Systems and Computing
Volume 2, Issue 1, 47-59, 2026.
https://doi.org/10.64409/sycom.v2.i1.29

Research Article Open Access

Improving Arabic Dialect Processing in IoT Systems: A Comparative Study of Baseline and Dialect-Aware AI Models

Rania Djehaiche [1], **Yamena Djehaiche** [2]

1. *Department of Electronics. Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria*
2. *Laboratory of Linguistic and Literary Contemporary Studies, Department of Arabic Language and Literature. Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria*

Corresponding author: Rania Djehaiche (rania.djehaiche@univ-bba.dz)

Manuscript Review Record:
Submitted:
October 14, 2025

Accepted:
December 19, 2025

Published:
January 15, 2026

Cite This:
Rania Djehaiche, Yamena Djehaiche. "Improving Arabic Dialect Processing in IoT Systems: A Comparative Study of Baseline and Dialect-Aware AI Models". *Systems and Computing*. Volume 2, Issue 1, 47-59, 2026. https://doi.org/10.64409/sycom.v2.i1.29

Abstract- **Context:** Bringing voice-controlled interfaces into Internet of Things (IoT) systems has created fresh opportunities for smart environments. However, existing voice assistants often struggle with non-standardized languages, especially Arabic dialects. **Objective:** This research paper explores the challenges and potential of integrating five Arabic dialect variants, namely Modern Standard Arabic (MSA), known as Fusha (الفصحى), Egyptian, Levantine, Gulf, and Algerian dialects, into AI-driven IoT systems. **Method:** For each dialect, a comparative simulation was performed using two AI models: a baseline model and a dialect-aware model. Key simulated metrics included automatic speech recognition (ASR) accuracy, intention recognition, task success rate, and system response time. **Results:** The results consistently show that the dialect-aware model outperforms the baseline model in all metrics. It provides higher ASR and intention recognition accuracy, improved task success rates, and faster response times, especially for regional dialects. The Algerian dialect, while still challenging, benefited significantly from the dialect-aware adaptations of the improved model. These results highlight the potential of dialect-aware AI to close the performance gap caused by linguistic variation and code-switching. **Conclusions:** This study highlights the importance of considering linguistic diversity when developing accessible, culturally appropriate IoT interfaces that ensure a more inclusive and natural user interaction.

Keywords- Arabic dialects, IoT, Dialect-aware model, Baseline model, AI.

Acknowledgement

The author(s) would like to express sincere gratitude to the Department of Electronics and the Laboratory of Linguistic and Literary Contemporary Studies, Department of Arabic Language and Literature**,** at Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria**,** for their continuous support, guidance, and encouragement throughout the development of this research.

1. Introduction

The Internet of Things (IoT) has witnessed rapid growth, enabling the development of smart environments in which users can interact naturally with machines through voice-based interfaces [1], [2], [3]. Recent advances in artificial intelligence (AI) have significantly enhanced the capabilities of these systems across various domains, including smart homes, healthcare, transportation, and industrial applications [4]. Voice-driven interaction plays a crucial role in improving accessibility and usability within IoT ecosystems; however, its effectiveness depends largely on the system's ability to accurately process natural language.

Although state-of-the-art voice-based AI systems demonstrate high performance when operating on standardized languages such as English and Modern Standard Arabic (MSA), they continue to face considerable challenges when handling non-standard and highly diverse dialects [5]. This limitation is particularly evident in the Arabic-speaking world, where regional dialects including Egyptian, Levantine, Gulf, and Algerian Arabic exhibit substantial variations in phonetics, vocabulary, grammatical structure, and code-switching behavior. These linguistic differences pose a significant barrier to the widespread adoption of AI-powered IoT solutions in dialect-rich environments.

To address this challenge, this paper investigates the impact of dialect-aware artificial intelligence models on Arabic speech processing in IoT systems. A comparative evaluation is conducted across five representative Arabic varieties: MSA, Egyptian, Levantine, Gulf, and Algerian Arabic. The study adopts a simulation-based evaluation framework to compare two AI architectures: a baseline model that provides minimal support for dialectal variation and a dialect-aware model that explicitly incorporates dialect-specific information into the processing pipeline.

The evaluation focuses on key performance metrics, including automatic speech recognition (ASR) accuracy, intent recognition accuracy, task success rate, and system response time. The objective is to analyze performance variations across dialects and to demonstrate how dialect-aware modeling can improve robustness, responsiveness, and inclusivity in AI-driven IoT environments.

By examining this relatively underexplored research area particularly with respect to the underrepresented Algerian dialect, this work contributes to ongoing efforts toward linguistically inclusive and culturally adaptive AI systems. The main contributions of this paper are threefold: (i) the proposal of a conceptual dialect-aware framework for Arabic speech processing in IoT applications, (ii) a comprehensive comparative analysis across multiple Arabic dialects using simulated evaluation metrics, and (iii) a critical discussion of the advantages and limitations of dialect-aware modeling in low-resource and dialectally diverse contexts. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the methodology, Section 4 presents the simulation results, Section 5 discusses the findings and limitations, and Section 6 concludes the paper.

