

A Hijaiyah Letters Sign Language Recognition Approach utilizing Deep Learning

Dadang Priyanto

Faculty of Engineering, Department of Computer Science, Universitas Bumigora, Indonesia
dadang.priyanto@universitasbumigora.ac.id (corresponding author)

Hairani

Faculty of Engineering, Department of Computer Science, Universitas Bumigora, Indonesia
hairani@universitasbumigora.ac.id

Received: 18 January 2025 | Revised: 14 February 2025, 24 February 2025, and 7 March 2025 | Accepted: 9 March 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.10285>

ABSTRACT

Sign language was created as a means of communication for individuals with hearing impairments. However, learning sign language is not easy and requires significant effort and dedication, posing challenges for deaf children. Mastering sign language, particularly the Arabic alphabet (Hijaiyah), is particularly demanding and necessitates specialized learning tools to enhance comprehension. The limited number of sign language teachers and the complexity of the learning process pose significant challenges at Sekolah Luar Biasa Negeri (SLBN) 1 Mataram. This study explores the use of a Logistic Regression (LR) algorithm to assist deaf students in learning Hijaiyah sign language. The dataset comprises 28 Hijaiyah letters. The process begins with data preprocessing to detect hands using hand detection landmarks, followed by the classification of the 28 Hijaiyah sign language gestures using LR. The study's results indicate that the proposed model had an accuracy rate of 96% in recognizing Hijaiyah sign language, demonstrating the algorithm's effectiveness for this application.

Keywords: sign language; logistic regression; Hijaiyah letters; deaf; deep learning

I. INTRODUCTION

Education is a fundamental right for all Indonesian citizens, as mandated by the Education Law of the Republic of Indonesia. According to the National Education System Law No. 20 of 2003, Article 13, Point A, every student has the right to receive religious education aligned with their faith, taught by educators of the same religion [1]. Ensuring access to religious education for children with disabilities is essential for their personal development [2,3].

One form of child disability in Indonesia is hearing impairment. Deaf children often attend religious activities without fully understanding the content or the message delivered by the speaker or lecturer. Similarly, learning the Qur'an is challenging due to their hearing limitations, particularly when teachers or tutors do not use sign language. The lack of tutors or teachers who master sign language is also a major problem. Therefore, effective learning tools are needed to support the teaching of Hijaiyah letters through sign language [4].

Several studies have explored Hijaiyah sign language recognition using various machine learning techniques, including deep learning, supervised and unsupervised learning, classification models, machine translation, and image recognition [5–9]. Authors in [10] developed an application to recognize the Indonesian Sign Language Alphabet (SIBI) using

the Convolutional Neural Network (CNN) method and produced a web-based application for learning SIBI. The dataset consisted of 416 images and achieved a training accuracy of 90.05%. Authors in [11] introduced SIBI sign language recognition using a CNN-based approach with 616 datasets, achieving 75% precision, 62.45% accuracy, and 58% recall. Authors in [12] implemented Hybrid CNN-CatBoost for Hijaiyah writing classification, achieving 96.07% accuracy. Authors in [13] utilized YOLOv5 for Hijaiyah letter recognition, achieving 95% accuracy on 1,014 images. Authors in [14] employed YOLOv6 for static and dynamic sign recognition, reporting 96% accuracy for training data and 92% for testing data. Authors in [15] used YOLOv5 for SIBI sign recognition by comparing 26 classes in testing and obtained an accuracy of 77%. Authors in [16] conducted sign language recognition with Learning Vector Quantization (LVQ), achieving 40% accuracy, while authors in [17] improved it to 88.75% using the same method. Authors in [18] developed a deep learning model for oral translation of Arabic into Hijaiyah letters, achieving 86.85% accuracy. Authors in [19] used the chain code method for SIBI letter signal feature detection, reaching 91% accuracy.

While deep learning models have demonstrated high accuracy in Hijaiyah letter recognition, their computational demands pose challenges for real-time applications. To address this, our study introduces a more computationally efficient

approach using MediaPipe and Logistic Regression (LR). MediaPipe is employed for hand landmark detection, while LR performs classification of Hijaiyah letters.

This study utilizes a dataset of 7,658 images labeled with 28 Hijaiyah letters. LR was chosen for its simplicity, interpretability, and efficiency in classification tasks, especially with relatively small datasets. Unlike deep learning models such as CNNs, LSTMs, and transformer-based architectures, which require extensive computational resources and large datasets, LR strikes a balance between accuracy and computational efficiency [20]. By introducing sign language recognition for Hijaiyah letters, this study aims to facilitate effective learning for deaf children, particularly in environments where qualified sign language teachers are scarce.

