

Accent-aware arabic speech processing using CNN–BiLSTM for dialect classification

Tarik Abdel-Kareem AbuAin¹, Waleed Abdel Karim Abuain²

¹College of Computing and Informatics, Prince Musaid bin Abdulrahman University, Riyadh, Saudi Arabia

²College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia

Article Info

Article history:

Received Sep 27, 2025

Revised Aug 18, 2026

Accepted Aug 28, 2026

Keywords:

Arabic accents

BiLSTM

CNN

Dialect classification

Speech processing

ABSTRACT

The Arabic language presents considerable dialectal diversity, which can pose challenges for speech technologies, particularly automatic speech recognition (ASR). Traditional models trained primarily on modern standard Arabic (MSA) do not account for regional accents, leading to poor performance on underrepresented dialects. However, this paper addresses this gap by introducing a hybrid CNN-BiLSTM model designed for automatic Arabic dialect classification. The dataset was divided into five categories gathered from various YouTube videos: Levantine (Jordan, Lebanon, Syria, Palestine), Gulf (Saudi Arabia, Yemen, Qatar, UAE), Nile Valley (Egypt, Sudan), Mesopotamian (Iraq, Kuwait), and North African (Algeria, Libya, Mauritania, Morocco). Therefore, a systematic augmentation was applied to address this imbalance, including pitch shifting, time stretching, and amplitude scaling, leading to 4,000 balanced utterances standardized to 16 kHz and 8 seconds. Convolutional neural network (CNN) layers captured localized spectral features, while bidirectional long short-term memory (BiLSTM) units modeled long-range temporal dependencies. Evaluation metrics, including classification report, confusion matrix, and learning curves, showed effective performance, with an overall accuracy of 88%.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Tarik Abdel-Kareem AbuAin

College of Computing and Informatics, Saudi Electronic University

Riyadh 11673, Saudi Arabia

Email: t.aboain@seu.edu.sa

1. INTRODUCTION

Arabic, spoken by 420 million people worldwide [1], has a variety of dialects across regions, with some dialects often substantially different from others [2]. Arabic accents classification is complex because the language encompasses many dialects throughout the Arab world. While modern standard Arabic (MSA) serves as the formal written and spoken standard used in media and education, native speakers mainly communicate in vernaculars that vary significantly in pronunciation, vocabulary, and syntax.

These are dialectal variations, normally rooted in geography, history, and culture, resulting in a range of accents that can prevent mutual understanding, even among native speakers. For instance, pronunciation differences between Gulf dialects and Maghrebi Arabic can significantly affect clarity and meaning. This linguistic diversity poses a challenge for speech-based technologies, especially automatic speech recognition (ASR), which struggles to perform well across such varied language groups [3].

Standard ASR models, primarily developed for native speech [4] and trained on MSA or limited-accent varieties, often struggle to accurately interpret or detect speech from less-common dialects. Accent classification aims to tackle this issue by automatically identifying the speaker's accent, enabling ASR and other speech systems to adapt in real time.

Apart from achieving higher recognition accuracy, this process is needed to create inclusive, culturally conscious AI systems that better serve Arabic-speaking populations. Specifically, our model uses label smoothing, a fixed 24×32 spectrogram input, and adaptive learning rate scheduling, design choices that address the acoustic variability of Arabic dialects. Unlike single-stream convolutional neural networks (CNNs) or BiGRU models, CNN-BiLSTM jointly captures local spectral patterns and long-range temporal dependencies, thereby achieving superior dialect differentiation. In this paper, we introduce a hybrid CNN-BiLSTM method for classifying light Arabic accents. The database is split into five categories: Levantine (Jordan, Lebanon, Syria, Palestine), Gulf (Saudi Arabia, Yemen, Qatar, UAE), Nile Valley (Sudan, Egypt), Mesopotamian (Iraq and Kuwait), and North African (Algeria, Libya, Mauritania, Morocco). The results obtained will be analyzed in the provided sections.

2. RELATED WORK

In recent years, many researchers have conducted studies on Arabic dialect identification. In this section, we will review some of these studies to gain insight into the methods and models previously used in this area. Aliwy *et al.* [5] conducted a study that applied preprocessing, clustering, and ensemble classifiers to the NADI dataset, reaching an F-measure of as high as 52.38%. Notably, clustering played a key role in significantly enhancing the accuracy of Arabic dialect classification.

