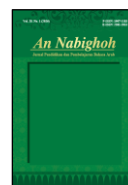




AN NABIGHOH

Journal homepage: <http://e-journal.metrouniv.ac.id/an-nabighoh>

ENHANCING ARABIC PHONETIC LEARNING THROUGH YOLO-BASED GRAPHEME RECOGNITION AND AUDIO FEEDBACK

Anisa Dwi Nurchayati ^a, Eric Kunto Aribowo ^{b,*}, Akhmad Fadjeri ^c

^{a, b} Department of Master of Arabic Language and Literature, Faculty of Adab and Cultural Sciences, Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia

^c Department of Informatics Engineering, Faculty of Engineering, Universitas Ma'arif Nahdlatul Ulama Kebumen, Kebumen, Indonesia

Article Info

*Corresponding Author:

Name:

Eric Kunto Aribowo

Email:

eric.aribowo@uin-suka.ac.id

Article History:

Received: April 12, 2026

Revised: May 19, 2026

Accepted: June 27, 2026

Published: June 30, 2026

Abstract

Background: Islamic religious education in Indonesia introduces Arabic script (*hijaiyah*) as the foundational step of Quranic literacy, primarily within non-formal settings such as *Madrasah Diniyah*. Students at *Madrasah Diniyah* in Kebumen face persistent difficulties distinguishing and pronouncing similar Arabic letters, such as ع (*ain*) and غ (*ghain*), due to phonological interference from their native Javanese and Indonesian languages, which lack equivalent guttural phonemes. Traditional methods, such as *Iqra'* books and letter cards, fail to provide real-time phonetic feedback, limiting the accuracy of pronunciation, which is essential for Quranic recitation. **Research Objectives:** In this study, a computer vision application was developed that detects Arabic letters from uploaded images, displays Indonesian phonetic transliteration, and generates automatic audio pronunciation (image-to-text-to-voice) using the You Only Look Once Version 11 (YOLOv11) algorithm within a Waterfall development model. **Methodology:** Employing a research and development (R&D) approach, the study followed five waterfall stages: requirements analysis, system design, implementation, verification, and maintenance. Verification involved expert judgment, end-user evaluation, and confusion matrix-based algorithm assessment. A dataset of 360 annotated Arabic letter images was used for training in Google Colab. **Results:** The resulting application, Arabic as a Foreign Language (AFL), detects Arabic letters from images, displays phonetic transliteration, and generates automatic audio pronunciation. The YOLOv11 model achieved 99.5% accuracy (mAP50) and 91.12% accuracy (mAP50-95). Expert Black-Box testing yielded an average Likert score above 4.0 across all functional criteria, with the audio feature rated highest (4.6). **Unique Contribution:** This study provides empirical evidence that computer vision-based pronunciation support effectively bridges grapheme recognition and accurate phoneme production, addressing a critical gap in AFL pedagogy within Indonesian Islamic education. **Conclusion:** The study successfully developed a YOLOv11-based application (AFL) with high accuracy and positive expert evaluations, effectively addressing pronunciation difficulties among *Madrasah Diniyah* students in Kebumen and demonstrating the potential of computer vision to bridge grapheme recognition and phoneme production in Arabic learning. **Recommendations:** Future work should improve accuracy for letter pairs *Kaf-Tsa* (ك-ث), *Nun-Tsa/Nun-Ta* (ن-ث, ن-ت), and *Alif-Ghayn/Za-Ghayn* (أ-ز, أ-غ), as they possess fundamentally distinct Unicode encodings and vector embeddings, not merely diacritical dot variations.

Copyright © 2026, Anisa Dwi Nurchayati et al.
This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Keywords:

Arabic computer vision; YOLO algorithm; Arabic phonetics; Grapheme detection; Arabic language learning; Computer-Assisted Language Learning (CALL).

Introduction

Islamic religious education in Indonesia introduces the Arabic script known as *huruf hijaiyah* as the foundational step of Quranic literacy, most commonly within non-formal settings such as *Madrasah Diniah*. In Indonesia, *Madrasah Diniah* is implemented from elementary school to university.¹ Mastery of the *hijaiyah* letters, encompassing both written recognition and correct pronunciation, is considered essential from early childhood to ensure accurate Quranic recitation.² The transfer of sound patterns from the learner's first language into the target language is consistently identified as the primary obstacle for Indonesian Arabic learners, whose native phonological inventory lacks equivalents for several Arabic phonemes.³ This obstacle often becomes a major problem because it replaces familiar Indonesian or local sounds with Arabic sounds, thus disrupting the accuracy of the learner's pronunciation and the effectiveness of their Arabic.⁴ In Javanese, phonological interference was found in the pronunciation of the letter ث (*tsa*) at the end of the word الترات "al turaats" at the end of the word *mengumah*, which becomes the sound "s".⁵ There is interference in the Sundanese dialect. The mother tongue tends to replace the phoneme ف (*fa*) with "pa," and in the Banyumas dialect, speakers often change the vowel sound "o" to "a" and the consonants ب (*ba*), د (*dal*), pronounced firmly, firmly, and straightforwardly.⁶ It is also found in the Gowa and Makassar dialects that there is a unique phonological interference in the form of additional phonemes, especially in the consonant "n" which becomes the addition of "ng" in the word الإنسان "al insaanu" as well as the addition of other consonant phonemes such as the word نكاح (*nikaahun*) to the sound نكاح "nikkahun" the addition of the consonant "k".⁷

¹ Ngainun Naim et al., "Integration of Madrasah Diniyah Learning Systems for Strengthening Religious Moderation in Indonesian Universities," *International Journal of Evaluation and Research in Education (IJERE)* 11, no. 1 (2022): 108, <https://doi.org/10.11591/ijere.v11i1.22210>.

² Damanhuri Daman and Jamiluddin Yacub, "Mengenal Huruf Hijaiyah Pada Anak Usia Dini," *Azzahra: Jurnal Pendidikan Anak Usia Dini* 3, no. 2 (2022), <https://ejournal.staidarussalamlampung.ac.id/index.php/azzahra/article/view/374>.

³ Torkis Lubis et al., "Enhancing Arabic Pronunciation Learning through Contrastive Analysis: An Empirical Study with Indonesian University Students," *International Journal of Special Education* 41, no. 2 (2026), <https://www.internationalsped.com/index.php/ijse/article/view/3331>.

⁴ Fitma Nailurrahmi and Lina Marlina, "Analisis Interferensi Fonetik Bahasa Ibu Terhadap Pelafalan Fonem Bahasa Arab Dalam Pembelajaran Kosakata," *Moderasi : Journal of Islamic Studies* 5, no. 1 (2025): 279–90, <https://doi.org/10.54471/moderasi.v5i1.113>.

⁵ Mohamad Zaka Al Farisi et al., "Investigation of the Pronunciation of the Voiceless Fricative Non-Sibilant Phoneme /θ/ in Arabic: An Acoustic Phonetics Comparative Analysis," *Dirasat: Human and Social Sciences* 52, no. 6 (2025): 5447, <https://doi.org/10.35516/Hum.2025.5447>.

⁶ Hilyah Ahya Muthmainnah and Lina Marlina, "Kesalahan Fonologi Dalam Membaca Al-Qur'an Pada Siswi Penutur Bahasa Sunda Di Tahfiz Qur'an Al-Marjan," *Al-Lahjah : Jurnal Pendidikan, Bahasa Arab, Dan Kajian Linguistik Arab* 8, no. 1 (2025), <https://ejournal.unwaha.ac.id/lahjah/article/view/5936>; Ratih Rahayu and Sri Munawarah, "Investigation of Yogyakarta Dialect's Vocabulary in Areas of Banyumas' Ngapak Dialect," in *Sociolinguistics and Dialectological Studies in Indonesia* (Nova Science Publishers, Inc., 2021), <https://www.scopus.com/pages/publications/85116743336>.