II. RESEARCH METHODS

This study follows several key stages: data collection, preprocessing, model development, and performance evaluation. The model is built using the Logistic Regression (LR) method, with hand landmark detection values serving as classification features.

A. Logistic Regression (LR)

LR is a classification algorithm that finds the relationship between discrete or continuous features (input) and the probability of certain discrete output results. [4]. It is widely used in machine learning, closely related to linear regression, but specifically designed for classification rather than numerical predictions. Equations (1-2) are the formula for the LR method. Equation (3) represents the sigmoid function formula.

$$y = \text{sigmoid}(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n) \quad (1)$$

$$g(X) = \text{sigmoid}(\alpha + \beta X) \quad (2)$$

$$y = \text{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

where y is the response variable (target variable), x_1, x_2, \dots, x_n are the explanatory variables (feature variables), α is the intercept (point of intersection of straight lines), β is the slope (the degree of slope of the line), $g(X)$ is the representation of a set of features, and the sigmoid function is a mathematical function that produces an output in the range 0-1. Exp the exponential number.

B. Landmark Value Detection

The process of detecting or retrieving landmark values can be seen in Figure 1. The flow of the model development process starts with preparing a dataset in the form of video and images of Hijaiyah letter signs. Then it is carried out by preprocessing the data, detecting hand detection landmark values, and continuing with retrieving the landmarks in the form of data in numerical format (coordinate values) from landmarks that serve as features for the classification process.

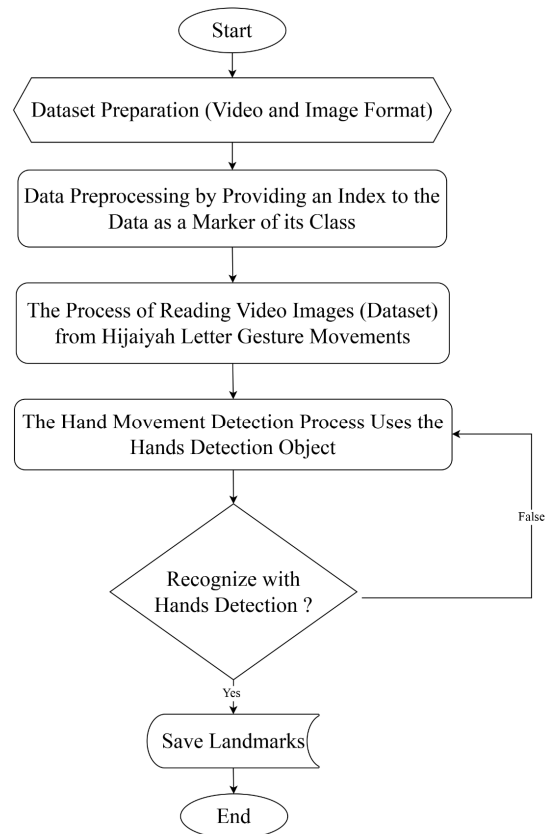


Fig. 1. Landmark hand detection.

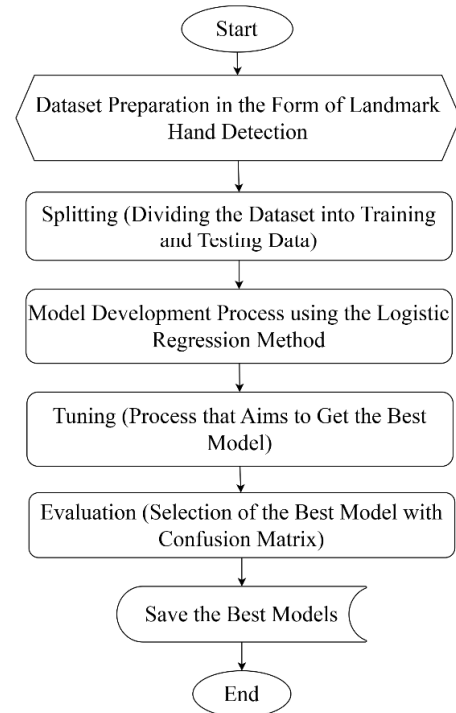


Fig. 2. Model development methods.

