IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

The Development of Machine Learning-Based Arabic Pronunciation Learning Classifier using Teachable Machine

Siti Sara Haji Ahmad[1], Dayang Hajah Tiawa binti Awang Haji Hamid[2], Abdul Hafidz bin Haji Omar[3], Muhamadul Bakir bin Yaakob[4], Achmad Yani[5], Muhamad Alif bin Haji Sismat[6], Nur Almira Ufairah binti Salim[7], and Nur Basirah binti Hj Rosmin[8]

[1,4,5,6,8]Faculty of Arabic Language, Universiti Islam Sultan Sharif Ali, Negara Brunei Darussalam

[2,3,7]Faculty of Islamic Technology, Universiti Islam Sultan Sharif Ali, Negara Brunei Darussalam

Abstract

Voice recognition technology has emerged as a valuable tool in language learning, offering real-time pronunciation feedback often missing in traditional classrooms. With the global voice recognition market valued at over USD 12.6 billion in 2022 and expected to exceed USD 50 billion by 2030, its integration into education is becoming increasingly significant, particularly in language apps where over 70% now use speech input to support speaking skills. In Brunei, students learning Arabic face challenges in mastering pronunciation due to limited speaking practice and a lack of individualized corrective feedback. This study aims to address the local need by developing a machine learning-based voice recognition model for evaluating Arabic pronunciation accuracy. The machine learning model was developed using Google's Teachable Machine, trained on a custom dataset of four Arabic words spoken by both adult and child voices. Each word class included approximately 30 audio samples in total. After training, the model was exported and embedded into a Unity-based game environment designed to guide

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

"The Development of Machine Learning-Based Arabic Pronunciation Learning Classifier using Teachable Machine"

learners through pronunciation tasks, with real-time visual feedback triggered by the model's classification. This study showed the system was able to recognize pronunciation patterns with moderate accuracy and deliver feedback through an intuitive user interface. These results demonstrate the potential of using accessible AI tools and small, diverse datasets to develop interactive pronunciation aids, especially in regions where Arabic is taught as a second language.

Keywords: Voice Recognition Technology, Arabic Pronunciation, Automated feedback system

1.0 Introduction

Voice recognition technology has become an important tool across many fields, including language learning. Traditionally, learning a new language has relied heavily on classroom instruction, textbooks, and teacher-guided pronunciation practice (Sholihah & Kholis, 2025). However,for second-language learners, especially those aiming to communicate fluently with native speakers, traditional methods often lack real-time support for pronunciation correction (Lázaro, 2024). This limitation has encouraged the growing use of machine learning technologies to fill the gap.

Machine learning, a branch of artificial intelligence (AI), has expanded rapidly in recent decades alongside the digital revolution. It involves designing algorithms that can learn and improve from experience without needing explicit programming. These algorithms are trained on sample data to detect patterns and make predictions or decisions when encountering new information. Some of the most common machine learning techniques include neural networks, decision trees, and support vector machines.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

Recent data reflects the rapid rise of voice recognition in both daily life and education. The global speech and voice recognition market was valued at USD 12.6 billion in 2022 and is expected to reach over USD 50 billion by 2030, driven largely by its adoption in mobile apps, virtual assistants, and educational platforms (Markets and Markets, 2023). In the realm of language learning, speech recognition technology is now integrated into more than 70% of the most-used language learning apps, offering learners real-time pronunciation feedback and speaking practice features (Holon IQ, 2023).

One significant application of machine learning is in speech or voice recognition, a technology that can process and interpret human speech into text or machine-readable commands. Advances in artificial neural networks (ANNs) have notably strengthened the performance of voice recognition systems, making them more accurate and responsive (Goel, Goel & Kumar, 2023). Today, many language learning apps incorporate speech recognition to help users practice speaking skills by identifying errors and offering feedback on pronunciation (Eshankulovna, 2021; Kholis, 2021).

Automatic pronunciation error detection systems have evolved from computer-assisted language learning (CALL) technologies. These systems aim to replicate the role of a language tutor by listening to a learner's spoken input, detecting errors, and providing immediate corrective feedback (Kalyanov, 2024). Research suggests that such systems can boost learners' motivation and improve their pronunciation, especially when paired with multimedia content and game-based learning experiences (Chen, 2023).