2. Literature Review

Arabic dialects pose a major challenge for ASR systems due to their high variability, lack of standardization, and limited availability of linguistic resources. Despite these challenges, prior research has made notable progress in addressing dialectal speech processing from both linguistic and technological perspectives.

Several studies have specifically focused on low-resource Arabic dialects, such as Algerian Arabic. In [6], the authors investigate ASR for the Algerian dialect and highlight the difficulties arising from limited annotated data and strong linguistic divergence from MSA. Their findings demonstrate that leveraging data from MSA and French can significantly enhance ASR performance, emphasizing the importance of dialect-aware and cross-lingual modeling strategies.

This observation aligns closely with the objectives of the present work, which also aims to reduce performance disparities caused by insufficient dialectal representation.

Broader investigations into Arabic dialectal ASR, such as [7], examine the role of dialect coverage during pre-training, dialect-specific fine-tuning, and multidialect modeling. The results indicate that incorporating multiple dialects generally improves ASR accuracy, particularly for low-resource dialects. These findings underscore the necessity of explicitly modeling dialectal variation to achieve robust and inclusive speech recognition systems.

Dialect identification has also been recognized as a crucial component of dialect-aware ASR pipelines. Studies such as [8] explore machine-learning approaches for Arabic dialect identification and highlight the challenges of recognizing dialects directly from acoustic features. The reported performance variability across models further supports the need for dialect-aware system designs, especially in voice-driven IoT applications where accuracy and responsiveness are critical.

Beyond Arabic-specific research, multilingual voice assistant studies provide valuable insights into inclusive speech technologies. The work presented in [9] focuses on multilingual assistants for underrepresented languages, including Hindi, Punjabi, and Bengali, and identifies key technical challenges related to speech recognition and language diversity. Although not centered on Arabic, these findings reinforce the importance of adaptable and language-aware architectures for intelligent systems.

Finally, [10] discusses the broader challenges of dialectal Arabic in speech recognition, including morphological complexity and the absence of standardization. The authors argue for continued research efforts to improve dialectal ASR accuracy, particularly in applications requiring natural human–machine interaction.

In summary, while substantial progress has been achieved in modeling Arabic dialects, many existing approaches prioritize large-scale training or offline evaluation and are not explicitly tailored to real-time or resource-constrained IoT environments. In contrast, the present study adopts a simulation-based, architectural perspective to evaluate the impact of dialect awareness on key performance metrics ASR accuracy, intent recognition accuracy, task success rate, and response time across five Arabic dialects. By comparing baseline and dialect-aware models, this work contributes to ongoing efforts to improve the robustness and inclusivity of voice-controlled smart systems.

3. Methodology

3.1 Simulation environment

All experiments were conducted within a controlled simulation environment designed to model the interaction between IoT systems and voice-based user commands across multiple Arabic dialects. The objective of this environment is to provide a descriptive and comparative evaluation of system behavior under different dialectal conditions, rather than to deploy or train production-level models. The simulation framework was implemented using Python 3.10, with TensorFlow 2.14 and PyTorch 2.0 employed to define and emulate deep learning model components. Librosa 0.10 was used for audio signal processing tasks, including feature extraction and spectrogram generation. All datasets were processed using a unified feature extraction pipeline to ensure consistency in input representation and enable a fair comparison across dialects.

The simulation pipeline consisted of the following sequential stages:

1) **Audio Input Generation:** Pre-recorded or synthetically simulated voice commands reflecting typical IoT interactions in the target dialects.

2) **Automatic Speech Recognition:** Speech transcription performed using either the Baseline Model or the Dialect-Aware Model.
3) **Intent Recognition Module:** A simulated Natural Language Understanding (NLU) component that extracts user intent from transcribed text. Intent recognition accuracy was modeled based on reported benchmarks in dialectal Arabic NLU literature.
4) **Task Execution Layer (IoT Emulator):** A virtual IoT interface emulating common smart environment actions (e.g., switching lights, adjusting temperature, or playing media).
5) **Latency and Response Time Estimation:** Network and processing delays were simulated to approximate system response time as a function of input complexity and ambiguity.

All simulations were executed on a high-performance workstation configured as follows:

- **CPU:** AMD Ryzen 9 7950X (16 cores, up to 5.7 GHz)
- **GPU:** NVIDIA RTX 3090 (24 GB VRAM)
- **RAM:** 128 GB DDR5
- **Operating System:** Ubuntu 22.04 LTS
- **Storage:** High-performance SSD

This setup enabled efficient execution of multiple simulation runs with GPU-accelerated inference, ensuring realistic latency modeling relevant to IoT responsiveness.

