Impact of Emojis Exclusion on the Performance of Arabic Sarcasm Detection Models

Ghalyah Aleryani[1], Wael Deabes[2], Khaled Albishre[3], Alaa E. Abdel-Hakim[4]
Department of Computer Science in Jamoum
Umm Al-Qura University, Makkah, 25371, Saudi Arabia[1]
Department of Computational, Engineering Mathematical Sciences (CEMS),
Texas A&M University-San Antonio, San Antonio, 78224, USA[2]
Computers and Systems Engineering Department, Mansoura University, Mansoura, 35516, Egypt[2]
Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah, 25371, Saudi Arabia[3]
Department of Computer Science in Jamoum, Umm Al-Qura University, Makkah, 25371, Saudi Arabia[4]

***Abstract*—The complex challenge of detecting sarcasm in Arabic speech on social media is exacerbated by the language's diversity and the nature of sarcastic expressions. There is a significant gap in the capability of existing models to effectively interpret sarcasm in Arabic, necessitating more sophisticated and precise detection methods. In this paper, we investigate the impact of a fundamental preprocessing component on sarcasm detection. While emojis play a crucial role in mitigating the absence of body language and facial expressions in modern communication, their impact on automated text analysis, particularly in sarcasm detection, remains underexplored. We examine the effect of excluding emojis from datasets on the performance of sarcasm detection models in social media content for Arabic, a language with a super-rich vocabulary. This investigation includes the adaptation and enhancement of AraBERT pre-training models by specifically excluding emojis to improve sarcasm detection capabilities. We use AraBERT pre-training to refine the specified models, demonstrating that the removal of emojis can significantly boost the accuracy of sarcasm detection. This approach facilitates a more refined interpretation of language, eliminating the potential confusion introduced by non-textual elements. The evaluated AraBERT models, through the focused strategy of emojis removal, adeptly navigate the complexities of Arabic sarcasm. This study establishes new benchmarks in Arabic natural language processing and offers valuable insights for social media platforms.**

***Keywords*—Arabic language; AraBERT; sarcasm detecting; data preprocessing; emojis impact; social media content**

I. Introduction

The evolution of social media platforms has transformed them into spaces for free speech, allowing users to express their ideas and opinions openly. While this open environment encourages meaningful discussions, it can also lead to problems when individuals use expressions or statements that may offend others due to differences in beliefs, backgrounds, gender, or race. Although such speech is protected under the latest version of the Communications Decency Act (CDA 230) [1], many social media platforms are actively working to enhance user protection against hateful and offensive content. A significant challenge, however, arises from their reliance on human monitoring and user reports [2], [3].

In recent years, there has been a significant increase in the prevalence of sarcastic speech on social media, giving rise to serious social problems, including social conflicts, racist crimes, and the spread of negative social influence. To mitigate this issue, many social media platforms have implemented word filters based on NLP techniques. Sarcastic speech includes exchanges of sarcastic or offensive remarks between individuals and extends beyond text comments to include multimedia content such as videos and audio [4]. Existing research efforts have primarily focused on proposing models for the automatic detection of sarcastic speech within text comments on platforms like Twitter or YouTube [5], [6]. These models typically involve several stages, including data processing, representation, and classification. However, these stages present several challenges, particularly concerning the Arabic language [7], [8]:

1) Its rich vocabulary, and dialectical variations.
2) Arabic sarcasm often relies heavily on contextual indications and cultural references.
3) Collecting and annotating a large dataset for Arabic sarcasm detection produces unique difficulties.
4) Time-consuming and requires extensive data learning.
5) The multimodal nature of social media content is integrating information from various sources such as text, images, videos, and emojis.

Generally, sarcasm detection heavily relies on textual data classification. However, textual data lacks crucial expressive features such as facial expressions, body language, and tone variations. To address this limitation, social media communities have introduced emojis as non-traditional vocabulary to compensate for this deficiency in information. Emojis have demonstrated their effectiveness in partially bridging the gap between textual and vocal/visual communication [9].

Nevertheless, rich languages like Arabic possess their own tools that can better address this gap than emojis, including a vast and diverse lexicon. Table I provides a comparison of the number of roots between Arabic and other languages. Additionally, Arabic speakers use numerous dialects with significant dialectical variations. Moreover, sarcasm in Arabic heavily relies on contextual cues, puns, and euphemisms. This results in an extensive textual information space generated solely using textual vocabulary, decreasing the contribution of the limited emojis dictionary.

