Jurnal Galaksi (Global Knowledge, Artificial Inteligent and Information System) E-ISSN: 3048-2399
Vol 3, No 2, August 2026 , pp. 59-70
https://doi.org/10.70103/galaksi.v3i2.132

# Comparison of Grayscale and Binary Image Representations for Balinese Script Classification Using EfficientNetV2

I Made Dwi Putra Asana[1*], I Komang Anom Widya Pratama[2], Putu Surya Wedra Lesmana[3], Made Leo Radhitya[4]

[1*,2,3,4]*Department of Informatics, Faculty of Technology and Informatics, Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia*

[*]*dwiputraasana@instiki.ac.id*

ARTICLE INFO

*Article history:*
Received 16 June 2026
Revised 20 July 2026
Accepted 17 August 2026
Available Online 31 August 2026

***Keywords***:
Balinese Script;
Efficientnetv2;
Transfer Learning;
Image Classification;
Image Representation

**ABSTRACT**

*Balinese script is a cultural heritage whose preservation increasingly depends on digital technology, yet automatic recognition of its characters remains difficult because many glyphs share similar strokes. This study compares two image representations, grayscale and binary, as input for classifying 28 classes of Balinese script consisting of eighteen basic characters (wreastra), six pengangge suara and four pengangge tengenan. A dataset of 1,337 images was resized to 224x224 pixels and augmented into 8,022 images, then classified using EfficientNetV2B0 with transfer learning, a batch size of 32, a learning rate of 0.001, a dropout rate of 0.3 and 30 epochs followed by fine tuning of the last thirty layers. Robustness was also examined through three hold-out splitting ratios of 80:10:10, 70:15:15 and 60:20:20, which produced accuracies of 90.55%, 89.55% and 88.81%. The grayscale representation reached a test accuracy of 91.92% with a macro F1-score of 92.11%, slightly higher than the binary representation with 91.67% and 91.78%. Deployment on new handwritten images revealed a domain gap that lowered recognition performance, indicating that data diversity is more decisive than the choice of representation.*



## 1. Introduction

Balinese script is a cultural heritage of considerable historical and linguistic value. It is still found on palm-leaf manuscripts, inscriptions, public signage, and in the regional language curriculum of schools in Bali. Its everyday use, however, has narrowed considerably, and younger generations increasingly read and write the script with difficulty. Digitising Balinese script is therefore no longer a matter of convenience alone: it supports the preservation of historical documents and widens access to technology-assisted learning.

Optical character recognition (OCR) is one of the technologies that makes such digitisation possible. While Latin script is supported by a mature ecosystem of OCR libraries, regional scripts remain comparatively neglected. Balinese script poses particular difficulties: its glyphs are dominated by curved strokes, stroke thickness varies with writing style, and several characters differ only in small details. The pengangge suara taleng and the pengangge tengenan adeg-adeg, for instance, are visually close enough that even human readers rely on context to separate them.

Convolutional neural networks (CNNs) address this problem by learning hierarchical visual features directly from data, removing the need for hand-crafted feature engineering (Archana & Jeevaraj, 2024). EfficientNetV2 is a recent CNN family that combines Fused-MBConv and MBConv blocks with compound scaling, achieving faster training and better parameter efficiency than earlier architectures (Tan & Le, 2021). It has performed well on the Baybayin script (Campit & Fontillas, 2023) and on Balinese mask image classification (Yuniari & Kumara, 2024).

Most work on regional script recognition, however, concentrates on the choice of architecture and treats pre-processing as a fixed procedure whose influence is rarely tested. This is a meaningful omission for script images, because the decision to binarise an image or to retain it in grayscale determines what visual information survives into training. Binarisation sharpens the underlying shape of a character and removes the influence of illumination, but it simultaneously discards stroke thickness and edge gradation, precisely the cues that may separate characters of similar shape.

This study therefore compares grayscale and binary image representations as input to an EfficientNetV2 model for the classification of 28 Balinese script classes. Two further questions are addressed: whether the model remains stable when the proportion of training data is reduced, and how it behaves when confronted with newly written handwritten images that lie outside the training distribution. The findings are intended as practical guidance for pre-processing decisions in the development of OCR systems for Balinese script.

