International Journal of Social Science and Human Research

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 08 Issue 08 August 2025

DOI: 10.47191/ijsshr/v8-i8-04, Impact factor- 8.007

Page No: 5761-5768

Evaluating the Accuracy of a Developed Machine Learning-Based Arabic Pronunciation Model for Beginner Level

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 Simat[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: The use of artificial intelligence (AI) in language learning has expanded rapidly, offering new methods to support speaking skill development. In Brunei, Arabic is widely taught as a second language, but learners often struggle with accurate pronunciation due to limited exposure to native speech environments. Recent advances in speech recognition and machine learning (ML) provide opportunities to bridge this gap, particularly through the use of lightweight, game-integrated pronunciation models. This study evaluates the accuracy performance of a machine learning-based Arabic pronunciation model developed using Google's Teachable Machine and implemented within a Unity-based educational game. Drawing on prior work that emphasizes the importance of evaluating speech models with real-world learner input, this study focuses specifically on pronunciation accuracy among beginner-level Arabic learners. The system was tested with 35 non-native Arabic learners in Brunei, each of whom completed a 20-word pronunciation task. Accuracy score (%) and average attempts per word were calculated. Descriptive statistics revealed a mean accuracy of 34.86% (SD = 14.22), and a mean attempt count of 2.17 (SD = 0.57). A boxplot highlighted variation in learner performance, while Pearson correlation analysis showed a weak, non-significant negative relationship between accuracy and number of attempts (r = –0.215, p = 0.214). Findings suggest that while the model demonstrates functional classification capability, further refinement is needed to improve reliability. This research contributes to the growing field of AI-assisted language learning by offering a foundation for developing accurate, accessible pronunciation feedback tools for Arabic learners.

KEYWORDS: Artificial Intelligence Model, Speech Recognition, Usability Testing

1.0 INTRODUCTION

The integration of artificial intelligence (AI) into language learning continues to gain prominence, particularly for enhancing oral proficiency in languages that are less accessible in non-native environments (Chen et al., 2020; Li & Hegelheimer, 2013). In Brunei, Arabic is widely taught in religious and academic institutions; however, despite structured instruction in vocabulary, grammar, and reading comprehension, students often struggle with accurate pronunciation (Ahmad & Salleh, 2018). This persistent challenge is largely attributed to the absence of native Arabic-speaking environments and limited exposure to spoken Arabic in daily life (Al-Jarf, 2007; Al-Mahrooqi & Denman, 2015).

Recent advancements in educational technology, specifically game-based learning and AI-powered feedback systems, offer promising solutions to bridge this gap (Vesselinov & Grego, 2012; Li et al., 2021). These technologies have been explored across various language learning contexts and have shown potential in making pronunciation practice more engaging, interactive, and accessible for learners (Burston, 2014; Wang et al., 2020). Recognizing this potential, the present study introduces and evaluates a game-based learning system integrated with a machine learning (ML) model specifically designed to assess Arabic pronunciation performance in real time (Teachable Machine, 2023; Unity Technologies, 2023).

Developed using Google's Teachable Machine and implemented within the Unity game engine, the system enables learners to pronounce selected Arabic vocabulary items in an interactive environment, while receiving immediate visual feedback from the embedded AI model (Google, 2023; Unity Technologies, 2023). This design aims to simulate one-on-one pronunciation practice and enhance learner engagement through gamification (Hamari et al., 2016; Sykes et al., 2018).

This study targets non-native Arabic learners in Brunei, where Arabic is taught as a second language (Ministry of Education Brunei, 2022). By focusing on the accuracy of the system's pronunciation classification and analyzing how students interact with the tool, the research seeks to contribute to the growing field of AI-enhanced language education, particularly in contexts with limited access to native speakers and personalized oral feedback (Li & Hegelheimer, 2013; Zhang et al., 2021).