These factors raise questions regarding the added value of emojis to classifier performance. While it is intuitive that

TABLE I. COMPARISON BETWEEN ARABIC AND SOME OTHER LANGUAGES IN TERMS OF THE NUMBER OF ROOTS

Language	Approximate Number of Roots	Notes
Arabic	23090 [11]	Roots of 3 letters By applying 73 triliteral patterns and 18 affixes produced around 27.6M words.
English	8,400 [12]	This study highlights the significant growth in root word vocabulary during the primary school years.
Russian	450 [13]	The study highlights the significance of understanding root words for mastering Russian vocabulary.

adding redundant data should not degrade performance, excess data has been shown to harm classifiers' performance in certain instances [10]. Hence, we hypothesize that using emojis for sarcasm detection in rich languages like Arabic may either reduce accuracy or offer no improvement. The rationale behind this approach is that focusing purely on the textual elements allows the models to concentrate on linguistic and semantic analysis without the potential confounds introduced by emojis.

The successful development of an Arabic sarcasm detection model, enhanced by transfer learning methods and refined by the strategic removal of emojis from the dataset, will have a significant social impact. This approach will enable social media platforms to more effectively detect and moderate sarcastic speech, thereby mitigating its potentially harmful effects and fostering a more positive and respectful online discourse. By focusing on the textual content and reducing noise in the data, the model's precision in identifying sarcasm will be improved, making the moderation process more efficient and reliable. Additionally, this research will lay the groundwork for further advancements in the creation of language-specific models for sarcasm detection, equipping diverse language communities with the tools needed to address such speech more effectively.

In this work, we investigate the following research questions: 1.

1) Does including emojis in the data improve pre-trained models' ability to detect sarcasm in the Arabic language?
2) How can the performance of the AraBERT pre-trained models for sarcasm detection in the Arabic language on social media platforms be improved by removing emojis from the data?
3) How accurately can it identify and classify sarcasm in Arabic speech on social media?

In the next section, we discuss related work on detecting textual aggressions on social media. In Section III, we describe the methodology used in this study. Section IV presents the experiments and analysis of the results. Finally, Section V concludes the study and discusses the future direction of sarcasm speech in the Arabic language.

II. RELATED WORK

User-generated content online, especially on social media platforms, can sometimes contain harmful language and hate speech, which can have detrimental effects on the online community and potentially lead to hate crimes. Recently, there has been a marked interest in smart algorithms designed to automatically identify and flag such offensive language and hate speech. Nonetheless, NLP research concerning the Arabic language is typically limited [14], as is the investigation into sarcasm detection. In this section, we aim to highlight previous efforts focused on detecting Arabic sarcasm on social media.

In the field of NLP, large-scale pre-trained models have become the standard approach for a wide range of tasks. Models like Bidirectional Encoder Representations from Transformers (BERT) are trained on massive datasets, allowing them to generalize effectively to various downstream tasks [6]. The BERT model has been employed for extensive Arabic datasets, resulting in the creation of AraBERT [15]. In [16], an automated approach to detect offensive language and detailed hate speech in Arabic tweets is proposed. The BERT model is utilized alongside two traditional machine learning methods: (i) Support Vector Machine (SVM) and (ii) Logistic Regression (LR). Additionally, the authors explore the integration of sentiment analysis and emojis descriptions as supplementary features to the textual content of tweets. Experimental results indicate that the BERT-based model outperforms existing benchmark systems in three key areas: (a) detecting offensive language with an F1-score of 84.3%, (b) identifying hate speech with an 81.8% F1-score, and (c) discerning detailed categories of hate speech (such as race, religion, social class, etc.) with a 45.1% F1-score. While sentiment analysis marginally boosts the model's efficiency in detecting offensive language and hate speech, it does not enhance the model's capability in classifying specific hate speech types.

Moreover, a universal language-agnostic approach is proposed in [17] to gather a substantial proportion of tweets with offensive and hate content, regardless of their subjects or styles. The authors gathered a significant collection of offensive tweets by leveraging the non-verbal cues in emojis. They then applied the proposed methodology to Arabic tweets and compared the results with English tweets, highlighting notable cultural variances. A consistent pattern emerged with these emojis signifying offensive content over various periods on Twitter. They hand-labeled and made publicly available the most extensive Arabic dataset encompassing offensive language, detailed hate speech, profanity, and violent content. As a result, the authors found that even advanced transformer models can overlook cultural contexts, backgrounds, or the precision inherent in authentic data, such as sarcasm.