C. Model Development

Figure 2 illustrates the model development process, which consists of several stages. The first stage involves loading the dataset, which contains landmark hand detection values. The dataset is then split into training (70%) and testing (30%) subsets. Next, a classification model is developed using the Logistic Regression (LR) algorithm. The goal is to optimize model performance through hyperparameter tuning, ensuring the most effective classification model is obtained. Finally, the model undergoes an evaluation phase, where its performance is assessed using a confusion matrix.

D. Performance Evaluation

The confusion matrix is a fundamental evaluation tool for assessing classification models in both binary and multi-class classification tasks. It compares predicted labels with actual labels in a structured table, consisting of four key components:

- True Positive (TP): Correctly predicted positive instances.
- True Negative (TN): Correctly predicted negative instances.
- False Positive (FP): Incorrectly predicted positive instances.
- False Negative (FN): Incorrectly predicted negative instances.

To evaluate the effectiveness of the LR-based Hijaiyah sign language recognition model, this study employs accuracy as a primary performance metric. The accuracy is computed using (4) [21]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

III. RESULTS AND DISCUSSION

A. Data Processing

This research uses several datasets in video format, processed by applying the LR algorithm to classify Hijaiyah letters in sign language. The dataset was generated through a landmark hand detection process, which extracts landmarks from video data. The dataset comprises 7,658 instances, split into 70% training data and 30% testing data. Data collection was conducted by recording seven students from Sekolah Luar Biasa Negeri (SLBN) 1 Mataram. The recorded videos were then converted into image frames, and the dataset distribution can be seen in Table I. At this stage, the dataset, comprising landmarks and labels, is imported. The video data is stored in a structured folder system, categorized into 28 classes, corresponding to the 28 Hijaiyah letters. A filtering process was applied to remove duplicate or redundant instances. Each dataset entry was assigned an index or marker to facilitate the classification process.

B. Landmark Hand Detection

After importing the video dataset, the next step involves extracting landmark hand detection data from Hijaiyah sign language movements, covering letters from "Alif" to "Zay", followed by machine learning classification. The extracted landmark coordinates serve as features in the classification process. The detailed detection process is as follows: i) hand landmarks are detected using the MediaPipe library, ii) each

detected hand provides landmark coordinates (x,y,z). A looping method retrieves these coordinates for movements corresponding to the Hijaiyah letters "Alif" to "Zay", iii) the extracted landmark values are stored in numerical format; iv) The landmark data is classified using the Logistic Regression (LR) method with default hyperparameters, which can be seen in Table II.

TABLE I. DATASET OF HIJAIYAH LETTERS

No	Label / Class	Amount of Data
1	Alif	368
2	Ayn	289
3	Ba	300
4	Dal	268
5	Dhad	254
6	Dzal	295
7	Dzha	283
8	Fa	280
9	Ghayn	254
10	Ha	274
11	Ha kecil	246
12	Jim	295
13	Kaf	247
14	Kha	270
15	Lam	240
16	Mim	253
17	Nun	269
18	Qaf	270
19	Ra	263
20	Shad	266
21	Sin	262
22	Syin	264
23	Ta	297
24	Tha	300
25	Tsa	309
26	Waw	229
27	Ya	250
28	Zay	262
Total Datasets		7658

TABLE II. HYPERPARAMETER OF LR

Hyperparameter	Settings
Splitting	Train: 70%, Test: 30%
tol	0.0001
C	1
solver	lbfgs
Max_iter	100

Next, the dataset is split into training and testing sets using the *train_test_split* function. To evaluate system performance, 30% of the data (X_{test} , y_{test}) is used for testing ($test_size=0.3$), while the remaining 70% (X_{train} , y_{train}) is used for training. Out of 7,658 total samples, 2,298 (30%) are assigned as test data via Stratified Shuffle Split, while the remaining 5,360 samples undergo a 5-fold cross-validation process:

- Training data: 4,288 samples (across 5 scenarios).
- Validation data: 1,072 samples (across 5 scenarios).
- Test data = 2,298 data (1 scenario split from the beginning).

The dataset has dimensions (7,658, 63), which represent $(n_row, n_col) = (n_data, n_features)$. The 63 features derive from 21 landmarks per frame consisting of 3 coordinates (x, y, z), resulting in the feature set: $x_1, y_1, z_1, x_2, y_2, \dots, x_{21}, y_{21}, z_{21}$.