Based on these gaps, this study investigates the effect of grayscale and binary image representations on Balinese script classification using EfficientNetV2B0 under a controlled experimental setting. The novelty of this study lies not in proposing a new classification architecture, but in systematically examining the contribution of input image representation and extending model evaluation to independently written handwritten characters. The study makes three specific contributions. First, it provides an empirical comparison between grayscale and binary representations for the same EfficientNetV2B0-based classification framework across multiple data-splitting scenarios. Second, it emphasizes a leakage-free experimental procedure by separating the original images into training, validation, and testing subsets before applying augmentation to the training data. Third, it complements conventional test-set evaluation with an independent handwritten-image evaluation and character-level error analysis to identify visually similar character classes that remain difficult to distinguish. These contributions provide a more comprehensive assessment of both classification performance and practical generalization in Balinese script recognition.

## 2. Literature Review

Research on Balinese script recognition has followed several lines. Sidqy Alfarisi and Subandi (2022) combined direction feature extraction with a K-nearest neighbour classifier, while Pradika and Rahning Putri (2023) applied zoning and direction features with backpropagation. Both rely on manually designed features, which makes their performance sensitive to variation in writing style. Indrawan and Asroni (2022) adopted the Tesseract framework for mobile recognition of Balinese script; being template-based, this approach adapts poorly to variation in character shape.

Deep learning has since become the dominant approach. Mustafa and Nur (2024) reported 91.6% accuracy for Lontara script using a conventional CNN, and Nurahdika and Sutopo (2025) analysed CNN performance on handwritten Lontara characters. Abdiansah and Sumarno (2025) applied CNNs to handwritten Javanese script, and Prasetiadi and Saputra (2023) surveyed deep learning approaches to OCR for Nusantara scripts more broadly. For Balinese script specifically, Mas Diyasa and Wijaya (2025) combined a CNN with an extreme learning machine for handwriting recognition.

Modern CNN architectures have produced stronger results still. Campit and Fontillas (2023) evaluated several EfficientNetV2 variants on Baybayin characters and reached a peak validation accuracy of 95.85% after hyperparameter tuning. Yuniari and Kumara (2024) compared EfficientNetV2 with Xception for Balinese mask classification and obtained 99% against 97%. Table 1 summarises these studies.

Table 1. Summary of related work

| Author | Object | Method | Result |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Sidqy Alfarisi & Subandi (2022) | Balinese script | Direction feature + KNN | Dependent on manual features |
| Indrawan & Asroni (2022) | Balinese script | Tesseract (mobile) | Limited to template patterns |
| Pradika & Rahning Putri (2023) | Balinese script | Zoning, direction, backpropagation | Sensitive to writing variation |
| Campit & Fontillas (2023) | Baybayin script | EfficientNetV2 | Accuracy 95.85% |
| Mustafa & Nur (2024) | Lontara script | Conventional CNN | Accuracy 91.6% |
| Yuniari & Kumara (2024) | Balinese masks | EfficientNetV2 vs Xception | Accuracy 99% vs 97% |
| Mas Diyasa & Wijaya (2025) | Balinese script | CNN + ELM | Handwriting recognition |
| Yang & Zuo (2024) | Document images | Review of binarisation | Information loss at character edges |

Two gaps emerge from this body of work. First, existing studies compare architectures or classification methods, whereas the effect of the input image representation on Balinese script recognition has not been examined in its own right, even though Yang and Zuo (2024) note that binarisation always carries a risk of information loss at character edges. Second, evaluation typically stops at an internal test set drawn from the same distribution as the training data, so the behaviour of these models on newly written characters remains largely undocumented. The present study addresses both gaps by comparing grayscale and binary representations under an EfficientNetV2 architecture and by examining model behaviour on data outside the training distribution.

## 3. Research Method

This study follows a quantitative experimental design comprising five stages: data collection, pre-processing, data splitting, model training, and evaluation, followed by a deployment test on unseen handwritten images. The complete workflow is shown in Fig. 1. All experiments were implemented in Python using TensorFlow, Keras, and OpenCV (Castro & Bruneau, 2024; Sinha & Kumar, 2025).

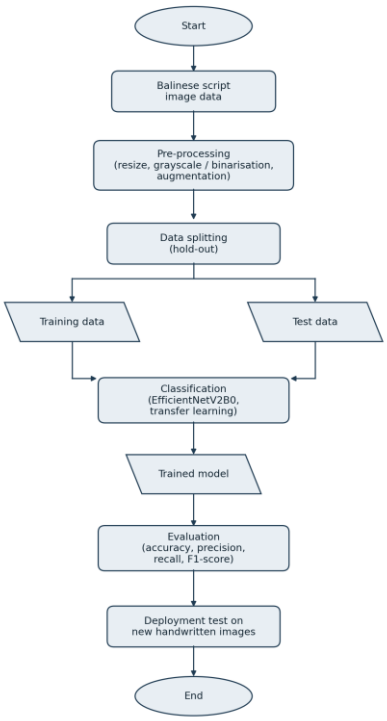


Fig. 1. Proposed method

### 3.1 Research dataset

The data consist of digital images of single Balinese characters obtained from a publicly available dataset on the Kaggle platform. Only the classes relevant to this study were retained, yielding 28 classes: 18 basic characters (wreastra), 6 pengangge suara, and 4 pengangge tengenan, for a total of 1,337 images. The number of images per class ranges from 31 to 58, so the dataset is imbalanced. The distribution is presented in Table 2.