Elmahdy *et al.* [6] present a novel cross-lingual adaptation approach (MLLR/MAP) that leverages MSA speech resources to enhance ASR for dialectal Arabic. The MSA dataset comprises 33 hours, ECA has 0.5 hours, and LCA includes 26 hours. Results indicate a 41.8% relative reduction in WER on ECA with modified phonemic models. The model achieved 91.34% accuracy, surpassing baselines such as RNN (76.41%) and CNN (81.43%). Table 1 provides an overview of recent studies on the classification of written Arabic dialects. Alsadhan [7] applied AraBERT and FastText/DBSCAN to social media data, attaining 88.9% dialect classification accuracy and effective emotion detection.

Table 1. Recent studies of written Arabic dialect classification

Ref.	Method	Datasets	Findings
Yafooz [8]	RNNs (LSTM/GRU), Transformers (AraBERT, and MARBERTv2), and LLMs (Gemini, Qwen)	QADI (440K Arabic tweets, 18 dialects)	The achieved accuracy is - 65% with MARBERTv2, and 52% with RNNs
Elsafty <i>et al.</i> [9]	ArDiaBERT/ArDiaGPT with rare-dialect balancing	ArDia (903k texts, 6 dialects, long-form TikTok/Wikipedia dominance)	The method obtained 86.55% Macro-F1
Saleh <i>et al.</i> [10]	hybrid stacked transformer models	Shami and IADD	The achieved accuracy is - Shami: 89.73% and ADD: 93.062%
Abdullah <i>et al.</i> [11]	Voice Conversion (VC)-Enhanced Training	MGB-3 ADI-5 (53.6 hrs, 14.6k samples) for training and MADIS-5 (curated 12-hr cross-domain set) for evaluation	The approach yielded consistent improvements of +34.1%
Alharbi [12]	A novel Multimodal Framework, called MF-Saudi	Saudi Dialect Audio-Text	The accuracy achieved is 96.5%
Alsawaylimi [13]	Hybrid CAMELBERT/ALBERT + BiLSTM; ML/DL/transformer	38,394 annotated social media comments (4 dialects)	The achieved accuracy is - CAMELBERT: 87.67% and ALBERT: 86.51%

Tibi and Messaoud [14] created a system that detects Arabic dialects by applying multi-scale product analysis (MPA) to extract features for a hamilton neural network (HNN). It was trained on eight LDC and eight ADI dialects, achieving an accuracy of 91.34%, which is better than the RNN (76.41%) and CNN (81.43%) models.

Research proposed by Ali [15] developed an end-to-end CNN-RNN-based model with a 1,400-hour multi-dialect corpus (MGB-2, Tashkeela, and so on). It had a word error rate (WER) of 14%, which is higher than that of previous systems.

A study by Al-Ayyoub *et al.* [16] used wavelet features and a Bayesian Network classifier to differentiate between Jordanian and Egyptian Arabic dialects. It was trained on a small dataset (a custom dataset of 35 samples) and, using these features, achieved 97% accuracy. Table 2 presents recent studies on the classification of Arabic spoken dialects. Al-Anzi and Thankaleela [17] used CNNs and MFCC on 4,071 Arabic audio clips across 5 dialects, achieving 94.28% test accuracy.

Table 2. Recent studies of Arabic spoken dialects classification

Ref	Method	Datasets	Findings
Alsaheel <i>et al.</i> [18]	BiGRU and BiLSTM	SADA corpus of dialectal Arabic	- 97% training accuracy and 90% test accuracy
Al-Shwayyat and Gerek [19]	Hybrid MFCC + CNN and Wavelet + RNN	Common Voice Arabic 12.0	- Achieved accuracy: - 91.2% for MFCC + CNN vs. - 66.5% for Wavelet + RNN
Talafha <i>et al.</i> [20]	Pre-trained models (Whisper, w2v-BERT, Conformer)	Speech corpus (Casablanca) covering 8 Arabic dialects for ID/ASR; multi-source collection for diacritization	The best achieved accuracy is 79.8%
Barakat <i>et al.</i> [21]	A deep learning classification model	More than 66,000 audio files from the YouTube ADI-17 MIT dataset.	The accuracy achieved is 86% on test data, and 95% when considering the top 3 predicted dialects
Nahar <i>et al.</i> [22]	Machine learning models (K-Nearest Neighbor (KNN), Random Forest (RF), Multi-Layer Perceptron (MLP), and Artificial Neural Networks (ANN))	The dataset was collected from YouTube videos and audio segments labeled by human experts	- MFCC Features KNN: 62%, RF: 60% MLP: 33%, ANN: 42% - TFCC Features KNN: 76%, RF: 64% MLP: 34%, ANN: 41%
Hamidi <i>et al.</i> [23]	HMM-GMM with MFCC	24 speakers (Moroccan / Algerian)	- Best accuracy: 78.8% with mixed dictionary - Speaker-independent: text outperforms audio with F1 score of 63.61%
Al-Azani and El-Alfy [24]	Combination of audio and textual features for Arabic dialect identification	524 opinion utterances extracted from 63 YouTube videos	- Speaker-dependent: both text and audio performed similarly, with an F1 score of 85.52%
Malmasi and Zampier [25]	Ensemble classifiers	Egyptian, Gulf, Levantine, North African, and MSA	- The achieved accuracy is about 68%