C. Object Motion Prediction and Evaluation with Confusion Matrix

Object detection is used to identify an object or its boundaries within an image. Object detection can be divided into soft detection and hard detection. Soft detection only detects the presence of an object, while hard detection detects the presence and location of an object. The object detection process typically involves searching different parts of the image to localize photometric or geometric features corresponding to the target object in the training database. This can be done by scanning the object template at different locations in the image, scales, and rotations, and the detection is declared if the similarity between the template and the image exceeds a certain threshold. The similarity between the template and the image area can be measured by correlation. In this study, object detection is used to track the movement of Hijaiyah sign language using a webcam, as shown in Figure 3. Figure 3 also shows an example of the results of the program testing process carried out by students at the SLBN 1 Mataram. The results were able to detect and predict hand movements by displaying landmark coordinate output.

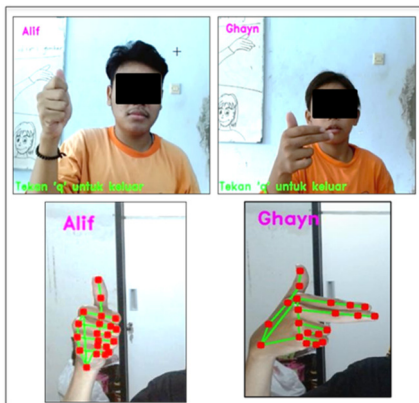


Fig. 3. Program testing on SLBN 1 Mataram students and determining landmark hand detection coordinates.

Figures 4 and 5 present the model's performance, based on the confusion matrices. In Figure 4, the confusion matrix for the LR model shows that the model data training process achieved a score of 0.981 (98%), indicating excellent performance in classification and prediction. In Figure 5, the confusion matrix for the LR model reveals a prediction test score of 0.958 or 96% on the testing data, demonstrating strong performance after model development, tuning, and evaluation.

Figure 5 also details the recognition performance of the LR method for Hijaiyah sign language. The accuracy of each class is as follows:

- Alif: 109 correct, 1 misclassified.
- Ayn: 80 correct, 7 misclassified.

- Ba: 89 correct, 0 misclassified (100% accuracy).
- Dal: 76 correct, 4 misclassified.
- Dhad: 76 correct, 0 misclassified (100% accuracy).
- Dzal: 86 correct, 3 misclassified.
- Dzha: 78 correct, 7 misclassified.
- Fa: 71 correct, 7 misclassified.
- Ghayn: 75 correct, 1 misclassified.
- Ha: 75 correct, 7 misclassified.
- Ha_kecil: 64 correct, 10 misclassified.
- Jim: 88 correct, 0 misclassified (100% accuracy).
- Kaf: 74 correct, 0 misclassified (100% accuracy).
- Kha: 78 correct, 0 misclassified (100% accuracy).
- Lam: 70 correct, 0 misclassified (100% accuracy).
- Mim: 76 correct, 0 misclassified (100% accuracy).
- Nun: 79 correct, 2 misclassified.
- Qaf: 71 correct, 9 misclassified.
- Ra: 77 correct, 2 misclassified.
- Shad: 77 correct, 3 misclassified.
- Sin: 75 correct, 4 misclassified.
- Syin: 77 correct, 2 misclassified.
- Ta: 89 correct, 0 misclassified (100% accuracy).
- Tha: 85 correct, 5 misclassified.
- Tsa: 91 correct, 2 misclassified.
- Waw: 62 correct, 7 misclassified.
- Ya: 75 correct, 0 misclassified (100% accuracy).
- Zay: 78 correct, 0 misclassified (100% accuracy).

A comparative test was also conducted using several methods or algorithm models shown in Table III. The comparative test shows that the LR model displays better prediction accuracy (96%) results than the other models.

TABLE III. COMPARATIVE MODEL TEST RESULTS

Model	Accuracy Test Results
Proposed Method	96%
CNN [11]	62.45%
YOLOv5 [13]	95%
Adaptive LVQ [17]	88.75%
YOLOv6 [14]	92%
VGGD + AGUM [22]	93.05%

This high accuracy is likely due to the relatively small dataset and the model's efficient ability to learn decision boundaries based on the feature extraction from MediaPipe. In contrast, deep learning models like CNN and YOLOv5 require

larger datasets to generalize effectively. Furthermore, LR's computational efficiency makes it well-suited for real-time sign language recognition on low-power devices, unlike deep learning models, which demand extensive training time and GPU resources.

