IJOLAE Indonesian Journal on Learning and Advanced Education

http://journals.ums.ac.id/index.php/ijolae

Deep Learning-Enhanced Kitabah Application for Inclusive and Adaptive Quranic Sign Language Education

Aris Rakhmadi[1]✉, Anton Yudhana[2], Sunardi[3], Yuli Rahmawati[4]
[1]Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Indonesia
[1,2,3]Faculty of Industrial Technology, Universitas Ahmad Dahlan, Indonesia
[4]Education and Cultural Attaché, Embassy of the Republic of Indonesia in Canberra, Australia

DOI: 10.23917/ijolae.v8i2.15390
Received: February 4th, 2026. Revised: April 29th, 2026. Accepted: May 2nd, 2026
Available Online: May 14th, 2026. Published Regularly: May, 2026

Abstract

People with hearing and speech disabilities (PHSD) continue to face barriers in accessing Quranic literacy education due to the dominance of auditory–verbal instructional approaches and the limited availability of adaptive digital learning environments. Although sign language recognition (SLR) technologies have advanced significantly, most existing systems are not aligned with pedagogically and theologically grounded Quranic learning frameworks. This study aims to develop and evaluate a deep learning-enhanced Kitabah application to support inclusive, adaptive, and technology-enhanced Quranic sign language education for PHSD learners. The study employed two approaches: (1) the development of a Kitabah-based mobile learning application integrating interactive visual–motor learning features, and (2) the implementation of a deep learning-based SLR model using ResNet-18 with transfer learning for static Hijaiyyah gesture recognition. The mobile application was evaluated through black-box testing and the System Usability Scale (SUS), while the SLR model was assessed using accuracy, precision, recall, and F1-score metrics. Results showed that all application functionalities operated successfully, with the application achieving a SUS score of 78.06, indicating good usability and accessibility. The SLR model achieved 98% classification accuracy across 31 Hijaiyyah sign classes, demonstrating strong recognition performance. These findings indicate that integrating the Kitabah method with deep learning and mobile learning technology can support progressive, inclusive, and adaptive Quranic literacy education through AI-assisted and learner-centered educational experiences for PHSD learners.

Keywords: advanced education and learning, AI-assisted learning, deep learning, inclusive learning, innovative learning design, Kitabah method, mobile learning technology, progressive education, enhanced learning.

This is an open access article under the CC-BY license

✉***Corresponding Author:***
Aris Rakhmadi, Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Indonesia
Email: aris.rakhmadi@ums.ac.id

1. Introduction

People with hearing and speech disabilities (PHSD) encounter persistent barriers in developing Quranic literacy because mainstream pedagogies emphasize auditory–verbal instruction. These limitations not only affect religious literacy acquisition but also restrict opportunities for visual literacy, digital literacy, and inclusive participation in technology-mediated learning environments. Mastery of the Hijaiah (Hijaiyyah) letters underpins Quranic recitation, prayer, and broader religious practice, yet visual-first pathways that match PHSD learning needs remain

under-provisioned in digital, scalable forms (Ali & Codina, 2024).

The Kitabah method issued by the Indonesian Ministry of Religious Affairs provides a culturally and theologically grounded framework by mapping letters, diacritics, and punctuation to visual hand gestures that mirror their written forms. In the context of advanced education and learning, the integration of Kitabah with AI-supported mobile learning environments enables a more adaptive, personalized, technology-enhanced, and learner-centered educational experience for PHSD learners. Furthermore, this integration promotes digital religious literacy and supports progressive educational transformation through intelligent learning technologies. Such an approach aligns with the growing development of technology-enhanced learning (TEL) and intelligent educational systems that support inclusive and equitable access to religious education. While Kitabah aligns with visual–motor learning and has proven practical in classroom settings, its dissemination largely relies on manual delivery and static media, limiting interactivity, personalization, and timely feedback capabilities that modern learners, especially PHSD, critically need (Dzulkifli, 2022).

Internationally, research on American Sign Language (ASL) has driven much of the technical progress in sign language recognition (SLR). Studies span isolated and continuous recognition using convolutional networks, 3D CNNs for spatiotemporal cues, recurrent models for sequence modeling, and more recently transformer-based architectures (Alsharif et al., 2023; Meitantya et al., 2024). These efforts demonstrate that deep learning can robustly classify handshapes and movements from images or video, sometimes aided by body or hand keypoints. However, ASL corpora and benchmarks are designed for conversational communication and general literacy; they do not encode Quranic rules, Arabic-script visual structures, or religiously specific gestural conventions.

In the Indonesian context, SIBI (Sistem Isyarat Bahasa Indonesia) has been widely explored for assistive technologies, often focusing on static handshapes for alphabet or word-level recognition (Arthur Limantara & Tristianto, 2024). Many prototypes rely on camera-based classification with conventional machine learning or shallow CNNs; some leverage sensor gloves to stabilize input variation (Rakhmadi et al., 2025a). These systems validate feasibility for everyday Indonesian-language communication but are not designed to represent Arabic orthography, Quranic diacritics, or recitation-specific movement patterns.

BISINDO (*Bahasa Isyarat Indonesia*), a natural sign language used by Deaf communities, has inspired research focused on dynamic, context-rich expressions (Fadlilah et al., 2022a). Work on BISINDO commonly addresses phrase-level or sentence-level communication, sometimes with skeleton-based pose estimation to capture motion (Agustin et al., 2023;). Yet, as with SIBI and ASL, BISINDO-oriented systems rarely address the distinctive visual–orthographic mapping required by Quranic learning, and standardized datasets for Arabic-script religious content remain scarce.

Across ASL, SIBI, and BISINDO studies, the prevailing design objective is general communication (Fadlilah et al., 2019). Datasets, annotation schemes, and evaluation protocols prioritize conversational semantics over orthographic fidelity, diacritic awareness, and movement conventions necessary for Quranic pedagogy. Consequently, even high-performing models in these domains may fail to capture the granularity and theological specificity demanded by Quranic sign representations.

Within Arabic-script SLR, widely used resources such as the Arabic Sign Language Alphabet (ArSLA) emphasize static letters and do not natively incorporate Kitabah's structure, which includes both static and movement-based signs tied to Quranic reading (Al-Khuraym & Ismail, 2022). This structural mismatch constrains direct transfer: models trained purely on generic Arabic alphabet signs cannot reliably encode Kitabah's movement semantics or its letter–diacritic–punctuation mappings that are integral to correct Quranic practice.

Taken together, the literature reveals a twofold gap. First, mobile learning tools that adopt visual pedagogy for Quranic literacy seldom integrate AI-driven SLR capable of adaptive feedback, personalization, and real-time recognition aligned with Kitabah. Second, SLR research that achieves strong results on ASL, SIBI, BISINDO, or generic Arabic alphabets has not been reconciled with culturally and theologically validated frameworks for Quranic learning (Darmawan et al., 2023). The absence of a unifying approach means PHSD learners are not benefiting from the synergy of established pedagogy, innovative learning design, and state-of-the-art AI-driven educational technologies capable of supporting adaptive and inclusive learning experiences.

Positioned against this backdrop, the present study supports prior advances in sign recognition and inclusive mobile learning while explicitly correcting their misalignment with Quranic requirements and extending them into a unified solution. By grounding recognition targets in Kitabah and situating them within an Android-based learning environment, the study debates the sufficiency of generic datasets for religious instruction and instead advocates a culturally anchored, pedagogy-first modeling strategy.

