Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Study of Baseline and CBAM-Enhanced ResNet50 and MobileNetV2 for Indonesian Rupiah Banknote Classification

Alvin[1)*], Robet[2)], Feriani Astuti Tarigan[3)]
[1,2)] Informatics Engineering, STMIK Time, Medan, Indonesia
[3)] Information Systems, STMIK Time, Medan, Indonesia
[1)*]alvindeveloper25@gmail.com, [2)]robertdetime@gmail.com, [3)]ferianiastutitime@gmail.com

Submitted :Nov 11, 2025 | **Accepted** : Dec 8, 2025 | **Published** : Jan 04, 2026

Abstract: This study investigates the performance of Convolutional Neural Network (CNN) architectures enhanced with Convolutional Block Attention Module (CBAM) for Indonesian banknote classification. Although attention mechanisms have shown strong potential in improving fine-grained visual recognition, their effectiveness for the classification of banknotes with fine textures and similar color patterns remains underexplored, forming a key research gap addressed in this work. Four architectures, ResNet50, ResNet50+CBAM, MobileNetV2, and MobileNetV2+CBAM, were evaluated using K-Fold cross-validation on a dataset of 1,281 images representing seven banknote denominations. Experimental results show that ResNet50 achieves strong baseline performance with a weighted Train accuracy of 99.14% and a Val accuracy of 96.72%, while the integration of CBAM further improves feature discrimination, with ResNet50+CBAM obtaining the highest average accuracy across all folds with a weighted Train accuracy of 100% and a Val accuracy of 99.45%. MobileNetV2 showed lower performance due to its lightweight capacity with a Train accuracy of 91.88% and a decrease in Val accuracy of 85.71%. However, the addition of CBAM provided measurable improvements and greater stability with a Train accuracy of 99.61% and Val accuracy of 92.82%. Overall, CBAM improved CNN's ability to focus on spatial information and salient channels, resulting in more reliable classification. ResNet50+CBAM emerged as the best-performing model, offering the best balance between accuracy and consistency. These findings support the development of reliable computer vision systems for financial technology applications, including automatic banknote recognition, counterfeit detection, and secure transaction verification.

Keywords: CNN; CBAM; ResNet50; MobileNetV2; Banknote Classification; Attention Mechanism

INTRODUCTION

In recent years, the rapid development of computer vision and deep learning has significantly enhanced image-based object recognition tasks, including the recognition of banknote rupiah (Pratap & Sardana, 2022; Rewina et al., 2024). Banknote classification plays a crucial role in financial automation systems such as vending machines, cash counters, and counterfeit detection devices (Nugroho et al., 2025; Ratnasri & Sharmilan, 2021). In Indonesian, the recognition of Rupiah banknotes remains a relevant challenge due to diverse currency designs, similar color tones, and varying conditions such as folding, lighting variations, or partial damage (Hanif et al., 2024). These visual inconsistencies often lead to misclassification when using conventional machine learning models that rely heavily on handcrafted features (Riski Rahmadan et al., 2025).

Convolutional Neural Network (CNN) has demonstrated outstanding performance in image classification tasks due to their ability to learn spatial hierarchies and extract complex features directly from image data (Turahman et al., 2024). Models such as ResNet50 and MobileNetV2 have been widely applied in various object recognition domains due to their high efficiency and accuracy. ResNet50 leverages residual learning to address the vanishing gradient problem in deep architectures, while MobileNetV2 introduces depth wise separable convolutions to improve computational efficiency for applications on lightweight systems (Hermanto et al., 2024).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

However, these architectures may still struggle to capture subtle local differences needed to distinguish visually similar rupiah denominations.

To address these limitations, attention mechanisms have been employed to improve feature discrimination. One widely adopted method is the Convolutional Block Attention Module (CBAM) (Rakha et al., 2024; Zhang et al., 2023), which selectively highlights salient spatial and channel information (Ismail et al., 2025). The integration of CBAM into architectures such as ResNet50 and MobileNetV2 has shown promising improvements in domains like medical imaging and defect detection. However, its use in banknote classification particularly for Indonesia Rupiah remains limited (Rakha et al., 2024; Zhang et al., 2023).

Existing research on banknote classification has primarily focused on either standard CNN architectures or lightweight models optimized for real-time applications. Although studies have shown the effectiveness of MobileNetV2 for banknote recognition (Aprillia et al., 2024), many still overlook attention-based enhancements that could improve performance under complex backgrounds of degraded note conditions. Moreover, only a few works have provided a direct comparison between baseline CNN models and their CBAM-enhanced variants using the same dataset (Ibrahim et al., 2023; Maulana Azhar et al., 2021; Rissa Ilmia Agustin et al., 2024).