⁷ Anugrah Anugrah et al., "Analysis of Regional Language Phonological Interference on the Pronunciation of Hijaiyah Letters of Madrasah Tsanawiyah Students in Gowa Regency," *Pinisi Journal of*

Field observations at two *Madrasah Diniyah* in Kebumen, *Madrasah Diniyah* Al-Barokah Kejawang and *Madrasah Diniyah* TPQ Masjid Baitul Muttaqin, confirmed persistent student difficulties in distinguishing and pronouncing similar Arabic letter pairs, including ع (*ain*) and غ (*ghain*), ج (*jim*) and خ (*kha*), and س (*sin*) and ش (*shin*).⁸ This challenge is compounded by the inherent complexity of Arabic graphemes: depending on its location within a word, each of the 28 Arabic letters may take up to four distinct positional forms, and multiple letters share identical base shapes differentiated solely by the placement of diacritical dots.⁹ A systematic literature review further confirmed that Arabic consonants can be categorized into three articulatory difficulty tiers for new learners, with phonemes absent from Indonesian such as ض, ظ, غ, ث, and ط consistently posing the greatest challenge.¹⁰

These linguistic challenges are compounded by pedagogical constraints. Traditional tools, such as the *Iqra'* book and handwritten letter cards, remain predominant in *Madrasah Diniyah* settings but cannot provide real-time corrective feedback on pronunciation or grapheme recognition. The current use of applications in instructional pedagogical guidance is still lacking. For example, the Tarteel application can detect errors but does not provide detailed rule-based explanations.¹¹ The heterogeneity of student cohorts ranging from children to adults within a single session further limits the effectiveness of uniform instructional approaches. Recent advances in deep learning have opened up new possibilities for automated Arabic text recognition. The You Only Look Once (YOLO) algorithm, originally introduced by Redmon et al., reframed object detection as a unified single-pass regression problem that simultaneously predicts bounding boxes and class probabilities from a complete image.¹² Subsequent iterations, most notably YOLOv4 by Bochkovskiy et al., introduced systematic improvements through bag-of-freebies techniques (Mosaic data augmentation, Ciou loss, DropBlock regularization) and bag-of-specials modules (Mish activation, cross-stage partial connections, and path aggregation network), achieving a state-of-the-

Art, Humanity, and Social Studies 4, no. 4 (2024), <https://journal.unm.ac.id/index.php/PJAHSS/article/view/2882>.

⁸ Zahra Nida Al Medina et al., "The Arabic Adventure; Media Board Game Untuk Keterampilan Berbicara Bahasa Arab Siswa Kelas VII MTs Di Purbalingga," *Lisanul Arab: Journal of Arabic Learning and Teaching* 8, no. 1 (2019): 1, <https://journal.unnes.ac.id/sju/laa/article/view/32541>.

⁹ Salah Alghyaline, "A Printed Arabic Optical Character Recognition System Using Deep Learning," *Journal of Computer Science* 18, no. 11 (2022): 1038–50, <https://doi.org/10.3844/jcssp.2022.1038.1050>.

¹⁰ Yuslina Mohamed et al., "A Systematic Review Of Arabic Phonetic: Hijaiyyah's Pronunciation Among New Learners," *Ijaz Arabi Journal of Arabic Learning* 7, no. 2 (2024), <https://doi.org/10.18860/ijazarabi.v7i2.24173>.

¹¹ Mehmet Birgün, "Integrating AI into Qur'an Learning: Technical Advances and Pedagogical Gaps," *Social Sciences & Humanities Open* 13 (June 2026): 102499, <https://doi.org/10.1016/j.ssaho.2026.102499>.

¹² Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, 779–88, <https://doi.org/10.1109/CVPR.2016.91>.

art accuracy of 43.5% AP on MS COCO while remaining trainable on a single conventional GPU.¹³

The evolution of YOLO from its initial version to its latest iteration demonstrates continuous improvements in speed, accuracy, and adaptability to diverse datasets. This research focuses on a comparative analysis of YOLOv5, YOLOv8, and YOLOv11 with a specific application to image recognition.¹⁴ The issue of image recognition of handwritten Arabic text using deep learning approaches is increasingly being discussed. A review specifically highlights CNN, RNN, and transformer-based methods for recognizing Arabic text in historical manuscripts. Their study emphasizes that the variability of calligraphic styles, the complexity of script structures, and the physical deterioration of documents are significant obstacles still faced by Arabic literary text recognition systems.¹⁵

Traditional optical character recognition (OCR) systems often fail to address these complexities, especially when applied to historical manuscripts or handwritten text.¹⁶ Therefore, modern object detection frameworks, such as YOLO, offer a promising alternative, leveraging convolutional neural networks and more sophisticated feature extraction mechanisms. YOLOv5, which is widely used in both research and industry, represents a mature stage in the YOLO family. This model is known for its ability to balance computational efficiency and detection accuracy. YOLOv5 introduces modular improvements to its backbone architecture and training strategy, making it suitable for lightweight applications and real-time detection. However, its performance on highly complex datasets, such as Arabic manuscripts with high levels of noise, remains limited compared with that of later versions.¹⁷

YOLOv8 presents a major improvement by adopting an anchor-free detection mechanism, a more robust backbone architecture, and a more optimized training strategy. These innovations enable YOLOv8 to achieve higher accuracy while maintaining reasonable inference speed. In the context of Arabic literature, YOLOv8 demonstrates better generalization capabilities across handwritten and printed text styles, both handwritten and printed. Its robustness to variations in letter size, text orientation, and background noise makes it a strong candidate for large-scale

¹³ Alexey Bochkovskiy et al., "YOLOv4: Optimal Speed and Accuracy of Object Detection," version 1, preprint, arXiv, 2020, <https://doi.org/10.48550/ARXIV.2004.10934>.

¹⁴ Ayşe Aybilge Murat and Mustafa Servet Kiran, "A Comprehensive Review on YOLO Versions for Object Detection," *Engineering Science and Technology, an International Journal* 70 (October 2025): 102161, <https://doi.org/10.1016/j.jestch.2025.102161>.

¹⁵ Bilal Abdulrahman Tuama and Farhan Mohamed, "A Systematic Literature Review of Deep Learning Methods for Handwritten Text Recognition in Historical Arabic Manuscripts," *Engineering, Technology & Applied Science Research* 15, no. 4 (2025): 25772–82, <https://doi.org/10.48084/etasr.12123>.

¹⁶ Lamia Mosbah et al., "ADOCRNet: A Deep Learning OCR for Arabic Documents Recognition," *IEEE Access* 12 (2024): 55620–31, <https://doi.org/10.1109/ACCESS.2024.3379530>.

