SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
TUWAIQ ACADEMY

T5

Arabic Recognition and Correct Pronunciation

Abstract

Automatic Arabic speech recognition implementing a sophisticated framework employing deep learning techniques, specifically convolutional neural networks (CNN). The CNN algorithms are adept at autonomously extracting hidden patterns from data, amalgamating them, and constructing more effective decision rules. Addressing a prevalent issue among students struggling with Arabic letter pronunciation, especially with the surge in e-learning, the model aims to assist users in recognizing and correctly pronouncing Arabic letters. By inputting a phonetically accurate sound for a letter, the system verifies whether the pronunciation is correct. Additionally, it includes features for reading written text and their corresponding pronunciations to recognize complete words using optical character recognition (ocr). Furthermore, the model incorporates a chatbot functionality for engaging in conversations and answering questions in the Arabic language. The investigation into Arabic alphabet recognition is carried out using deep convolutional neural networks, employing two distinct models: First, the alphabet classifier based on custom-built CNN layers, achieve an impressive accuracy rate of 94%. Second, the automatic Arabic speech recognition named Klaam, utilizes a transformer technique.

Introduction

Arabic, with its 290 million native speakers and 132 million non-native speakers, holds the distinction of being the most widely spoken language. This ancient language has two main dialects: Classical Arabic (CA) and modern standard Arabic (MSA). CA is the language of the Quran, while MSA is its modified version, commonly used in everyday communication.

In the pursuit of educational goals and the empowerment of individuals, self-directed learning has emerged as a powerful tool. With the advent of artificial intelligence (AI), this approach to learning takes on a new dimension. AI becomes a crucial partner in our journey towards development, offering personalized and innovative learning experiences. One such example is the use of voice user interfaces, which can be employed in educational settings and everyday life. These interfaces, based on speech recognition technology, not only aid in second language acquisition but also help learners improve their pronunciation and speaking fluency. Information and communication technology (ICT) plays an essential role in education. It not only increases student awareness but also enables teachers to employ more effective teaching methods. By harnessing the power of ICT, educators can create engaging and interactive learning environments that cater to the diverse needs of their students. By embracing AI and ICT, we can pave the way for a future where education is accessible to all, tailored to individual needs, and driven by innovation.

Literature Review:

- Nishmia Ziafat ,Hafiz Farooq Ahmad ,Iram Fatima ,Muhammad Zia ,Abdulaziz Alhumam andKashif Rajpoot. “Correct Pronunciation Detection of the Arabic Alphabet Using Deep Learning,” 2021. https://www.mdpi.com/2076-3417/11/6/2508.
- May Hanoon, Dalal Yassin, and Hatem Ratrout. “Arabic Letter Recognition and Pronunciation Evaluation”. https://github.com/HatemRatrout/Arabic-Letter-Recognition-and-Pronunciation-Evaluation

Data Description and Structure :

The primary data source for this project is the Arabic Alphabet short audio dataset, which was obtained from Kaggle. The dataset consists of audio recordings of native Arabic speakers pronouncing each letter of the Arabic alphabet. The recordings were collected from various sources and contributors, ensuring a diverse range of voices. The Arabic Alphabet short audio dataset is organized in a structured format, facilitating easy access and analysis. The data is primarily stored as individual audio files, with each file representing a specific letter of the Arabic alphabet. These audio files are organized into a hierarchical folder structure, where each folder corresponds to a particular letter. This organization allows for efficient retrieval and management of the data. Preprocessing steps, before utilizing the data for modeling, some preprocessing steps were undertaken to ensure its quality and usability. These steps included:

First, Data cleaning: the audio files were carefully reviewed to remove noise and silence from audio.
Second, Standardization: to maintain consistency across the dataset, all audio files were normalized to a standardized audio format. This involved adjusting the audio channel and sampling rates to ensure uniformity. Third, Labeling: Each audio file was labeled with the corresponding letter of the Arabic alphabet it represents. This labeling process was essential for supervised learning tasks and allowed for accurate classification and analysis.

Methodology

In the development of the Arabic Alphabet Classifier, we employed a Convolutional Neural Network (CNN) as the foundational model. The input for our CNN model consisted of mel spectrograms derived from audio data, each with dimensions of 128x32. Our CNN architecture comprised two convolutional layers, two max-pooling layers, and a Flatten Layer. The output layer featured 28 labels and employed the softmax aggregation function. This specific configuration was chosen for its effectiveness in extracting pertinent features from the input data. To assess the model's performance, we employed metrics such as accuracy, which gauges the proportion of correctly classified instances relative to the total instances. In our project, the model demonstrated an accuracy of 94%, underscoring its efficacy in accurately classifying Arabic alphabet characters. In the context of Arabic Speech Recognition, we utilized A Klaam as a transformer due to its proficiency in recognizing the Arabic language, even with the addition of Tashkeel. In our educational chatGpt, named " "مساعدي التعليمي)Educational Assistant), we customized it with Arabic language lessons content for responsive interactions. The framework employed for this purpose is LongChain.

Discussion and Results:

Our project was centered around Arabic alphabet Recognition and Pronunciation Evaluation. Through extensive research and analysis, we collected valuable data, leading to meaningful conclusions. The significance of our project's results lies in several aspects. Firstly, our findings align with previous research in the field, providing credibility to our approach. This consistency contributes to the existing body of knowledge and expands our understanding of the subject. Despite encountering limitations and challenges, such as the difficulty in finding a representative dataset with diverse speakers, impacting our model's bias towards one voice type (male voices), we addressed this by incorporating Arabic speech recognition into the pronunciation checker's conditional statement. Acknowledging the encountered challenges, it's crucial to mention our initial attempt to use an audio vector with RNN. However, due to the extended processing time and inadequate accuracy, we faced constraints with the available GPU and couldn't add more epochs. Despite this setback, our project exceeded initial expectations and aligns with Saudi Vision 2030 objectives. The discussion and analysis of our findings offer valuable insights into the project's implications, emphasizing the potential for further advancements in line with the vision for Saudi Arabia's future.

Conclusion and Future Work

When looking at the "Vision 2030" in the ever-changing world of education and technology, adding Arabic speech recognition to a special learning platform for kids is a great move. It helps correct pronunciation and assists in learning phrases, making language education more up-to-date. By using technology for teaching, this initiative not only helps kids learn better but also preserves and promotes the Arabic language and culture. This creative approach shows a dedication to giving young learners a well-rounded and effective educational experience.

Future Work:

- Connecting the system to the Ministry of Education's Madrasati platform.
- Building a follow-up system for the child's education and linking it with the school and parents.
- Mobile and tablet application.
- Developing a version to teach the Arabic language to non-native speakers.

Team Members:

Mahad Aljohani
Ebtesam Alansari
Wejdan Ahmed
Waad Ghos
Renad Alnajaei

Supervised by:

Ali H. El-kassas