This gap underscores the need for a systematic investigation into how attention mechanisms, such as the Convolutional Block Attention Module (CBAM) influence classification performance on Indonesia banknote images. Therefore, this study evaluates the performance of ResNet50, MobileNetV2, CBAM-based models, and the hybrid variants ResNet50+CBAM and MobileNetV2+CBAM in classifying Rupiah denominations (Raja et al., 2024). The study further examines the extent to which CBAM enhances accuracy, precision, and feature extraction effectiveness compared to baseline models (Zhang et al., 2023).

Thus, this study addresses these research gaps by comparing baseline CNN architectures with their CBAM-enhanced variants and analyzing the effect of CBAM on feature representation and classification performance in Indonesia Rupiah banknote images. Overall, the findings of this study are expected to provide clearer insights into how attention mechanisms—specifically CBAM—enhance feature representation and classification performance in Indonesia Rupiah banknote images. The outcomes are anticipated to support the development of more accurate and efficient banknote recognition models for various financial technology systems and automated transaction devices in Indonesia.

LITERATURE REVIEW

Deep learning has become a dominant paradigm in image classification tasks due to its ability to extract hierarchical features from raw image data automatically. Among deep learning architectures, the Convolutional Neural Network (CNN) has demonstrated remarkable performance in various computer vision domains, including object detection, facial recognition, and medical imaging (Kohsasih et al., 2022). The strength of Convolutional Neural Network (CNN) lies in its convolutional layers, which are capable of learning spatial and structural patterns without manual feature engineering. According to (Robet et al., 2025), the ResNet architecture introduced connections that mitigate the vanishing gradient problem, enabling the construction of deeper and more accurate models. ResNet50, a 50-layer variant, has been widely adopted due to its balance between depth and computational efficiency.

Another notable Convolutional Neural Network (CNN) architecture is MobileNetV2, introduced by (Amaludin et al., 2025), which focuses on model efficiency through depth wise separable convolutions. This design significantly reduces the number of parameters and computation costs, making it ideal for real-time and mobile (Nur Hidayat et al., 2023). Previous research has demonstrated the effectiveness of MobileNet-based modes for lightweight image recognition systems, including facial detection and currency recognition. However, the trade-off between efficiency and feature richness remains a limitation, as lightweight models tend to lose fine-grained spatial detail critical for distinguishing visually similar objects.

To overcome this limitation, researchers have proposed integrating attention mechanisms within Convolutional Neural Network (CNN) architectures. The Convolutional Block Attention Module (CBAM), introduced by (Robet et al., 2025), enhances feature representation by sequentially applying channel and spatial attention. This mechanism allows the network to focus on more informative regions of an image, improving discrimination between classes with subtle visual differences. Studies have shown that attention-based models outperform traditional Convolutional Neural Network (CNN) in medical imaging and detection tasks, where local texture and region-specific information are crucial (Rissa Ilmia Agustin et al., 2024).

Several studies have attempted to apply Convolutional Neural Network (CNN) models to currency classification. For instance, (Zhang et al., 2023) utilized a CNN-based model to classify Chinese banknotes, achieving high accuracy under controlled lighting conditions. Similarly, Amaludin et al. (2025) explored MobileNetV2 for currency recognition but reported performance degradation when banknotes were folded, damaged, or photographed under inconsistent illumination. These findings underscore the importance of enhancing feature focus and robustness against visual noise. However, few studies have explicitly examined the role of

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

attention modules such as CBAM in currency classification, particularly in the context of the Indonesian Rupiah, which presents unique challenges due to its design similarities and diverse color schemes (Rakha et al., 2024).

Table 1. Summary of Rupiah Banknote Classification Studies

Authors	Year	Datasets	Methods	Accuracy
(Kohsasih et al., 2022)	2022	White blood cell image dataset	CNN architectures: AlexNet, VGG16, VGG19, ResNet50	AlexNet: 98%, VGG16: 95%, VGG19: 94%, ResNet50: 99%
(Nur Hidayat et al., 2023)	2023	7 nominal classes of rupiah banknotes, 190 images per class (total 1330 images)	CNN (VGG16), input: banknotes images, trained on dataset, implemented on Android	83%
(Zhang et al., 2023)	2023	8241 CT slices from 972 patients	ResNet50 with CBAM, input: ROI of pulmonary nodules (64x64); optional morphological and clinical features, transfer learning (ImageNet) + fine-tuning	Accuracy 0.898, AUC 0.957
(Rakha et al., 2024)	2024	Ultrasound images from two breast-cancer ultrasound databases (public + private)	Transfer-learning using MobileNetV2 backbone + CBAM, classification, with visualization via Grad-CAM for interpretability	93% Test Accuracy
(Amaludin et al., 2025)	2025	Dataset ISIC 2019, 8 skin diseases classes	Transfer-learning with MobileNetV2, hyperparameter tuning (learning rate = 0.0001, batch size = 16, epochs = 70, data split 80 training 20 test)	96.63%
(Robet et al., 2025)	2025	Leaf and pod images of cacao (public sources, field-like conditions), multiple classes (5 diseases + healthy = 6 classes)	CNN backbone: ResNetX50 with both SE and CBAM, preprocessing, resize 224x224, normalization, data augmentation (flips, rotations, color jitter, random crops), training with Adam optimizer + early stopping	97% Test Accuracy Macro-F1 0.97