¹⁷ Rahima Khanam and Muhammad Hussain, "What Is YOLOv5: A Deep Look into the Internal Features of the Popular Object Detector," version 1, preprint, arXiv, 2024, <https://doi.org/10.48550/ARXIV.2407.20892>.

digitization projects.¹⁸ YOLOv11, the latest iteration, further pushes the boundaries of object detection capabilities. YOLOv11 achieves state-of-the-art accuracy on various benchmark datasets with a refined architecture, optimized training algorithms, and advanced feature extraction mechanisms. The performance of YOLOv11 in complex and cluttered environments demonstrates its exceptional suitability for Arabic literature recognition, where manuscripts often contain irregularities, faded ink, and overlapping characters. Its ability to capture fine details ensures that even small variations in the Arabic script can be accurately detected and classified.¹⁹

A comparative study of YOLO5, YOLOv8, and YOLOv11 is crucial for identifying the most effective model for Arabic literature applications.²⁰ YOLO5 offers efficiency and accessibility, YOLOv8 strikes a balance between speed and accuracy, and YOLOv11 delivers state-of-the-art performance with higher computational demands.²¹ This study aims to highlight the strengths, limitations, and suitability of each model for various application needs by systematically evaluating the models on Arabic text datasets. The integration of Arabic linguistics and machine learning has become a growing research field, addressing the morphological, syntactic, and phonological complexities of Arabic using advanced computational approaches.²² Research shows that Arabic, with its diverse dialects and complex morphological structure, presents unique challenges for NLP systems, necessitating specially tailored models.²³ Modern approaches have employed a variety of techniques, ranging from traditional machine learning to deep learning and large language models (LLMs), to address tasks such as text classification, sentiment analysis, and dialect identification.²⁴ For example, studies on Arabic dialect identification using hybrid machine and deep learning models with attention mechanisms have successfully classified Egyptian, Gulf, Jordanian, and Yemeni dialects, while other research has shown that transformer models such as AraBERT improve

¹⁸ Muhammad Yaseen, "What Is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," version 1, preprint, arXiv, 2024, <https://doi.org/10.48550/ARXIV.2408.15857>.

¹⁹ Rahima Khanam and Muhammad Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," version 1, preprint, arXiv, 2024, <https://doi.org/10.48550/ARXIV.2410.17725>.

²⁰ Wafa Khelif, "Multi-Lingual Scene Text Detection Based on Convolutional Neural Networks" (Doctorat ès Informatique et applications, PhD, La Rochelle, 2022), <https://doi.org/10.70675/cb8b6b05z423ez4574z8799z6408ef415b50>.

²¹ Qian Zhao and Junhao Zhu, "An Improved YOLOv11 Architecture with Multi-Scale Attention and Spatial Fusion for Fine-Grained Residual Detection," *Results in Engineering* 27 (September 2025): 107061, <https://doi.org/10.1016/j.rineng.2025.107061>.

²² Wael M. S. Yafooz, "Enhancing Arabic Dialect Detection on Social Media: A Hybrid Model with an Attention Mechanism," *Information* 15, no. 6 (2024): 316, <https://doi.org/10.3390/info15060316>.

²³ Abdullah Y. Muaad et al., "Artificial Intelligence for Text Analysis in the Arabic and Related Middle Eastern Languages: Progress, Trends, and Future Recommendations," *International Journal of Intelligent Systems* 2025, no. 1 (2025): 6091900, <https://doi.org/10.1155/int/6091900>.

²⁴ Mafaza Chabane et al., "Advancing Low-Resource Dialect Identification: A Hybrid Cross-Lingual Model Leveraging CAMELBERT and FastText for Algerian Arabic," *Expert Systems with Applications* 284 (July 2025): 127816, <https://doi.org/10.1016/j.eswa.2025.127816>.

performance in contextual information retrieval.²⁵ At the phonological level, short vowel and dialect sound recovery efforts using the Dialectal Sound and Vowelization Recovery (DSVR) framework utilize self-supervised learning representations to recognize dialect acoustic units with limited data, and adversarial training can improve the system's robustness to perturbations by up to 9%, as demonstrated by a wav2vec 2.0 model trained to evaluate the pronunciation of isolated Arabic letters.²⁶ In addition, research on morphophonological mapping using transformer models to correct residual errors from hand-crafted rules in Egyptian Arabic shows that even under low-resource conditions, integrating linguistic knowledge with machine learning can effectively predict spoken forms from morphological representations.²⁷

The digitization and computational analysis of the Arabic script have significant cultural and academic value. Manuscript preservation efforts and automated text recognition contribute to historical knowledge accessibility and digital humanities studies development. The integration of YOLO-based models into this domain not only enhances technical capabilities but also supports interdisciplinary research that bridges computer science with language studies.²⁸

Arafat and Iqbal's research on cursive script detection confirmed that deep learning pipelines achieve high precision even for scripts with complex positional dependencies, such as Urdu, which shares structural characteristics with Arabic.²⁹ Most recently, Bouchal et al. demonstrated that a YOLOv5-based pipeline for historical Arabic manuscripts achieves an F1-score of 97.98% in text-line detection, further demonstrating the generalizability of YOLO to Arabic script tasks.³⁰

Despite these advances, a critical gap remains: existing Arabic OCR systems focus exclusively on recognition or detection tasks for technical or archival applications and, do not address the bilingual phonetic-grapheme dimension necessary for Arabic language education in the Indonesian *Madrasah Diniyah* context. This study offers a context-specific implementation for Indonesian *Madrasah Diniyah* learners that integrates Arabic letter detection with automatic Indonesian transliteration and audio output within a single pedagogical application.

²⁵ Amany A. Slamaa, "AI-Driven Semantic Information Retrieval for Arabic Language- Systematic Literature Review," *Computer Science Review* 60 (May 2026): 100920, <https://doi.org/10.1016/j.cosrev.2026.100920>.

²⁶ Hadi Zaatiti et al., "Towards Stable AI Systems for Evaluating Arabic Pronunciations," *NLP and Machine Learning Trends 2025*, August 23, 2025, 23–33, <https://doi.org/10.5121/csit.2025.151603>.

²⁷ Salam Khalifa et al., "Picking Up Where the Linguist Left Off: Mapping Morphology to Phonology through Learning the Residuals," *Proceedings of The Second Arabic Natural Language Processing Conference*, 2024, 258–64, <https://doi.org/10.18653/v1/2024.arabicnlp-1.22>.

²⁸ Lallouani Bouchakour and Nadjla Bettayeb, "Arabic Text Detection and Recognition in Video Using Deep Learning," *Signal, Image and Video Processing* 19, no. 8 (2025): 685, <https://doi.org/10.1007/s11760-025-04295-1>.

²⁹ Syed Yasser Arafat and Muhammad Javed Iqbal, "Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning," *IEEE Access* 8 (2020): 96787–803, <https://doi.org/10.1109/ACCESS.2020.2994214>.

³⁰ Hakim Bouchal et al., "Towards Accurate Recognition of Historical Arabic Manuscripts: A Novel Dataset and a Generalizable Pipeline," *ACM Transactions on Asian and Low-Resource Language Information Processing* 24, no. 10 (2025): 1–30, <https://doi.org/10.1145/3744243>.

Therefore, this study aims to: (1) develop a computer vision application that detects Arabic letters from images, displays their Indonesian phonetic transliteration, and produces automatic audio pronunciation (image-to-text-to-voice); (2) evaluate the trained YOLO model's accuracy using standard machine learning metrics; and (3) assess system usability through expert black-box testing.

Method

This study employed a Research and Development (R&D) approach to produce and validate a functional computer vision application for Arabic letter detection.³¹ The development process followed the Waterfall model, selected for its linear and systematic structure, which ensures that each phase is fully completed and verified before progressing to the next.³² Five stages were applied: requirements analysis, system design, implementation, verification, and maintenance.

System requirements were identified through structured interviews with *ustadz/ustadzah* (Islamic teachers) at both *Madrasah Diniyah* institutions in Kebumen, supplemented by direct observation and a pre-test administered to students. The data collected consisted of 17 phonetic error samples from *Madrasah Diniyah*. These errors involved the pronunciation of letters such as ز, ذ, خ, ح, ث, ق, ش, غ, ع, ط, ض, and ص. These phonetic errors occurred at four levels: sound, word, phrase, and sentence, with the main cause being difficulty in pronouncing Arabic letters that have no equivalents in Indonesian.