3.2 Data collection

Voice command data were collected for five Arabic varieties to ensure diversity and representativeness, with the dataset compiled from multiple complementary sources. Crowdsourcing platforms were utilized to gather naturally spoken, user-generated voice commands reflecting common IoT use cases, while open-source speech corpora, including VoxArabica and Mozilla Common Voice, provided standardized recordings for Modern Standard Arabic (MSA) and several regional dialects. Due to the limited availability of resources for Algerian Arabic, a locally collected dataset was developed, comprising recordings from 25 native speakers (13 male and 12 female) across five distinct regions, covering both urban and rural areas. This dataset captures regional lexical variation, frequent French–Arabic code-switching, and dialect-specific phonetic characteristics. For the remaining dialects MSA, Egyptian, Levantine, and Gulf speaker diversity was ensured through the integration of crowdsourced data and publicly available datasets, including contributions from speakers of different age groups, genders, and regional backgrounds. To ensure reliability, recordings were made under heterogeneous acoustic conditions, ranging from clean indoor environments to moderately noisy settings, with low-quality or unclear segments excluded or processed through a pipeline of noise reduction, resampling, and segmentation. This multi-source approach provided broad dialectal coverage while explicitly addressing the underrepresentation of Algerian Arabic in existing ASR datasets, resulting in a dataset that reflects realistic usage scenarios for voice-controlled IoT systems. The following table summarizes speaker distribution and datasets used across the five Arabic dialects.

Table 1. Speaker distribution and dataset characteristics across Arabic dialects.

Dialect	Number of Speakers	Gender Distribution	Regional Coverage	Recording Conditions
MSA (Fusha)	~40	Balanced (M/F)	Multiple Arab regions	Clean to moderate noise
Egyptian Arabic	~35	Balanced (M/F)	Urban & rural Egypt	Moderate background noise
Levantine Arabic	~30	Balanced (M/F)	Levant region	Mostly indoor recordings
Gulf Arabic	~28	Balanced (M/F)	Gulf countries	Clean to moderate noise
Algerian Arabic	25	13M / 12F	5 regions (urban & rural)	Code-switching, moderate noise

3.3 Data preprocessing

To ensure consistency and enhance robustness across dialects, all audio samples underwent the same preprocessing pipeline:

- **Noise Reduction:** Background noise suppression using a spectral gating algorithm.
- **Sample Rate Normalization:** Resampling of all audio signals to 16 kHz.
- **Segmentation:** Division of long recordings into short command-level utterances (1–3 seconds).
- **Feature Extraction:** Conversion of audio segments into Mel-spectrogram representations suitable for ASR modeling.

This standardized preprocessing ensured comparability across datasets and minimized bias introduced by recording conditions.

3.4 Models Compared

To evaluate the impact of explicit dialect modeling on Arabic speech recognition in IoT environments, two ASR models were conceptually designed and compared within a controlled simulation framework: a Baseline Model and a Dialect-Aware Model. Both models follow an end-to-end architecture and operate on identical preprocessed inputs to ensure a fair comparison.

1) Baseline Model

The Baseline Model represents a conventional Arabic ASR system optimized for MSA. It is implemented using an end-to-end Transfots of six Transformer encoder layers, each with a model dimension of 512, eight self-attention heads, and a feed-forward network size of 2048. The decoder mirrors this configuration with six Transformer decoder layers and an output softmax layer for token prediction. Positional encoding is applied to preserve temporal information. This model is trained exclusively on MSA speech data and does not incorporate any dialectal adaptation mechanisms, dialect identification modules, or specialized embeddings. As a result, it serves as a reference system for assessing performance degradation when applied to non-MSA dialects.

2) Dialect-Aware Model

The Dialect-Aware Model extends the baseline architecture by explicitly modeling dialectal variability. It adopts a multi-task, dialect-conditioned architecture composed of three main components: a shared acoustic encoder, a dialect classification module, and a dialect-conditioned ASR decoder. The shared encoder follows the same Transformer configuration as the Baseline Model (six layers, model dimension 512, eight attention heads), ensuring architectural comparability. On top of the encoder, a Dialect Classification Layer is introduced, consisting of a fully connected layer with a softmax activation, which predicts one of the five target dialects (MSA, Egyptian, Levantine, Gulf, or Algerian). The predicted dialect label is used to condition the decoding process through dialect-specific phoneme embeddings. These embeddings are concatenated with the encoder representations before being passed to the decoder, allowing the model to adapt its acoustic–phonetic modeling to dialect-specific characteristics. To enhance robustness across dialects, a domain adaptation strategy is conceptually modeled, in which the majority of model parameters are shared across dialects, while a small subset of dialect-dependent parameters is fine-tuned per dialect. Regularization mechanisms, including dropout with a rate of 0.1, are incorporated to mitigate overfitting in the simulated setup. It is important to emphasize that the Dialect-Aware Model is evaluated within a simulation-based descriptive framework. The architecture and hyperparameters reflect commonly adopted design choices in recent dialect-aware ASR literature, and the reported results represent expected performance trends rather than outcomes of fully trained production systems.

Figure 1 presents the conceptual architecture of the proposed dialect-aware ASR system, highlighting the integration of dialect classification, dialect-specific embeddings, and intent-driven IoT task execution.