From the reviewed literature, it can be concluded that while Convolutional Neural Networks (CNNs), such as ResNet50 and MobileNetV2, have achieved considerable success in visual recognition, their performance can still be enhanced through attention mechanisms that refine feature selection and spatial emphasis. Nonetheless, a significant research gap remains in the comparative evaluation of Convolutional Neural Network (CNN) and CBAM-augmented CNN architectures for the specific task of Rupiah banknote classification. This study addresses that gap by analyzing and comparing the performance of ResNet50, MobileNetV2, and their Convolutional Block Attention Module (CBAM)-enhanced variants, contributing new insights into the application of attention mechanisms for improving currency recognition accuracy.

METHOD

The research flow is shown in Figure 1. The process begins with dataset collection, followed by setup and data preprocessing, including normalization and augmentation. Four Convolutional Neural Network (CNN) models were then used in the training process: ResNet50, ResNet50+CBAM, MobileNetV2, and MobileNetV2+CBAM. The training results were then evaluated using several performance metrics, including accuracy, precision, recall, F1-score, and support. Furthermore, graphical analysis of accuracy, loss, and confusion matrices was performed to assess the overall model performance.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

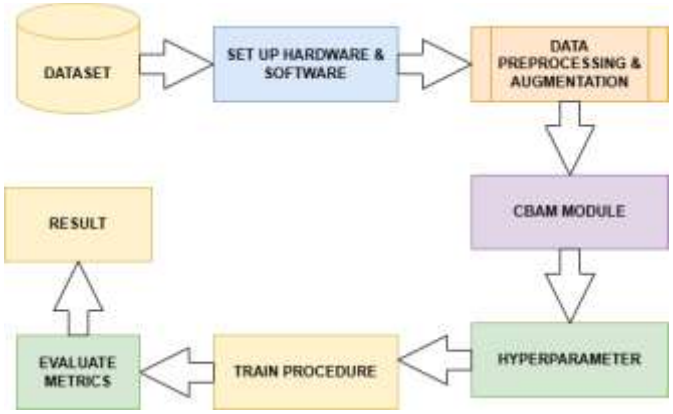


Fig. 1 Research Methodology

Dataset

The dataset used in this study is sourced from the Kaggle platform. Titled "Uang Emisi 2022 Baru" and can be accessed via the link: https://www.kaggle.com/datasets/fannyzahrahramadhan/uang-emisi-2022-baru. This dataset contains a collection of images of the latest issued rupiah banknotes in various denominations, including Rp1.000, Rp2.000, Rp5.000, Rp10.000, Rp20.000, Rp50.000, and Rp100.000. Each class represents a single denomination, allowing the model to learn visual characteristics such as the color, pattern, and texture of each denomination.

Fig. 2 Samples Dataset Currency Rupiah

The sample dataset of banknotes in Rupiah can be divided into classes, with image files in each class, as shown in Table 1.

Table1. Files Images per Class

Class	Number of Images
Rp1.000	183 Files
Rp2.000	227 Files
Rp5.000	184 Files
Rp10.000	214 Files
Rp20.000	75 Files
Rp50.000	171 Files
Rp100.000	227 Files
Total	1281 Files

Setup Hardware & Software

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The research was conducted using a hybrid computing environment combining local and cloud resources. The local system, an Asus X515 laptop equipped with an Intel® Core™ i3-1115G4 processor and 4 GB of RAM, was utilized for code development and file management. However, the limited capabilities of the CPU, such as the low core count, absence of parallel acceleration, and restricted memory, made it inefficient for training large deep learning models. To address these constraints, the primary training process was executed on Google Colab Pro, which provides access to an NVIDIA T4 GPU through Google Compute Engine. The GPU-enabled environment significantly accelerates tensor computations, increases training throughput, and reduces overall training time. Google Drive served as research storage, and datasets were sourced from Kaggle.