Table 1. Samples of Arabic Letter Pronunciation Errors in the Recitation of the Quran by Students at *Madrasah Diniyah* Al Barokah and *Madrasah Diniyah* TPQ Baitul Muttaqin

Letter Type	Example of Recitation in the Qur'an	Types of Pronunciation Errors	Number of students with incorrect pronunciation
(ع) <i>ain</i>	أَعُوذُ بِاللَّهِ	pronounced like the letter (ء) <i>hamzah</i> and also as " <i>ngain</i> "	14 students (82,35%)
(ث) <i>tsa</i>	الْكَوْثَرُ	sounds like the letter (س) <i>sin</i>	7 students (41,17%)
(ش) <i>syin</i>	الْفَهْمُ رَحْلَةُ السَّيَاءِ	sounds like the letter (س) <i>sin</i>	4 students (23,52%)
(ذ) <i>dzal</i>	الَّذِي يَكْدِبُ	sounds like the letter (د) <i>dal</i> or like the letter (ز) <i>za</i>	9 students (52,94%)
(ح) <i>ha</i>	الْحَمْدُ	sounds like the letter (هـ) <i>ha</i>	5 students (29,41%)
(غ) <i>ghoin</i>	الْمَغْضُوبِ	pronounced only as " <i>go</i> "	3 students (17,64%)
(خ) <i>kho</i>	لَقَدْ خَلَقْنَا	pronounced only as " <i>ho</i> "	3 students (17,64%)
(ق) <i>qof</i>	لَا يُلْفِ قَرِيشٍ	pronounced as the letter (ك) <i>kaf</i>	4 students (23,52%)
(ض) <i>dad</i> and (ظ) <i>dho</i>	وَلَا الضَّالِّينَ	pronounced as (د) <i>dal</i>	15 students (88,23%)

³¹ Nicola O'Connell, "Research and Development," *Nurse Prescribing* 5, no. 5 (2007): 226–29, <https://doi.org/10.12968/npre.2007.5.5.23745>.

³² Ari Kusuma Wardana et al., "Developing Martial Art Championship Scheduling System Using the Waterfall Model," *Proceeding of the 7th International Conference of Science, Technology, and Interdisciplinary Research (IC-STAR 2021)* (Bandar Lampung, Indonesia), 2023, 040006, <https://doi.org/10.1063/5.0105697>.

The analysis confirmed three principal system requirements: (1) accurate recognition of Arabic letter images with correct text output; (2) automatic Arabic audio pronunciation to support phonetic learning; and (3) an interface accessible to users of diverse ages and technical backgrounds without specialized training.

The dataset comprised 360 images of Arabic hijaiyah letters sourced from the Roboflow repository.³³ Images were manually annotated using the Labellmg tool, generating bounding box coordinates stored as txt files compatible with the YOLO training format.

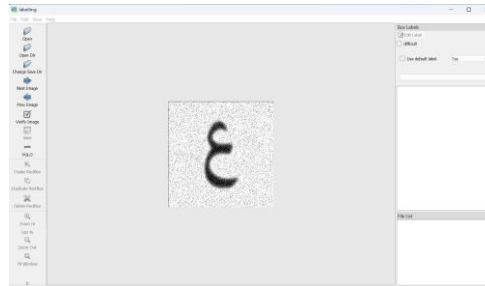


Figure 1. The Labellmg annotation interface, generating txt coordinates with YOLO training format

The annotation interface Labellmg was used for the manual labeling of Arabic hijaiyah boxes (Figure 1). Each letter image was individually annotated with its corresponding class label, generating.txt coordinate files compatible with the YOLO training format. The dataset was divided into 315 training images and 45 validation images following multiple experimental configurations. The YOLOV11 model was trained in Google Colab with the following hyperparameters: 500 epochs, a learning rate of 0.001, a batch size of 16, and an Adam optimizer. The application interface was designed to execute the full image-to-text-to-voice workflow: the user uploads an Arabic letter image, the YOLO model detects and classifies the letter, the corresponding Indonesian transliteration is displayed, and automatic audio pronunciation is generated. The system design flowchart and interface prototype are presented below.

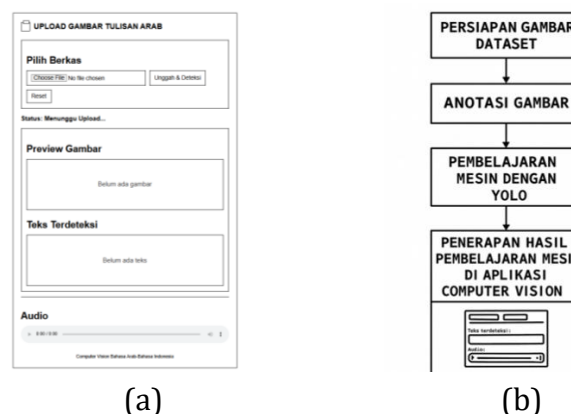


Figure 2. (a) Application interface prototype displayed. (b) Flowchart of system development

³³ Politeknik Negeri Batam, "Hijaiyah Letters Computer Vision Model," Roboflow, accessed June 11, 2026, <https://universe.roboflow.com/politeknik-negeri-batam-dtarv/hijaiyah-letters>.

Application interface prototype displaying the four integrated components: file upload panel, image preview, detected text display, and audio player. Figure 2 shows a system development flowchart illustrating the sequential pipeline from dataset preparation and image annotation through YOLO machine learning training to application deployment. The implementation phase encompassed the complete machine learning pipeline and application deployment. Bochkovski et al. established the YOLO architecture incorporating bag-of-freebies enhancements such as mosaic augmentation and CIoU loss, which processed Arabic letter images by dividing each input into a grid, with each grid cell simultaneously predicting bounding boxes, confidence scores, and class probabilities. The trained model weights were embedded into the web-based application, which executes the full inference pipeline from image upload through letter detection to text rendering and audio output. Figure 3 shows a screenshot of the application in operation, demonstrating the complete workflow on a real Arabic letter image.

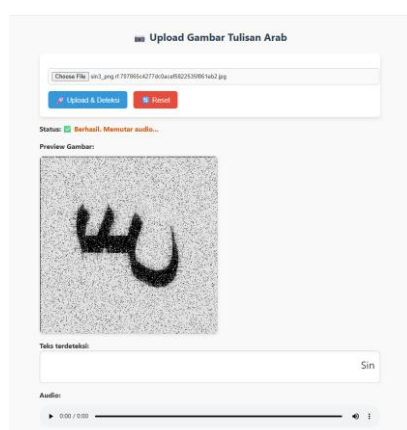


Figure 3. The computer vision application in operation, demonstrating the complete image-to-text-to-voice pipeline.

An uploaded image of the Arabic letter *Sin* (س) is displayed in the preview panel. The YOLO model correctly detects and classifies the letter, rendering the Indonesian transliteration 'Sin' in the text output field. The system status reads 'Berhasil. Memutar audio...' (Success. Playing audio...), confirming that automatic pronunciation has been triggered simultaneously. The verification was conducted in two stages. In the first stage, machine learning performance was evaluated using two standard detection metrics: (1) mAP50³⁴ and (2) mAP50-95³⁵, which provide a stricter assessment of bounding box localization precision. The model prediction patterns were further examined using a confusion matrix, which maps actual versus predicted letter classes and identifies systematic misclassification tendencies.³⁶ The

³⁴ Mean Average Precision at an Intersection over Union (IoU) threshold of 0.50, which assesses overall detection and classification accuracy.

³⁵ Averaging precision across IoU thresholds from 0.50 to 0.95 in 0.05-step increments.

³⁶ Erick Burhaein et al., "Application of Naive Bayes Algorithm for Physical Fitness Level Classification," *International Journal of Disabilities Sports and Health Sciences* 7, no. 1 (2024): 178–87, <https://doi.org/10.33438/ijdshts.1330745>.

second stage evaluated application functionality through Black-Box Testing, administered to one AI practitioner and one AI researcher. Evaluators independently rated five system dimensions on a 1–5 Likert scale: (a) ease of use across diverse age groups, (b) upload and detection workflow stability, (c) detection accuracy relative to uploaded images, (d) interface clarity, and (e) audio feature utility for pronunciation learning.