Popular platforms like Duolingo and Babbel have integrated speech recognition for language practice. However, these systems are typically optimized for widely spoken languages such as English and Spanish, and tend to perform less effectively when applied to languages

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH
(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *
https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

like Arabic. Arabic, with its rich phonetic structure, complex morphology, and wide range of dialects, presents particular challenges for speech recognition technology (Alsayadi, et. Al, 2022; Rahman et al., 2024). Developing robust, adaptable models for Arabic speech requires addressing these linguistic complexities, which many mainstream systems are not designed to handle.

In response to this need, there has been a growing push to create specialized Arabic pronunciation tools. Some researchers have attempted to build Arabic automatic speech recognition (ASR) models using deep learning and large-scale datasets (Alsayadi et al., 2022; Besdouri & Belguith, 2024). However, these models often require significant resources and are not easily adapted for real-time, interactive use in educational settings.

This study aims to contribute to this emerging area by developing a game-based learning tool that integrates machine learning for Arabic pronunciation practice. The system is designed to provide real-time feedback, encourage active participation through gamification, and offer a practical solution for learners and educators working with the challenges of Arabic speech acquisition.

2.0 Background of Study

In Brunei, Arabic language education holds a central place in the curriculum of Islamic religious schools, where it is primarily taught for religious understanding and academic advancement. Despite its importance, many students face significant challenges in mastering Arabic, particularly in speaking and pronunciation skills (Siti et al., 2024). Research has consistently shown that learners in Brunei struggle with Arabic phonology, fluency, and speaking confidence

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

(A.H.A et al., 2005; Yani et al., 2020; Siti et al., 2024). These difficulties are often attributed to limited exposure to native Arabic speakers, as well as a lack of opportunities for consistent oral practice outside the classroom environment.

While traditional classroom instruction effectively covers grammar, reading, and vocabulary acquisition, it often falls short in supporting the development of speaking proficiency. Pronunciation correction tends to be limited by factors such as large teacher-to-student ratios and tight classroom schedules, meaning that students rarely receive the individualized feedback necessary to improve their pronunciation accuracy (O'Brien, 2021; Gallegos, 2023). As a result, many learners progress through their studies with persistent speaking errors that are difficult to correct later on.

With the increasing availability of AI-driven educational tools, there is a promising opportunity to complement traditional teaching methods by introducing technology-based support systems (Annuš, 2024; Moemeke, 2024; Radif & Hameed, 2024). In particular, machine learning and speech recognition technologies can provide real-time pronunciation feedback, enabling students to practice independently and receive immediate responses tailored to their spoken input (Saadia, 2023; Sun, 2023). Furthermore, incorporating game-based learning elements into educational tools has been shown to boost learner engagement and motivation by creating interactive, goal-oriented learning experiences (Rye, Sousa & Sousa, 2025).

This study introduces the development of a game-based Arabic pronunciation learning system aimed at supporting students in Brunei. The system utilizes Google's Teachable Machine to train an audio classification model based on selected Arabic vocabulary items. It is then implemented within the Unity platform to create an engaging, interactive learning environment. By offering real-time feedback on learners' pronunciation performance through a

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

gamified experience, the system seeks to help students overcome common challenges in speaking Arabic more confidently and accurately.

This first part of the research paper focuses on detailing the system's development process, including the creation of the machine learning model and the technical integration with Unity. The second part will report on the usability testing results and analyze the speech recognition accuracy based on actual learner interactions with the system.

3.0 Research Objectives

The research objective of this study is to develop a machine learning-based voice recognition model for Arabic word pronunciation accuracy.

4.0 Literature Review

4.1 Machine Learning and Voice Recognition in Language Learning

Voice recognition systems powered by machine learning algorithms have become increasingly popular in language learning, particularly for providing automatic feedback on pronunciation. These systems typically employ supervised learning techniques, where models are trained on labeled audio samples to recognize both correct and incorrect pronunciation patterns (Bogach el al., 2021; Jing, 2025). Research has shown that integrating voice recognition into language learning platforms can significantly support oral skill development by offering learners immediate, actionable feedback and opportunities for self-correction (Shadiev and Feng, 2024; Zhang, 2024).

Modern voice recognition technologies often leverage deep learning approaches, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to extract audio features and perform classification tasks (Al-Fraihat et al., 2024). These methods have proven more effective than traditional approaches at handling the variability in speech patterns, including different accents and non-native pronunciations. However, the accuracy of these systems is highly dependent on the quality and diversity of the training data, underscoring the importance of carefully designed datasets (Lenson & Airlangga, 2023).