Table 2. Distribution of the research data

| Character group | Number of classes | Number of images |
|---|---|---|
| Basic characters (wreastra) | 18 | 845 |
| Pengangge suara | 6 | 296 |
| Pengangge tengenan | 4 | 196 |
| Total | 28 | 1,337 |

### 3.2 Pre-processing and data augmentation

All images were resized to 224 x 224 pixels using bilinear interpolation, matching the input dimension of EfficientNetV2 while preserving the shape and contour of each character; the choice of resizing method is known to affect downstream recognition performance (Talebi & Milanfar, 2021). Each image was then converted into the two representations under comparison. The first is a grayscale image, which retains pixel intensity information. The second is a binary image, in which only two pixel values remain, so that the principal shape of the character is emphasised and the influence of illumination and background is suppressed.

To address class imbalance and the limited volume of data, augmentation was applied in the form of rotation, translation, and zooming, increasing the dataset from 1,337 to 8,022 images with an equal number of images per class (Zeng, 2024; Hwang & Kim, 2025). Augmentation was used solely to enrich the variation of character shapes and did not alter class identity.

### 3.3 Data splitting and model configuration

The data were divided using the hold-out method under three ratios of training, validation, and test data: 80:10:10, 70:15:15, and 60:20:20. These scenarios were used to test the stability of the model as the volume of training data decreases. The resulting sample sizes are given in Table 3.

Table 3. Data splitting scenarios

| Scenario | Ratio | Training | Validation | Test |
|---|---|---|---|---|
| Scenario 1 | 80:10:10 | 6,414 | 804 | 804 |
| Scenario 2 | 70:15:15 | 5,610 | 1,206 | 1,206 |
| Scenario 3 | 60:20:20 | 4,812 | 1,608 | 1,608 |

The classifier was built on the EfficientNetV2B0 architecture using transfer learning from weights pre-trained on ImageNet, an approach that has proved effective when the target dataset is small (Li & Chang, 2022). The classification head was adapted to the 28 Balinese script classes. The hyperparameter configuration is listed in Table 4. After the classification head had been trained for 30 epochs, a fine-tuning stage was carried out in which the final 30 layers were unfrozen, allowing the feature representation to adapt to the stroke patterns of Balinese script.

Table 4. Hyperparameter configuration

| Parameter | Value |
|---|---|
| Architecture | EfficientNetV2B0 |
| Input size | 224 x 224 pixels |
| Batch size | 32 |
| Learning rate | 0.001 |
| Dropout | 0.3 |

| | |
|---|---|
| Epochs | 30 (+ fine-tuning of the final 30 layers) |
| Output activation | Softmax |

### 3.4 Model evaluation

Evaluation was based on a confusion matrix, from which accuracy, precision, recall, and F1-score were derived. Accuracy, given in Equation (1), is the proportion of correct predictions over the entire test set. Precision, in Equation (2), expresses the reliability of predictions for a given class; recall, in Equation (3), expresses the ability of the model to retrieve all instances of that class; and F1-score, in Equation (4), is their harmonic mean. Because the number of classes is large, results are reported as both macro and weighted averages.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

In addition to internal testing, a deployment test was performed using newly written handwritten images that formed no part of the training, validation, or test sets. These images were cropped, converted to grayscale, resized to 224 x 224 pixels, and binarised before being passed to the model.

## 4. Results and Discussion

### 4.1 Results across the three splitting scenarios

The three splitting scenarios were evaluated to determine how far performance depends on the volume of training data. The results are presented in Table 5 and Fig. 2.