The novelty of this work lies in fusing the Kitabah method with deep-learning-based SLR within a mobile application specifically designed for inclusivity, adaptivity, and progressive digital learning. The proposed framework integrates deep learning, mobile learning technology, and learner-centered pedagogy to support meaningful and equitable Quranic literacy education. Beyond technological integration, this study reflects the principles of progressive education by promoting learner-centered interaction, self-paced learning, and equitable educational participation for PHSD learners through accessible digital environments (Putrada et al., 2023). The approach is not merely a port of generic SLR into a religious context; it re-specifies the recognition problem around Kitabah's representational units, incorporates movement-aware signing, and embeds the technology within an accessible learning experience that aligns with theological guidance and community practices.

Although the integration of the Kitabah method with deep-learning-based SLR is presented as a novel approach, it is important to situate this work within prior research on gesture-based learning and Arabic-alphabet SLR applications. Previous studies have explored various gesture-based learning systems, including those using the ASL alphabet and ArSLA, as well as those employing convolutional neural networks (CNNs) for gesture recognition (Fadlilah et al., 2021). However, these approaches typically focus on generic alphabetic gestures and lack the specific focus on the theological and orthographic requirements of Quranic education. While SLR systems for Arabic alphabet recognition have made significant advancements, they generally do not incorporate the distinctive visual mappings and diacritical precision required for Quranic recitation, as seen in the Kitabah method. This study builds upon existing work

by integrating Kitabah's unique visual–motor pedagogy with a deep learning model tailored to Quranic literacy, ensuring that both the static and dynamic gestures used in Quranic recitation are appropriately addressed.

To address the objectives of this study in a measurable way, the following research questions have been formulated: (1) How effective is the Kitabah-based mobile learning application in improving PHSD learners' ability to recognize Hijaiyyah letters? This will be measured through pre- and post-tests, as well as quizzes, to assess learners' ability to recall and recognize letters. (2) What is the accuracy of the deep learning-based sign language recognition (SLR) model in classifying static Hijaiyyah signs? The model's performance will be evaluated using classification metrics such as accuracy, precision, recall, and F1-score. (3) Does the integration of visual and motoric cues in the Kitabah mobile app enhance the learning experience of PHSD learners compared to traditional methods? This will be assessed through usability testing and learner feedback. (4) How does the performance of the deep learning-based SLR system compare to existing sign language recognition systems? This will be determined by benchmarking the model against other systems on relevant datasets, with a focus on accuracy and real-time recognition capabilities. These questions aim to ensure that the study's outcomes are both focused and quantifiable, providing clear insights into the effectiveness of the integrated mobile app and SLR system.

Accordingly, the objective of this research is to design and evaluate a deep learning–enhanced Kitabah application for inclusive and adaptive Quranic sign language education. The study seeks to operationalize Kitabah as the pedagogical backbone, integrate AI-based recognition to enable feedback-rich, personalized learning, and articulate an inclusive mobile design that can be credibly adopted by educators, learners, and stakeholders in Quranic education for PHSD.

2. Method

This study employs two distinct methodological approaches that are purposefully kept independent. The first develops a Kitabah-based mobile learning application and validates its software quality and usability. The second builds and evaluates a deep-learning sign language recognition (SLR) model aligned with Quranic letter representations. Each approach has its own objectives, materials, and evaluation procedures; any cross-use of components is intentionally deferred to later sections of the paper.

a. Kitabah-Based Mobile Learning Application

The pedagogical framework follows the Kitabah guideline issued by the Ministry of Religious Affairs, which links Arabic letters, diacritics, and punctuation to visual hand gestures that resemble the written Arabic script. The framework was further designed using principles of innovative learning design, multimodal learning, and technology-enhanced education to support meaningful, adaptive, and interactive learning experiences for PHSD learners. A structured waterfall process is employed to guide the development cycle, from needs analysis to maintenance (Al Fajar et al., 2022). The overall flow of this process is illustrated in Figure 1, providing a visual reference to help orient the reader before delving into further details.

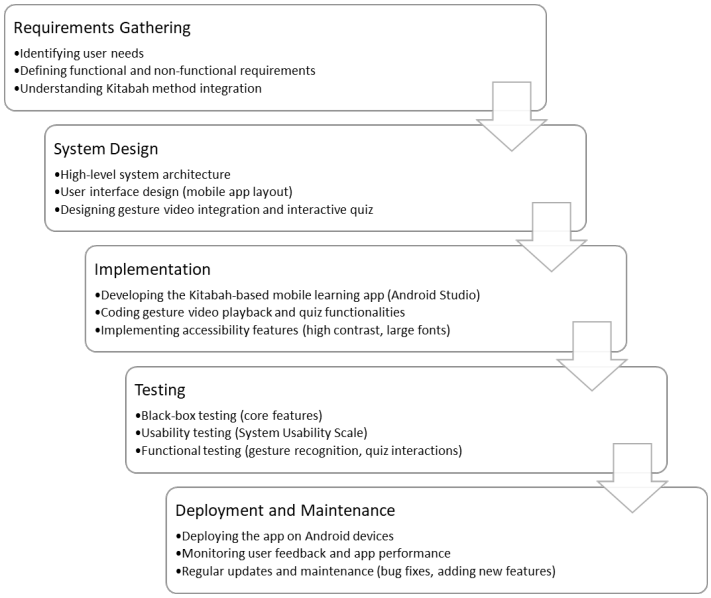

Figure 1. Waterfall Process for the Kitabah Application

During the needs analysis phase, the development team conducted a structured identification of pedagogical, technical, and accessibility requirements to ensure that the proposed application aligned with the learning characteristics of People with Hearing and Speech Disabilities (PHSD). The functional requirements were formulated around three core learner-oriented capabilities (Farouk et al., 2024). First, the application was designed to provide structured access to 32 Hijaiyyah learning entities based on the Kitabah method, enabling learners to recognize Quranic letters through visually grounded representations. Second, curated gesture-based instructional videos were integrated to strengthen multimodal learning experiences by combining visual observation with motoric imitation, thereby supporting deeper cognitive engagement and retention. Third, an interactive quiz mechanism was incorporated to facilitate formative assessment and immediate feedback, allowing learners to monitor their understanding and reinforce letter recognition progressively.

In addition to the functional aspects, several non-functional requirements were established to ensure usability, inclusivity, and technical reliability within a mobile learning environment. Particular emphasis was placed on visual accessibility through the implementation of high-contrast interfaces, large typography, and simplified layouts to accommodate diverse learner needs and reduce cognitive load. Sab & Biradar berpendapat bahwa Sab & Biradar (2023) argue that consistent navigation structures and intuitive interaction patterns were also prioritized to improve user orientation and minimize confusion during learning activities. Furthermore, the application was designed to maintain compatibility across Android devices to ensure broader accessibility and scalability, particularly in contexts where low- to mid-range smartphones remain dominant.

These requirement specifications served as the foundational basis for subsequent system design and implementation decisions, ensuring that pedagogical objectives, accessibility principles, and technical feasibility were systematically integrated throughout the development process. To improve methodological transparency, facilitate reproducibility, and strengthen traceability between identified needs and

implemented features, a concise summary of both the functional and non-functional requirements is presented in Table 1, entitled "Functional and Non-Functional Requirements for the Kitabah Application."