4.2 Accuracy of ML Models: Sample Size and Variation

The accuracy of machine learning models in audio classification is closely linked to both the quantity and variation of the training data. Studies suggest that increasing the number of samples per class improves model generalization and helps reduce overfitting, particularly in speech recognition tasks (Kheddar et al., 2023; Santos & Papa, 2022). For instance, Zhou et al. (2022) found that models trained with 40 to 60 samples per class achieved significantly better classification performance than those trained with fewer than 20 samples.

Equally important is the inclusion of diverse voice samples. Variability in gender, age, and speaking styles increases a model's ability to generalize across different user profiles (Zellou, Cohn & Ferenc, 2021). This is especially crucial in language learning, where learners' pronunciations often deviate considerably from native speaker norms. Researchers have emphasized the importance of balanced datasets to promote fairness and robustness in AI-driven speech recognition systems (Stecker & D'Onofrio, 2022).

4.3 System Design and Architecture in Language Learning Applications

The design of AI-powered language learning systems typically adopts a modular architecture that separates input capture, machine learning inference, and feedback output. Recent studies

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ *Volume 08 || Issue 07|| July, 2025 ||*

have recommended frameworks optimized for real-time audio processing using lightweight neural networks, particularly in mobile or browser-based platforms (Zhang, Kutscher & Cui, 2924). Key design priorities include minimizing latency and ensuring intuitive user interactions to maintain learner engagement.

Best practices in system design also highlight the importance of separating the machine learning model from the user interface, enabling easier updates and scalability (Huyen, 2022). Many educational platforms now rely on cloud-based inference or lightweight local deployments (e.g., TensorFlow.js, ONNX) to integrate speech models into interactive environments without the need for heavy back-end systems.

4.4 Feedback Mechanism and User Interface Design

An effective feedback mechanism is critical to the success of computer-assisted pronunciation training (CAPT) systems. Research consistently shows that timely, clear, and actionable feedback helps learners identify errors and improve their pronunciation more efficiently (Alam, 2025). Visual cues such as color-coded indicators, progress bars, or animated avatars are commonly used to enhance learners' understanding and maintain their motivation (Diller, Scheuermann & Wiebel, 2022).

Additionally, user interface (UI) design plays a significant role in the overall effectiveness of educational tools. Studies in educational game design emphasize the importance of intuitive controls, clear task instructions, and low cognitive load (Skulmowski & Xu, 2022). In pronunciation training, interfaces that gently guide learners through practice tasks while minimizing frustration are associated with better learning outcomes and higher retention rates (Taeza, 2025).

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696)
Achmad Yani5 *
https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

4.5 Challenges in Arabic Voice Recognition

Despite advancements in speech technology, Arabic voice recognition continues to present unique challenges. Arabic's linguistic complexity, such as its rich inventory of phonemes, including pharyngeal and emphatic sounds, poses difficulties for automated systems that often depend on textual representations for training (Rahman et al., 2024). The absence of short vowels (harakat) in most written Arabic further complicates pronunciation modeling, as it removes essential phonetic cues.

Another major obstacle is Arabic's diglossic nature. Modern Standard Arabic (MSA) differs significantly from the various regional dialects spoken across the Arab world (Alqadasi et al., 2023). These dialects vary in vocabulary, syntax, and phonology, making it difficult for machine learning models to generalize across Arabic-speaking populations. Most existing datasets focus on MSA, while regional dialects remain underrepresented, leading to decreased performance in dialectal speech recognition.

In addition, Arabic speech recognition research suffers from a lack of large, publicly available datasets. Unlike English or Mandarin, Arabic lacks extensive open-source corpora such as LibriSpeech or AISHELL (Zouhair, 2021). Consequently, many projects rely on small-scale or proprietary datasets, limiting their ability to develop robust, generalizable models.

Speaker variability, including differences in age, gender, and language proficiency, further impacts model performance (Jahan et al., 2025). Pronunciation training tools must account for these differences to effectively support non-native learners. Studies show that

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

models trained exclusively on adult native speakers often struggle with recognizing child or beginner pronunciations (Graham & Roll, 2024).

To address these challenges, recent efforts have focused on developing lightweight, domain-specific models tailored for educational and mobile environments. These models often prioritize targeted pronunciation feedback rather than full transcription and rely on smaller datasets augmented through techniques like data augmentation. Research shows that even with as few as 30-50 samples per class, usable accuracy can be achieved for specific tasks like emotion or keyword recognition (Sahu et al, 2021; Zhang et al., 2021). This supports the viability of using small, diverse datasets for lightweight systemssuch as those built with Teachable Machine and deployed within Unity-based learning platforms.