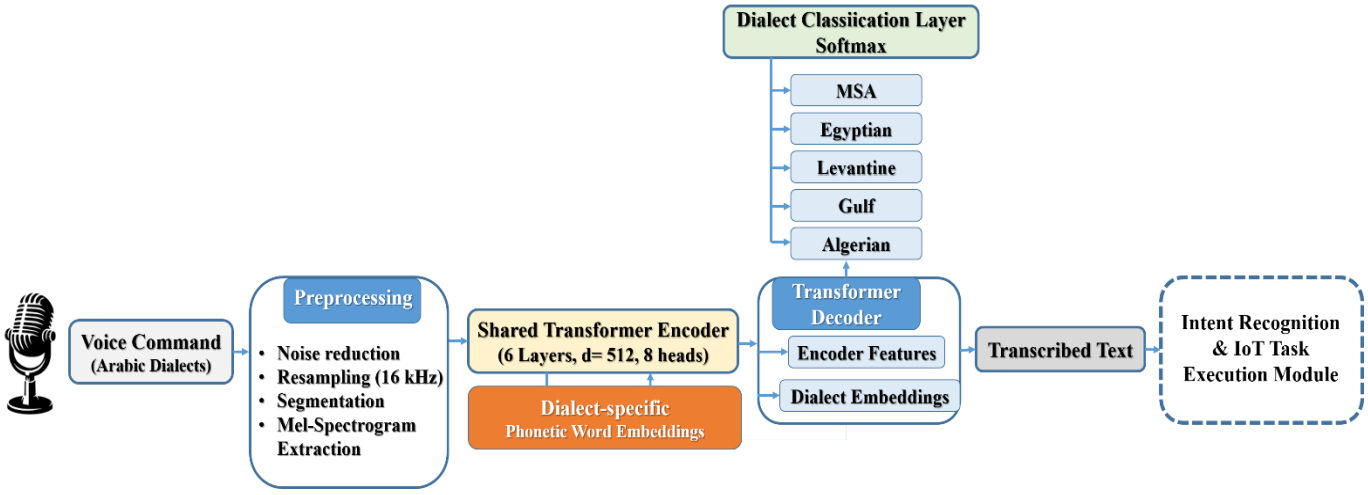

Figure 1. Architecture of the Proposed Dialect-Aware ASR Model for Arabic Voice-Controlled IoT Systems.

3.5 Evaluation Metrics

The system performance was assessed using four complementary metrics reflecting both linguistic accuracy and IoT responsiveness:

1. **ASR Accuracy (%):** Ratio of correctly transcribed words to the total number of spoken words.
2. **Intent Recognition Accuracy (%):** Percentage of correctly identified user intents.
3. **Task Success Rate (%):** Proportion of IoT actions successfully executed based on the interpreted command, representing end-to-end system effectiveness.
4. **Average Response Time (s):** Simulated latency measured from voice input initiation to task completion.

These metrics collectively enable a comprehensive comparison of baseline and dialect-aware approaches across different Arabic dialects.

4. Simulation results

This section presents the results obtained from the simulated evaluation of the Baseline and Dialect-Aware models across five Arabic dialects: MSA, Egyptian, Levantine, Gulf, and Algerian Arabic. The objective is to analyze how dialectal variation impacts system performance in voice-controlled IoT environments and to assess the relative benefits of incorporating dialect-aware mechanisms. All reported results are derived from a controlled simulation framework, designed to reflect performance trends commonly observed in dialectal Arabic ASR and NLU systems, rather than from real-world deployment or full-scale training.

4.1 Automatic Speech Recognition Accuracy

To evaluate the effectiveness of dialect-aware AI in speech recognition, we compared the ASR accuracy of two AI models across the five Arabic dialects (see Table 2).

Table 2. Comparison of ASR accuracy in different Arabic dialects using baseline and dialect models.

Dialect	Baseline Model (%)	Dialect-Aware Model (%)
MSA (Fusha)	91.3	92.7
Egyptian Arabic	72.5	87.4
Levantine Arabic	74.1	85.9
Gulf Arabic	69.8	84.2
Algerian Arabic	58.2	80.6

As Figures 2 and 3 show, the baseline model performs well on MSA (91.3%), but on other dialects, especially Algerian Arabic, its performance deteriorates considerably. The Dialect-Aware Model, on the other hand, achieves a remarkable increase in accuracy in all dialects, especially for low-resource variants such as Algerian (+22.4%) and Gulf Arabic (+14.4%). This shows the importance of adapting the AI models to take into account the phonetic and lexical differences in dialectal Arabic in order to improve speech recognition performance.