As highlighted by [18], identifying offensive language within Arabic content is intricate. Several challenges arise, such as: (i) The colloquial language frequently used on social media platforms. These posts often contain abbreviations and slang, making it challenging for classifiers to understand and process them semantically. (ii) The diverse dialects and versions of the Arabic language add another layer of complexity to discerning offensive content. The text may require extensive preprocessing before being fed into a classification model. To combat the issue of colloquialism, the researchers processed each tweet by translating emoticons and emojis into their Arabic textual equivalents and breaking down hashtags into individual words separated by spaces. To address the issue of dialect variation, the texts were transformed from regional dialects to Modern Standard Arabic (MSA). The study tested various classifiers, including traditional machine learning models such as SVM, LR, Decision Tree (DT),

Bagging, AdaBoost, and Random Forest (RF). The results show that, among traditional machine learning models, SVM topped the list with an F1-score of 82%, followed by LR at 81% and DT at 69%. For ensemble models, Bagging led with an F1-score of 88%, followed by RF at 87%, and AdaBoost at 86%.

In the study [19], features were derived from textual descriptions of emojis. Depending on the class size and the specific emojis, these features either enhanced or hindered the model's performance. The authors introduced a transformer-based technique to tackle the problem of detecting offensive language. Their approach utilized variations of the CAMeL-BERT model and was tested using a combination of four benchmark Arabic Twitter datasets, all annotated for hate speech detection, including the dataset from the OSACT5 2022 workshop shared task. The model demonstrated proficiency in identifying offensive content in Arabic tweets, achieving an accuracy of 87.15% and an F1-score of 83.6%.

In [20], the word embedding (word2vec) feature was applied in tandem with part-of-speech and/or emojis to detect such language in Indonesian tweets on Twitter. They also experimented with combining unigrams with part-of-speech and/or emojis. Classification for this study was performed using a Support Vector Machine, Random Forest Decision Tree, and Logistic Regression methods. The highest accuracy attained was 79.85%, with an F-Measure of 87.51%, using the fusion of unigram features, part-of-speech, and emojis. Furthermore, the work in [21] investigates the impact of combining emojis-based elements, which are increasingly common on social media, with multiple textual factors for the sentiment analysis of casual Arabic tweets. The authors employed four methods to extract textual features: Bag-of-Words (BoW), Latent Semantic Analysis, and two Word Embedding variants. The study evaluates the impact of merging emojis with these textual elements using the SVM classifier, considering both scenarios: with and without feature selection. Results indicate that models incorporating emojis with word embedding and optimal feature selection produce better outcomes. The implications of amalgamating emojis-based aspects from Arabic tweets with diverse textual elements are explored.

In this study, we hypothesize that removing emojis from the training dataset will enhance the ability of AraBERT pre-trained models to discern the subtleties of sarcastic language in Arabic text, as it will encourage the model to focus more on context.

III. Proposed Work

We use transfer learning with three pre-trained models: AraBERT-v2, AraBERTv02-twitter, and Multi-dialect-BERT-base-Arabic. The models are evaluated using datasets from three different sources: SemEval 2020, YouTube, and L-HSAB, preprocessing. The evaluation of the models is performed with and without emojis. This suggests that the role of emojis in understanding the text is considered a variable in the model's performance. Finally, the performance of these models is evaluated using standard performance metrics: accuracy, precision, recall, and F1-score.

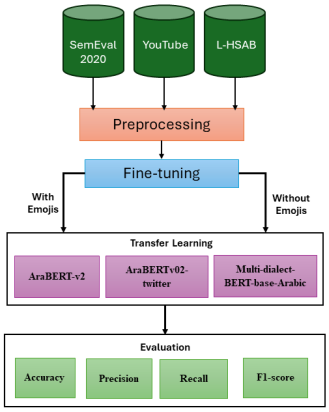

Fig. 1. Workflow of the methodology.

A. Classification Model

The core of this study revolves around the selection and application of advanced machine learning models, specifically tailored for processing Arabic text, the workflow of this study is presented in Fig. 1.

These models are critical in accurately detecting sarcasm in Arabic social media content. The selected models are:

1) Arabert_v2: This model is a BERT-based framework specifically adapted for Arabic text. AraBERT's architecture allows it to understand the contextual nuances of the Arabic language, making it an ideal choice for tasks like sarcasm detection where context plays a crucial role. Moreover, it comprises 12 transformer layers and 768 hidden units in each layer [22].