Table 5. Results Across The Three Data Splitting Scenarios

| Scenario | Accuracy | Macro precision | Macro recall | Macro F1-score |
|---|---|---|---|---|
| 80:10:10 | 90.55% | 0.9116 | 0.9073 | 0.9071 |
| 70:15:15 | 89.55% | 0.8999 | 0.8912 | 0.8885 |
| 60:20:20 | 88.81% | 0.8940 | 0.8904 | 0.8895 |

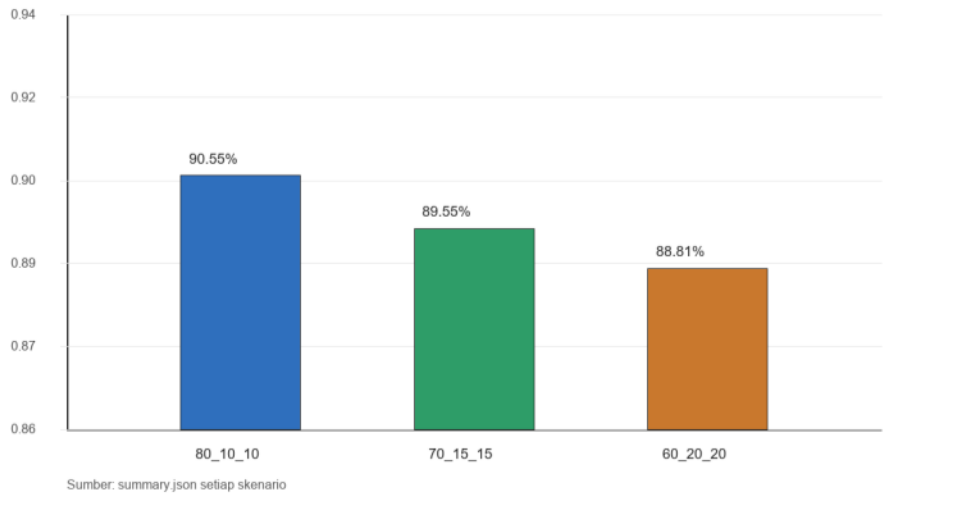


Fig. 2. Model accuracy across the three data splitting scenarios

The highest accuracy, 90.55%, was obtained under the 80:10:10 scenario, followed by 89.55% for 70:15:15 and 88.81% for 60:20:20. The gap between the first and third scenarios is less than two percentage points, and all precision, recall, and F1-score values remain above 88%. The proximity of the macro and weighted averages indicates that performance is distributed fairly evenly across classes rather than being driven by the larger ones. EfficientNetV2 thus retains its classification performance as training data are reduced, which means that the representation experiments described below are not materially confounded by the choice of splitting ratio.

### 4.2 Comparison of Grayscale and Binary Representations

Both representations were trained under an identical configuration, including the fine-tuning stage, so that any difference in outcome is attributable to the input representation alone. The results are compared in Table 6.

Table 6. Comparison of grayscale and binary representations

| Metric | Grayscale | Binary |
|---|---|---|
| Test accuracy | 91.92% | 91.67% |
| Test loss | 0.1759 | 0.2150 |
| Macro precision | 92.27% | 92.21% |
| Macro recall | 92.20% | 91.87% |
| Macro F1-score | 92.11% | 91.78% |
| Weighted precision | 91.99% | 92.04% |
| Weighted recall | 91.92% | 91.67% |
| Weighted F1-score | 91.82% | 91.60% |
| Best validation accuracy | 94.40% | 94.28% |

The training process for both representations is shown in Fig. 3 and Fig. 4. The two curves follow a similar pattern: a steep rise in accuracy over the first ten epochs, a plateau up to epoch 30, and a further improvement once fine-tuning begins. Best validation accuracy reached 94.40% for the grayscale representation and 94.28% for the binary one. The gap between the training and validation curves remains narrow in both cases, giving no indication of severe overfitting.