The maintenance stage encompasses ongoing model retraining with expanded and more diverse datasets, bug resolution, user feedback-responsive interface updates, and systematic version documentation. Periodic functional and performance evaluations ensure that the system maintains consistent detection accuracy and relevance to evolving pedagogical contexts.

Results and Discussion

Phonetics and Arabic Graphemes

Phonetics is the branch of linguistics concerned with the physical properties of speech sounds. Phonetic analysis in Arabic is inseparable from the concept of *makhraj*. Arabic possesses fifteen *makhraj* positions producing 28 consonantal letters, each carrying distinctive phonological properties such as *hams* (voiceless) or *jahr* (voiced) articulation.³⁷ Phonological interference is the most prevalent source of pronunciation errors among Indonesian Arabic learners.³⁸

A grapheme is the smallest contrastive unit in a writing system. In Arabic, the 28 *hijaiyah* letters are graphemes, each potentially assuming up to four positional variants depending on the word position. Critically, multiple Arabic letters share identical base forms differentiated solely by the number and placement of diacritical dots: for example, ب (*ba*), ت (*ta*), ث (*tsa*), and ن (*nun*) share the same fundamental curved shape and differ only in dot configuration.³⁹ This graphemic similarity constitutes both a linguistic challenge for learners and a computational challenge for automated recognition systems, directly explaining the misclassification patterns Confusion matrix results.

Computer Vision and YOLO Algorithm

Through training on large labeled datasets, computer vision enables machines to extract meaningful information from digital images. When applied to text recognition, computer vision transforms character identification into an object detection task that localizes and classifies each letter as a distinct visual object. YOLO (You Only Look Once), introduced by Redmon et al., reframed object detection as a

³⁷ Eiman Alsharhan and Allan Ramsay, "Improved Arabic Speech Recognition System through the Automatic Generation of Fine-Grained Phonetic Transcriptions," *Information Processing & Management* 56, no. 2 (2019): 343–53, <https://doi.org/10.1016/j.ipm.2017.07.002>.

³⁸ Ardhan Tirta Kusumawan and Fetty Poerwita Sary, "The Influence of Job Characteristics on Job Application Intention Through Employer Attractiveness as a Mediating Variable on Generation Z in the Indonesian Pharmaceutical Manufacturing Industry," *Enrichment: Journal of Multidisciplinary Research and Development* 3, no. 3 (2025): 400–413, <https://doi.org/10.55324/enrichment.v3i3.389>.

³⁹ Thomas Milo and Alicia González Martínez, "A New Strategy for Arabic OCR: Archigraphemes, Letter Blocks, Script Grammar, and Shape Synthesis," *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, May 8, 2019, 93–96, <https://doi.org/10.1145/3322905.3322928>.

single regression problem: a convolutional neural network divides the input image into an $S \times S$ grid and simultaneously predicts B bounding boxes, confidence scores, and C class probabilities for each cell in one forward pass.⁴⁰ This unified architecture enables real-time inference at 45 frames per second (base model) and exceeds 150 frames per second in its fast variant, making it exceptionally well-suited for interactive applications.

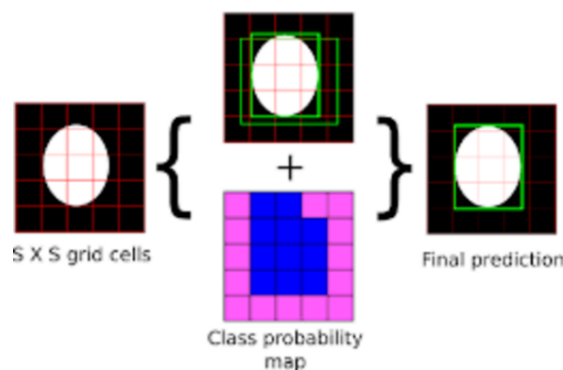


Figure 4. YOLO detection mechanism.

The input image is divided into an $S \times S$ grid; each cell independently predicts bounding boxes with confidence scores and a class probability map, which are combined to produce the final detection output (Figure 4). This unified single-pass architecture enables real-time object detection without the multistage pipeline required by region-proposal methods such as R-CNN.⁴¹ YOLOv4 substantially extended these capabilities. Its bag-of-freebies techniques, including mosaic augmentation (combining four training images per batch), Ciou loss for precise bounding box regression, and DropBlock regularization, increase training accuracy without adding inference cost. Its bag-of-specials modules Mish activation, CSP connections, and PAN path-aggregation neck deliver significant accuracy gains at modest additional inference cost. The resulting architecture (CSPDarknet53 + SPP + PAN + YOLOv3 head) achieved an AP of 43.5% on MS COCO at ~65 FPS, which was trainable on a single GTX 1080Ti or RTX 2080Ti GPU.²²

Evaluation Metrics

The confusion matrix maps predicted class labels against actual ground-truth labels, revealing patterns of correct classification (true positives), false alarms (false positives), missed detections (false negatives), and correct rejections (true negatives). The Confusion Matrix is particularly informative for identifying systematic confusion between graphemically similar letter pairs in the context of Arabic letter detection. The mAP50 metric aggregates the average precision across all classes at an IoU threshold of 0.50, reflecting detection and classification accuracy. The stricter mAP50-95 metric averages precision across IoU thresholds from 0.50 to 0.95 in 0.05 increments, severely penalizing imprecise bounding box placement.

⁴⁰ Redmon et al., "You Only Look Once."

⁴¹ Akhmad Fadjari, *Computer Vision Untuk Pemula* (CV. Rumah Kreatif Wadas Kelir, 2024).

Machine Learning Evaluation Results

The evaluation framework for the machine learning stage of this study is built on the confusion matrix. Table 2 presents the four outcome categories that underpin all accuracy measurements in this study.

Table 2. Confusion Matrix Framework for Model Evaluation

Predicted/Actual	Actual Positive (1)	Actual Negative (0)
Predicted Positive (1)	True Positive (TP)	False Positive (FP)
Predicted Negative (0)	False Negative (FN)	True Negative (TN)

Based on this framework, the accuracy proportion of all correct predictions over total predictions is defined in Equation 1.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Equation 1. Accuracy formula derived from the Confusion Matrix.

In the object detection context, accuracy is further operationalized through mean average precision (mAP). mAP50 measures the average precision at an intersection over union (IoU) threshold of 0.50, reflecting detection and classification accuracy. The stricter mAP50-95 averages precision across IoU thresholds from 0.50 to 0.95 in 0.05-step increments, providing a more rigorous assessment of bounding box localization precision. The YOLOV11 model achieved its best performance at epoch 297 of a 500-epoch training run. From the perspective of computational complexity and resource limitations (computational budget), this research was designed as a proof-of-concept study that aims to demonstrate the feasibility of the YOLOv11 architecture in recognizing Arabic letters, not to achieve state-of-the-art performance on large datasets. Initial validation with limited datasets is a justifiable scientific step in many computer vision studies before expanding to a larger scale. The mAP50 value of 99.5% on a limited dataset indicates that the model has a high learning capacity, which is a positive indication for further development on larger datasets. The results are presented in Table 3.