5.0 Methodology

5.1 Research Design

This study adopts a qualitative, design-based research approach focusing on the development and system integration of a machine learning-based Arabic pronunciation learning tool. The aim is to create a functional prototype that merges speech recognition capabilities with interactive game-based learning to support Arabic language learners.

The study emphasizes iterative development, prototyping, and system testing through observational and technical evaluation rather than hypothesis testing. The process involves designing the system architecture, collecting voice samples, training the machine learning model, and integrating the classifier into an educational game environment built with Unity.

5.2 Development Tools and Platforms

Teachable Machine by Google was used to develop the voice recognition model. This platform allows users to create audio classification models without deep programming knowledge, making it ideal for rapid prototyping.Unity Game Engine was used to design and implement the interactive learning environment. Unity's flexibility allows the integration of custom machine learning models and user interface elements to facilitate real-time feedback.

5.3 Data Collection for Model Training

To train the audio classification model, voice samples were collected from four individuals:

- 1 adult male
- 1 adult female
- 1 male child
- 1 female child

Each speaker contributed approximately 30 to 40 samples per class, resulting in a dataset of moderate size with a total of around 160 samples per class (across speakers). The dataset was designed to reflect common Arabic pronunciation tasks encountered by beginner learners.

5.4 Integration and Testing

Once the model was trained and exported, it was integrated into Unity. The testing phase focused on:

- Functionality: Ensuring smooth input-processing-feedback flow.
- Responsiveness: Confirming that feedback is provided in real-time.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

- User experience: Observing whether users could intuitively interact with the game-based interface.

6.0 System Architecture and Design

The system developed in this study consists of two primary components:

1. A machine learning-based audio classifier built using Google's Teachable Machine, and
2. An interactive game-based learning environment developed in Unity.

The integration of these two components results in a responsive and engaging learning experience, in which users are prompted to pronounce Arabic words and receive real-time visual feedback based on recognition accuracy.

6.1 Machine Learning Model Flowchart and Processing Pipeline

The machine learning component of the system is developed using Google's Teachable Machine, which simplifies the process of creating a custom audio classifier through a no-code, browser-based interface. The following flowchart (Figure 4.1) illustrates the core pipeline used to train and implement the model:

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

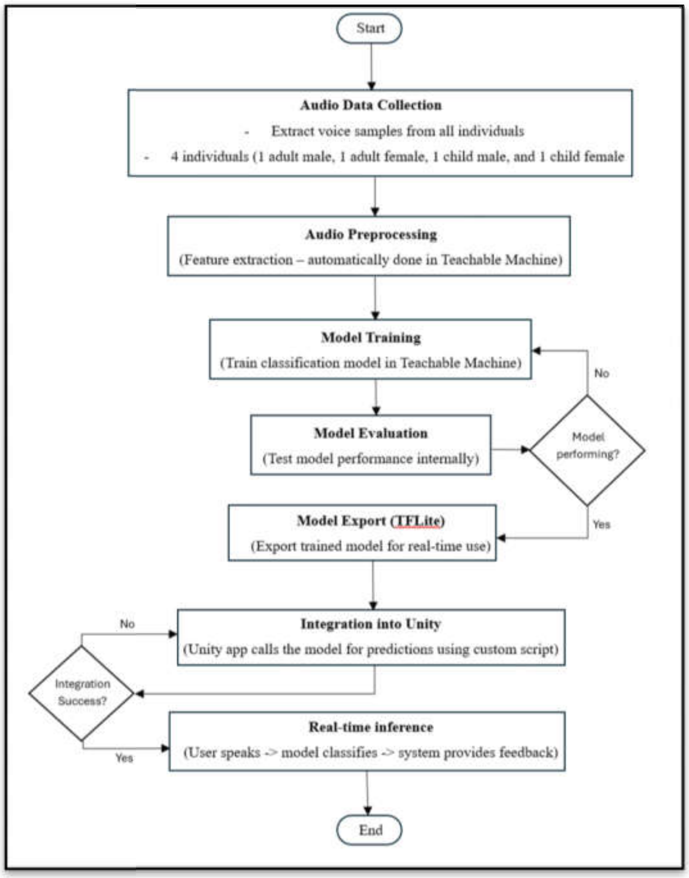

Figure 4.1: Machine Learning Model Flowchart and Processing Pipeline

Model Development Process:

1. Audio Data Collection

   Audio samples were recorded from four individuals, each person providing 5 to 10 voice samples per word. This produced a dataset with moderate class balance and variation to train the classification model.