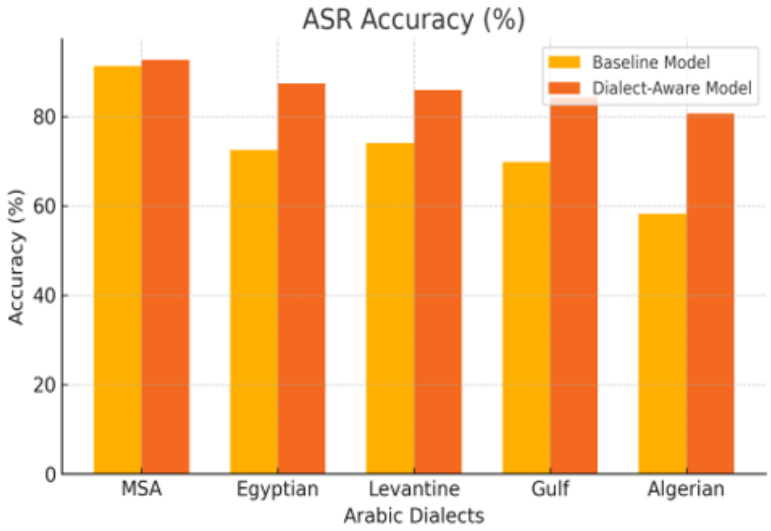

Figure 2. Bar chart of ASR accuracy across Arabic dialects: Baseline vs. dialect-aware model performance.

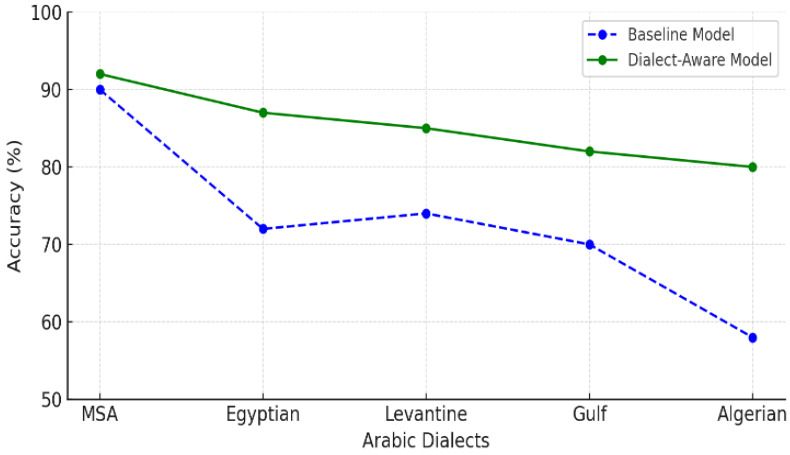

Figure 3. Line chart of ASR accuracy variation across Arabic dialects comparing Baseline and dialect-aware models.

4.2 Intent Recognition Accuracy

Table 3 compares intent recognition accuracy across five Arabic dialects using baseline and dialect-aware models.

Table 3. Dialect-wise intent recognition performance: baseline vs. Dialect-Aware Models.

Dialect	Baseline Model (%)	Dialect-Aware Model (%)
MSA (Fusha)	93.5	94.6
Egyptian Arabic	76.3	89.2
Levantine Arabic	77.5	88.0
Gulf Arabic	71.1	86.7
Algerian Arabic	60.4	83.4

The results show that the dialect-aware model consistently outperforms the baseline model for all dialects. There are particularly notable improvements for underrepresented dialects, such as Algerian Arabic (improving from 60.4% to 83.4%) and Gulf Arabic (improving from 71.1% to 86.7%). Even for MSA, which usually achieves high accuracy, the dialect-aware model shows a modest improvement (from 93.5% to 94.6%). These results highlight the effectiveness of incorporating dialect-specific features in enhancing intent recognition performance across diverse Arabic dialects (see Figures 4 and 5).

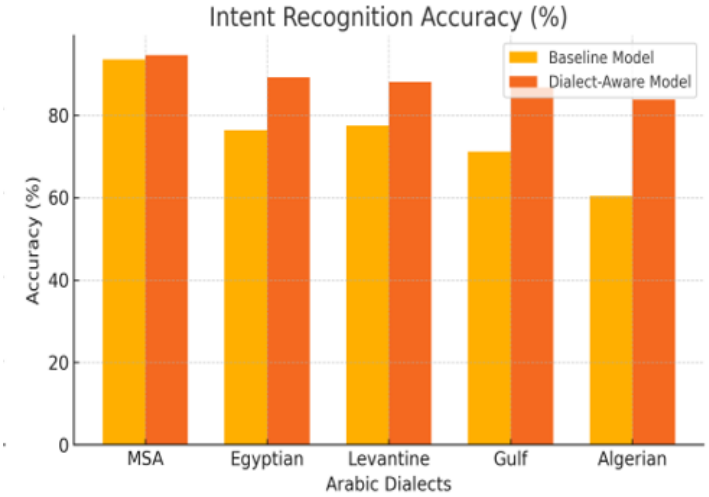

Figure 4. Bar Chart of Intent Recognition Accuracy Across Arabic Dialects: Baseline vs. Dialect-Aware Models.