2) AraBERTv02-twitter: Building upon the foundation laid by AraBERT, Arabert_v2 is an advanced version that offers improved capabilities. Its enhanced features include a better understanding of dialectal variations and more refined contextual analysis. This version is particularly effective in handling the intricate aspects of sarcasm in various forms of Arabic language and dialects, maintaining the same architecture layer of Arabert_v2 [23].

3) Multidialect Bert base AraBERT: Recognizing the diversity of the Arabic language, this model is designed to handle multiple dialects. Its architecture is tailored to adapt to the linguistic variations found across different Arabic-speaking regions and the model is trained on the entire Wikipedia for each language, which is crucial for a comprehensive sarcasm detection tool that can operate effectively across diverse social media platforms and content. The architecture is composed of 12 encoder blocks, each equipped with 12 self-attention heads, and is followed by hidden layers that have a size of 768 [24], [23].

Table II presents the key hyperparameters for the three models, which have been determined based on empirical.

B. Datasets

The success of transfer learning models in NLP tasks heavily relies on the quality and relevance of the datasets used

TABLE II. MAIN HYPERPARAMETERS FOR FINE-TUNING THE THREE MODELS

Parameter	Value
Adam optimizer	1e-8
Learning Rate	5e-5
Batch Size	16
Maximum Sequence Length	256
Epochs	15

Fig. 2. Snapshot of SemEval dataset.

for training and testing. In this study, three distinct datasets have been carefully selected to ensure comprehensive coverage of the various facets of Arabic sarcasm and related nuances in social media contexts. These datasets are:

1) SemEval 2020 task 12 dataset: This dataset is specifically designed for sarcasm detection, making it highly relevant for this research. Fig. 2 shows a sample of the SemEval dataset. It comprises a collection of social media posts annotated for sarcasm, providing a foundational basis for training and testing the models. Including this dataset is crucial as it offers direct insights into the textual characteristics and markers indicative of sarcasm in social media content [25].

2) YouTube dataset for anti-social behavior detection: Recognizing the importance of context in sarcasm detection, this dataset from YouTube, comprising comments and posts in Arabic, is employed to understand and identify patterns of anti-social behavior, as shown in Fig. 3 [26]. This dataset provides a broader perspective on how sarcasm might be intertwined with or differentiated from other forms of communication, particularly those that are anti-social or negative.

3) L-HSAB (Levantine hate speech and abusive language) dataset: The L-HSAB dataset is instrumental in understanding the landscape of hate speech and abusive language in Arabic,

Fig. 3. Snapshot of YouTube dataset.

Fig. 4. Snapshot of L-HSAB dataset.

particularly in the Levantine dialect. Fig. 4 shows a snapshot of the L-HSAB dataset. Its inclusion aids in refining the detection algorithms to discern sarcasm from potentially similar but contextually different expressions like hate speech or abusive remarks.

C. Preprocessing

Preprocessing plays a crucial role in the machine learning workflow for classification tasks. This phase requires the modification and improvement of input data to make it suitable and beneficial for use with transfer learning models. The study implemented the following steps:

1) Data cleaning is a pivotal step in data analysis, emphasizing the importance of removing noise and irrelevant information to enhance data quality. Data cleaning is a pivotal step in data analysis, emphasizing the importance of removing noise and irrelevant information to enhance data quality [27]. This step was applied to the three datasets by removing unnecessary columns, handling missing values, and text data normalization
2) Text data preprocessing is a fundamental step in NLP, enabling the transformation of raw text into a clean, organized format suitable for analysis. This process involves: (i) using regular expressions to identify and remove unwanted characters, spaces, or patterns within text data, (ii) tokenization is the process of breaking down the text into smaller units, such as words or phrases, (iii) removal of stop words, and (iv) performing Stemming which is reducing words to their root form.
3) Splitting the dataset into training, validation, and test sets in two distinct versions: one version includes emojis in the text data, and the other version with emojis excluded.

D. Performance Metrics

The evaluation of the models is conducted using a set of standard performance metrics in transfer learning:

1) Accuracy: this metric measures the proportion of correctly identified instances (both sarcastic and non-sarcastic) among the total instances. It provides an overall effectiveness of the model.