2. Audio Preprocessing

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

The Teachable Machine platform automatically extracts features from audio clips using Mel-frequency cepstral coefficients (MFCCs) or similar spectral features to prepare data for model training.

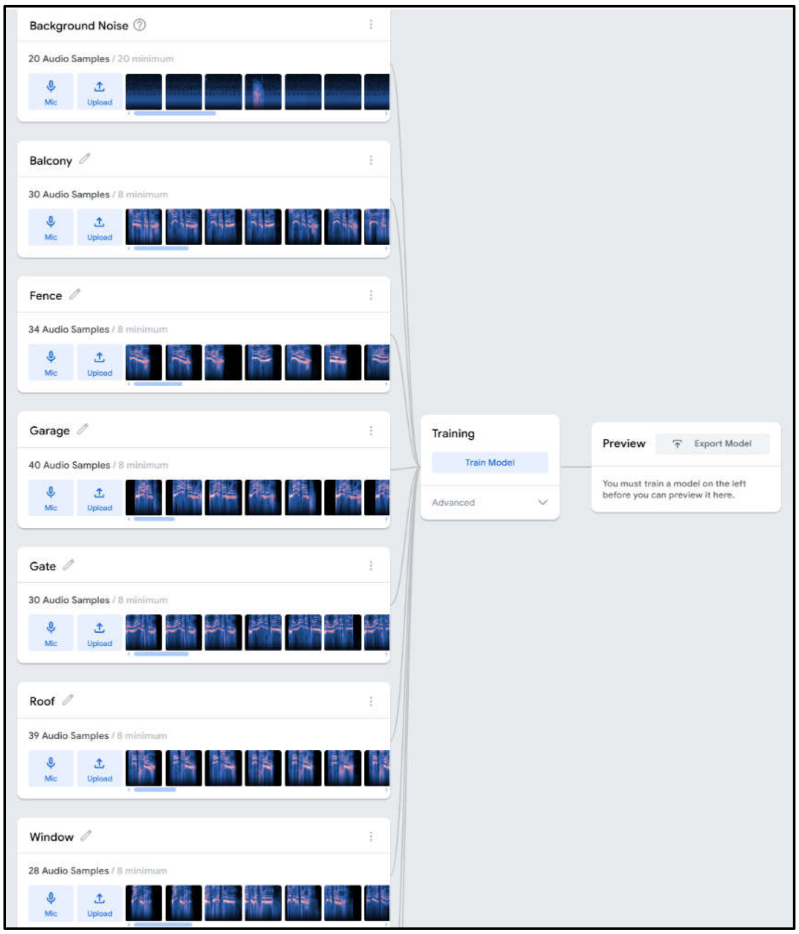

Figure 4.2: Audio Preprocessing and model training using Teachable Machine

3. Training the Model

The preprocessed audio is used to train a neural network classifier. The model learns to associate patterns in the audio data with specific class labels (Arabic words). Hyperparameters such as the number of epochs, learning rate, and training-validation

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH
(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *
https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

split are handled automatically by the platform, making it suitable for educational and rapid prototyping contexts.

4. Model Testing and Export

   Once trained, the model is tested internally on unseen samples and exported as a TensorFlow Lite (TFLite) model. This format ensures that the classifier can run efficiently in real-time within the Unity-based application.

