

# Scientists develop machine learning tool to accurately identify Arabic dialects in 22 Arabic-speaking countries

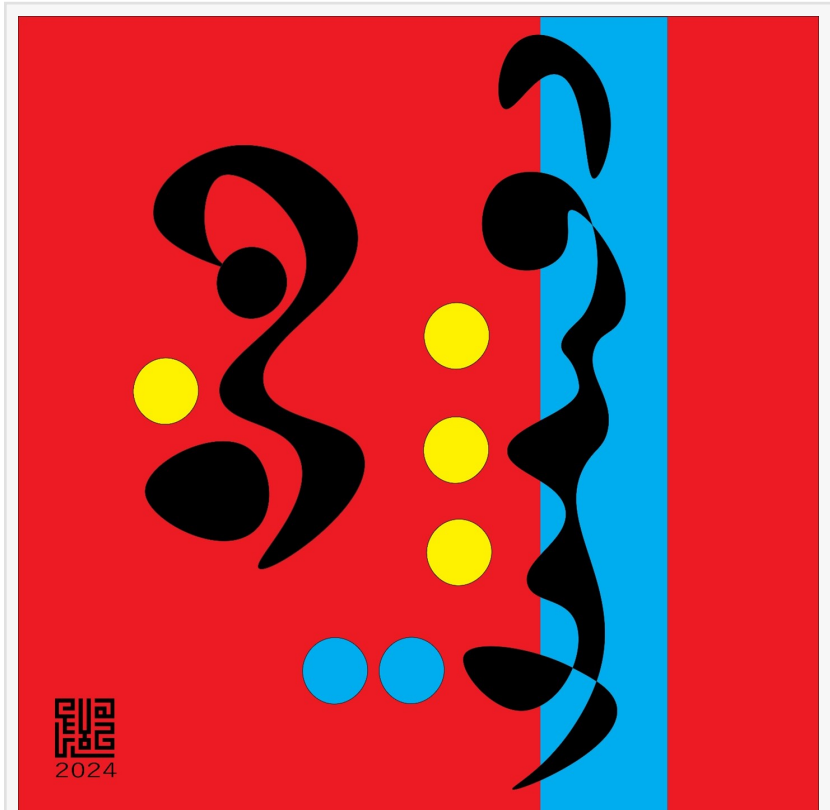
*Sharjah University scientists design AI tool to identify 19 different Arabic dialects spoken in 22 countries in which Arabic is the official language.*

SHARJAH, EMIRATE OF SHARJAH, UNITED ARAB EMIRATES, October 8, 2024 /EINPresswire.com/ -- Scientists from the University of Sharjah believe they have created an artificial intelligence system that can automatically identify which Arabic dialect someone is speaking.

They say their system unravels the rich and complex tapestry of Arabic dialects which hitherto conventional speech systems fall short of accurately interpreting and identifying.

“Arabic is a rich language with many regional dialects, and each one has its own unique vocabulary, expressions, and pronunciation. This diversity makes it challenging for technology to accurately understand and differentiate between them,” said Ashraf Elnagar, Professor of Computer Science and Intelligence Systems. “To address this, we developed a system that can automatically identify which Arabic dialect someone is speaking.”

The official language in 22 countries spanning the Middle East, North Africa and the Arabian Peninsula, Arabic is one of the most spoken languages globally with more than 370 million people having it as their mother tongue. It is also one of the world’s most immersed languages in culture and those having it as a mother tongue or learning it as a second or foreign language



The official language in 22 countries spanning the Middle East, North Africa and the Arabian Peninsula, Arabic is one of the most spoken languages globally. Credit: Dr. Hala Georges, College of Fine Arts and Design, University of Sharjah.

find themselves learning about Islam and its culture as well.

With a totally different alphabet than English, the language has numerous sounds that are specific to its phonology. The charm of its sounds and characters bewilders countless foreign learners who aspire to speak it fluently. Though most learning of the Arabic language occurs in the standard formal variety, many foreign learners opt for colloquial or daily versions, particularly the spoken forms in currency in Egypt and Syria.