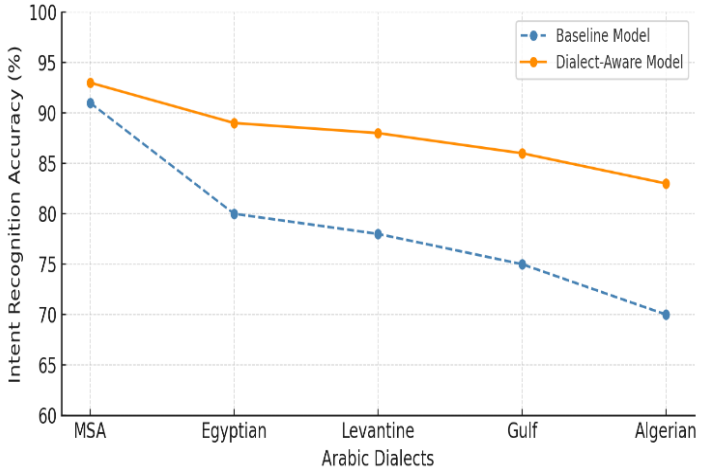

Figure 5. Line Chart of Intent Recognition Accuracy Trends: Comparison Between Baseline and Dialect-Aware Models Across Arabic Dialects.

4.3 Task Success Rate

The task success rate measures the system's ability to correctly execute voice commands from input to final action. As can be seen in Table 4, the Dialect-Aware Model outperforms the Baseline Model by a significant margin across all five Arabic dialects.

Table 4. Task success rate across Arabic dialects for baseline vs. Dialect-Aware AI Models.

Dialect	Baseline Model (%)	Dialect-Aware Model (%)
MSA (Fusha)	89.9	91.8
Egyptian Arabic	70.4	85.0
Levantine Arabic	71.2	84.6
Gulf Arabic	66.5	82.4
Algerian Arabic	52.6	78.1

As Figures 6 and 7 show, Algerian Arabic has seen the most dramatic improvement, with task success rising from 52.6% to 78.1%. This highlights the importance of addressing complex code-switching and phonetic variation. Gulf and Egyptian Arabic also demonstrate significant improvements of around 17–19 percentage points. Even MSA (Fusha), which is already well supported, shows a modest improvement. These results reinforce the idea that modelling specific to a given dialect is critical not only for recognition, but also for effective end-to-end interaction in AI-powered IoT environments.

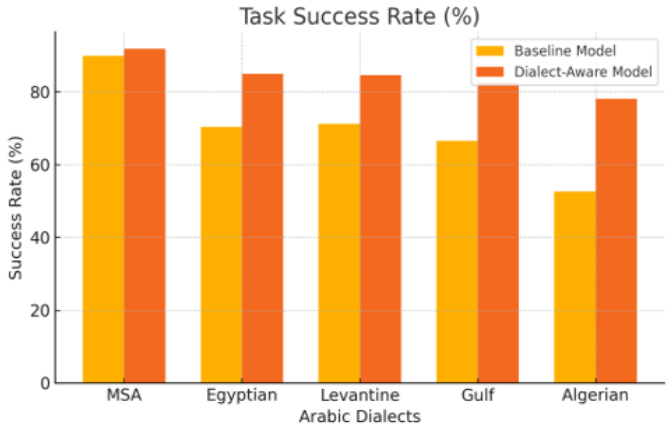

Figure 6. Bar Chart of task success rate across dialects using two AI models.

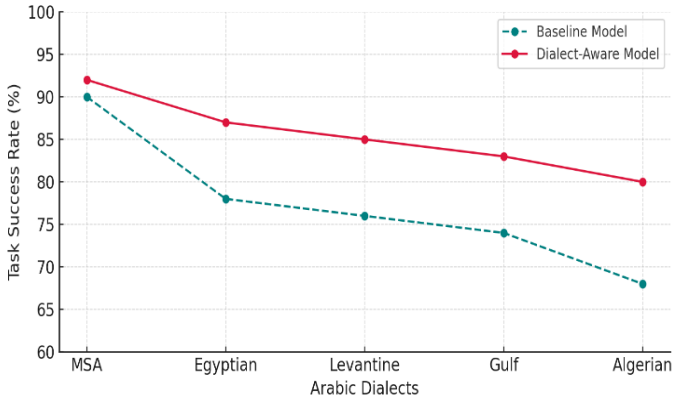

Figure 7. Line Chart of task success rate across dialects using two AI models.

4.4 System Response Time

The average system response time for each dialect using two AI models is presented in Table 5.

Table 5. System response time for baseline vs. Dialect-Aware Models Across Arabic Dialects.

Dialect	Baseline Model (s)	Dialect-Aware Model (s)
MSA (Fusha)	1.2	1.1
Egyptian Arabic	1.4	1.2
Levantine Arabic	1.5	1.3
Gulf Arabic	1.6	1.4
Algerian Arabic	1.9	1.5

The following graphs (Figs 8 and 9) clearly show that the response time when using AI models for processing voice commands in IoT systems varies depending on the Arabic dialect. The Dialect-Aware Model consistently outperforms the Baseline Model and provides faster responses in all dialects.

For example, MSA shows the fastest response at 1.1 seconds in the Dialect-Aware Model, compared to 1.2 seconds in the Baseline Model. In contrast, the Algerian dialect shows the highest latency in both models, with 1.9 seconds for the baseline model and 1.5 seconds for the dialect-aware model. This discrepancy highlights the greater challenges in processing regional dialects, which are likely due to linguistic variability and underrepresentation in traditional models. These results highlight the importance of dialect-specific optimizations to ensure real-time responsiveness and inclusivity in Arabic-speaking intelligent environments.