Table 3. YOLO Model Training Configuration and Best Performance Metrics

Training Configuration	mAP50 (B)	mAP50-95 (B)
500 epochs LR: 0.001 Batch: 16 Adam optimizer Best epoch: 297	0.995 (99.5%)	0.911 (91.12%)

Table 2 confirms that the model operates at a very high level of accuracy across both metrics. To provide a more granular view of the prediction quality at the individual letter class level, which reveals which specific letter pairs the model confuses and why, Figure 5 presents the normalized confusion matrix generated from the 45-image validation set. The matrix maps each actual letter class (x-axis) against the predicted label (y-axis) of the model, enabling the direct identification of systematic misclassification patterns across all 28 Arabic letter classes.

small physical dimensions of individual Arabic characters, which lack color features to aid discrimination, further complicate the precise placement of bounding boxes. Therefore, the 8.38-percentage-point gap between mAP50 and mAP50-95 reflects bounding box localization imprecision rather than letter misclassification, a distinction directly supported by the Confusion Matrix analysis.

Confusion matrix analysis revealed that the model performs correctly for the most letter classes. Systematic misclassifications were concentrated among graphemically similar letter pairs: *Kaf* (ك) was occasionally predicted as *Tsa* (ث); *Nun* (ن) generated partial confusions with *Tsa* (ث) and *Ta* (ت); *Alif* (ا) was occasionally predicted as *Ghain* (غ); and *Za* (ز) was misclassified as *Ghain* (غ) in a small number of cases. These error patterns directly reflect the graphemic structure of Arabic: all confused letter pairs either share a base form differentiated only by diacritical dots or share similar stroke proportions and dimensions. This finding corroborates Alghyaline's identification of systematic YOLO confusions among letters with high inter-class visual similarity and underscores the importance of large, diverse, and precisely annotated training datasets for achieving higher precision on these specific pairs.⁴⁴

Application Testing: Results of Black-Box Testing

Black-Box testing was conducted with two expert evaluators, one AI practitioner (teacher of Islamic boarding school) and one AI researcher, who independently rated the application across five functional criteria on a 1–5 Likert scale. The results are presented. This important to note that this testing phase focused exclusively on expert judgment and did not involve end-users (students/*santri*), whose direct experiences and needs will be explored in subsequent user-acceptance evaluations in Table 3.

Table 3. Black-Box Testing Results by an Expert Evaluator

Evaluation Criterion	AI Practitioner	AI Researcher	Mean
System ease of use across diverse age groups	4.2	4.6	4.4
Upload and detection workflow runs without technical errors	4.0	4.4	4.2
Detection results accurately correspond to the uploaded letter	4.2	4.8	4.5
The user interface is clear and easy to understand	4.0	4.6	4.3
Audio features support learning Arabic pronunciation	4.5	4.7	4.6

All five evaluation items received mean Likert scores above 4.0, indicating a strong positive expert assessment. The audio pronunciation feature received the highest mean score (4.6), reflecting its perceived value as a direct tool for addressing the primary learning objective: accurate Arabic phonetic production. Detection accuracy (item 3, mean 4.5) and ease of use (item 1, mean 4.4) also received high ratings, confirming that the YOLO-based detection pipeline performs reliably and that the application is accessible to users without technical expertise. Similarly, interface clarity (item 4, mean 4.3) and technical workflow stability (item 2, mean 4.2) were positively evaluated, indicating that the system design effectively serves its pedagogical use case.

⁴⁴ Alghyaline, "A Printed Arabic Optical Character Recognition System Using Deep Learning."

Based on Table 3, the five sub-items of the Black-Box Testing evaluation criteria focus on three main aspects. The first aspect is user experience, which covers sub-items 1, 4, and 5, with the assessment focusing on ease of use across diverse age groups, user interface clarity, and audio feature support for Arabic pronunciation learning, all three of which address the comfort and educational benefits of the application for students (*santri*). The second aspect is Technical Reliability, found in sub-item number 2, which focuses on the upload and detection workflow's stability to ensure it runs without technical issues, such as crashes or system failures. The third aspect is AI accuracy, which is contained in sub-item number 3, which focuses on the accuracy of detection results in matching the uploaded letter as the core functionality of the application. Overall, the assessment places greater emphasis on user experience and pedagogical aspects (60% of the total criteria) than on technical and accuracy aspects, which aligns with the application's goal of supporting the teaching and learning process in the Islamic boarding school environment.

The primary novelty of this research lies in the integration of YOLO-based Arabic character detection with a bilingual phonetic-grapheme output system combining Indonesian transliteration text and automatic Arabic audio pronunciation within a single, accessible web application. This distinction distinguishes this work from prior studies in three important respects. First, existing YOLO applications to Arabic text focus exclusively on OCR or detection tasks without connecting visual recognition to phonetic output.⁴⁵ Second, studies addressing Arabic phonetics for Indonesian learners operate purely within the domain of linguistics and pedagogy without computational implementation.⁴⁶ Third, this application targets a specific underserved educational context rather than archival digitization or commercial OCR.

By bridging computational Arabic OCR with phonetic-grapheme linguistics and situating the technology within the context of the Indonesian *Madrasah Diniyah*. This study offers a context-specific implementation for Indonesian *Madrasah Diniyah* learners that integrates Arabic letter detection with automatic Indonesian transliteration and audio output within a single pedagogical application. The image-to-text-to-voice pipeline extends the state-of-the-art by providing a full multimodal learning intervention that supports both identification and phonetic production. The achieved mAP50 of 99.5% demonstrates that YOLO's generalization capability documented for natural images by Redmon et al. and for Arabic script by Alghyaline and Khan et al. extends effectively to *hijaiyah* letter detection even with a relatively small training dataset of 360 images, provided annotations are precise and data augmentation is applied.⁴⁷

⁴⁵ Khan et al., "Robust Arabic and Pashto Text Detection in Camera-Captured Documents Using Deep Learning Techniques"; Alghyaline, "A Printed Arabic Optical Character Recognition System Using Deep Learning."

⁴⁶ Mohamed et al., "A Systematic Review Of Arabic Phonetic."

⁴⁷ Redmon et al., "You Only Look Once"; Alghyaline, "A Printed Arabic Optical Character Recognition System Using Deep Learning"; Khan et al., "Robust Arabic and Pashto Text Detection in Camera-Captured Documents Using Deep Learning Techniques."

Conclusion

This study successfully developed a computer vision application for Arabic-Indonesian linguistic learning based on a phonetics-grapheme framework using the YOLO object detection algorithm within a Waterfall development model. The trained YOLO model achieved 99.5% accuracy (mAP50) and 91.12% accuracy (mAP50-95), substantially surpassing comparable benchmarks in the Arabic character recognition literature, including the 95.7% CRR reported by Alghyaline and the 91.30% mAP reported by Khan et al. The image-to-text-to-voice pipeline was validated through expert Black-Box testing, with all five functional evaluation items receiving mean Likert scores above 4.0 and the audio feature receiving the highest rating (4.6). Confusion matrix analysis confirmed that residual misclassification errors were concentrated among graphemically similar letter pairs that shared base forms that were differentiated only by diacritical dots.

The study's primary limitation is the relatively small and homogeneous training dataset (360 images from a single source), which may restrict the model's generalizability to diverse handwriting styles, varied lighting conditions, and complex image backgrounds. Expert testing was also limited to two evaluators. For future validation, broader user testing with actual *Madrasah Diniyah* students and teachers is recommended. Future research directions include: expanding the dataset with handwritten and calligraphic Arabic samples, integrating the application with Learning Management Systems (LMS) for formal *Madrasah Diniyah* adoption, adding word- and sentence-level detection, and exploring the integration of vision-language model components to support morphological disambiguation and contextual understanding of Arabic script.

Acknowledgment

The authors express sincere gratitude to the *ustadz* and *ustadzah* of *Madrasah Diniyah* Al-Barokah Kejawang Kebumen and *Madrasah Diniyah* TPQ Masjid Baitul Muttaqin Kebumen for their cooperation during the needs analysis and field observation phases. The authors would like to thank the AI expert evaluators who participated in the Black-Box testing. The authors would also like to thank the editor and reviewers for their constructive feedback, and the thorough assessment of the manuscript. This study received no specific external funding.