IV. CONCLUSION

This study proposes a model that uses MediaPipe and Logistic Regression (LR) for the recognition of Hijaiyah sign language letters. MediaPipe detects hand movement landmarks, and the resulting features are extracted as numeric data. The numeric data is then classified using LR to recognize Hijaiyah letters. The results demonstrate that the combination of MediaPipe and LR achieves an accuracy of 96% in the recognition of Hijaiyah sign language letters. This high

accuracy suggests that using MediaPipe and LR together is suitable for real-time detection of Hijaiyah sign language letters. The proposed method outperforms state-of-the-art approaches, showing a 1% accuracy improvement over YOLOv5, a 4% improvement over YOLOv6, and a 3% improvement over VGGD + AGUM. This contribution advances Hijaiyah sign language detection, particularly for aiding deaf children to learn sign language. However, the study has limitations: it does not assess computation time, real-time detection accuracy, the impact of noise under varying lighting conditions, or the issue of unbalanced data, which affects the generalization of accuracy across classes. These areas should be addressed in future research. Additionally, future studies could conduct hypothesis testing using the t-test to evaluate the statistical significance of the results.

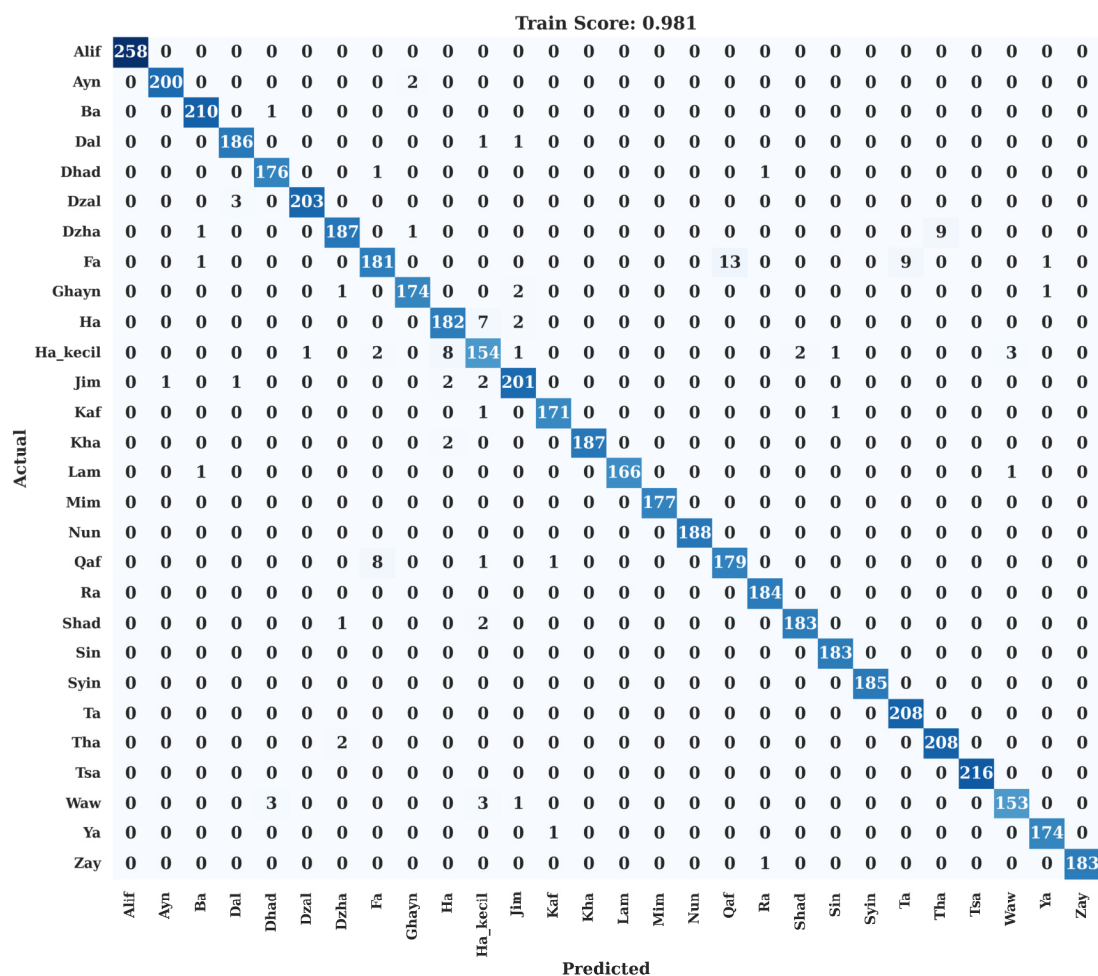


Fig. 4. Confusion matrix of data training.