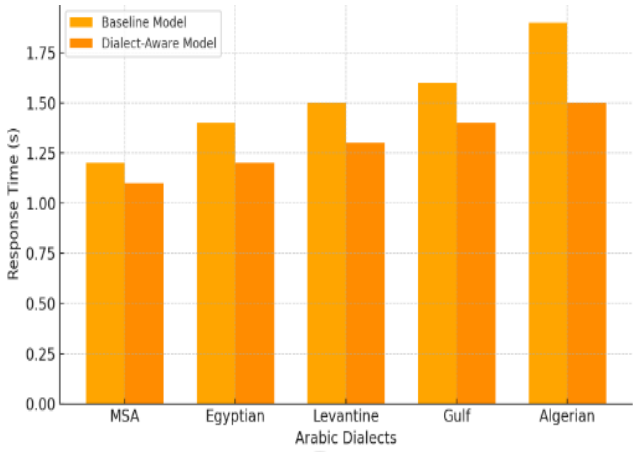

Figure 8. Bar Chart of Average response time for Baseline vs. Dialect-Aware Models across Arabic dialects.

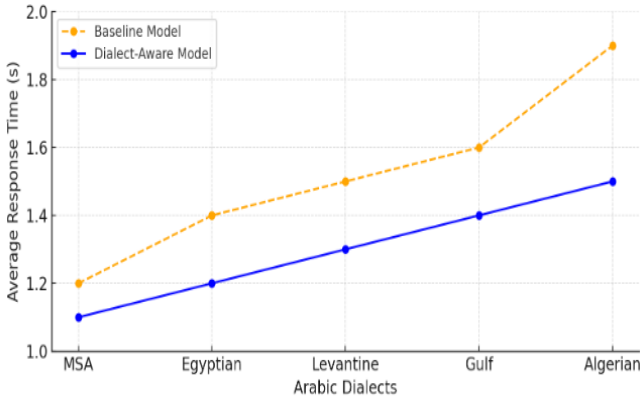

Figure 9. Line Chart of Average response time for Baseline vs. Dialect-Aware Models across Arabic dialects.

5. Discussion

5.1 Impact of Dialect Awareness on ASR and NLU Performance

The simulation results indicate that explicitly modeling dialectal variation yields consistent improvements across all evaluated metrics, including ASR accuracy, intent recognition accuracy, and task success rate. These improvements are most pronounced for non-standard and low-resource dialects, particularly Algerian Arabic, which exhibits significant lexical variation, phonetic diversity, and frequent code-switching. These findings are aligned with recent studies emphasizing the importance of dialect coverage and multi-dialect training in Arabic ASR systems. For instance, Ali et al. [5] and Djanibekov et al. [7] demonstrate that models trained with broader dialectal exposure achieve better generalization and robustness. Similarly, VoxArabica [8] highlights the role of dialect-aware representations in mitigating performance degradation across regional varieties. The observed gains in intent recognition further suggest that improvements at the ASR level propagate positively to downstream Natural Language Understanding (NLU) components. This is particularly relevant for IoT applications, where accurate intent extraction directly determines task execution success.

5.2 Relevance to Real-Time IoT Systems

From an IoT systems perspective, the task success rate and response time are critical indicators of practical usability. The results show that the Dialect-Aware Model achieves a more balanced task success rate across dialects, reducing the disparity between MSA and regional varieties. This suggests that dialect-aware ASR can contribute to more inclusive and user-friendly smart environments. Although the Dialect-Aware Model introduces a marginal increase in response time due to additional processing stages, this overhead remains within acceptable limits for real-time IoT interactions. Similar trade-offs between accuracy and latency have been reported in recent multilingual voice assistant studies [9], where modest increases in processing time are justified by substantial gains in interaction reliability.

5.3 Handling Linguistic Variability and Code-Switching

One of the key challenges addressed in this study is the handling of linguistic variability, particularly in dialects characterized by frequent code-switching, such as Algerian Arabic. While the simulation framework does not explicitly model code-switching at the token level, the improved performance of the Dialect-Aware Model suggests that dialect-specific embeddings and joint training strategies can partially mitigate its effects. This observation is consistent with prior work highlighting the benefits of multilingual and multidialect training for handling mixed-language inputs [6], [9]. However, fully addressing code-switching remains an open research challenge and warrants dedicated modeling strategies in future work.

5.4 Limitations and External Validity

Despite the promising results, several limitations must be acknowledged. First, the evaluation was conducted entirely within a simulated environment, without real-world user testing or deployment. As a result, the findings should be interpreted as indicative performance trends rather than empirically validated improvements. Second, no statistical significance testing was performed, and the reported results are based on modeled accuracy values derived from prior benchmarks. Future work should incorporate real datasets, multiple experimental runs, and statistical validation to strengthen the reliability of the conclusions. Third, the study focuses on five Arabic dialects, which, although representative, do not capture the full diversity of Arabic varieties such as Sudanese or Moroccan Arabic. Therefore, the generalizability of the results to other dialects remains an open question.