2.0 LITERATURE REVIEW

Advancements in artificial intelligence (AI) and machine learning (ML) have significantly influenced language learning, particularly in the area of automated speech recognition (ASR). In the context of Arabic pronunciation, evaluating how accurately ML models can recognize spoken input from non-native learners is critical to ensuring the effectiveness of AI-driven learning tools. This section reviews relevant literature on accuracy evaluation of speech-based ML models in educational contexts, with an emphasis on challenges associated with non-native Arabic pronunciation.

2.1 Accuracy Testing of Machine Learning Models in Educational Contexts

Evaluating the accuracy of ML models in educational settings requires more than standard computational metrics. In pronunciation learning, the primary concern is whether the model can reliably assess non-native speech and provide meaningful feedback that supports improvement. This is especially important for languages like Arabic, which feature complex phonetic structures and subtle sound distinctions that are often difficult for second-language learners to master (Sahu et al., 2021; Zhang et al., 2021).

Models trained on small or moderate-sized datasets, such as the classifier developed in this study, often encounter challenges related to accent variability, pronunciation errors, and inconsistent speech patterns. Despite these challenges, several studies have shown that carefully constructed small datasets particularly those incorporating diverse speakers can still yield acceptable levels of classification accuracy in controlled applications (Zhou et al., 2022).

One of the key considerations in pronunciation evaluation is the balance between model sensitivity and tolerance. Overly strict models may misclassify acceptable pronunciation variations as errors, while overly lenient models may fail to detect incorrect articulation (Rahman et al., 2024). As such, pronunciation accuracy should not only be evaluated in terms of correct-versus-incorrect classifications but also in how well the model reflects real learner performance.

Common practices in accuracy testing include calculating overall recognition rates, analyzing per-word accuracy, and identifying patterns of misclassification. These approaches help determine whether the model generalizes effectively across learners and word types. For Arabic in particular, accuracy analysis must take into account the presence of emphatic consonants, short vowels (harakat), and pharyngeal sounds, all of which contribute to the linguistic complexity of the language.

3.0 RESEARCH OBJECTIVES

The objective of this study is to evaluate the accuracy level of a developed machine learning model for Arabic speaking skill practice within a game-based learning application developed for beginner-level learners.

4.0 METHODOLOGY

The following sections explain the research design, participant selection, data collection instruments, procedures, analysis methods, and ethical considerations guiding this study.

4.1 Research Design

This study used a quantitative correlational research design to evaluate the usability of a speech recognition AI model within a game-based learning application. The study aimed to identify relationships between students' pronunciation accuracy and the number of attempts required for correct recognition. Group comparisons were also conducted to examine differences in performance based on microphone types.

4.2 Research Sample

A convenience sampling method was used to recruit 35 beginner-level Arabic learners from a single institution in Brunei.

4.3 Research Instrument

A score sheet was used to record and organize data for each participant. The score sheet included the following fields:

- Student ID
- Number of words tested (fixed at 20)
- Number of words correctly recognized by the AI
- Total number of attempts

Calculated fields:

- Accuracy Score (%)
- Average Attempts per Correct Word

The AI system itself served as a built-in instrument to log the number of attempts and correctness.

4.4 Data Collection Procedure

Experimental participants were guided to use the developed game-based learning application in a semi-controlled classroom environment. Each student was tasked with pronouncing 20 Arabic vocabulary words, during which the AI model provided real-time feedback after every attempt. To assess the impact of hardware variability, students were assigned one of three different microphone types. All interaction data was digitally recorded using a structured score sheet. Performance was evaluated using Accuracy Score and Average Attempts Per Word, calculated using standard formulas commonly applied in machine learning and usability testing (Zhang et al., 2021; Nielsen, 1993):

$$\textbf{Accuracy Score (\%)} = \left(\frac{\text{Number of Correctly Classified Words}}{\text{Total Words Tested}}\right) \times 100$$