3. METHODS AND MATERIALS

3.1. Datasets

This study collected a dataset from various YouTube videos, categorizing it into five Arabic dialect groups: Levantine (Jordan, Lebanon, Syria, Palestine), Gulf (Saudi Arabia, Yemen, Qatar, UAE), Nile Valley (Egypt, Sudan), Mesopotamian (Iraq, Kuwait), and North African (Algeria, Libya, Mauritania, Morocco), as shown in Figure 1. YouTube was selected because it offers natural, spontaneous speech samples, and these groups align with widely accepted regional dialect classifications.

Each category contains labeled audio samples in .wav format and allows supervised dialect classification. The structure of the dataset facilitates regional generalization and aids in effective model training and evaluation. Additionally, the dataset is divided into training and testing sets (80/20) for model development.

3.2. Preprocessing data augmentation

In this paper, a systematic preprocessing and augmentation method for Arabic speech data was developed to balance and enrich the Arabic speech corpus across five dialect groups: Levantine, Gulf, Nile Valley, Mesopotamian, and North African. All sound files were downsampled to 16 kHz and time-normalized to 8 seconds (128,000 samples) to provide temporal consistency during model training.

The trimmings were performed at 20 dB, and padding or deletions were employed to maintain uniform length. In order to address dataset imbalance and generalization, each dialectal group was balanced to 800 utterances. When the available recordings were insufficient relative to the target, augmentation was used.

An augmentation step involved pitch shifting (increasing or decreasing the pitch by 1–2 semitones), speed variation (slowing down or speeding up the audio by up to 10%), and amplitude scaling (reducing or amplifying the volume by up to 20%), which was applied randomly to generate diverse yet realistic speech samples. For example, a file pool of 520 utterances for Levantine required an additional 280 synthetically generated samples, which is 35% of the group's data, to achieve the target. This augmentation not only balanced group sizes but also expanded acoustic variability, effectively simulating speaker and recording diversity. The result was 4,000 processed files using a uniform procedure, that provides a strong basis for accent classification with hybrid CNN-BiLSTM-based models. The resulting spectrograms are shown in Figure 2 (Spectrograms of Augmented Audio Signals).

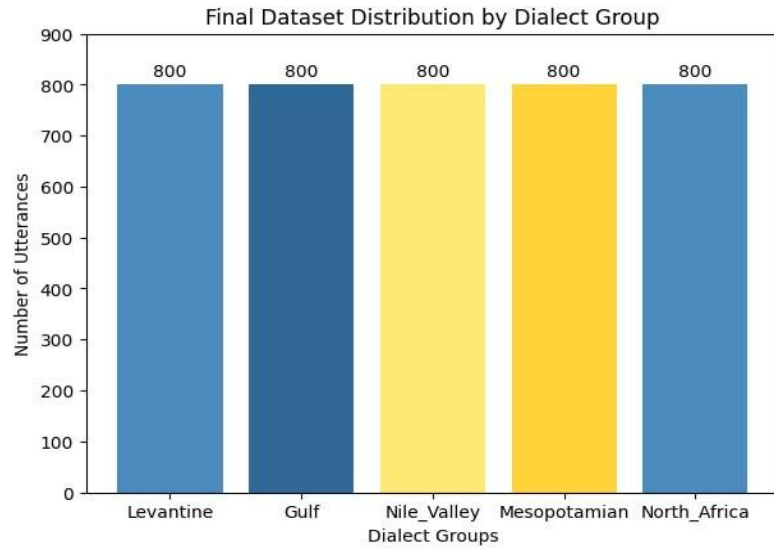


Figure 1. Distribution of the final Arabic speech dataset among five dialect groups: Levantine, Gulf, Nile Valley, Mesopotamian, and North African