5.5 Implications for Future Research

The findings of this study highlight several directions for future research. Integrating real-world speech data and evaluating the models under noisy and dynamic IoT environments would significantly enhance external validity. Additionally, incorporating state-of-the-art pretrained models, such as dialectal AraBERT variants or multilingual systems like Whisper, could provide a more comprehensive comparison with current state-of-the-art approaches. Furthermore, extending dialect coverage and explicitly modeling code-switching phenomena would contribute to more robust and culturally adaptive voice-enabled IoT systems.

6. Conclusion and future research work

This research paper investigated the effectiveness of Arabic dialect processing in voice-enabled IoT systems through a comparative, simulation-based evaluation of a baseline ASR model and a dialect-aware model. The results demonstrate consistent improvements in ASR accuracy, intent recognition accuracy, task success rate, and system response time across five representative Arabic varieties: MSA (الفصحى), Egyptian, Levantine, Gulf, and Algerian dialects. The dialect-aware model exhibited superior robustness, particularly for linguistically complex and low-resource dialects such as Algerian Arabic, which are characterized by phonetic variability and frequent code-switching. These findings highlight the importance of explicitly modeling dialectal variation when designing AI-driven voice interfaces for Arabic-speaking users. Despite these promising results, the study remains subject to certain limitations. The evaluation was conducted within a controlled simulation environment, which may not fully capture the acoustic variability, spontaneous speech patterns, and contextual complexity of real-world IoT deployments. Additionally, the analysis was limited to five dialects, and the dialect-aware model relied on existing corpora that may not encompass the full linguistic diversity present in everyday usage.

Future work will focus on empirical validation in real IoT environments involving native speakers from diverse Arabic dialect regions. This includes the collection of real-world voice commands, the development of annotated datasets for underrepresented dialects, and the integration of noise-robust and domain-adaptive learning strategies. Furthermore, transfer learning and domain-specific fine-tuning will be explored to enhance generalization, reduce response-time variability, and improve cultural and linguistic inclusiveness in AI-powered smart environments.

References

[1] R. Djehaiche, S. Aidel, M. Belazzoug, and N. Saeed, “Design, implementation, and deployment of IoT/M2M smart city applications based on MCNs,” in *Advanced Computational Techniques for Renewable Energy Systems*, M. Hatti, Ed., Lecture Notes in Networks and Systems, vol. 591. Cham, Switzerland: Springer, 2023, pp. 55–67. https://doi.org/10.1007/978-3-031-21216-1_5

[2] R. Djehaiche, S. Aidel, and N. Saeed, “Implementation of M2M-IoT smart building system using Blynk app,” in *Artificial Intelligence and Heuristics for Smart Energy Efficiency in Smart Cities*, M. Hatti, Ed., Lecture Notes in Networks and Systems, vol. 361. Cham, Switzerland: Springer, 2022, pp. 451–460. https://doi.org/10.1007/978-3-030-92038-8_44

[3] R. Djehaiche, *Deployment and Convergence of M2M Networks and 4G/5G Mobile Networks*, Ph.D. dissertation, Faculty of Science and Technology, Univ. of Bordj Bou Arréridj, Algeria, 2023.

[4] R. Djehaiche, S. Aidel, A. Sawalmeh, N. Saeed, and A. H. Alenezi, "Adaptive control of IoT/M2M devices in smart buildings using heterogeneous wireless networks," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7836–7849, Apr. 2023. https://doi.org/10.1109/JSEN.2023.3247007

[5] A. Ali, S. Chowdhury, M. Afify, W. El-Hajj, H. Hajj, M. Abbas, and A. Alqudah, "Connecting Arabs: Bridging the gap in dialectal speech recognition", *Communications of the ACM*, vol. 64, no. 4, pp. 124–129, Apr. 2021.

[6] M. A. Menacer and K. Smaïli, "Investigating data sharing in speech recognition for an under-resourced language: The case of Algerian dialect" in *Proc. 7th Int. Conf. on Natural Language Processing (NATP)*, 2021.

[7] A. Djanibekov, H. O. Toyin, R. Alshalan, A. Alitr, and H. Aldarmaki, "Dialectal coverage and generalization in Arabic speech recognition" *arXiv preprint arXiv:2411.05872*, 2024.

[8] A. Waheed, B. Talafha, P. Sullivan, A. Elmadany, and M. Abdul-Mageed, "VoxArabica: A robust dialect-aware Arabic speech recognition system" *arXiv preprint arXiv:2310.11069*, 2023.

[9] S. Sabharwal and R. Sahni, "Tackling the problem of multilingualism in voice assistants", *International Journal of Electrical, Electronics and Computers*, Vol-9, Issue-5 2024. https://dx.doi.org/10.22161/eec.95.1

[10] H. A. Alsayadi, A. A. Abdelhamid, I. Hegazy, B. Alotaibi, and Z. T. Fayed, "Deep investigation of the recent advances in dialectal Arabic speech recognition" *IEEE Access*, vol. 10, pp. 57063–57079, 2022. https://doi.org/10.1109/ACCESS.2022.3177191