Figure 4.1: Accuracy Score Percentage Equation (Zhang, Y., Chen, D., & Zhao, L., 2021)

$$\textbf{Avg Attempts Per Word} = \frac{\text{Total Number of Attempts}}{\text{Total Words Tested}}$$

Figure 4.2: Average Attempts Per Word Equation (Nielsen, J., 1993)

4.5 Data Analysis

The data collected were analyzed using IBM SPSS Statistics to evaluate the performance of the developed machine learning-based pronunciation model. Descriptive statistics, including the mean, standard deviation, minimum, and maximum values, were used to summarize the accuracy scores and the average number of attempts per correct word. Pearson correlation analysis was conducted to examine the relationship between accuracy and average attempts, aiming to determine whether a higher number of attempts was associated with lower accuracy performance. In addition, a boxplot was generated to visualize the distribution and spread of accuracy scores across participants. These statistical procedures were selected to identify usability trends and evaluate the consistency of the model's performance across all participants.

4.6 Ethical Consideration

All participants were briefed about the purpose of the study and gave informed consent prior to participation. No personal identifiers were collected to ensure anonymity and confidentiality. The testing environment was kept non-invasive, and students had the right to withdraw at any time without penalty. The study was conducted in accordance with the ethical standards of educational research and was approved by the institutional review board of the hosting institution.

5.0 RESEARCH FINDINGS

This section presents the findings from the posttest evaluation of the game-based Arabic pronunciation system, focusing on the students' pronunciation accuracy, interaction behavior, and usability outcomes.

5.1 Descriptive Statistics

Descriptive statistics were conducted to evaluate the overall accuracy performance of the developed machine learning-based Arabic pronunciation model. As shown in Table 5.1, the mean accuracy score achieved by the 35 participants was 34.86% (SD = 14.22), with scores ranging from 15% to 75%. This indicates a moderate level of recognition accuracy, with considerable variability across users.

The average number of attempts per word was 2.17 (SD = 0.57), suggesting that most learners required more than one attempt to receive a correct classification. This finding reflects the expected challenge of pronunciation tasks at the beginner level, as well as the limitations of the model in recognising less clear inputs.

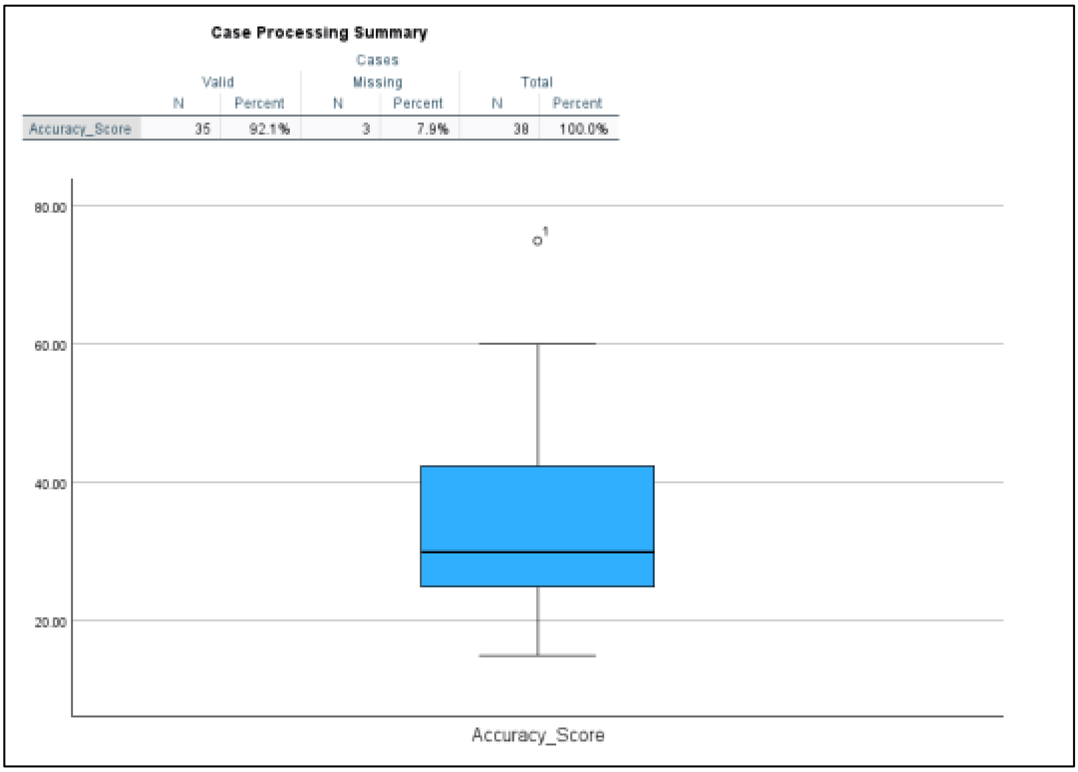
Table 5.1: Descriptive statistics for Accuracy Score and Average Attempts per Word