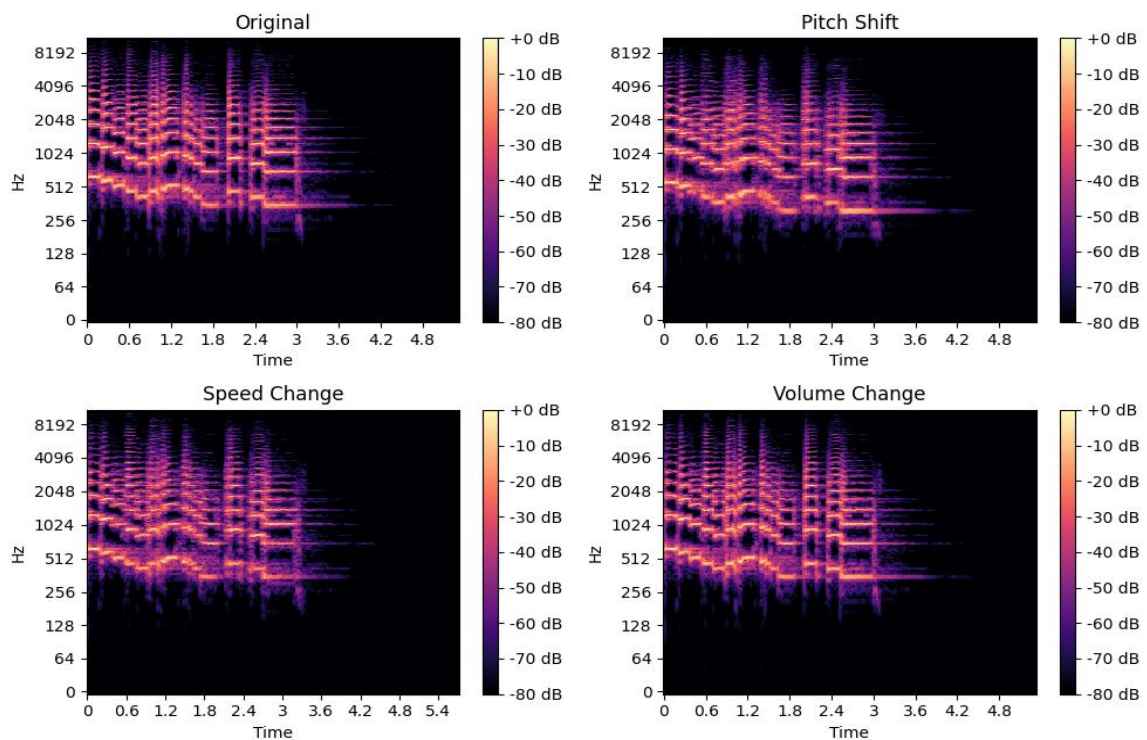


Figure 2. Spectrograms of augmented audio signals

3.3. Model architecture

The proposed architecture includes convolutional and recurrent layers to detect local and sequential patterns in the input embeddings. The input of each instance, which is in the form of a sequence of 24-time steps with 32-dimensional features, becomes subject to processing by a one-dimensional convolutional block in the form of a building block. This convolutional block contains a convolution layer, a normalization layer, an activation function, and a pooling layer, which together perform local pattern discovery and reduce data length while improving computational efficiency.

The resulting feature maps are input into a bidirectional long short-term memory (BiLSTM) network that captures long-range dependencies in both directions. By combining the forward and backward hidden states, the model creates a 512-dimensional representation of the sequence. This representation is then fed into a fully connected feedforward network with ReLU activation, dropout regularization, and an output-dimension mapping to its final space.

To reduce overfitting in predictions and improve generalization, the model uses a label smoothing cross-entropy loss. The training process uses adaptive learning-rate scheduling and early stopping to ensure robust convergence.

All in all, CNN-BiLSTM integrates local feature extraction with a sequential modeling approach, making it comparable to methods that rely on structured representations and sequence-based classification. The model was trained using Adam ($lr=1 \times 10^{-3}$), batch size 32, up to 50 epochs with early stopping (patience=5), 128 CNN channels, 256 BiLSTM hidden units, and 0.1 label smoothing. The overall architecture is illustrated in Figure 3.