Table 1. Functional and Non-Functional Requirements for the Kitabah Application

Requirement Type	Requirement Description	Purpose
Functional	Display 32 Hijaiah letters according to the Kitabah method	Enables learners to visually recognize each letter
	Playback of gesture videos for each letter	Reinforces learning through visual and motoric cues
	Interactive quizzes to test knowledge	Provides formative feedback to learners
	Simple and intuitive navigation between letters, videos, and quizzes	Ensures learners can access content easily
Non-Functional	Visual accessibility: high contrast, large fonts, clear layout	Supports learners with visual or cognitive challenges
	Consistent interface across screens	Improves usability and reduces confusion
	Android device compatibility (Android 5.0+)	Ensures broad usability on commonly available devices
	Stable performance without crashes or lag	Provides reliable learning experience

The system design phase was conducted using Unified Modeling Language (UML) to establish a structured, intuitive, and accessible learning architecture for the Kitabah-based mobile application. UML modeling was employed not only to document the functional structure of the system but also to ensure that the interaction flow remained simple, consistent, and appropriate for People with Hearing and Speech Disabilities (PHSD) (Fadlilah et al., 2022b). By utilizing visual modeling techniques, the development process was able to systematically translate pedagogical requirements and accessibility considerations into a coherent application design framework.

The use-case diagram was developed as the primary representation of the interaction between users and the system. This diagram illustrates the complete learning journey, beginning from the main navigation interface and extending to the application's two central learning modules, namely *Learn Letters* and *Interactive Quiz*. The use-case model defines the boundaries of the system and clarifies how learners interact with each feature within the application environment (Kinanti & Maulana, 2024). Through this representation, the diagram provides a high-level overview of user behavior, helping readers understand the relationship between learning activities, system responses, and instructional objectives.

To complement the use-case perspective, activity diagrams were designed to describe the internal workflow and sequential logic of each major feature in greater detail. For the *Learn Letters* module, the activity flow includes character selection, search processes, retrieval of corresponding Kitabah representations, and playback of gesture-based instructional videos. This sequence was intentionally designed to support multimodal learning by combining textual, visual, and motoric learning cues within a single interaction cycle (Khwuta et al., 2023). Meanwhile, the *Interactive Quiz* module was modeled to illustrate the step-by-step assessment process, starting from the presentation of image-based stimuli, learner

response selection, system validation, immediate feedback delivery, and concluding with a performance recap. Such a workflow enables learners to engage in formative assessment activities that reinforce retention and self-directed learning.

Furthermore, the UML-based design process contributed to the implementation of accessibility-oriented interaction principles, including simplified navigation structures, reduced cognitive complexity, and consistent screen transitions. These considerations are particularly important in inclusive digital learning environments, where interface clarity and predictable interaction patterns significantly influence learner engagement and usability (Rumetna et al., 2022). By clearly visualizing both user interactions and process flows, the UML models also served as communication tools between the pedagogical and technical development stages, ensuring alignment between instructional goals and software implementation.

Overall, the system design framework provides a conceptual and technical foundation for the subsequent implementation phase, enabling the application to integrate educational functionality, accessibility, and usability within a unified mobile learning environment. The complete design representation, including the use-case and activity diagrams, is presented in Figure 2.

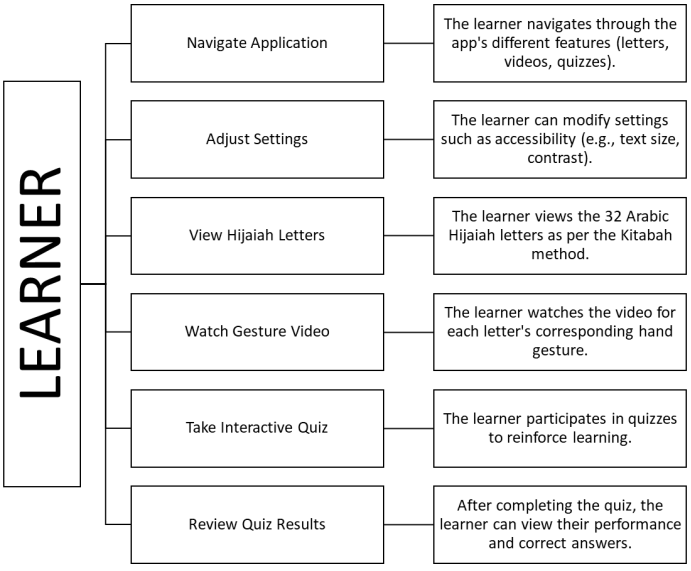

Figure 2. Use Case for the Kitabah Application

Implementation was conducted in Android Studio using Java. The Learn Letters module provides an on-screen Hijaiyyah keyboard, an input field, and a gesture-video player tied to Kitabah assets; the Quiz module presents image-based items with immediate feedback. Accessibility principles such as high contrast, large interaction targets, and single-task-per-screen rhythm were applied across screens to lower cognitive and motor load. These design decisions align with the principles of Universal Design for Learning (UDL), emphasizing progressive and inclusive educational practices that support flexible, accessible, and technology-mediated learning environments for diverse learners.

Software verification employed black-box testing across critical user flows, with scenarios and expected outcomes organized in a verification matrix placed at the outset of the testing subsection to foreground coverage before narrative discussion (see Table 2, “Black-

box test matrix for core flows"). Usability was assessed with the System Usability Scale (SUS) following standard procedures; to ensure transparency, the instrument, respondent profile, data-collection procedure, and scoring method are summarized before any results are reported elsewhere. Ethical and theological review accompanied content curation and interface decisions throughout this approach.

Table 2. Black-box Test Matrix for Key Functionality

Step	Action	Expected Result	Outcome
1	Open the app and check the main menu	The main screen appears, showing options for "Learn Letters" and "Interactive Quiz"	Passed
2	Choose "Learn Letters"	The app displays all 32 Hijaiyyah letters along with corresponding gesture videos	Passed
3	Watch the corresponding gesture video	The selected video plays without any interruptions	Passed
4	Participate in the interactive quiz	The quiz interface appears, and questions are shown for the user to answer	Passed
5	Select an answer in the quiz	After an answer is chosen, feedback is provided, and the quiz continues	Passed
6	Review the results after completing the quiz	Displays the user's total score, along with feedback on correct and incorrect answers	Passed
7	Modify accessibility settings	Opens the settings menu, where the user can adjust options such as font size and contrast	Passed

b. Deep-Learning Sign Language Recognition (SLR)

The second approach focuses on developing a baseline Sign Language Recognition (SLR) model for static Quranic letters, which serves as the foundation for future Quran-specific recognition tasks. While commonly used Arabic sign datasets primarily focus on static alphabetic handshapes, the Kitabah method incorporates additional movement-based signs; the current model only addresses static classes, with plans to extend to dynamic gestures in subsequent research. To provide a clearer understanding of the data, Figure 3 presents representative samples from the Arabic sign alphabet corpus along with the preprocessing pipeline, illustrating how the dataset is prepared for the model.

The dataset used in this study comprises 7,856 RGB images, each representing one of the 31 static hand gestures of the Arabic Sign Language Alphabet (ArSLA). These images serve as the foundation for training the deep learning model to accurately recognize Arabic sign language letters. Each image in the dataset corresponds to a specific letter of the Arabic alphabet, captured in different hand gestures to represent each letter in sign language. The dataset is diverse, covering a range of visual characteristics including variations in hand shapes, skin tones, and lighting conditions, which helps ensure the model can generalize well across different scenarios.