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Accuracy_Score	35	15.00	75.00	34.8571	14.21917
Average_Attempts_per_Word	35	1.00	3.00	2.1714	.56806
Valid N (listwise)	35				

5.2 Boxplot of Accuracy Score

A boxplot of the accuracy scores (Figure 5.1) illustrates the distribution and spread of performance. The median score falls just below the mean, and one outlier was identified above the 60% mark, representing a high-performing user. The interquartile range reflects that most scores clustered between 25% and 42.5%, while the lower whisker extends toward the minimum score of 15%, indicating a few learners with lower performance.



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Accuracy_Score	35	92.1%	3	7.9%	38	100.0%

Figure 5.1: Boxplot showing distribution of Accuracy Scores for all participants

These results suggest that while the model demonstrates functional classification ability, improvements in accuracy and user-specific adaptability may be required for broader educational application.

5.3 Pearson Correlation Result

A Pearson correlation analysis was conducted to examine the relationship between learners' pronunciation accuracy scores and the average number of attempts required per word. The analysis revealed a weak negative correlation, r = –0.215, p = 0.214, indicating that as the number of attempts per word increased, the accuracy score tended to decrease. However, this relationship was not statistically significant at the 0.05 level.

This suggests that while there is a slight trend indicating that students who needed more attempts generally achieved lower accuracy, the association was not strong enough to rule out random variation. These findings imply that other factors such as individual pronunciation ability, microphone quality, or model sensitivity may have played a role in influencing accuracy scores than the number of attempts alone.

Correlations

		Accuracy_Score	Average_Attempts_per_Word
Accuracy_Score	Pearson Correlation	1	-.215
	Sig. (2-tailed)		.214
	N	35	35
Average_Attempts_per_Word	Pearson Correlation	-.215	1
	Sig. (2-tailed)	.214	
	N	35	35

Figure 5.2 Pearson Correlation Analysis showing the relationship between Accuracy Scores and Average Attempts per Word

5.4 Discussion of Result

The evaluation of the developed machine learning-based Arabic pronunciation model for beginner learners was based on three core analyses: descriptive statistics, boxplot visualization, and Pearson correlation.

The descriptive analysis indicated that the average accuracy score among the 35 participants was 34.86% (SD = 14.22), with scores ranging from 15% to 75%. This suggests a moderate recognition capability by the model, although performance varied considerably across learners. Additionally, the average number of attempts per word was 2.17 (SD = 0.57), reflecting that most users required more than one attempt to receive a correct classification. This aligns with expectations for a beginner-level cohort, where pronunciation clarity and confidence are still developing.

The boxplot visualization of accuracy scores further highlighted the distributional characteristics of model performance. The majority of learners achieved accuracy within the interquartile range of 25% to 42.5%, with a median slightly below the mean, indicating a slight skew towards lower performance. A single outlier was identified above the 60% mark, suggesting that at least one participant achieved notably higher results, potentially due to clearer articulation, better recording conditions, or closer pronunciation match with the training data.