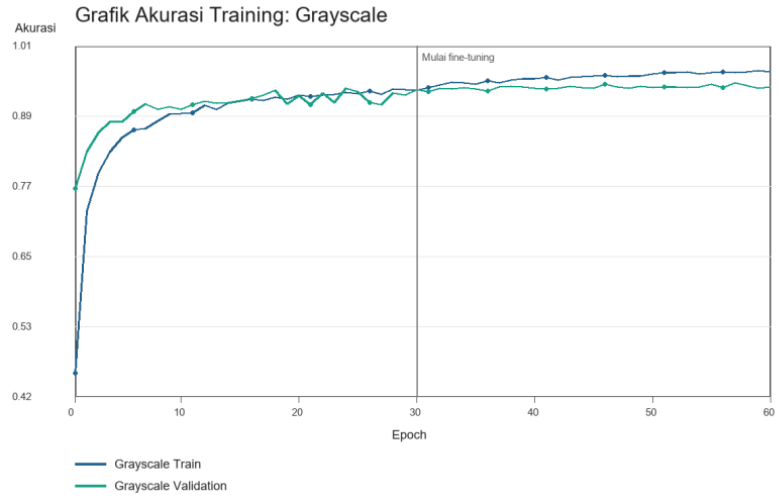


Fig. 3. Training accuracy for the grayscale representation


Fig. 4. Training accuracy for the binary representation

The grayscale representation achieved a test accuracy of 91.92% with a macro F1-score of 92.11%, against 91.67% and 91.78% for the binary representation. A difference of 0.25 percentage points is small enough that the two may be regarded as practically equivalent in accuracy. The clearer distinction lies in test loss: 0.1759 for grayscale against 0.2150 for binary. A lower loss indicates a more confident probability distribution over the predicted classes, so although the two representations misclassify a comparable number of images, the grayscale model is more certain when it is correct.

This advantage can be traced to the visual information that grayscale preserves. Rather than learning character shape as a two-valued pattern, the model also has access to variation in stroke thickness, edge sharpness, and grey-level gradation. Such cues matter because several Balinese characters are distinguished only by fine detail. Binarisation, by contrast, simplifies the image and discards part of that detail, consistent with the observation of Yang and Zuo (2024) that document binarisation invariably risks losing information at character edges.

Binarisation is not without merit, however. For certain difficult classes, notably the pengangge suara taleng and the pengangge tengenan adeg-adeg, the binary representation achieved marginally higher F1-scores than grayscale. Emphasising the underlying shape therefore appears to help with characters formed from simple strokes, although the benefit is too small to outweigh grayscale overall.

### 4.3 Error analysis

The confusion matrices for both representations are shown in Fig. 5 and Fig. 6, and reveal a consistent pattern of error. The largest source of confusion is the pair pengangge suara taleng and pengangge tengenan adeg-adeg: under the grayscale representation, 21 taleng images were predicted as adeg-adeg and 12 adeg-adeg images as taleng, with 21 and 11 respectively under the binary representation. Further errors occur among characters with similar stroke patterns, such as La predicted as Ha, Ka predicted as Na, and Ba predicted as Nga.