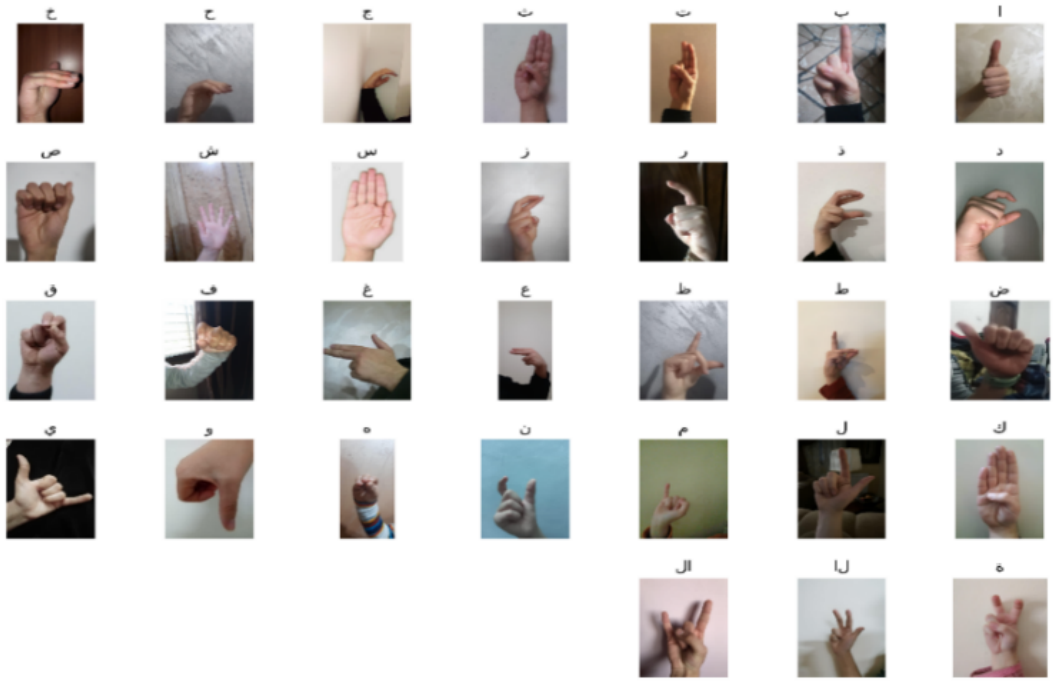

Figure 3. Sample Images from the RGB Arabic Alphabet Sign Language Dataset, Showing Variations in Hand Shapes and Background Configurations Across Different Selected Classes

Preprocessing of the dataset involved several essential steps to prepare the images for training the deep learning model (Al-Barham et al., 2022). Initially, the images were resized to a uniform 224×224 pixels to ensure consistent input size, which is crucial for maintaining data quality. After resizing, normalization was applied to scale the pixel values to a standardized range, ensuring that the model would not be biased by large variations in pixel intensities. Additionally, light data augmentation techniques, such as small rotations and horizontal flips, were performed on the dataset. These techniques were chosen to enhance the model's ability to generalize by introducing slight variations in the data, making the model more robust to different angles and orientations of the hand gestures, while maintaining the discriminability of the Hijaiah letters. This augmentation helped reduce the risk of overfitting and improved the model's ability to recognize letters in various real-world scenarios.

After preprocessing, the dataset was divided into three distinct subsets: training (70%), validation (15%), and test (15%). This stratified partitioning method was employed to ensure that each subset contained a representative proportion of each class, thereby preserving the dataset's balance (Al-Barham et al., 2023). Stratification was particularly important because it ensured that every subset whether for training, validation, or testing had a similar distribution of the 31 Arabic letters, preventing potential bias toward any particular letter during the model's training or evaluation phases. This careful data partitioning enabled robust training, ensuring the model learned effectively from a diverse set of examples. It also provided a reliable framework for validating and testing the model's performance on unseen data, ensuring consistency and minimizing the risk of overfitting. The balanced distribution across the three subsets provided a strong foundation for developing a model that generalizes well to unseen sign language gestures.

The model was trained using a transfer-learning approach, where the ResNet-18 architecture was initialized with weights from

ImageNet (Alnuaim et al., 2022). The final classification layer was modified to fit the number of static letter classes being recognized. The training process employed the Adam optimizer, starting with a cautious learning rate that was adjusted downwards when the validation performance plateaued. The batch size, the number of training epochs, and the early stopping criteria were carefully selected to ensure stable training. These parameters were selected to prevent overfitting and ensure smooth model convergence. To evaluate the model's performance, key metrics such as accuracy, precision, recall, F1-score, and a confusion matrix were used to assess the model's effectiveness in recognizing sign language gestures. The results for these metrics are discussed later, but they are not included here to maintain focus on the methodological details. To help readers visualize the overall workflow of the SLR model, a brief schematic was provided at the end of this section, outlining the entire training process from data input, through preprocessing and optimization, to model evaluation.

Ethical and inclusivity considerations were applied to dataset use and interpretation, with attention to the balance of representation and the theological implications of mapping letter forms to hand configurations intended for religious learning. Any prospective linkage to Kitabah movement-based categories is treated as a separate, subsequent stage and is not conflated with the static-class modeling reported in this methods strand.

3. Result and Discussion

The results and findings from the two approaches to the study are presented and discussed, focusing on the development of the Kitabah-based mobile learning application and the deep-learning-enhanced sign language recognition (SLR) system. The first approach aimed to create an accessible learning tool for PHSD learners to master Quranic literacy using visual gestures, while the second employed deep learning techniques to recognize static Hijaiah letters. The evaluation covers aspects such as system functionality, usability, model performance, and pedagogical effectiveness, as well as the challenges encountered and potential directions for future improvements in both the mobile application and SLR model to enhance Quranic education.

a. Mobile Learning Application Utilizing the Kitabah Method

The Kitabah-based mobile learning application was developed to support PHSD learners in acquiring Quranic literacy by using the Kitabah method's visual representation of Arabic letters. The application featured 32 Arabic letters and corresponding hand gestures, with each letter's gesture represented through videos. An interactive quiz allowed learners to reinforce their understanding through active participation. Functional testing was conducted to ensure the application worked as intended, with no major failures in features such as letter input, gesture-video playback, and quiz interactions. The black-box testing confirmed that all core features were functional (Khusnani et al., 2026). Figure 4 illustrates the developed application's interface, showing how the learning content and features are presented to users.

Figure 4. User Interface of the Kitabah-Based Mobile Learning Application

The application also underwent usability testing, using the System Usability Scale (SUS), which yielded a positive outcome. The app received an average SUS score of 78.06, indicating that it is user-friendly and intuitive for PHSD learners (Castilla et al., 2024). The results underscore the application's potential as an effective tool for teaching Quranic literacy to learners who rely on visual and motoric cues. From an educational perspective, the application promotes active engagement, self-directed learning, digital literacy development, and technology-supported participation, which are fundamental characteristics of advanced, inclusive, and progressive learning ecosystems. Despite these successes, the app is limited to teaching the basic Hijaiah letters and lacks the capability to accommodate more advanced features such as dynamic gestures for Tajweed and Quranic recitation rules (Almubarak et al., 2025). Further work should incorporate these dynamic elements to provide a more comprehensive learning experience.