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

2) Precision and recall: these metrics evaluate the accuracy of the model in identifying sarcasm. Precision measures the proportion of correctly identified sarcastic instances among all instances identified as sarcastic, while recall measures the proportion of correctly identified sarcastic instances among all actual sarcastic instances.

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

3) F1 score: it is the harmonic mean of precision and recall, providing a single metric that balances both. It is particularly useful when the class distribution is imbalanced.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

IV. Results and Analysis

In this work, the investigation primarily focuses on the influence of emojis in classifying offensive language in Arabic. Additionally, it presents a comprehensive evaluation and interpretation of the outcomes, highlighting which machine learning model is most effective in accurately identifying offensive speech with or without emojis. The results are critical for understanding the nuances and accuracy of language processing techniques in a linguistically diverse and complex region like the Arab world.

The comparative analysis, which is presented in Fig. 5v to 7, assesses the influence of emojis on the performance of Arabic language models in classifying offensive content within the SemEval, YouTube, L-HSAB datasets, respectively. This evaluation is segmented into four distinct metrics: accuracy, recall, precision, and F1-score. The performance of the three considered, AraBERT_v2 (v2), AraBERT_v2_Twitter (TW), and multi_dialect_bert_base_arabert (MD) models are investigated for all of these datasets.

A. Accuracy

In Fig. 5(a), 6(a), and 7(a), the accuracy metrics are evaluated across multiple models, including AraBERT_v2 (v2), AraBERT_v2_Twitter (TW), and multi_dialect_bert_base_arabert (MD). Across these figures, the accuracy rates for each model are displayed both with and without the inclusion of emojis in the dataset. In all cases, except for L-HSAB with the TW model, the accuracy tends to be higher when emojis are excluded. This trend suggests that emojis may introduce ambiguity or noise that degrades the models' ability to accurately classify text.

B. Recall

The recall results across the three models: V2, TW, and MD, are examined through Fig. 5(b), 6(b), and 7(b), respectively. For all the datasets, it observed that excluding emojis leads to an increase in recall for TW and MD, while in the V2 models with the YouTube and SemEval dataset, recall shows a slight improvement when emojis are included. This suggests that emojis may enhance detection for some models but could potentially disrupt others, resulting in missed detection. The V2 model, however, exhibits negligible differences in both cases.

C. Precision

Analyzing precision across different models in Fig. 5(c), 6(c), and 7(c) yields the following observations. In Fig. 5(c), the precision is higher when emojis are excluded for all three models. Particularly, there is a significant increase in precision observed for the V2 and TW models when emojis are excluded, suggesting that emojis may have a notable impact on lowering precision, especially for these models. For Fig. 6(c), the precision outcomes show improvements for all models when emojis are removed. The V2 and MD models demonstrate significant enhancement in precision without emoji, indicating a reduction in false positives. The TW model shows no significant difference in both cases. In Fig. 7(c), the MD model shows improvement in precision without emoji, indicating potential complications that emojis introduce in precision tasks across different models. However, the V2 and TW models show a minor increase in precision with emoji, suggesting a better ability to identify positive instances when emojis are present.

D. F1-score

In Fig. 5(d) and 6(d), a consistent trend is observed where excluding emojis benefits the F1-score across all models. This indicates that emojis do not significantly contribute to the classification process and could potentially even restrict it. The same thing almost exists in Fig. 7(d). For V2, the F1-score slightly drops when emojis are excluded. The difference in the MD case is almost negligible. These findings imply that emojis exclusion mitigates the trade-off between precision and recall for this particular model.

Based on these results, the proposed framework concludes that excluding emojis consistently improves accuracy, recall, precision, and F1-score across various models, indicating its beneficial impact on sarcasm speech classification. Emojis introduce ambiguity or noise that degrades classification accuracy, while their exclusion leads to increased recall, particularly for the TW and MD models. Precision notably increases, especially for the V2 and MD models, when emojis are removed, reducing false positives. Similarly, the F1-score improves across all models when emojis are excluded, except for a slight drop in V2's case, implying a better balance between precision and recall. Therefore, these results support our main hypothesis (RQ1, RQ2, and RQ3) that removing emojis during the preprocessing stage has a positive, or at least non-negative, impact on improving sarcasm speech classification performance.

Reflecting these findings, we believe that excluding emojis in sarcasm detection increases focus on an important aspect of NLP. Although emojis are commonly used in digital communication, their influence on language processing can vary significantly depending on the linguistic context. In the case of Arabic, a language rich in vocabulary and dialects, emojis may introduce more noise than benefit, as demonstrated by the performance improvements observed in our models when these non-textual elements were omitted.