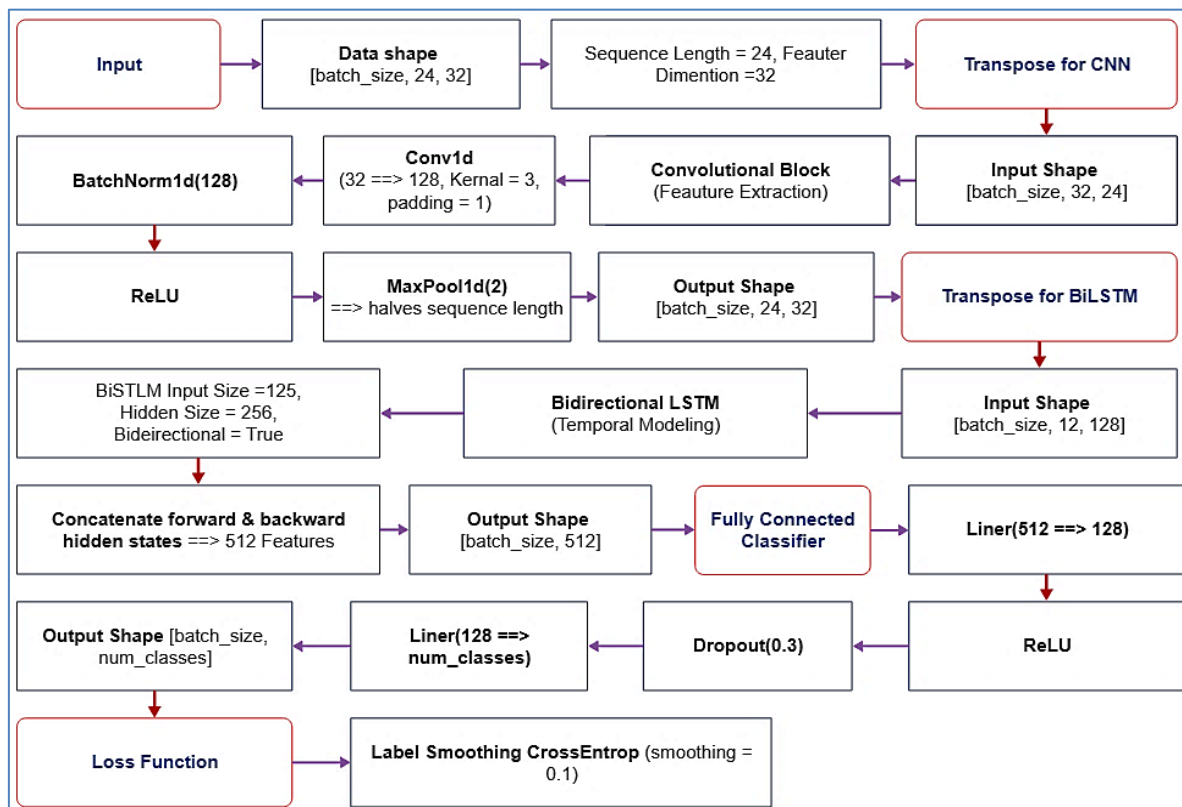


Figure 3. Research model architecture

4. RESULTS AND DISCUSSION

As shown in Figure 4, the confusion matrix shows only a few misclassifications, mostly among linguistically similar dialects, indicating the model's ability to detect subtle linguistic differences. Moreover, it also shows that Levantine and Nile Valley dialects are frequently confused with the Gulf dialect, with 13 and 11 misclassified samples, respectively, probably due to phonetic similarities across neighboring regions. Since the model relies on spectrograms rather than textual data, these errors mainly stem from acoustic similarity. Incorporating phoneme-level features or lexical cues in future iterations could help reduce these dialect misclassifications.

The classification report Table 3 confirms the model's robustness, exhibiting an overall accuracy of 88%. Both macro- and weighted-averaged metrics for precision, recall, and F1 Score are identical at 0.88. The Mesopotamian dialect achieves the highest F1-score of 0.92, reflecting excellent precision (0.93) and recall (0.91). Furthermore, the Gulf and North Africa dialects performed strongly, with F1-scores of 0.85 and 0.88, respectively.

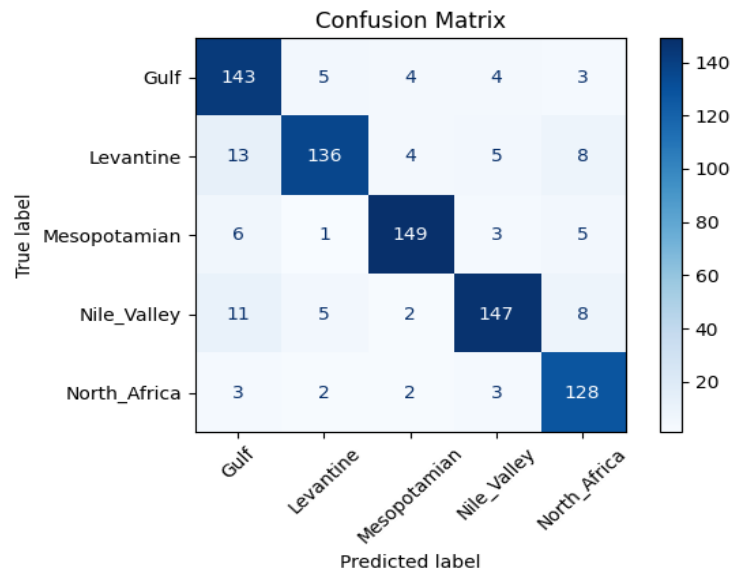


Figure 4. Confusion matrix of the dialect classification results

Table 3. Classification report showing precision, recall, and F1-scores by dialect