In future iterations, the addition of auditory feedback could significantly enhance the app's accessibility, particularly for learners with additional sensory impairments, such as those who are both deaf and blind. Furthermore, while the Kitabah method is effective in teaching basic literacy, integrating dynamic gestures for recitation would enrich the user experience, providing a more complete learning system.

b. Deep Learning-Enhanced Sign Language Recognition for Quranic Literacy

In the second approach, a deep learning-based sign language recognition (SLR) system was developed to recognize static Arabic letters in the context of the Quran, using the Arabic Sign Language Alphabet (ArSLA) dataset. The model was built on the ResNet-18 architecture, chosen for its lightweight structure and strong performance in image classification tasks (Alsaadi et al., 2022). Transfer learning was applied by initializing the model with pre-trained weights from ImageNet and replacing the final classification layer with a fully connected layer tailored to classify 31 output classes. The model was trained for 30 epochs using the Adam optimizer with an initial learning rate of 0.001 and categorical cross-entropy as the loss function. To enhance convergence and generalization, a learning rate scheduler was employed that reduced the learning rate when the validation loss plateaued, improving performance during training.

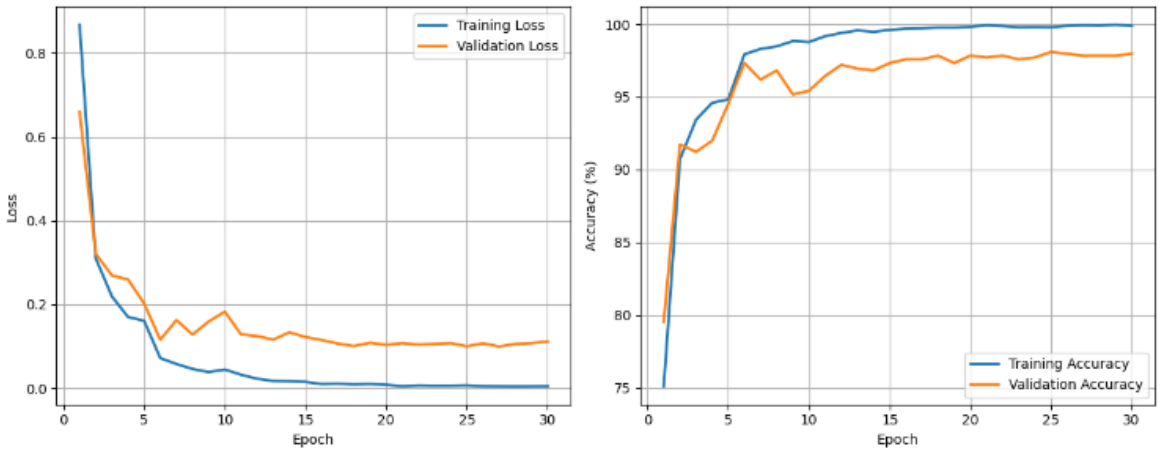

Figure 5. Training and Validation Performance of the ResNet-18 Model Over 30 Epochs

Figure 5 presents the training and validation performance of the ResNet-18 model across 30 training epochs, illustrating the effectiveness of the transfer-learning strategy in learning discriminative visual representations for Hijaiyyah sign language recognition. Both the training and validation loss curves exhibit a steep decline during the initial epochs, indicating that the model rapidly adapted to the dataset and successfully extracted meaningful visual features from the Arabic sign language images (Podder et al., 2023). This rapid reduction in loss demonstrates the suitability of the pretrained ResNet-18 architecture for the classification of Quranic sign language gestures, particularly in scenarios where the dataset size is moderate and efficient feature transfer is required.

A significant decrease in training loss was observed within the first six epochs, where the value fell below 0.08, while the validation loss simultaneously approached approximately 0.12. This convergence pattern suggests that the model achieved stable optimization early in the training process and was capable of generalizing effectively to unseen validation samples. Unlike models that exhibit unstable fluctuations or widening gaps between training and validation losses, the curves in Figure 5 remain closely aligned throughout the training phase. Such behavior indicates that the regularization strategy, data preprocessing pipeline, and augmentation techniques successfully minimized overfitting while preserving learning efficiency.

From epochs 11 to 30, the validation loss remained relatively stable within the range of 0.10–0.13, further confirming the robustness of the trained model. The stability of the validation curve implies that the model maintained consistent predictive capability even after prolonged training iterations, without experiencing substantial degradation in performance. This finding is particularly important in sign language recognition tasks, where slight variations in hand posture, lighting conditions, and background configurations may introduce classification complexity. The model's ability to sustain low validation loss across later epochs demonstrates strong resilience against such variations and indicates that the extracted features possess high discriminative power (Ajid et al., 2025).

The accuracy curves shown in Figure 5 further reinforce these findings. Training accuracy increased rapidly and surpassed 99% by epoch 11, eventually stabilizing near 99.9% at the end of the training process. Meanwhile, validation accuracy also

improved consistently, reaching 97.33% by epoch 6 and peaking at 98.09% around epoch 25 before experiencing only minor fluctuations. The close correspondence between training and validation accuracy suggests that the model did not merely memorize the training samples, but instead learned generalized representations that could be reliably applied to unseen data. This balanced learning behavior reflects the effectiveness of the transfer-learning framework and indicates that ResNet-18 is capable of achieving high recognition performance without excessive model complexity.

Moreover, the learning dynamics observed in Figure 5 highlight the practical suitability of the proposed model for mobile and educational deployment scenarios. The relatively fast convergence rate reduces computational training costs and enables efficient model optimization, which is advantageous for adaptive learning systems intended for real-world educational applications (Rifanka et al., 2025). In the context of Quranic sign language education, achieving stable and highly accurate recognition is particularly crucial because incorrect interpretation of Hijaiyyah gestures may affect orthographic understanding and theological accuracy. Therefore, the high convergence stability and strong generalization performance demonstrated by the model provide important evidence that deep learning can effectively support inclusive, AI-assisted Quranic literacy education for PHSD learners.

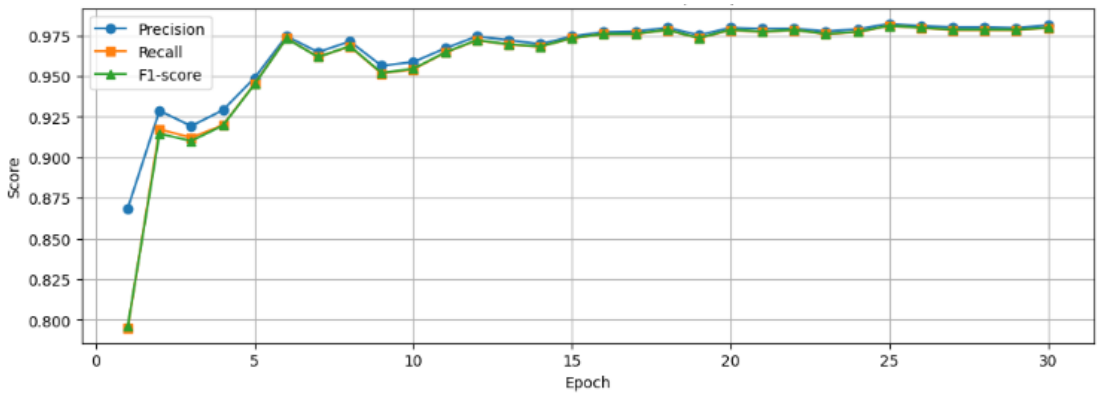

Figure 6. Classification Performance of the ResNet-18 Model on 31 Hijaiyyah Sign Language Classes

The ResNet-18 model demonstrated strong performance in classifying 31 types of Hijaiyyah sign language characters using validation data, achieving an overall accuracy of 98%. The model showed high precision, recall, and F1 Score, with all averages at 0.98. Most classes performed excellently, with perfect scores (1.00) across these metrics, showcasing the model's consistency in recognizing both commonly and rarely represented characters. Several classes, such as labels 2, 4, 6, 11, 12, 13, 19, 20, 21, 22, 26, and 27, exhibited perfect precision and recall, with no false positives or negatives. However, a few classes, including class 5 (precision: 0.82), class 16 (recall: 0.91), and class 25 (recall: 0.90), showed slightly lower scores, suggesting minor misclassifications possibly due to visual similarities between some signs or fewer training samples. Despite these small discrepancies, the F1-scores remained high, with the lowest at 0.90 and most above 0.95, which highlights the model's robustness. This high level of accuracy is crucial for

applications such as Quranic recitation, where precise recognition is essential to maintain theological and linguistic integrity. Figure 6 shows the detailed classification results.