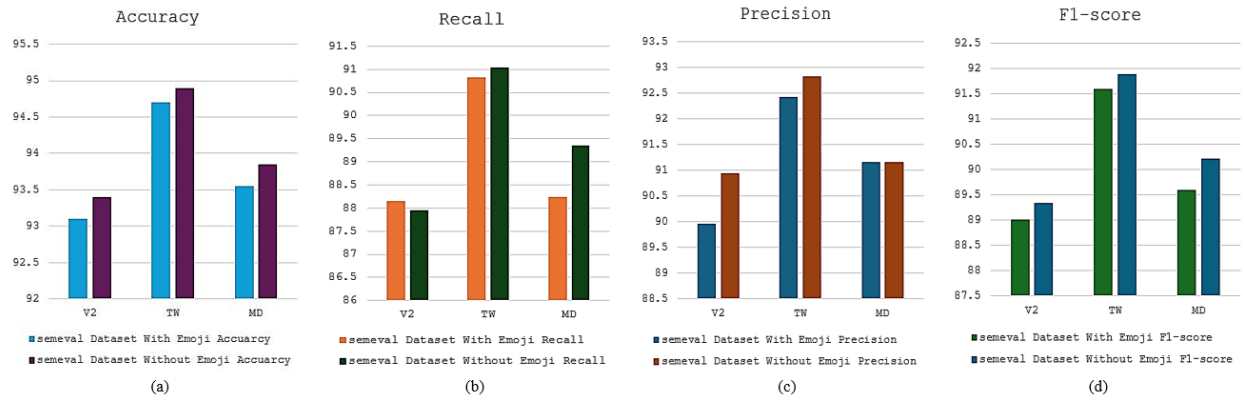

Fig. 5. Comparison results of classification with and without emojis on the SemEval dataset.

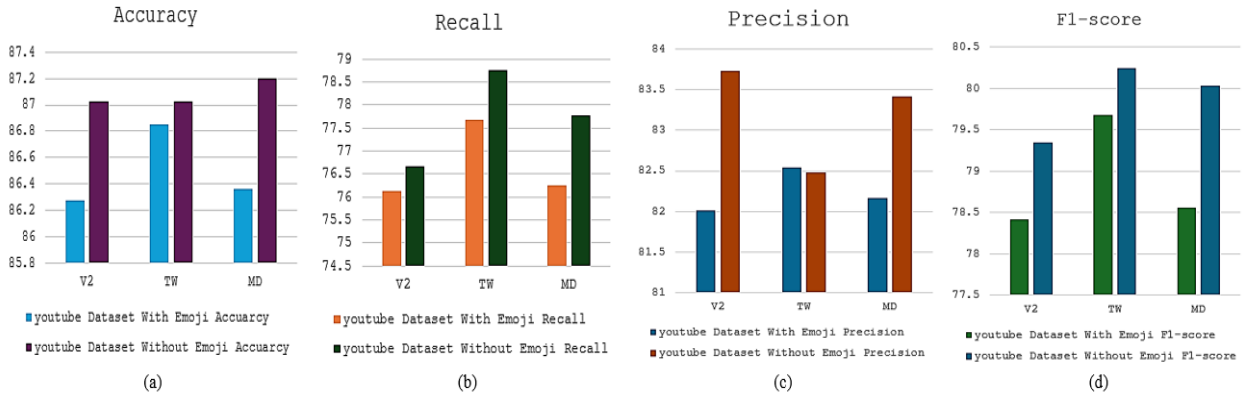

Fig. 6. Comparison results of classification with and without emojis on the YouTube dataset.