Finally, a Pearson correlation analysis was conducted to examine the relationship between accuracy score and average number of attempts per word. The result revealed a weak negative correlation ($r = -0.215$, $p = 0.214$), suggesting that as the number of attempts increased, the accuracy score tended to decrease. However, this relationship was not statistically significant, indicating that the number of attempts alone may not reliably predict performance. The absence of a strong correlation implies that other factors such as speaker variability, recording environment, or word-specific phonetic difficulty may have influenced the accuracy of the model's predictions.

6.0 CONCLUSION AND FUTURE RECOMMENDATIONS

This study examined the accuracy performance of a machine learning-based Arabic pronunciation model developed using Google's Teachable Machine and implemented within a game-based learning environment. The primary aim was to assess how effectively the model could classify spoken Arabic words by beginner-level, non-native learners in Brunei.

6.1 Conclusion

Descriptive analysis revealed that the model achieved a moderate average accuracy score of approximately 34.86%, with considerable variation across learners. The average number of attempts per word was 2.17, indicating that multiple tries were often required before a correct classification was achieved. A boxplot visualization further supported these findings, showing that the majority of accuracy scores clustered between 25% and 42.5%, with a single high-performing outlier. Pearson correlation analysis showed a weak negative relationship between the number of attempts and accuracy score ($r = -0.215$), though this relationship was not statistically significant ($p = 0.214$). These results suggest that while the model demonstrates functional classification ability, it does not yet exhibit strong predictive reliability in its current form.

The findings highlight both the potential and the limitations of using lightweight, low-resource machine learning models for pronunciation evaluation. Although the model was able to provide real-time feedback and recognize patterns in learner speech, the variability in accuracy suggests the need for further refinement particularly in terms of training data diversity and robustness to different speech profiles.

6.2 Future Recommendations

To enhance the model's effectiveness and scalability in real-world educational settings, the following recommendations are proposed:

1. Expand the Training Dataset
   Incorporate a larger and more diverse dataset including speakers of different ages, genders, and accents to improve generalization and reduce bias.
2. Include More Phonetically Complex Arabic Words
   Add vocabulary items that reflect a broader range of Arabic phonetic structures, including short vowels (harakat) and pharyngeal sounds.
3. Introduce Noise-Handling and Environment Adaptation
   Train the model with audio samples recorded in various conditions to increase its robustness in classroom or home environments.
4. Conduct Longitudinal Studies
   Evaluate the impact of the model on learner improvement over time to determine its long-term educational value.
5. Integrate a Confidence Scoring System
   Implement a scoring feature that provides learners with probabilistic feedback (e.g., high, medium, low confidence) to support self-awareness and targeted correction.

In conclusion, while the current model shows promise as a prototype for AI-supported pronunciation learning, further development and testing are necessary to optimize its reliability and educational impact. The insights from this study contribute to the growing field of AI-enhanced language learning and offer a foundation for future innovations in Arabic speech recognition for non-native learners.