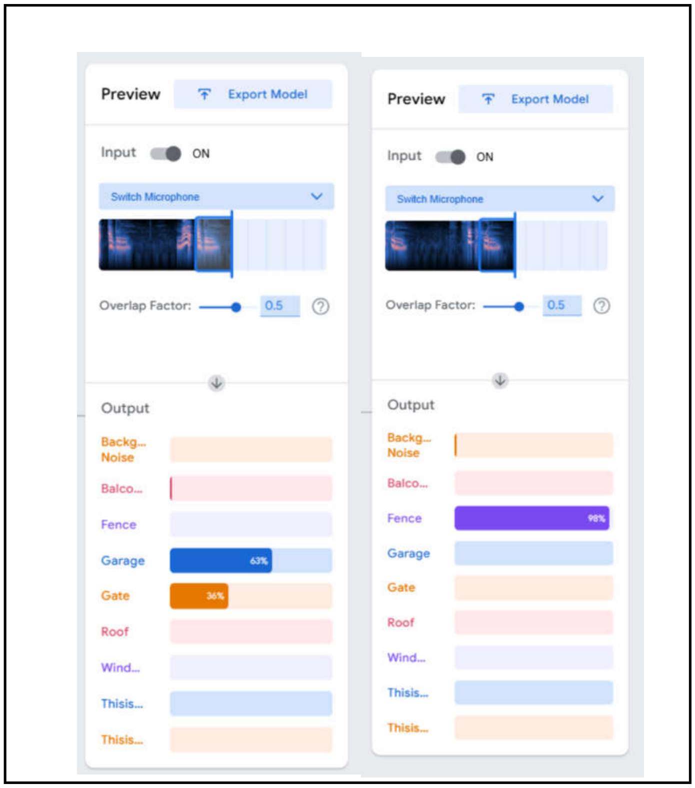

Figure 4.3: Model classification accuracy testing before exporting

5. Integration and Inference

   During gameplay, user speech input is passed through the model, which returns prediction probabilities for each class. If the highest confidence score exceeds a defined

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

threshold (above 60%), the system considers pronunciation correct and provides positive feedback to the learner.

6.2 System Workflow

The system architecture follows a straightforward input-processing-feedback loop, designed for usability and responsiveness:

1. User Input: The learner is presented with a target Arabic word on the screen and prompted to pronounce it.
2. Audio Capture: The spoken input is recorded using the device's built-in microphone.
3. Audio Classification (ML Model): The captured audio is processed by a pre-trained audio classification model created using Teachable Machine. The model analyzes the pronunciation and determines whether it closely matches the intended word.
4. Feedback Generation: Depending on the model's classification, the system provides visual feedbacksuch as a green indicator for correct pronunciation or a grey icon if the pronunciation is unrecognized or incorrect.
5. Gamification Layer: To encourage continued participation, learners interact with a game interface where successful pronunciation unlocks additional Arabic words, enhancing engagement and motivation.

This modular architecture is intentionally designed to be lightweight and easily deployable in educational settings with limited technological infrastructure. The separation of components also allows for future scalability, such as adding more Arabic vocabulary, adjusting feedback mechanisms, or enhancing the game environment in Unity.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH
(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *
https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

7.0 Dataset Limitations and Future Recommendations

One of the main limitations encountered during the development phase was the size and diversity of the dataset used to train the machine learning model. The audio samples were collected from only four speakers, which are two adults (male and female) and two children (male and female). Although each speaker contributed approximately 30 to 40 samples per class, resulting in a total of around 160 samples per word, the limited speaker variation restricts the generalizability of the model.

Previous research (Sahu et al, 2021; Zhang et al., 2021) has shown that models trained on small but diverse datasets can perform reasonably well in constrained or educational contexts. However, to improve robustness and reduce speaker bias, it is important to include samples from a larger pool of speakers, incorporating different accents, tones, speech speeds, and background conditions (Jahan et al., 2025).

Furthermore, the model was trained in a controlled environment, where audio quality was consistent and background noise minimal. In realworld classroom or home environments, such ideal conditions may not always be achievable. Therefore, the current model may require future data augmentation or retraining with noisier and more varied samples to better reflect authentic usage scenarios.

To address these limitations in future iterations, it is recommended to:

- Increase the number of speakers contributing to the dataset.
- Include dialectal and regional pronunciation variations of Arabic.
- Expand the vocabulary set to cover a broader range of common Arabic terms.
- Explore synthetic data generation or noise injection techniques to simulate real-world conditions.

Despite these constraints, the current dataset was sufficient to develop a functioning prototype and demonstrate the feasibility of integrating machine learning into a game-based Arabic pronunciation tool.

8.0 Conclusion

This study presented the design and development of an interactive Arabic pronunciation learning tool that integrates machine learning-based voice recognition with a game-based educational interface. By combining Google's Teachable Machine for speech classification with the Unity game engine, the system offers real-time feedback to learners, encouraging active participation and self-correction in a dynamic learning environment.

The system architecture follows a modular design that supports lightweight deployment and future scalability, making it suitable for implementation in resource-constrained educational settings, such as Arabic language schools in Brunei. Although the prototype was trained using a limited dataset, it successfully demonstrates the feasibility of using small datasets for targeted language learning applications. The recognition accuracy achieved during internal testing highlights the potential of using simple yet effective models in supporting pronunciation learning, particularly when traditional classroom methods fall short.However, limitations related to speaker diversity, environmental variability, and vocabulary scope must be addressed in future development stages. Expanding the dataset and refining the model through further testing will improve recognition performance and ensure greater inclusivity.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

References:

Alam, S. (2025). Impact of Mobile-Facilitated Peer Feedback Platform on Improving the Accuracy of Spoken English: An Experimental Study. International Journal of Information and Education Technology, 15(2).