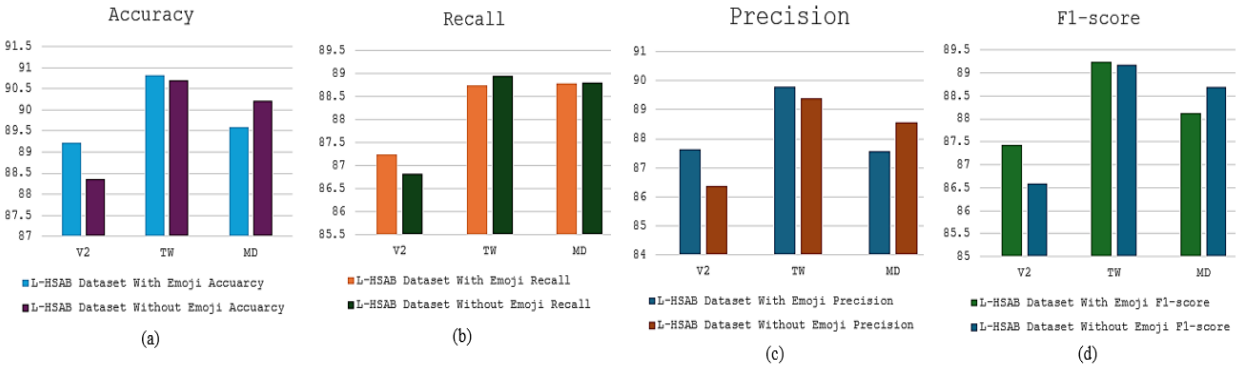

Fig. 7. Comparison results of classification with and without emojis on the L-HSAB dataset.

V. Conclusions and Future Work

In this paper, we investigated the impact of excluding emojis during the preprocessing stage on the performance of Arabic sarcasm detection models. Detecting sarcasm in Arabic social media speech presents a multifaceted challenge due to linguistic diversity and the intricacies of sarcastic expressions. By evaluating the accuracy, recall, precision, and F1-score metrics across various models and datasets, we demonstrated that emojis exclusion enhances sarcasm detection accuracy. This research provides a more precise interpretation of language by eliminating potential confusion introduced by non-textual elements, ultimately contributing to the advancement of language processing techniques in linguistically diverse regions. Moreover, our findings offer valuable insights for social media platforms and natural language processing research. By highlighting the positive impact of emojis exclusion on sarcasm detection model performance, new benchmarks are established in Arabic natural language processing.

Nevertheless, there are some limitations to this study, which need to be mentioned. First, such research specifies the

focus of sarcasm detection on textual information rather than considering that social media content often contains data in the form of images, videos, and so on in addition to text data. Second, although the given datasets offered a wide variety of texts, it should be noted that they might not include all the different types, styles, and details of the Arabic language used across Arabic-speaking countries. This limitation implies that when generalizing the findings of this study, the conclusions may be restricted to the contexts captured by the datasets.

Future research could focus on developing advanced machine-learning tools that interpret emojis alongside the text, reducing confusion from emojis misuse or overuse. This study provides the way for further advancements, particularly in enhancing the robustness and applicability of sarcasm detection models for Arabic and other languages, as follows:

- Integrating text with other data types like images, videos, and audio could enhance the accuracy of sarcasm detection. For instance, facial expressions in images accompanying sarcastic text might provide additional context that is not captured by textual analysis alone. The ability to process multiple data formats in real time could significantly improve the performance of the model.
- Real-time monitoring tools on social media platforms could be significantly enhanced to detect and report sarcasm in live conversations, which can help prevent the propagation of harmful comments and improve automated customer service responses. Integrating these models into platforms like Twitter and Facebook could also offer deeper insights into public opinion trends.
- A key challenge in Arabic sarcasm detection is the availability of datasets where the sarcasm is annotated. Future efforts could enhance this by using more effective schemes of data annotation like crowdsourcing or semi-supervised learning, as well as in collecting datasets that include various dialects and cultural settings to improve the model's transferability and reliability.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (24UQU4350534DSR02).

REFERENCES

[1] V. Dumas, "Enigma machines: Deep learning algorithms as information content providers under section 230 of the communications decency act," *Wis. L. Rev.*, p. 1581, 2022.

[2] A. O. Marwa Khairy, Tarek M. Mahmoud and T. A. El-Hafeez, "Comparative performance of ensemble machine learning for arabic cyberbullying and ofensive language detection," *Language Resources and Evaluation, Springer*, 2023.

[3] V. Sukhavasi and V. Dondeti, "Effective automated transformer model based sarcasm detection using multilingual data," *Multimedia Tools and Applications*, pp. 1–32, 2023.

[4] S. Mihi, B. Ait Ben Ali, I. El Bazi, S. Arezki, and N. Laachfoubi, "Automatic sarcasm detection in dialectal arabic using bert and tf-idf," in *The Proceedings of the International Conference on Smart City Applications*. Springer, 2021, pp. 837–847.