REFERENCES

1) Ahmad, N., & Salleh, M. (2018). Challenges in Arabic pronunciation among Bruneian learners. *Journal of Language Teaching*, 12(2), 45-60.
2) 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).
3) 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.
4) Al-Jarf, R. (2007). The role of input in improving Arabic pronunciation. *Language Learning Journal*, 35(1), 23-32.
5) Al-Mahrooqi, R., & Denman, C. J. (2015). Arabic language learning in non-native contexts: Challenges and strategies. *International Journal of Language Studies*, 9(3), 55-70.
6) 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.
7) 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.
8) Annuš, N. (2024). EDUCATIONAL SOFTWARE AND ARTIFICIAL INTELLIGENCE: STUDENTS'EXPERIENCES AND INNOVATIVE SOLUTIONS. Information Technologies and Learning Tools, 101(3), 200.
9) Besdouri, F. Z., Zribi, I., & Belguith, L. H. (2024). Challenges and progress in developing speech recognition systems for Dialectal Arabic. Speech Communication, 103110.
10) bin Haji Maila, A. H. A., & bin Ampuan, A. D. H. B. Arabic Schools in Negara Brunei Darussalam (1941–2005): Development and Challenges.
11) 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.
12) 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.
13) Burston, J. (2014). Mobile-assisted language learning: A selected annotated bibliography of implementation studies 1994–2012. *Language Learning & Technology*, 18(3), 1-23.
14) Chen, C.-M., Li, Y.-L., & Chen, M.-C. (2020). Personalized pronunciation training system using AI. *Computers & Education*, 150, 103837.
15) 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.
16) 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.
17) Eshankulovna, R. A. (2021). Modern technologies and mobile apps in developing speaking skill. Linguistics and Culture Review, 1216-1225.
18) 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.
19) 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.
20) Google. (2023). Teachable Machine. Retrieved from https://teachablemachine.withgoogle.com/
21) Graham, C., & Roll, N. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. JASA Express Letters, 4(2).
22) Hamari, J., Koivisto, J., & Sarsa, H. (2016). Does gamification work? – A literature review of empirical studies on gamification. *Proceedings of the 47th Hawaii International Conference on System Sciences*.
23) HolonIQ. (2023). *Language Learning Apps and the Global EdTech Market*. Retrieved from https://www.holoniq.com
24) Huyen, C. (2022). Designing machine learning systems. " O'Reilly Media, Inc.".
25) 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 Gender, Age, Accent, and More. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

26) 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.
27) 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).
28) 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.
29) 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.
30) 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).
31) 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.
32) Li, X., Wang, Y., & Zhang, L. (2021). AI-powered feedback for language learning: A review. *Educational Technology Research and Development*, 69, 123-145.
33) Li, Z., & Hegelheimer, V. (2013). Mobile-assisted pronunciation training: Effects on ESL learners' pronunciation. *Language Learning & Technology*, 17(3), 117-137.
34) MarketsandMarkets. (2023). *Speech and Voice Recognition Market – Global Forecast to 2030*. Retrieved from https://www.marketsandmarkets.com
35) Ministry of Education Brunei. (2022). Arabic language curriculum overview. Bandar Seri Begawan: MOE Publications.
36) 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.
37) Nielsen, J. (1993). *Usability Engineering*. Academic Press.
38) O'Brien, M. G. (2021). Ease and difficulty in L2 pronunciation teaching: A mini-review. Frontiers in Communication, 5, 626985.
39) Radif, M., & Hameed, O. M. (2024). AI-driven innovations in e-learning: Transforming educational paradigms for enhanced learning outcomes. Arts Educa, 38.
40) 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.
41) 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.
42) 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.
43) 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.
44) 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.
45) 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.
46) Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. Interactive learning environments, 32(6), 2538-2566.
47) 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.
48) 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.
49) 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.
50) 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.
51) Sykes, J. M., Reinhardt, J., & Thorne, S. L. (2018). Gamification in language learning: A systematic review. *Language Learning & Technology*, 22(3), 1-14.
52) 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.

53) 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.
54) 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.
55) 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.
56) Zhang, S., Liu, R., Tao, X., & Zhao, X. (2021). Deep cross-corpus speech emotion recognition: Recent advances and perspectives. Frontiers in neurorobotics, 15, 784514.
57) Zhang, Y., Chen, D., & Zhao, L. (2021). *Evaluation metrics in speech recognition systems: A comparative study*. Journal of Intelligent Systems, 30(2), 142–156.
58) Zhang, Y., Kutscher, D., & Cui, Y. (2024). Networked metaverse systems: Foundations, gaps, research directions. IEEE Open Journal of the Communications Society.
59) 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.
60) Zouhair, T. (2021). Automatic Speech Recognition for low-resource languages using Wav2Vec2: Modern Standard Arabic (MSA) as an example of a low-resource language.