Classes	Precision	Recall	F1-score	Support
Gulf	0.81	0.90	0.85	159
Levantine	0.91	0.82	0.86	166
Mesopotam-ian	0.93	0.91	0.92	164
Nile_Valley	0.91	0.85	0.88	173
North_Africa	0.84	0.93	0.88	138

Though Levantine recorded a slightly lower recall (0.82), it maintained high precision (0.91), indicating that predictions for this class were conservative yet reliable. The Nile Valley dialect showed consistent results with an F1-score of 0.88, further confirming balanced performance across classes.

The learning curves in Figure 5 show smooth convergence, with accuracy steadily improving and loss declining throughout training. The close alignment between training and validation trajectories indicates the absence of overfitting, thereby validating the stability, reliability, and generalizability of the proposed model.

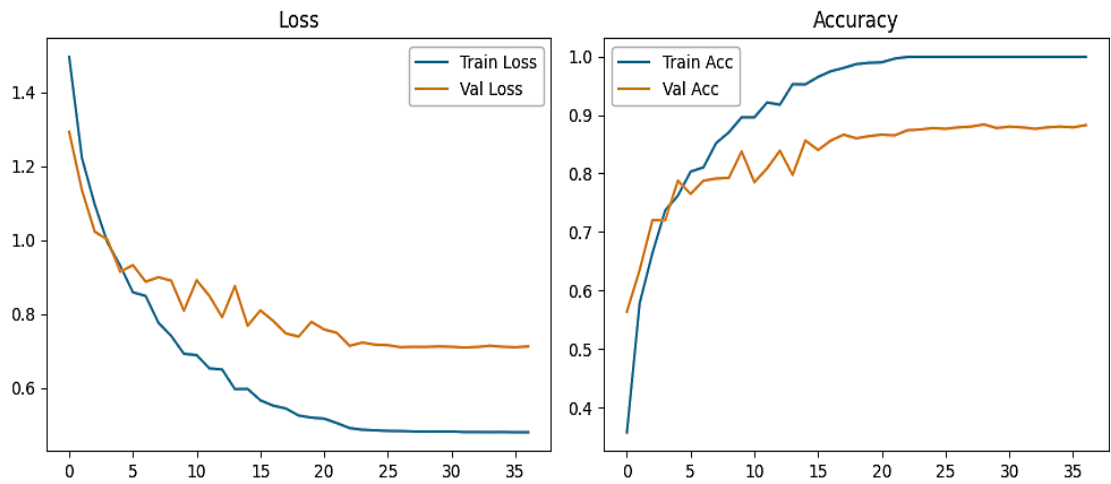


Figure 5. Training and validation loss and accuracy graphs

Initially, the systematic augmentation method tackled the problem of uneven class distribution and introduced acoustic diversity by simulating the physical differences between speakers. This enabled the model to generalize more effectively across dialects rather than overfitting to the most frequent classes. Moreover, the hybrid CNN-BiLSTM model was found to be advantageous since it combined the extraction of local spectral features with the capacity to capture long-range dependencies between consecutive observations, a feature which is important for distinguishing subtle vocal differences in closely related dialects.

The main reason for the misclassifications in the confusion matrix is the presence of similar terms, which shows that the task itself is difficult, rather than pointing to any shortcomings in the methodology. Moreover, the stability of the learning curves confirms the robustness of the optimization process and shows that there is no overfitting. When compared with previous studies that used shallow classifiers or single-stream neural networks, the present experiment achieves a better balance between accuracy and generalization. This suggests that CNN-BiLSTM models have promise as a valuable way for future development in accent-sensitive Arabic speech technologies. Table 4 demonstrates a comparative summary of our method in comparison to existing approaches that also used similar YouTube-based datasets.