The confusion matrix, Figure 7, shows the ResNet-18 model's classification performance across 31 classes of Arabic sign language characters, specifically designed for Quranic recitation. Most of the values in the matrix are concentrated along the diagonal, indicating that the model made correct predictions for the majority of classes, which aligns with the previously reported overall accuracy of 98%. Several classes, including Ṭā' (class 2), Jīm (class 3), Ṣād (class 11), and Nūn (class 19), achieved perfect classification with no misclassifications, demonstrating the model's ability to learn distinctive visual patterns. However, minor misclassifications occurred in a few instances, such as ʿAyn (class 16) being misclassified as Ghayn (class 17), and Dhāl (class 25) being misclassified as Ṭā' (class 26), suggesting some visual similarities in the hand gestures of adjacent letters. Additionally, a single misclassification occurred between Ḍād (class 10) and Yā' (class 30). These minor errors highlight the need for more refined feature representations and suggest that additional training data or motion-based cues could improve recognition for visually similar signs (Ajid et al., 2025). Despite these slight discrepancies, the confusion matrix overall reinforces the model's robustness and its suitability for Quranic sign language recognition tasks.

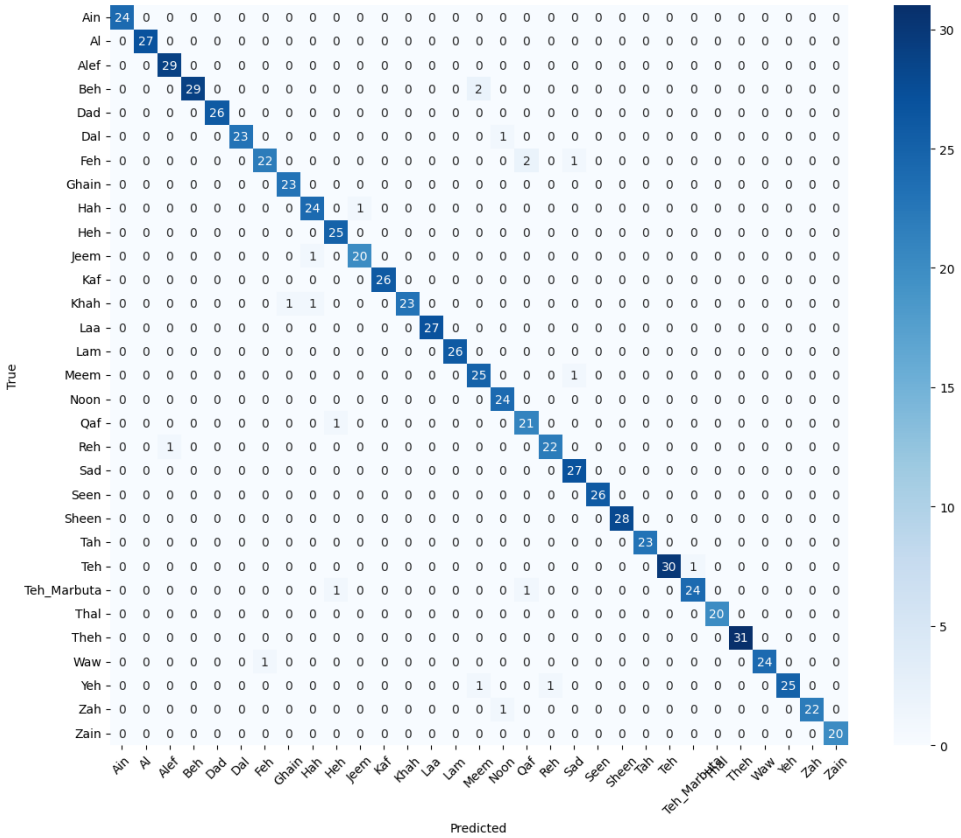

Figure 7. Confusion Matrix Showing Classification Performance of ResNet-18 Model

c. Comparative Performance with Other Models

To address the gap in comparing the proposed model's performance, we explicitly compare it with other well-established models in the Arabic Sign Language (ArSL) domain. This study benchmarks the performance of our ResNet-18 model against the VGG16 and

EfficientNetB0 models, both previously used to classify static Arabic Sign Language gestures. The VGG16 model achieved a test accuracy of 97.07%, comparable to ResNet-18's 98.0% (Rakhmadi et al., 2025b). While both models show strong performance, the VGG16 architecture, being simpler, is more suitable for environments with limited computational resources, making it an appealing choice for real-world applications in the deaf and hard-of-hearing community, where efficiency is crucial.

Furthermore, we compare the ResNet-18 model with the EfficientNetB0 model developed in a previous study. The EfficientNetB0 model achieved 99.2% training accuracy, 97.8% validation accuracy, and 95.2% test accuracy, with a macro-averaged F1-score of 0.95 (Rakhmadi, et al., 2025). This model demonstrated excellent performance, especially in environments with computational constraints. While the EfficientNetB0 model outperformed ResNet-18 in test accuracy, ResNet-18 offers a balance between accuracy and computational efficiency, providing faster convergence during training and greater adaptability for mobile applications (Setiawan et al., 2026). Thus, both models are highly suitable for use in educational contexts where efficiency and accuracy are equally important.

In comparing all three models (ResNet-18, VGG16, and EfficientNetB0), it is evident that each model has its strengths depending on the deployment environment. The VGG16 model, with its simpler architecture, performs well in resource-constrained scenarios, whereas EfficientNetB0 achieves the highest accuracy but requires more computational resources. The ResNet-18 model strikes a balance between both aspects, offering a robust solution that can be deployed in various settings (Nugraha et al., 2024). These comparisons highlight the trade-off between model complexity and resource efficiency, demonstrating that lightweight CNNs, while simpler, can achieve competitive performance with fewer resources, and reinforcing the flexibility of deep learning models for inclusive religious education.

d. Measurement of Learning Outcomes through SUS

While this study did not directly measure learning outcomes such as increased ability to recognize letters or long-term retention, the effectiveness of the Kitabah-based mobile learning application was evaluated using the System Usability Scale (SUS) (Almubarak et al., 2025). SUS is a widely used tool for assessing system usability, and in this study, it was employed to gauge learners' overall satisfaction with the application. The SUS score for the app was 78.06, indicating that the app is perceived as user-friendly and efficient. This score suggests that learners found the application easy to navigate, engaging, and effective for its intended purpose of teaching the Hijaiyyah letters.