- [10] M. Sholawati, K. Auliasari, and Fx. Ariwibisono, "Pengembangan Aplikasi Pengenalan Bahasa Isyarat Abjad Sibi Menggunakan Metode Convolutional Neural Network (CNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 1, pp. 134–144, Mar. 2022, <https://doi.org/10.36040/jati.v6i1.4507>.
- [11] Y. A. Gerhana *et al.*, "Automatic Detection of Hijaiyah Letters Pronunciation using Convolutional Neural Network Algorithm," *Jurnal Online Informatika*, vol. 7, no. 1, pp. 123–131, Jun. 2022, <https://doi.org/10.15575/join.v7i1.882>.
- [12] D. Mulya, "Classification of Hijaiyah Letters Using Hybrid CNN-CatBoost," *Intelmatika*, vol. 3, no. 2, pp. 39–44, Aug. 2023, <https://doi.org/10.25105/itm.v3i2.17521>.
- [13] D. I. Mulyana, M. F. Lazuardi, and M. B. Yel, "Deteksi Bahasa Isyarat Dalam Pengenalan Huruf Hijaiyah Dengan Metode YOLOV5," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 2, pp. 145–151, Aug. 2022, <https://doi.org/10.32528/elkom.v4i2.8145>.
- [14] A. M. Buttar, U. Ahmad, A. H. Gumaedi, A. Assiri, M. A. Akbar, and B. F. Alkhamees, "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs," *Mathematics*, vol. 11, no. 17, Aug. 2023, Art. no. 3729, <https://doi.org/10.3390/math11173729>.
- [15] D. Permana and J. Sutopo, "Aplikasi Pengenalan Abjad Sistem Isyarat Bahasa Indonesia (sibi) Dengan Algoritma YOLOv5," *Jurnal Simantec*, vol. 11, no. 2, pp. 231–240, Jul. 2023, <https://doi.org/10.21107/simantec.v11i2.19783>.
- [16] A. A. P. Arnaz, "Implementasi Metode Learning Vector Quantization (lvq) Pada Pengenalan Bahasa Isyarat Yang Mengandung Kata Kerja," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 3, no. 1, pp. 1–8, Jun. 2022, <https://doi.org/10.46510/jami.v3i1.40>.
- [17] A. R. T. H. Ririd, Y. Yunhasnawa, and Y. G. Buata, "Sistem Pengenalan Huruf Bahasa Isyarat Menggunakan Adaptive Learning Vector Quantization," *Jurnal Informatika Polinema*, vol. 4, no. 2, Feb. 2018, Art. no. 145, <https://doi.org/10.33795/jip.v4i2.163>.
- [18] M. Almahmood *et al.*, "Translation from spoken Arabic digits to sign language based on deep learning approach," *Applied and Computational Engineering*, vol. 30, no. 1, pp. 25–31, Jan. 2024, <https://doi.org/10.54254/2755-2721/30/20230064>.
- [19] A. Nur and H. Nugroho, "Deteksi Fitur Huruf Sistem Isyarat Bahasa Indonesia menggunakan Metode Chain Code," *POSITIF: Jurnal Sistem dan Teknologi Informasi*, vol. 8, no. 1, pp. 36–40, Nov. 2022, <https://doi.org/10.31961/positif.v8i1.1133>.
- [20] T. Ayer, J. Chhatwal, O. Alagoz, C. E. Kahn, R. W. Woods, and E. S. Burnside, "Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation," *RadioGraphics*, vol. 30, no. 1, pp. 13–22, Jan. 2010, <https://doi.org/10.1148/rg.301095057>.
- [21] F. T. Kristanti, M. Y. Febrianta, D. F. Salim, H. A. Riyadh, and B. A. H. Beshr, "Predicting Financial Distress in Indonesian Companies using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17644–17649, Dec. 2024, <https://doi.org/10.48084/etasr.8520>.

AUTHORS PROFILE

Dadang Priyanto received a Dr (Doctorate) degree from the University of North Sumatra Indonesia in 2021 in the field of computer science. Currently active as a lecturer and researcher in the Master of Computer Science study program at Bumigora University, Mataram City, Indonesia. Born in 1974 and has been a researcher in the field of computers and lecturers for about 22 years. The focus of current research is related to the fields of intelligent systems and expert systems.

Hairani obtained his Master's degree from Universitas Gadjah Mada Indonesia in 2017 in the field of computers. He is currently active as a lecturer in the Bachelor of Computer Science study program at Bumigora University, Mataram City, Indonesia. Born in 1992 and has been a lecturer in the field of computer science for about 6 years. The current research focus is related to the field of intelligent systems