Table 4. Performance comparison with existing methods

Study	Model	Dataset	Accuracy (%)
Alshabanah and Annavaram [1]	A deep learning classification model	Over 66,000 audio files from the YouTube ADI-17 MIT dataset	Achieved 86% accuracy in test data
Keleg <i>et al.</i> [2]	Machine learning models (K-Nearest Neighbor (KNN), Random Forest (RF), Multi-Layer Perceptron (MLP), and Artificial Neural Networks (ANN))	The dataset was gathered from YouTube videos and audio segments labeled by human experts	- MFCC features KNN: 62%, RF: 60% MLP: 33%, ANN: 42%
The proposed method	Hybrid convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM) model	The dataset was collected from various YouTube videos and divided into five category groups, such as Levantine, Gulf, Nile Valley, Mesopotamian, and North African	- TFCC Features KNN: 76%, RF: 64% MLP: 34%, ANN: 41% Achieved 88% accuracy in test data

5. CONCLUSION

The paper shows that the CNN-BiLSTM framework deals with the long-standing problem of classifying Arabic dialects, achieving an overall accuracy of 88% and showing consistent performance within five dialect groups. What makes the model special is its hybrid architecture, this architecture combining convolutional layers that pick out local spectral features with BiLSTM units capable of capturing temporal dependencies. It is because of this design that the model is able to effectively tell the difference between closely related dialects. Moreover, the structured data augmentation strategy combated the problem of class imbalance by introducing acoustic variability, thus improving robustness in real-world situations.

Even though the dataset has been balanced using augmentation, the YouTube recordings include uncontrolled noise and variations between channels, which impede the model's ability to generalize to clear speech. The distinction between dialect groups such as Iraqi and Kuwaiti or between Algerian and Moroccan still remains unresolved, making it difficult to apply the model to minor dialect differences. Moreover, the train-test split was carried out at the sample level, which might result in excessively optimistic estimates of accuracy. Future work might study transformer-based audio models such as Whisper and Wav2Vec, as well as x-vector embeddings, in order to increase speaker robustness. Moreover, the development of multimodal ASR systems that integrate both acoustic and textual cues could also contribute to better dialect coverage.

ACKNOWLEDGMENTS

The research work was self-supported.

FUNDING INFORMATION

The research work was self-funded.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tarik Abdel-Kareem AbuAin	✓	✓			✓		✓		✓	✓	✓	✓	✓	✓
Waleed Abdel Karim Abuain			✓	✓		✓		✓						

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The author states no conflict of interest.

INFORMED CONSENT

Not Applicable – this research requires no human or individual consent.

ETHICAL APPROVAL

Not Applicable – this research did not involve any human participation.

DATA AVAILABILITY

Datasets were collected from various YouTube videos.

REFERENCES

- [1] A. Alshabanah and M. Annavam, "On using Arabic language dialects in recommendation systems," in *Findings of the Association for Computational Linguistics: NAACL 2025*, 2025, pp. 2178–2186, doi: 10.18653/v1/2025.findings-naacl.115.
- [2] A. Keleg, S. Goldwater, and W. Magdy, "Revisiting common assumptions about Arabic Dialects in NLP," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2025, pp. 3309–3327, doi: 10.18653/v1/2025.acl-long.166.
- [3] M. Habbash *et al.*, "Recognition of Arabic accents from English spoken speech using deep learning approach," *IEEE Access*, vol. 12, pp. 37219–37230, 2024, doi: 10.1109/access.2024.3374768.
- [4] K. Kashif, A. Alwan, Y. Wu, L. De Nardis, and M.-G. Di Benedetto, "MKELM based multi-classification model for foreign accent identification," *Heliyon*, vol. 10, no. 16, p. e36460, Aug. 2024, doi: 10.1016/j.heliyon.2024.e36460.
- [5] A. Aliwy, H. Taher, and Z. AboAltaheen, "Arabic dialects identification for all Arabic countries," *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pp. 302–307, 2020.
- [6] M. Elmahdy, R. Gruhn, and W. Minker, *Novel techniques for dialectal Arabic speech recognition*. Springer US, 2012.
- [7] N. Alsadhan, "A novel dialect-aware framework for the classification of Arabic Dialects and emotions," *Journal of Computer Science*, vol. 21, no. 1, pp. 88–95, Jan. 2025, doi: 10.3844/jcssp.2025.88.95.
- [8] W. M. S. Yafouz, "Enhancing Arabic dialect detection on social media: a hybrid model with an attention mechanism," *Information (Switzerland)*, vol. 15, no. 6, 2024, doi: 10.3390/info15060316.
- [9] H. Elsafty, B. Abdou, T. Deußer, M. Pielka, C. Bauckhage, and R. Sifa, "ArDia: improving arabic dialectal language classification using a novel dataset," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 19, no. 1cws, pp. 2413–2422, 2025, doi: 10.1609/icws.v19i1.35944.
- [10] H. Saleh *et al.*, "Advancing arabic dialect detection with hybrid stacked transformer models," *Frontiers in Human Neuroscience*, vol. 19, 2025, doi: 10.3389/fnhum.2025.1498297.
- [11] B. M. Abdullah, M. Baas, B. Möbius, and D. Klakow, "Voice conversion improves cross-domain robustness for spoken arabic dialect identification," 2025.
- [12] R. Alharbi, "MF-Saudi: A multimodal framework for bridging the gap between audio and textual data for Saudi dialect detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 6, p. 102084, 2024, doi: 10.1016/j.jksuci.2024.102084.
- [13] A. A. Alsawaylimi, "Arabic dialect identification in social media: A hybrid model with transformer models and BiLSTM," *Heliyon*, vol. 10, no. 17, p. e36280, 2024, doi: 10.1016/j.heliyon.2024.e36280.
- [14] N. Tibi and M. Anouar Ben Messaoud, "Arabic dialect classification using an adaptive deep learning model," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 2, pp. 1108–1116, Apr. 2025, doi: 10.11591/eei.v14i2.8165.
- [15] A. R. Ali, "Multi-Dialect Arabic speech recognition," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–7, doi: 10.1109/ijcnn48605.2020.9206658.
- [16] M. Al-Ayyoub, M. K. Rihani, N. I. Dalgamoni, and N. A. Abdulla, "Spoken Arabic dialects identification: The case of Egyptian and Jordanian dialects," in *2014 5th International Conference on Information and Communication Systems (ICICS)*, Apr. 2014, pp. 1–6, doi: 10.1109/iacs.2014.6841970.