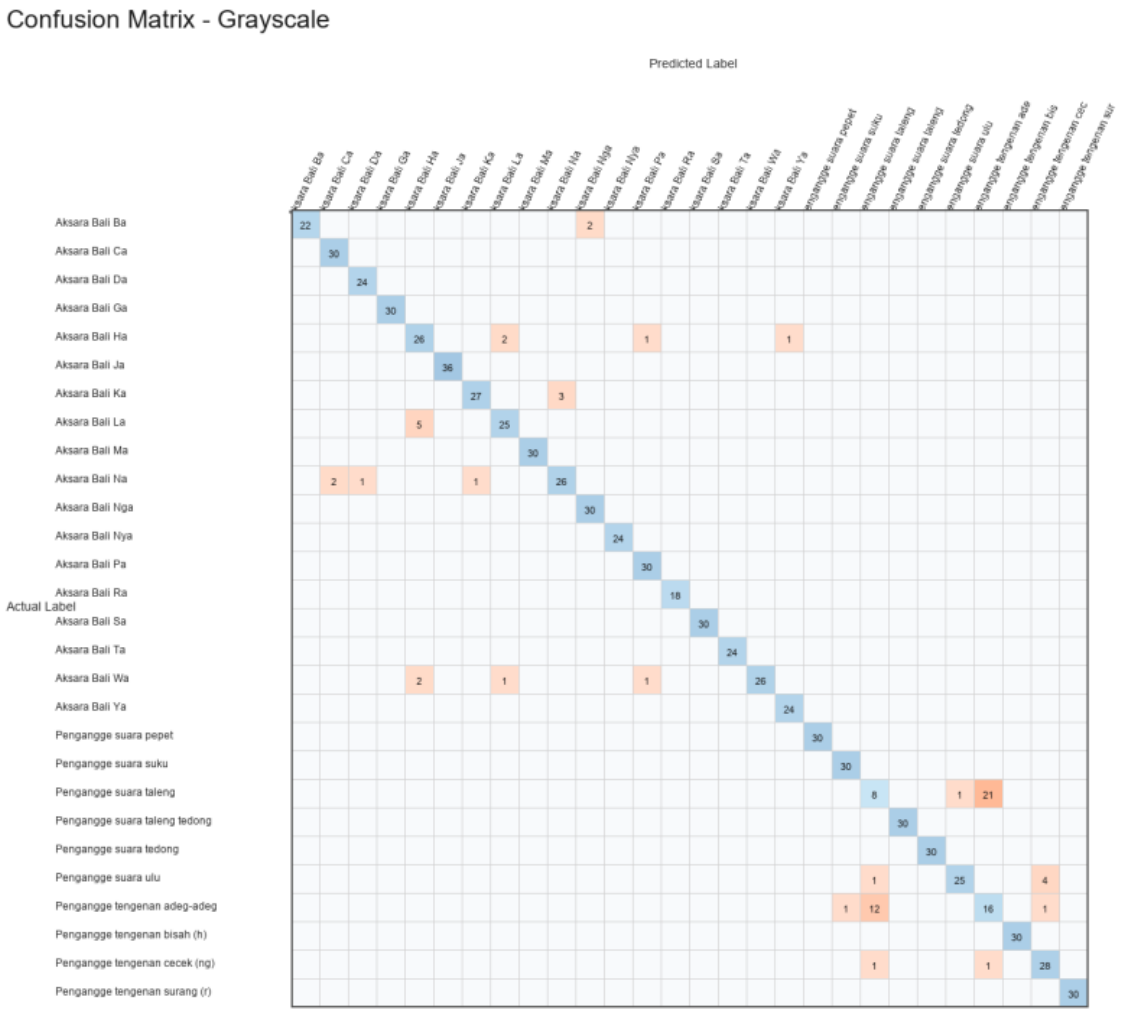


Fig. 5. Confusion matrix for the grayscale representation


Fig. 6. Confusion matrix for the binary representation

That the same errors recur under both representations indicates that their source is not the form of the input image but the structural similarity between classes, which places the feature representations of certain characters close together in feature space. Refining pre-processing alone will therefore not resolve the problem; additional strategies are required, such as targeted collection of data for the difficult classes or a loss function more sensitive to closely neighbouring classes.

### 4.4 Deployment test on new handwritten images

The deployment test used handwritten images produced independently of the research dataset. The model recognised a number of characters correctly, particularly those written intact, centred, on a clean background, and with a distinctive stroke pattern, among them Ya, pengangge suara pepet, pengangge suara taleng tedong, pengangge tengenan adeg-adeg, and pengangge tengenan cecek. Other characters failed: Ca and Da were predicted as pengangge tengenan bisah, Ga as pengangge suara pepet, and pengangge suara taleng as taleng tedong.

