, 2023.
[5] P. Saha et al., "Detecting Violent Extremism on Twitter Using NLP Techniques," *arXiv preprint arXiv:*2303.00080, 2023.
[6] R. Gupta, S. Kataria, and M. Imran, "Analyzing Online Extremism Using Language Models and Social Graphs," *arXiv preprint arXiv*:2311.09479, 2023.
[7] H. Alotaibi et al., "Transformer-Based Multilingual Extremism Classification on Social Media*," arXiv preprint arXiv*:2403.01208, 2024.
[8] K. Gligorić, A. Anderson, and R. West, "Adoption of Twitter's New Length Limit: Is 280 the New 140?," *arXiv preprint arXiv*:2009.07661, 2020.
[9] R. Ravi, "Ideological orientation and extremism detection in online environments," *Journal of Information Security and Applications,* vol. 82, Art. no. 103845, 2024.
[10] M. Fernández and H. Alani, "Contextual Semantics for Radicalisation Detection on Twitter," *CEUR Workshop Proceedings,* 2018.
[11] J. Fuhriman et al., "Extracting and Analyzing Geographical Perspectives of Terrorists," *Journal of Terrorism Studies, 2020*.
[12] T. Alshaabi et al., "Event Detection on Twitter Using Transformer-Based Models," *Procedia Computer Science,* 2021.
[13] C. Crampton et al., "MIWS: Multi-Ideology White Supremacist and Jihadist Dataset," 2021.
[14] S. Al-Hassan and A. Al-Dossari, "Arabic Extremist Content Classification Using AraBERT," 2021.
[15] A. Karimi et al., "A Longitudinal Dataset of ISIS-Related Tweets," 2022.
[16] A. Al-Qurishi et al., "Arabic Extremist Content Detection Using Machine Learning," 2022.
[17] R. Al-Taie et al., "Extremism Detection on Twitter Using RoBERTa," 2022.
[18] Y. Meng et al., "Text Classification Using Label Names Only*," Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL),* 2020.
[19] T. Schick and H. Schütze, "Exploiting Cloze Questions for Few-Shot Text Classification and Natural Language Inference," *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL),* pp. 255–269, 2021.
[20] M. Al-Zahrani et al., "RADAR#: Hybrid CNN-LSTM and AraBERT Model," 2025.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.
[22] T. Wolf, L. Debut, V. Sanh, et al., "Transformers: State-of-the-Art Natural Language Processing," *EMNLP (System Demonstrations*), pp. 38–45, 2020.
[23] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale," *ACL*, pp. 8440–8451, 2020.
[24]FifthTribe,"How ISIS Uses Twitter," *Kaggle*, 2015. Available: https://www.kaggle.com/datasets/fifthtribe/how-isis-uses-twitter
[25] Y. Zhang, X. Liu, J. Zhao, and Z. Wang, "Zero-Shot Text Classification via Natural Language Inference: A Survey and Benchmark," *arXiv preprint arXiv*:2211.15247, 2022.