Although SUS does not directly measure learning outcomes, the high usability score implies that the application likely contributed to a positive learning experience. Users' satisfaction and ease of use often correlate with increased motivation and engagement, both of which are essential for effective learning (Rifanka et al., 2025). Future studies could build upon this by incorporating more specific measures, such as pre- and post-test comparisons or tracking the retention of learned content over time. For now, the SUS results provide valuable insight into the app’s usability and user experience, offering a foundation for future improvements and more targeted measurements of learning outcomes.

e. Comparative Insights and Future Work

Both the Kitabah-based mobile app and the deep-learning-based SLR system have demonstrated that integrating traditional pedagogical methods with modern AI technologies can significantly enhance Quranic literacy among PHSD learners. The Kitabah app successfully facilitated visual learning of the Hijaiah letters, while the SLR system showed strong performance in static letter recognition.

The integration of Kitabah pedagogy with AI-powered SLR demonstrates the potential for using technology to enhance Quranic education for PHSD learners. This finding reinforces the growing role of learning technology, deep learning, and AI-assisted educational systems in creating scalable, adaptive, innovative, and inclusive learning ecosystems that support equitable access to religious and digital literacy education. However, while both approaches are promising, they are currently limited in scope. The Kitabah app's focus on basic letter learning needs to be expanded to include dynamic gestures for Tajweed. Addressing this challenge is essential for developing more comprehensive and context-aware educational technologies capable of supporting complex Quranic learning processes in inclusive educational settings.

Looking ahead, the next steps involve expanding both approaches to incorporate dynamic gestures. For the mobile app, integrating dynamic gesture recognition will be essential for teaching Quranic recitation beyond the basics. For the SLR system, incorporating a more comprehensive dataset that includes both static and dynamic gestures aligned with Quranic recitation rules will be crucial for achieving high performance and theological accuracy.

4. Conclusion

This study demonstrates that integrating the Kitabah method with deep learning-based sign language recognition (SLR) can significantly enhance Quranic literacy for people with hearing and speech disabilities (PHSD). The Kitabah-based mobile application successfully provides an accessible, visually oriented learning platform that aligns with the cultural and theological framework of Quranic education. Usability testing results indicate that the application is both intuitive and effective for teaching the Hijaiyyah letters, while functional testing confirms the reliability of its core features. By embedding visual–motor learning into a mobile platform, the study addresses a long-standing gap in accessible Quranic pedagogy. More importantly, the proposed approach contributes to the advancement of progressive, technology-enhanced, and inclusive education by integrating deep learning, adaptive mobile learning, and AI-assisted educational technology to support equitable participation and meaningful Quranic literacy development for PHSD learners.

In parallel, the deep learning-enhanced SLR model, developed using a ResNet-18 architecture with transfer learning, achieved a high classification accuracy of 98% on 31 static Hijaiyyah sign language characters. This high precision and recall across most classes underscores the model's robustness and its potential for real-world deployment in Quranic education contexts. Although some misclassifications occurred with visually similar signs, the results validate the feasibility of adapting state-of-the-art image classification techniques for Quranic-specific sign language recognition. Importantly, this approach not only preserves the orthographic and theological integrity of the script but also lays a strong technical foundation for incorporating more complex gesture recognition in the future.

Future work should focus on expanding both the mobile application and the SLR system to encompass dynamic gestures and

intelligent adaptive feedback mechanisms. Such developments may contribute toward the establishment of a more advanced AI-driven learning ecosystem that supports personalized, inclusive, and context-sensitive Quranic education. This will require the collection and annotation of an extended dataset that incorporates both static and movement-based signs grounded in the Kitabah framework. By bridging pedagogical authenticity with advanced AI capabilities, such developments have the potential to deliver a truly inclusive, adaptive, and contextually faithful digital learning solution for PHSD learners, thereby contributing to the broader goal of equitable access to Quranic education.

5. References

Agustin, R. R., Maulana, H., & Mandyartha, E. P. (2023). Detection of Actions BISINDO (Indonesian Sign Language) into Text-to-Speech using Long Short-Term Memory with MediaPipe Holistics. *Jurnal Teknik Informatika (Jutif)*, *5*(4), 1051–1061. https://doi.org/10.52436/1.jutif.2024.5.4.1492

Ajid, S. N., Kusumaningtyas, D. A., Ratih, K., & Lava, S. (2025). Strategies for Integrating Problem-Based Learning, Teaching Modules, and Formative Assessments to Enhance Learning Outcomes and Critical Thinking Skills. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, *7*(2), 218–232. https://doi.org/10.23917/ijolae.v7i2.8612

Al Fajar, M., Dar, M. H., & Rohani, R. (2022). Application of Waterfall model in development of family planning participants information system. *Sinkron*, *7*(2), 679–686. https://doi.org/10.33395/sinkron.v7i2.11387

Al-Barham, M., Alsharkawi, A., Al-Yaman, M., Al-Fetyani, M., Elnagar, A., SaAleek, A. A., & Al-Odat, M. (2023). *RGB Arabic Alphabets Sign Language Dataset*. http://arxiv.org/abs/2301.11932

Al-Barham, M., Sa'aleek, A. A., Al-Odat, M., Hamad, G., Al-Yaman, M., & Elnagar, A. (2022). Arabic Sign Language Recognition Using Deep Learning Models. *2022 13th International Conference on Information and Communication Systems, ICICS 2022*, 226–231. https://doi.org/10.1109/ICICS55353.2022.9811162

Ali, A., & Codina, G. (2024). Representations of Disability in Qur'anic Narratives. *Journal of Disability and Religion*. https://doi.org/10.1080/23312521.2024.2353603

Al-Khuraym, B. Y., & Ismail, M. M. Ben. (2022). Arabic Sign Language Recognition using Lightweight CNN-based Architecture. *International Journal of Advanced Computer Science and Applications*, *13*(4), 319–328. https://doi.org/10.14569/IJACSA.2022.0130438

Almubarak, A., Nahadi, N., Yuliani, G., Talib, C. A., & Arini, D. N. (2025). Steinbuch's Epistemological Model and the Reality of Particles: A Philosophical Perspective for Chemistry Education. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 37–51. https://doi.org/10.23917/ijolae.v7i1.23911

Alnuaim, A., Zakariah, M., Hatamleh, W. A., Tarazi, H., Tripathi, V., & Amoatey, E. T. (2022). Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/8777355

Alsaadi, Z., Alshamani, E., Alrehaili, M., Alrashdi, A. A. D., Albelwi, S., & Elfaki, A. O. (2022). A Real Time Arabic Sign Language Alphabets (ArSLA) Recognition Model Using Deep Learning Architecture. *Computers*, *11*(5). https://doi.org/10.3390/computers11050078

Alsharif, B., Altaher, A. S., Altaher, A., Ilyas, M., & Alalwany, E. (2023). Deep

Learning Technology to Recognize American Sign Language Alphabet. *Sensors*, *23*(18). https://doi.org/10.3390/s23187970

Arthur Limantara, M., & Tristianto, D. (2024). SIBI Alphabet Detection System Based on Convolutional Neural Network (CNN) Method as Learning Media. *Internet of Things and Artificial Intelligence Journal*, *4*(1), 143–161. https://doi.org/10.31763/iota.v4i1.716