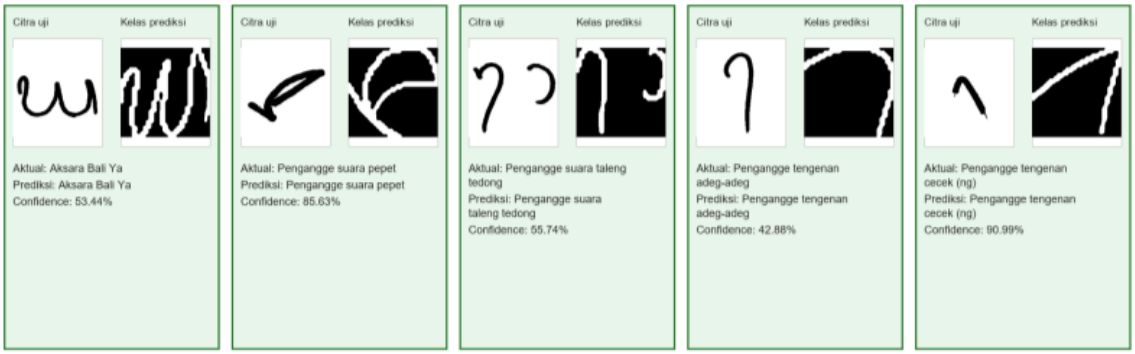


Fig. 7. Examples of handwritten images classified correctly

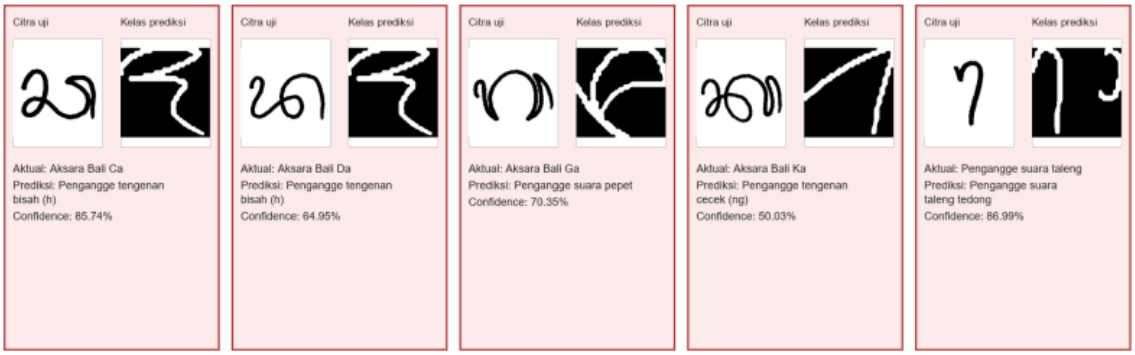


Fig. 8. Examples of handwritten images classified incorrectly

Representative results are shown in Fig. 7 and Fig. 8. Confidence values for the correctly recognised images mostly fall below 90%, and in several cases between 42% and 55%, whereas some misclassified images were assigned confidence as high as 86.99%. The model is therefore not merely mistaken on these images but confidently mistaken, which is the more serious failure mode in a deployed system.

This drop in performance reflects a domain gap between the training data and the new handwritten images, arising from differences in writing style, stroke thickness, lighting, character position, and the outcome of the cropping and binarisation steps. The finding matters because high accuracy on an internal test set evidently does not guarantee comparable performance under real conditions; models that perform well on curated benchmarks may rely on cues that do not transfer (Wichmann & Geirhos, 2023). Within the scope of this study, it also indicates that the diversity of the training data exerts a stronger influence on practical performance than the choice of input representation.

Compared with earlier work, the accuracy obtained here is in line with other studies on regional script recognition, such as the 91.6% reported by Mustafa and Nur (2024) for Lontara script, but lower than the 95.85% reported by Campit and Fontillas (2023) for Baybayin. The difference is unsurprising given that this study covers 28 classes, several of which are highly similar in shape, and that the volume of original images before augmentation was comparatively small.

## 5. Conclusion

This study compared grayscale and binary image representations as input to an EfficientNetV2B0 model for the classification of 28 classes of Balinese script. The grayscale representation achieved a test accuracy of 91.92%, a macro F1-score of 92.11%, and a test loss of 0.1759, against 91.67%, 91.78%, and 0.2150 for the binary representation. The 0.25 percentage point difference in accuracy is

marginal, but the lower prediction error of the grayscale model makes it the preferable input format. Across the three splitting scenarios, accuracies of 90.55%, 89.55%, and 88.81% differ by less than two percentage points, indicating that the model is stable with respect to the proportion of training data. Error analysis showed that misclassification concentrates on characters of high shape similarity, principally pengangge suara taleng and pengangge tengenan adeg-adeg, and that this pattern is consistent across both representations. The deployment test on new handwritten images revealed a domain gap that degrades recognition performance. Future work should therefore draw on handwriting datasets collected from many writers under varied acquisition conditions, and develop specific strategies for the classes that prove difficult to separate, for example targeted augmentation or attention mechanisms of the kind applied by Imane and Alae (2025) to handwritten Arabic.

This study has several limitations. First, the experiments were based on 1,337 original images covering 28 Balinese script classes. Although augmentation was used to increase the number of training samples, augmented images do not provide the same level of variability as independently collected observations. Second, the experimental dataset represents a specific image and writing distribution, and therefore the findings may not fully represent the diversity of Balinese script handwriting encountered in practical applications. Third, the comparison focused on grayscale and binary image representations using EfficientNetV2B0; consequently, the findings do not establish that one representation will produce the same relative performance when combined with other recognition architectures. Finally, the independent handwritten evaluation provides an indication of generalization behaviour but does not constitute a comprehensive benchmark across diverse writers, writing instruments, image acquisition devices, and environmental conditions

**Acknowledgment**

The authors thank Institut Bisnis dan Teknologi Indonesia for the academic support and facilities provided throughout this research.

**References**

Abdiansah, L., & Sumarno. (2025). Implementation of convolutional neural networks algorithm for Javanese handwriting recognition. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 5(2), 496-504.

Archana, R., & Jeevaraj, P. S. E. (2024). Deep learning models for digital image processing: A review. Artificial Intelligence Review, 57. https://doi.org/10.1007/s10462-023-10631-z

Campit, K. F., & Fontillas, S. A. (2023). Character recognition of handwritten Baybayin symbols using EfficientNetV2 convolutional network. IEEE Conference Proceedings, 1-6.