[5] J. A. García-Díaz, S. M. Jiménez-Zafra, M. A. García-Cumbreras, and R. Valencia-García, "Evaluating feature combination strategies for hate-speech detection in spanish using linguistic features and transformers," *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2893–2914, 2023.

[6] M. Koroteev, "Bert: a review of applications in natural language processing and understanding," *arXiv preprint arXiv:2103.11943*, 2021.

[7] I. A. Farha and W. Magdy, "From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020, pp. 32–39.

[8] A. Rahma, S. S. Azab, and A. Mohammed, "A comprehensive review on arabic sarcasm detection: Approaches, challenges and future trends," *IEEE Access*, 2023.

[9] J. Subramanian, V. Sridharan, K. Shu, and H. Liu, "Exploiting emojis for sarcasm detection," in *Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12*. Springer, 2019, pp. 70–80.

[10] A. E. Abdel-Hakim and W. A. Deabes, "Impact of sensor data glut on activity recognition in smart environments," in *2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*. IEEE, 2017, pp. 1–5.

[11] M. T. B. Othman, M. A. Al-Hagery, and Y. M. El Hashemi, "Arabic text processing model: Verbs roots and conjugation automation," *IEEE Access*, vol. 8, pp. 103 913–103 923, 2020.

[12] A. Biemiller and N. Slonim, "Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition." *Journal of educational psychology*, vol. 93, no. 3, p. 498, 2001.

[13] G. Z. Patrick, "Roots of the russian language," *(No Title)*, 1989.

[14] M. El-Melegy, A. Abdelbaset, A. Abdel-Hakim, and G. El-Sayed, "Recognition of arabic handwritten literal amounts using deep convolutional neural networks," in *Pattern Recognition and Image Analysis: 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, July 1–4, 2019, Proceedings, Part II 9*. Springer, 2019, pp. 169–176.

[15] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.

[16] M. J. Althobaiti, "Bert-based approach to arabic hate speech and offensive language detection in twitter: Exploiting emojis and sentiment analysis," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.

[17] H. Mubarak, S. Hassan, and S. A. Chowdhury, "Emojis as anchors to detect arabic offensive language and hate speech," *Natural Language Engineering*, vol. 29, no. 6, pp. 1436–1457, 2023.

[18] F. Husain and O. Uzuner, "Investigating the effect of preprocessing arabic text on offensive language and hate speech detection," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 4, pp. 1–20, 2022.

[19] S. Al-Dabet, A. ElMassry, B. Alomar, and A. Alshamsi, "Transformer-based arabic offensive speech detection," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2023, pp. 1–6.

[20] M. O. Ibrohim, M. A. Setiadi, and I. Budi, "Identification of hate speech and abusive language on indonesian twitter using the word2vec, part of speech and emoji features," in *Proceedings of the 1st International Conference on Advanced Information Science and System*, 2019, pp. 1–5.

[21] S. Al-Azani and E.-S. M. El-Alfy, "Combining emojis with arabic textual features for sentiment classification," in *2018 9th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2018, pp. 139–144.

[22] H. Elfaik *et al.*, "Combining context-aware embeddings and an attentional deep learning model for arabic affect analysis on twitter," *IEEE Access*, vol. 9, pp. 111 214–111 230, 2021.

[23] M. A. Humayun, H. Yassin, and P. E. Abas, "Dialect classification using acoustic and linguistic features in arabic speech," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, p. 739, 2023.

[24] A. S. Alammary, "Bert models for arabic text classification: a systematic review," *Applied Sciences*, vol. 12, no. 11, p. 5720, 2022.

[25] M. Zampieri, P. Nakov, S. Rosenthal, P. Atanasova, G. Karadzhov, H. Mubarak, L. Derczynski, Z. Pitenis, and Ç. Çöltekin, "Semeval-2020 task 12: Multilingual offensive language identification in social media (offenseval 2020)," *arXiv preprint arXiv:2006.07235*, 2020.

[26] A. Alakrot, L. Murray, and N. S. Nikolov, "Dataset construction for the detection of anti-social behaviour in online communication in arabic," *Procedia Computer Science*, vol. 142, pp. 174–181, 2018.

[27] T. Nguyen, C. Van Nguyen, V. D. Lai, H. Man, N. T. Ngo, F. Dernoncourt, R. A. Rossi, and T. H. Nguyen, "Culturax: A cleaned, enormous, and multilingual dataset for large language models in 167 languages," *arXiv preprint arXiv:2309.09400*, 2023.