- [17] F. S. Al-Anzi and B. S. S. Thankaleela, "Region-wise recognition and classification of Arabic dialects and vocabulary: a deep learning approach," *Applied Sciences*, vol. 15, no. 12, p. 6516, 2025, doi: 10.3390/app15126516.
- [18] A. Alsaheel, N. Almatrafi, N. Alhumaid, D. M. Ibrahim, and A. A. Al-Shargabi, "Enhancing Arabic dialect classification with deep learning techniques," in *2025 5th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Aug. 2025, pp. 1–6, doi: 10.1109/esmarta66764.2025.11132273.
- [19] G. Al-Shwayyat and O. N. Gerek, "Hybrid deep learning and signal processing for Arabic dialect recognition in low-resource settings," *arXiv preprint arXiv:2506.21386*, 2025, [Online]. Available: <http://arxiv.org/abs/2506.21386>.
- [20] B. Talafha *et al.*, "NADI 2025: the first multidialectal Arabic speech processing shared task," 2025.
- [21] A. Barakat, O. Al Hammadi, A. Aldhaheeri, and A. Elnagar, "Arabic dialect identification from speech," in *2024 15th Annual Undergraduate Research Conference on Applied Computing (URC)*, Apr. 2024, pp. 1–6, doi: 10.1109/urc62276.2024.10604557.
- [22] K. M. O. Nahar, O. M. Al-Hazaimah, A. Abu-Ein, and M. A. Al-Betar, "Arabic dialect identification using different machine learning methods," *Research Square*, pp. 1–21, 2022.
- [23] M. Hamidi *et al.*, "Mixed ASR system for amazigh and Arabic underresourced dialects in maghreb region," 2024, doi: 10.31219/osf.io/9sw54.
- [24] S. Al-Azani and E.-S. M. El-Alfy, "Audio-textual arabic dialect identification for opinion mining videos," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2019, pp. 2470–2475, doi: 10.1109/ssci44817.2019.9003031.
- [25] S. Malmasi and M. Zampieri, "Arabic dialect identification in speech transcripts," in *Proceedings of the third workshop on NLP for similar languages, varieties and dialects*, no. 2010, pp. 106–113, 2016.

BIOGRAPHIES OF AUTHORS



Tarik Abdel-Kareem AbuAin    earned his first degree in Computer Science from Yarmouk University in Jordan in 2003. He went on to obtain a master's degree in computer science from AMU University, also in Jordan, in 2007. In 2016, he received his PhD in Artificial Intelligence from the Artificial Intelligence Technology Center at the Department of Computer Science, National University of Malaysia (UKM). Since completing his education, he has been working as a computer science researcher and a full-time associate professor in the Department of Computer Science at SEU University in Saudi Arabia. His research interests include artificial intelligence, machine learning and optimization, computer vision, deep learning, and the Internet of Things. He can be contacted at email: t.aboain@seu.edu.sa.



Waleed Abdel Karim Abuain    Waleed Abdel Karim Abuain received his PhD in artificial intelligence in 2016 from the Artificial Intelligence Technology Center of the National University of Malaysia, Department of Computer Science (UKM). Since 2016, he is currently working as a full-time associate professor in the Department of Computer Science at Taibah University, Applied College, Saudi Arabia. His research interests include artificial intelligence, machine learning and optimization, data science, computer vision, deep learning, and the internet of things. He can be contacted at email: wabuain@taibahu.edu.sa.