The authors say they faced not an easy task in their attempt when teaching computers to recognize different Arabic dialects just by listening to spoken words. They write, "The primary challenge is the development of a machine learning model capable of accurately identifying a wide range of Arabic dialects from audio recordings.

"This task is compounded by the inherent diversity and complexity of Arabic dialects, coupled with the technical challenges of audio processing and machine learning model optimization."

The authors utilized datasets comprising more than 3000 hours of audio segments collected from YouTube. The data includes 19 different dialects spoken in Algeria, Egypt, Iraq, Jordan, Saudi Arabia, Kuwait, Lebanon, Libya, Mauritania, Tunisia, Morocco, Oman, Palestine, Qatar, Sudan, Syria, the United Arab Emirates (U.A.E.), Bahrain and Yemen.

“

“Our model correctly identified regional dialects 97.29% of the time and specific country dialects 94.92% of the time.”

*Prof. Ashraf Elnagar*

The results were impressive, said Prof. Elnagar, underscoring the model's high accuracy in Arabic dialect identification regionally and at country levels. "Our model correctly identified regional dialects 97.29% of the time and specific country dialects 94.92% of the time.

"What is remarkable is that we achieved this using only

29% of the training data typically required by other researchers. We have made our models publicly available so that other researchers and developers can use them to create better speech-related technologies for Arabic speakers."



The project has the potential to enhance communication and accessibility for millions of Arabic speakers worldwide. Credit: Dr. Hala Georges, College of Fine Arts and Design, University of Sharjah.

The project has the potential to enhance communication and accessibility for millions of Arabic speakers worldwide. Prof. Elnagar said the model's ability to correctly identify a dialect can "improve voice-activated technologies like virtual assistants, translation services, and automated customer support systems. "This not only bridges communication gaps between different Arabic-speaking regions but also contributes to making technology more inclusive and user-friendly for Arabic speakers."

Despite the astounding results, Prof. Elnagar noted, the project can still be improved. For this purpose, the authors have made their system publicly available "online on a platform called HuggingFace, so others can access and build upon our work to improve Arabic language technologies."



Arabic is a rich language with many regional dialects, and each one has its own unique vocabulary, expressions, and pronunciation. Credit: Dr. Hala Georges, College of Fine Arts and Design, University of Sharjah

The research, published in IEEE Xplore, is the outcome of collaboration between Prof. Elnagar and three of his undergraduate students as part of a project to build a deep learning model for Arabic dialect identification from speech. The initial research results were first presented at the 15th Annual Undergraduate Research Conference on Applied Computing (URC) in 2024. (Original Source URL: <https://ieeexplore.ieee.org/document/10604557>)

"Developed by our dedicated students, the technology behind our system integrates cutting-edge methodologies and deep learning techniques. Expanding its functionality from text to audio signals sets it apart, providing a multi-modal approach to understanding and processing the Arabic language," Prof. Elnagar said.

For student researcher Amr Barakat the project "bridges a critical gap in language technology, enabling more inclusive and accurate communication for Arabic speakers worldwide. By leveraging advanced machine learning, we have created a model that not only excels in performance but also paves the way for future innovations in speech recognition."

Another student researcher Abdulla Aldhaheeri reported wide interest from the industry in the project, as it "holds the potential for widespread adoption, offering numerous benefits and improvements to various AI-driven language applications and services."

Besides its high accuracy, the tool the authors have developed, unlike currently available models, requires less data and computational resources, rendering it accessible for wider use. This feature, according to the authors, was behind the industry's interest in their work. They cited tech corporations like Microsoft and governmental bodies in Sharjah in the U.A.E. as being particularly enthusiastic about their work.

LEON YOUSIF BARKHO  
University Of Sharjah  
+971 501654376  
[email us here](#)



With a totally different alphabet than English, the language has numerous sounds that are specific to its phonology. The charm of its sounds and characters bewilders countless foreign learners who aspire to speak it fluently. Credit: Dr. Hala Georges, Co

This press release can be viewed online at: <https://www.einpresswire.com/article/749915209>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2024 Newsmatics Inc. All Right Reserved.