Castilla, D., Jaen, I., Suso-Ribera, C., Garcia-Soriano, G., Zaragoza, I., Breton-Lopez, J., Mira, A., Diaz-Garcia, A., & Garcia-Palacios, A. (2024). Psychometric Properties of the Spanish Full and Short Forms of the System Usability Scale (SUS): Detecting the Effect of Negatively Worded Items. *International Journal of Human-Computer Interaction*, *40*(15), 4145–4151. https://doi.org/10.1080/10447318.2023.2209840

Darmawan, I. D. M. B. A., Linawati, L., Sukadarmika, G., Wirastuti, N. M. A. E. D., Pulungan, R., Mulyanto, & Hariyanti, N. K. D. (2023). Advancing Total Communication in SIBI: A Proposed Conceptual Framework for Sign Language Translation. *2023 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, 23–28. https://doi.org/10.1109/ICSGTEIS60500.2023.10424020

Dzulkifli, I. (2022). Quranic Education for Deaf Students in Malaysia; Implementation and Challenges. *Nadwa: Jurnal Pendidikan Islam*. https://doi.org/10.21580/nw.2022.16.1.10823

Fadlilah, U., Mahamad, A. K., & Handaga, B. (2021). The Development of Android for Indonesian Sign Language Using Tensorflow Lite and CNN: An Initial Study. *Journal of Physics: Conference Series*, *1858*(1). https://doi.org/10.1088/1742-6596/1858/1/012085

Fadlilah, U., Mahamad, A. K., Handaga, B., Saon, S., Ratih, K., Rahmawati, L. E., & Thamrin, H. (2022a). *Android Application Prototype of Basic BISINDO Introduction and Practice for General People*.

Fadlilah, U., Mahamad, A. K., Handaga, B., Saon, S., Ratih, K., Rahmawati, L. E., & Thamrin, H. (2022b, September). Android Application Prototype of Basic BISINDO Introduction and Practice for General People. *Proceeding of International Conference on Special Edcation in South East Asia Region (ICSAR 2022)*.

Fadlilah, U., Wismoyohadi, D., Mahamad, A. K., & Handaga, B. (2019). Bisindo information system as potential daily sign language learning. *AIP Conference Proceedings*, *2114*. https://doi.org/10.1063/1.5112492

Farouk, A. M., Zenhom, A. M., Abdelaleem, E. M., Fadel, S. A., Elsayed, K. K. A., Mohammed, M. S., Hassan, R. H., & Shedeed, H. A. (2024). A New Approach for Arabic Sign Language Recognition (ArSLR). *NILES 2024 - 6th Novel Intelligent and Leading Emerging Sciences Conference, Proceedings*, 545–550. https://doi.org/10.1109/NILES63360.2024.10753193

Hartanto, D., Fauziah, M., Moog, R. C., Rizal, Y., Diki Herdiansyah, & Hanum, N. C. (2025). Virtual Art Therapy for Adolescence Mental Health Education: Systematic Literature Review and Future Perspectives. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, *7*(2), 233–251. https://doi.org/10.23917/ijolae.v7i2.8200

Khwuta, Y. D. D. Y., Londa, M. A., & Wee, Y. A. (2023). Designing and building a business data collection application using the waterfall method. *MATRIX : Jurnal Manajemen Teknologi Dan Informatika*, *13*(1), 42–51. https://doi.org/10.31940/matrix.v13i1.42-51

Kinanti, A., & Maulana, D. (2024). Convolutional Neural Network Implementation in BISINDO Alphabet Sign Language Recognition System Using Flask.

International Journal of New Media Technology), *11*(1), 16.

Khusnani, A., Jufriansah, A., Wahab, D. S., Samana, F. R. A., Bahruddin, S. A., Anwar, Z., … Arifin, A. S. (2026). Temporal and Spatial Dynamics of Volcanic Aerosols: Absorbing Aerosol Index (AAI) Analysis During the Eruption of Mount Lewotobi Laki-laki. *Jurnal Penelitian Sains Teknologi*, *2*(1), 34–43. https://doi.org/10.23917/saintek.v2i1.15729

Meitantya, M. D., Sari, C. A., Rachmawanto, E. H., & Ali, R. R. (2024). VGG-16 Architecture on CNN for American Sign Language Classification. *Jurnal Teknik Informatika (Jutif)*, *5*(4), 1165–1171. https://doi.org/10.52436/1.jutif.2024.5.4.2160

Nugraha, M. F. A., Roemintoyo, R., Djono, D., & Al-Hakimi, H. (2024). Development of a Mobile Application for Occupational Health and Safety Education in Vocational High Schools: A Case Study in Construction and Housing Engineering. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 315–327. https://doi.org/10.23917/ijolae.v6i3.235921

Podder, K. K., Ezeddin, M., Chowdhury, M. E. H., Sumon, M. S. I., Tahir, A. M., Ayari, M. A., Dutta, P., Khandakar, A., Mahbub, Z. Bin, & Kadir, M. A. (2023). Signer-Independent Arabic Sign Language Recognition System Using Deep Learning Model. *Sensors*, *23*(16). https://doi.org/10.3390/s23167156

Putrada, A. G., Alamsyah, N., Fauzan, M. N., & Perdana, D. (2023). PCA-SVM for a Lightweight ASL Hand Gesture Image Recognition. *Proceedings of the International Conference on Electrical Engineering and Informatics*. https://doi.org/10.1109/ICEEI59426.2023.10346744

Rakhmadi, A., Khairunnisa, S., Adhantoro, M. S., Riyadi, S., Yudhana, A., & Sunardi. (2025). EfficientNetB0-Based Recognition of Arabic Sign Language for Quranic Recitation Support. *2025 12th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 137–143. https://doi.org/10.1109/EECSI67060.2025.11290662

Rakhmadi, A., Yudhana, A., & Sunardi. (2025a). A Study of Worldwide Patterns in Alphabet Sign Language Recognition Using Convolutional and Recurrent Neural Networks. *Jurnal Teknik Informatika (JUTIF)*, *6*(1), 187–204. https://doi.org/10.52436/1.jutif.2025.6.1.4202

Rakhmadi, A., Yudhana, A., & Sunardi, S. (2025b). VGG16-Based Feature Extraction for Arabic Alphabet Sign Language Classification to Support Qur'anic Tadarus Accessibility. *Jurnal Teknik Informatika (Jutif)*, *6*(4), 2602–2624. https://doi.org/10.52436/1.jutif.2025.6.4.4953

Rifanka, I. N., Panjaitan, R. G. P., Tenriawaru, A. B., & Li, L. (2025). Development and Validation of a Hemostatic Activity-Based Learning Video on the Cir-culatory System for High School Biology Education. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 69–87. https://doi.org/10.23917/ijolae.v7i1.23749

Rumetna, M. S., Lina, T. N., Rajagukguk, I. S., Pormes, F. S., & Santoso, A. B. (2022). Payroll Information System Design Using Waterfall Method. *International Journal of Advances in Data and Information Systems*, *3*(1). https://doi.org/10.25008/ijadis.v3i1.1227

Sab, I., & Biradar, A. (2023). ASL and Hand Gesture Emoji Translator Using Computer Vision. *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*. https://doi.org/10.1109/ICCCNT56998.2023.10308381

Setiawan, A., & Ulinnuha, M. (2026). Development and Feasibility Testing of Virtual Reality Based Learning Media for

Practicum in LAN and Fiber Optic Network Cabling Using the 4D Model. *Jurnal Penelitian Sains Teknologi*, *2*(1), 20–33. https://doi.org/10.23917/saintek.v2i1.12693