Author's Contribution Statement

ADN contributed to the conceptualization, data collection and annotation, model training, application development, and drafting of the original manuscript. EKA contributed to the development of the theoretical framework and manuscript review and editing. AF contributed to research conceptualization, methodology design, and manuscript revision. All authors have read and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

AI Writing Statement

During the preparation of this manuscript, the authors used Grammarly to assist with language editing, grammar checking and Notebook LM for organizing writing ideas. The authors carefully reviewed and edited the content generated with the help of AI and takes full responsibility for the final content of this manuscript.

References

- Al Farisi, Mohamad Zaka, Mad Ali, Zawawi Ismail, Hikmah Maulani, Nalahuddin Shaleh, and Shofa Mushtofa Khalid. "Investigation of the Pronunciation of the Voiceless Fricative Non-Sibilant Phoneme /θ/ in Arabic: An Acoustic Phonetics Comparative Analysis." *Dirasat: Human and Social Sciences* 52, no. 6 (2025): 5447. <https://doi.org/10.35516/Hum.2025.5447>.
- Alghyaline, Salah. "A Printed Arabic Optical Character Recognition System Using Deep Learning." *Journal of Computer Science* 18, no. 11 (2022): 1038–50. <https://doi.org/10.3844/jcssp.2022.1038.1050>.
- Alsharhan, Eiman, and Allan Ramsay. "Improved Arabic Speech Recognition System through the Automatic Generation of Fine-Grained Phonetic Transcriptions." *Information Processing & Management* 56, no. 2 (2019): 343–53. <https://doi.org/10.1016/j.ipm.2017.07.002>.
- Anugrah, Anugrah, Mantasiah Rivai, and Fatkhul Ulum. "Analysis of Regional Language Phonological Interference on the Pronunciation of Hijaiyah Letters of Madrasah Tsanawiyah Students in Gowa Regency." *Pinisi Journal of Art, Humanity, and Social Studies* 4, no. 4 (2024). <https://journal.unm.ac.id/index.php/PJAHSS/article/view/2882>.
- Arafat, Syed Yasser, and Muhammad Javed Iqbal. "Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning." *IEEE Access* 8 (2020): 96787–803. <https://doi.org/10.1109/ACCESS.2020.2994214>.
- Birgün, Mehmet. "Integrating AI into Qur'an Learning: Technical Advances and Pedagogical Gaps." *Social Sciences & Humanities Open* 13 (June 2026): 102499. <https://doi.org/10.1016/j.ssaho.2026.102499>.
- Bochkovski, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection." Version 1. Preprint, arXiv, 2020. <https://doi.org/10.48550/ARXIV.2004.10934>.
- Bouchakour, Lallouani, and Nadjla Bettayeb. "Arabic Text Detection and Recognition in Video Using Deep Learning." *Signal, Image and Video Processing* 19, no. 8 (2025): 685. <https://doi.org/10.1007/s11760-025-04295-1>.
- Bouchal, Hakim, Ahror Belaid, and Farid Meziane. "Towards Accurate Recognition of Historical Arabic Manuscripts: A Novel Dataset and a Generalizable

Pipeline.” *ACM Transactions on Asian and Low-Resource Language Information Processing* 24, no. 10 (2025): 1–30. <https://doi.org/10.1145/3744243>.

Burhaein, Erick, Akhmad Fadjeri, and Ibnu Prasetyo Widiyono. “Application of Naive Bayes Algorithm for Physical Fitness Level Classification.” *International Journal of Disabilities Sports and Health Sciences* 7, no. 1 (2024): 178–87. <https://doi.org/10.33438/ijdshts.1330745>.

Chabane, Mafaza, Fouzi Harrag, and Khaled Shaalan. “Advancing Low-Resource Dialect Identification: A Hybrid Cross-Lingual Model Leveraging CAMELBERT and FastText for Algerian Arabic.” *Expert Systems with Applications* 284 (July 2025): 127816. <https://doi.org/10.1016/j.eswa.2025.127816>.

Daman, Damanhuri, and Jamiluddin Yacub. “Mengenalkan Huruf Hijaiyah Pada Anak Usia Dini.” *Azzahra: Jurnal Pendidikan Anak Usia Dini* 3, no. 2 (2022). <https://ejournal.staidarussalamlampung.ac.id/index.php/azzahra/article/view/374>.

Fadjeri, Akhmad. *Computer Vision Untuk Pemula*. CV. Rumah Kreatif Wadas Kelir, 2024.

Khalifa, Salam, Abdelrahim Qaddoumi, Ellen Broselow, and Owen Rambow. “Picking Up Where the Linguist Left Off: Mapping Morphology to Phonology through Learning the Residuals.” *Proceedings of The Second Arabic Natural Language Processing Conference*, 2024, 258–64. <https://doi.org/10.18653/v1/2024.arabicnlp-1.22>.

Khan, Nisar, Riaz Ahmad, Khalil Ullah, et al. “Robust Arabic and Pashto Text Detection in Camera-Captured Documents Using Deep Learning Techniques.” *IEEE Access* 11 (2023): 135788–96. <https://doi.org/10.1109/ACCESS.2023.3336404>.

Khanam, Rahima, and Muhammad Hussain. “What Is YOLOv5: A Deep Look into the Internal Features of the Popular Object Detector.” Version 1. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2407.20892>.

Khanam, Rahima, and Muhammad Hussain. “YOLOv11: An Overview of the Key Architectural Enhancements.” Version 1. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2410.17725>.

Khelif, Wafa. “Multi-Lingual Scene Text Detection Based on Convolutional Neural Networks.” Doctorat ès Informatique et applications, PhD, La Rochelle, 2022. <https://doi.org/10.70675/cb8b6b05z423ez4574z8799z6408ef415b50>.

Kusumawan, Ardhian Tirta, and Fetty Poerwita Sary. “The Influence of Job Characteristics on Job Application Intention Through Employer Attractiveness as a Mediating Variable on Generation Z in the Indonesian Pharmaceutical

Manufacturing Industry." *Enrichment: Journal of Multidisciplinary Research and Development* 3, no. 3 (2025): 400–413. <https://doi.org/10.55324/enrichment.v3i3.389>.

Lubis, Torkis, Sounia Rabhi, and Rahmat Linur. "Enhancing Arabic Pronunciation Learning through Contrastive Analysis: An Empirical Study with Indonesian University Students." *International Journal of Special Education* 41, no. 2 (2026). <https://www.internationalsped.com/index.php/ijse/article/view/3331>.

Medina, Zahra Nida Al, Zukhaira Zukhaira, and Muhammad Yusuf Ahmad Hasyim. "The Arabic Adventure; Media Board Game Untuk Keterampilan Berbicara Bahasa Arab Siswa Kelas VII MTs Di Purbalingga." *Lisanul Arab: Journal of Arabic Learning and Teaching* 8, no. 1 (2019): 1. <https://journal.unnes.ac.id/sju/laa/article/view/32541>.

Milo, Thomas, and Alicia González Martínez. "A New Strategy for Arabic OCR: Archigraphemes, Letter Blocks, Script Grammar, and Shape Synthesis." *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, May 8, 2019, 93–96. <https://doi.org/10.1145/3322905.3322928>.

Mohamed, Yuslina, Zainurrijal Abd Razak, Tuan Haji Sulaiman Ismail, Nada Ibrahim Alribdi, and Mesbahul Hoque. "A Systematic Review Of Arabic Phonetic: Hijaiyyah's Pronunciation Among New Learners." *Ijaz Arabi Journal of Arabic Learning* 7, no. 2 (2024). <https://doi.org/10.18860/ijazarabi.v7i2.24173>.