Al-Fraihat, D., Sharrab, Y., Alzyoud, F., Qahmash, A., Tarawneh, M., & Maaita, A. (2024). Speech recognition utilizing deep learning: A systematic review of the latest developments. Human-centric computing and information sciences, 14.

Alqadasi, A. M. A., Abdulghafor, R., Sunar, M. S., & Salam, M. S. B. H. (2023). Modern standard Arabic speech corpora: A systematic review. Ieee Access, 11, 55771-55796.

Alsayadi, H. A., Abdelhamid, A. A., Hegazy, I., Alotaibi, B., & Fayed, Z. T. (2022). Deep investigation of the recent advances in dialectal arabic speech recognition. IEEE access, 10, 57063-57079.

Annuš, N. (2024). EDUCATIONAL SOFTWARE AND ARTIFICIAL INTELLIGENCE: STUDENTS'EXPERIENCES AND INNOVATIVE SOLUTIONS. Information Technologies and Learning Tools, 101(3), 200.

Besdouri, F. Z., Zribi, I., & Belguith, L. H. (2024). Challenges and progress in developing speech recognition systems for Dialectal Arabic. Speech Communication, 103110.

bin Haji Maila, A. H. A., & bin Ampuan, A. D. H. B. Arabic Schools in Negara Brunei Darussalam (1941-2005): Development and Challenges.

bin Haji Omar, A. H., bin Yaakob, M. B., Yani, A., bin Haji Sismat, M. A., & binti Salim, N. A. U. THE TECHNOLOGY GAP IN ARABIC LANGUAGE LEARNING ACROSS SCHOOLS IN BRUNEI DARUSSALAM.

Bogach, N., Boitsova, E., Chernonog, S., Lamtev, A., Lesnichaya, M., Lezhenin, I., ... & Blake, J. (2021). Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching. Electronics, 10(3), 235.

Chen, Y. (2023). The effect of using a game-based translation learning app on enhancing college EFL learners' motivation and learning experience. Education and Information Technologies, 28(1), 255-282.

Diller, F., Scheuermann, G., & Wiebel, A. (2022). Visual cue based corrective feedback for motor skill training in mixed reality: A survey. IEEE Transactions on Visualization and Computer Graphics.

Eshankulovna, R. A. (2021). Modern technologies and mobile apps in developing speaking skill. Linguistics and Culture Review, 1216-1225.

Gallegos, R. (2023). Training for Children's Online EFL Teachers: An Exploration of Strategies Developed by Those in the Field. University of Arizona Global Campus.

Goel, A., Goel, A. K., & Kumar, A. (2023). The role of artificial neural network and machine learning in utilizing spatial information. Spatial Information Research, 31(3), 275-285.

Graham, C., & Roll, N. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. JASA Express Letters, 4(2).

Huyen, C. (2022). Designing machine learning systems. " O'Reilly Media, Inc.".

HolonIQ. (2023). *Language Learning Apps and the Global EdTech Market*. Retrieved from https://www.holoniq.com

Jahan, M., Mazumdar, P., Thebaud, T., Hasegawa-Johnson, M., Villalba, J., Dehak, N., & Moro-Velazquez, L. (2025, April). Unveiling Performance Bias in ASR Systems: A Study on

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696)

Gender, Age, Accent, and More. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

Jing, Y. (2025). Error pattern recognition and correction methods in English oral learning process based on deep learning. Journal of Computational Methods in Sciences and Engineering, 14727978251318798.

Kalyanov, A. (2024). Software Developers' Experiences With CALL In the Context of the Four Language Competencies (Reading, Writing, Listening, and Speaking) and Teacher and Learner Fit: A Qualitative Descriptive Study (Doctoral dissertation, University of Massachusetts Global).

Kheddar, H., Himeur, Y., Al-Maadeed, S., Amira, A., & Bensaali, F. (2023). Deep transfer learning for automatic speech recognition: Towards better generalization. Knowledge-Based Systems, 277, 110851.

Kholis, A. (2021). Elsa speak app: automatic speech recognition (ASR) for supplementing English pronunciation skills. Pedagogy: Journal of English Language Teaching, 9(1), 01-14.

Lázaro, R. R. (2024). Exploring An 11th-grade Teacher's Understanding, Practices, and Perceptions of Corrective Feedback to Enhance Second Language Pronunciation: A Case Study (Doctoral dissertation, Universidad De Córdoba).