Castro, O., & Bruneau, P. (2024). Landscape of high-performance Python to develop data science and machine learning applications. ACM Computing Surveys, 56(3). https://doi.org/10.1145/3617588

Hwang, D., & Kim, J. J. (2025). Image augmentation approaches for building dimension estimation in street view images using object detection and instance segmentation based on deep learning. Applied Sciences, 15(5). https://doi.org/10.3390/app15052525

Imane, B., & Alae, A. (2025). Enhancing Arabic handwritten word recognition: A CNN-BiLSTM-CTC architecture with attention mechanism and adaptive augmentation. Discover Applied Sciences, 7(5). https://doi.org/10.1007/s42452-025-06952-z

Indrawan, G., & Asroni, A. (2022). Balinese script recognition using Tesseract mobile framework. Lontar Komputer: Jurnal Ilmiah Teknologi Informasi, 13(3), 160. https://doi.org/10.24843/lkjiti.2022.v13.i03.p03

Li, Y. Q., & Chang, H. S. (2022). Large-scale printed Chinese character recognition for ID cards using deep learning and few samples transfer learning. Applied Sciences, 12(2), 1-18. https://doi.org/10.3390/app12020907

Mas Diyasa, I. G. S., & Wijaya, P. A. (2025). Balinese script handwriting recognition using CNN and ELM hybrid algorithms. Jurnal Nasional Pendidikan Teknik Informatika, 14(1), 49-59. https://doi.org/10.23887/janapati.v14i1.87524

Mustafa, M., & Nur, N. (2024). Pengenalan huruf aksara Lontara menggunakan metode convolutional neural network. Journal of Computer and Information System, 7(1), 34-42. https://doi.org/10.31605/jcis.v7i1.3768

Nugraha, K. A. (2024). Penerapan optical character recognition untuk pengenalan variasi teks pada media presentasi pembelajaran. Jurnal Buana Informatika, 15(1), 69-78.

Nurahdika, N. P. I. P., & Sutopo, J. (2025). Analisa performa convolution neural network dalam pengenalan tulisan tangan aksara Lontara. INTECOMS: Journal of Information Technology and Computer Science, 8(1), 12-21. https://doi.org/10.31539/intecoms.v8i1.13627

Pradika, I. K. A. C., & Rahning Putri, L. A. A. (2023). Pengenalan pola karakter tulisan tangan aksara Bali menggunakan fitur zoning, direction, dan backpropagation. JELIKU: Jurnal Elektronik Ilmu Komputer Udayana, 11(3), 663. https://doi.org/10.24843/jlk.2023.v11.i03.p23

Prasetiadi, A., & Saputra, J. (2023). Deep learning approaches for Nusantara scripts optical character recognition. IJCCS: Indonesian Journal of Computing and Cybernetics Systems, 17(3), 325. https://doi.org/10.22146/ijccs.86302

Sidqy Alfarisi, & Subandi, S. (2022). Implementasi pengenalan aksara Bali menggunakan direction feature extraction dan K-nearest neighbor. Jurnal Ticom: Technology of Information and Communication, 11(1), 50-54. https://doi.org/10.70309/ticom.v11i1.71

Sinha, E., & Kumar, A. (2025). OpenCV for computer vision applications. International Journal for Multidisciplinary Research, 7(3), 1-8. https://doi.org/10.36948/ijfmr.2025.v07i03.44280

Swasono, N. E., & Himamunanto, A. R. (2024). Recognition of letter characters in handwritten images using convolutional neural network and K-means clustering algorithm. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 4(4), 1646-1656.

Talebi, H., & Milanfar, P. (2021). Learning to resize images for computer vision tasks. Proceedings of the IEEE International Conference on Computer Vision, 487-496. https://doi.org/10.1109/ICCV48922.2021.00055

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. Proceedings of Machine Learning Research, 139, 10096-10106.

Wichmann, F. A., & Geirhos, R. (2023). Are deep neural networks adequate behavioral models of human visual perception? Annual Review of Vision Science, 9, 501-524. https://doi.org/10.1146/annurev-vision-120522-031739

Yang, Z., & Zuo, S. (2024). A review of document binarization: Main techniques, new challenges, and trends. Electronics, 13(7). https://doi.org/10.3390/electronics13071394

Yuniari, W., & Kumara, S. (2024). Analisis klasifikasi citra penokohan topeng Bali menggunakan model EfficientNetV2 dan Xception. Jurnal Bangkit Indonesia, 13(2), 24-31. https://doi.org/10.52771/bangkitindonesia.v13i2.319

Zeng, W. (2024). Image data augmentation techniques based on deep learning: A survey. Mathematical Biosciences and Engineering, 21(6), 6190-6224. https://doi.org/10.3934/mbe.2024272