Mosbah, Lamia, Ikram Moalla, Tarek M. Hamdani, Bilel Neji, Taha Beyrouthy, and Adel M. Alimi. "ADOCRNet: A Deep Learning OCR for Arabic Documents Recognition." *IEEE Access* 12 (2024): 55620–31. <https://doi.org/10.1109/ACCESS.2024.3379530>.

Muaad, Abdullah Y., Md Belal Bin Heyat, Faijan Akhtar, et al. "Artificial Intelligence for Text Analysis in the Arabic and Related Middle Eastern Languages: Progress, Trends, and Future Recommendations." *International Journal of Intelligent Systems* 2025, no. 1 (2025): 6091900. <https://doi.org/10.1155/int/6091900>.

Murat, Ayşe Aybilge, and Mustafa Servet Kiran. "A Comprehensive Review on YOLO Versions for Object Detection." *Engineering Science and Technology, an International Journal* 70 (October 2025): 102161. <https://doi.org/10.1016/j.jestch.2025.102161>.

Muthmainnah, Hilyah Ahya, and Lina Marlina. "Kesalahan Fonologi Dalam Membaca Al-Qur'an Pada Siswi Penutur Bahasa Sunda Di Tahfiz Qur'an Al-Marjan." *Al-Lahjah : Jurnal Pendidikan, Bahasa Arab, Dan Kajian Linguistik Arab* 8, no. 1 (2025). <https://ejournal.unwaha.ac.id/lahjah/article/view/5936>.

- Nailurrahmi, Fitma, and Lina Marlina. "Analisis Interferensi Fonetik Bahasa Ibu Terhadap Pelafalan Fonem Bahasa Arab Dalam Pembelajaran Kosakata." *Moderasi: Journal of Islamic Studies* 5, no. 1 (2025): 279–90. <https://doi.org/10.54471/moderasi.v5i1.113>.
- Naim, Ngainun, Abdul Aziz, and Teguh Teguh. "Integration of Madrasah Diniyah Learning Systems for Strengthening Religious Moderation in Indonesian Universities." *International Journal of Evaluation and Research in Education (IJERE)* 11, no. 1 (2022): 108. <https://doi.org/10.11591/ijere.v11i1.22210>.
- O'Connell, Nicola. "Research and Development." *Nurse Prescribing* 5, no. 5 (2007): 226–29. <https://doi.org/10.12968/npre.2007.5.5.23745>.
- Politeknik Negeri Batam. "Hijaiyah Letters Computer Vision Model." Roboflow. Accessed June 11, 2026. <https://universe.roboflow.com/politeknik-negeri-batam-dtarv/hijaiyah-letters>.
- Rahayu, Ratih, and Sri Munawarah. "Investigation of Yogyakarta Dialect's Vocabulary in Areas of Banyumas' Ngapak Dialect." In *Sociolinguistics and Dialectological Studies in Indonesia*. Nova Science Publishers, Inc., 2021. <https://www.scopus.com/pages/publications/85116743336>.
- Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, 779–88. <https://doi.org/10.1109/CVPR.2016.91>.
- Slamaa, Amany A. "AI-Driven Semantic Information Retrieval for Arabic Language-Systematic Literature Review." *Computer Science Review* 60 (May 2026): 100920. <https://doi.org/10.1016/j.cosrev.2026.100920>.
- Tuama, Bilal Abdulrahman, and Farhan Mohamed. "A Systematic Literature Review of Deep Learning Methods for Handwritten Text Recognition in Historical Arabic Manuscripts." *Engineering, Technology & Applied Science Research* 15, no. 4 (2025): 25772–82. <https://doi.org/10.48084/etasr.12123>.
- Wardana, Ari Kusuma, Marti Widya Sari, and Kartikadyota Kusumaningtyas. "Developing Martial Art Championship Scheduling System Using the Waterfall Model." *Proceeding of the 7th International Conference of Science, Technology, and Interdisciplinary Research (IC-STAR 2021)* (Bandar Lampung, Indonesia), 2023, 040006. <https://doi.org/10.1063/5.0105697>.
- Yafooz, Wael M. S. "Enhancing Arabic Dialect Detection on Social Media: A Hybrid Model with an Attention Mechanism." *Information* 15, no. 6 (2024): 316. <https://doi.org/10.3390/info15060316>.

- Yaseen, Muhammad. "What Is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector." Version 1. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2408.15857>.
- Zaatiti, Hadi, Hatem Hajri, Osama Abdullah, and Nader Masmoudi. "Towards Stable AI Systems for Evaluating Arabic Pronunciations." *NLP and Machine Learning Trends* 2025, August 23, 2025, 23–33. <https://doi.org/10.5121/csit.2025.151603>.
- Zhao, Qian, and Junhao Zhu. "An Improved YOLOv11 Architecture with Multi-Scale Attention and Spatial Fusion for Fine-Grained Residual Detection." *Results in Engineering* 27 (September 2025): 107061. <https://doi.org/10.1016/j.rineng.2025.107061>.

Biography of the Authors



Anisa Dwi Nurchayati   is a Master's student in the Arabic Language and Literature at UIN Sunan Kalijaga Yogyakarta, Indonesia. She completed her Bachelor's degree in Islamic Education Management at UIN Prof. KH Saifuddin Zuhri Purwokerto. Her academic work spans Arabic linguistics, educational technology, and ICT-based learning management systems. Among her published works are a systematic literature review on Arabic phonetic interference (*Stilistika: Jurnal Pendidikan Bahasa dan Sastra*, 2026), a study on the effectiveness of Learning Management Systems in administrative services at madrasah level (*Jurnal Kridatama Sains dan Teknologi*, 2023), a collaborative paper on ICT-based education management development at MTs Plus Nurul Falah (*Jurnal Kridatama Sains dan Teknologi*, 2024), and a study on ICT-based learning media validity testing (*Jurnal Pendidikan Surya Edukasi*, 2022). She is actively involved in community service activities and applied research on the intersection of digital technology and Islamic education. She can be contacted at 25201011004@student.uin-suka.ac.id.



Eric Kunto Aribowo     is a lecturer at UIN Sunan Kalijaga Yogyakarta, specializing in linguistics, sociolinguistics, onomastics, and educational technology. His latest research spans Arabic onomastics, language technology, genre-aware language modelling, and technology-mediated language learning. His relevant publications include Genre Aware Language Modeling for Indonesian Academic Writing: Building and Evaluating IndoSciBERT (*Lingua: Journal of Linguistics and Language*, 2025), studies on Arabic naming practices and Arab-Javanese linguistic identity (*Arabi: Journal of Arabic Studies*, 2019; *Arabiyat*, 2019; *Langkawi*, 2020), and multiple works on digital educational tools including gamification, mobile learning, and AI-assisted academic writing. He has delivered invited lectures on AI in research and publication, bibliometric analysis, and Arabic language learning technology at over 100 events. He can be contacted at eric.aribowo@uin-suka.ac.id.



Akhmad Fadjeri     is a lecturer at Universitas Ma'arif Nahdlatul Ulama Kebumen, Indonesia, specialising in computer science, machine learning, and computer vision. He is the author of “*Computer Vision untuk Pemula*” (Rumah Kreatif Wadas Kelir, 2024), a foundational textbook on computer vision for Indonesian practitioners. His applied research focuses on YOLO-based object detection systems, safety helmet detection using YOLOv8, physical fitness classification using Naive Bayes algorithms, and the development of ICT-based educational management systems. He has supervised multiple undergraduate and postgraduate research projects at the intersection of deep learning and educational technology, and has contributed to community service programmes in digital innovation and Islamic education institution management. He can be contacted at fadjeri.akhmadfadjeri@gmail.com.