Lenson, A. K., & Airlangga, G. (2023). Comparative analysis of MLP, CNN, and RNN models in automatic speech recognition: dissecting performance metric. Buletin Ilmiah Sarjana Teknik Elektro, 5(4), 576-583.

MarketsandMarkets. (2023). *Speech and Voice Recognition Market - Global Forecast to 2030*. Retrieved from https://www.marketsandmarkets.com

Moemeke, C. D. (2024). Artificial Intelligence and Machine Learning in enhancing Science Learning Experiences: Exploring Possibilities and Concerns. NIU Journal of Educational Research, 10(2), 59-72.

O'Brien, M. G. (2021). Ease and difficulty in L2 pronunciation teaching: A mini-review. Frontiers in Communication, 5, 626985.

Radif, M., & Hameed, O. M. (2024). AI-driven innovations in e-learning: Transforming educational paradigms for enhanced learning outcomes. Arts Educa, 38.

Rahman, A., Kabir, M. M., Mridha, M. F., Alatiyyah, M., Alhasson, H. F., & Alharbi, S. S. (2024). Arabic speech recognition: Advancement and challenges. IEEE Access.

Rahman, A., Kabir, M. M., Mridha, M. F., Alatiyyah, M., Alhasson, H. F., & Alharbi, S. S. (2024). Arabic speech recognition: Advancement and challenges. IEEE Access.

Rye, S., Sousa, M., & Sousa, C. (2025). Introduction to Game-Based Learning. In Transformative Learning Through Play: Analogue Games as Vehicles for Educational Innovation (pp. 29-68). Cham: Springer Nature Switzerland.

Saadia, K. H. (2023). Assessing the Effectiveness of Text-to-Speech and Automatic Speech Recognition in Improving EFL Learner's Pronunciation of Regular Past-ed.

Sahu, B., Palo, H. K., & Mohanty, S. N. (2021, March). A performance evaluation of machine learning algorithms for emotion recognition through speech. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 13-17). IEEE.

Santos, C. F. G. D., & Papa, J. P. (2022). Avoiding overfitting: A survey on regularization methods for convolutional neural networks. ACM Computing Surveys (Csur), 54(10s), 1-25.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. Interactive learning environments, 32(6), 2538-2566.

Sholihah, A. F., & Kholis, A. (2025). Developing Electronic English Educational Cartoon (E-EduToon): A Culturally-Based Supplementary Material for Reading Instruction. Journal of General Education and Humanities, 4(2), 401-430.

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. Educational psychology review, 34(1), 171-196.

Stecker, A., & D'Onofrio, A. (2022). Variation in Evaluations of Gendered Voices: Individual Speakers Condition the Variant Frequency Effect. Journal of English Linguistics, 50(3), 281-314.

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. Frontiers in Psychology, 14, 1210187.

Taeza, J. (2025). The role of AI-powered chatbots in enhancing second language acquisition: An empirical investigation of conversational AI assistants. Edelweiss Applied Science and Technology, 9(3), 2616-2629.

Yani, A., Muritala, Y. T., Ismail, N. A. B. H., Abd, M. A. Z. B. H., & Bint Jawrami, N. A. (2020). The Content of the curriculum of Arabic language of the Arabic secondary schools of Brunei Darussalam: Class Eleven as a Case Study. European Journal of Education Studies.

Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age-and gender-related differences in speech alignment toward humans and voice-AI. Frontiers in Communication, 5, 600361.

IJO - INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH

(E.ISSN: 2805-413X) (P.ISSN: 2536-6696) Achmad Yani5 *

https://ijojournals.com/ Volume 08 || Issue 07|| July, 2025 ||

Zhang, M. (2024). Enhancing self-regulation and learner engagement in L2 speaking: exploring the potential of intelligent personal assistants within a learning-oriented feedback framework. BMC psychology, 12(1), 421.

Zhang, S., Liu, R., Tao, X., & Zhao, X. (2021). Deep cross-corpus speech emotion recognition: Recent advances and perspectives. Frontiers in neurorobotics, 15, 784514.

Zhang, Y., Kutscher, D., & Cui, Y. (2024). Networked metaverse systems: Foundations, gaps, research directions. IEEE Open Journal of the Communications Society.

Zhou, J., Huang, S., Wang, M., & Qiu, Y. (2022). Performance evaluation of hybrid GA-SVM and GWO-SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. Engineering with Computers, 1-19.

Zouhair, T. (2021). Automatic Speech Recognition for low-resource languages using Wav2Vec2: Modern Standard Arabic (MSA) as an example of a low-resource language.