The software environment included TensorFlow 2.19.0 and Keras 3.10.0 for deep learning development, NumPy 2.0.2 for numerical computation, and Matplotlib 3.10.0 along with Seaborn 0.13.2 for data and performance visualization. Scikit-Learn 1.6.1 was used for model evaluation and cross-validation, while standard Python libraries such as os and zipfile supported file management. This configuration provides an efficient, stable, and reproducible platform for conducting deep learning experiments, and the complete hardware and software specifications are summarized in Table 2.

Table2. Set up Hardware and Software

Technology	Description
Device Laptop Asus X515	Device
Intel® Core™ i3-1115G4	Device Processor
Google Colab Pro	Research Notebook
T4 GPU	Google Compute Engine
Google Drive	Research File Storage
Kaggle	Dataset Source
os, zipfile	File Management Library
TensorFlow 2.19.0	Deep Learning Framework
Seaborn 0.13.2	Data Visualization
NumPy 2.0.2	Numerical Computation
Keras 3.10.0	Neural Network Library
Matplotlib 3.10.0	Data Visualization Metrics
Scikit-Learn 1.6.1	Model Evaluation

Data Preprocessing

The dataset preprocessing stage is performed to prepare the image dataset before the model training process. The initial dataset is extracted from a compressed .zip file to ensure that all images are accessible and properly organized according to their class labels. After extraction, each class folder is scanned, and all image file paths are collected. Class labels are then mapped to numerical indices to enable consistent label encoding during the training process. Unlike the conventional approach that relies on a fixed training-validation split, this study employs k-fold cross-validation (with k=2) due to the relatively small dataset size. Through this method, the entire dataset is partitioned into two folds that alternately function as training and validation sets, ensuring that every image contributes to both stages. This strategy provides a more reliable estimation of model performance under limited data conditions. To optimize the efficiency of data handling during training, the image files are normalized and resized to the required 224x224-pixel format. Additional pipeline optimizations, such as caching, shuffling within each fold, and prefetching, are applied to accelerate data loading and minimize ordering bias. This preprocessing configuration ensures that the dataset is clean, consistent, and ready for the training of CNN models or CNN architectures enhanced with CBAM module.

Data Augmentation

During the training phase, data augmentation was implemented using the ImageDataGenerator to increase image diversity and reduce overfitting. Several stochastic transformations were applied, including minor rotations up to 5 degrees, horizontal and vertical shifts up to 5 percent, zoom variations up to 5 percent, and brightness adjustments ranging from 0.8 to 1.2. A shear transformation of 0.05 was also utilized. Both horizontal and vertical flips were disabled to maintain consistency with the original image characteristics. This augmentation strategy introduces meaningful variability while preserving the essential structure of the observed objects.

Convolutional Block Attention Module

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This process ensured that the ResNet50, MobileNetV2, ResNet50+CBAM, and MobileNetV2+CBAM models were well-prepared for the learning phase. The Convolutional Block Attention Module (CBAM) was integrated to enhance feature representation. CBAM sequentially applies Channel Attention and Spatial Attention to highlight informative features and important spatial regions within the image. The channel attention uses global average and max pooling followed by a shared MLP, while the spatial attention employs pooling across channels and a 7×7 convolution to refine spatial focus (Zakariah & Alnuaim, 2024). To ensure consistency and reproducibility across experiments, all models were trained using identical hyperparameter settings. The Adam optimizer was used with a learning rate of 0.0001, and all models were trained for 50 epochs using categorical cross-entropy loss. The batch size was set to 32 to balance computational efficiency with stability during training. Input images were resized to 224×224 pixels, pretrained weights from ImageNet were used for initialization, and no early stopping was applied due to limitations of CPU-based runtime. Data augmentation was applied dynamically at the batch level during training, while validation data remained unchanged. Figure 3 and Figure 4 illustrates the architecture, showing the integration of the attention module with the backbone CNN and the flow toward the classification output.

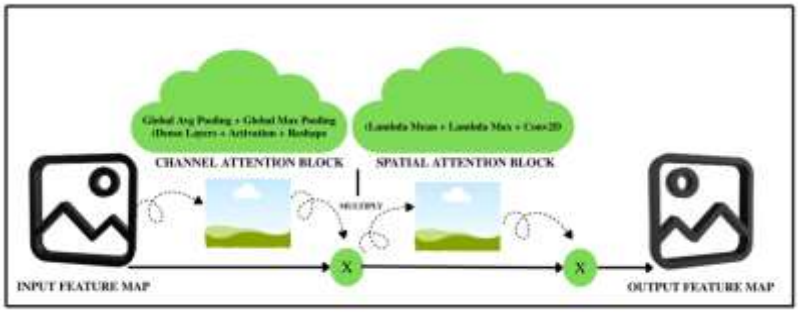


Fig. 3 CBAM Architecture (Channel Attention and Spatial Attention)

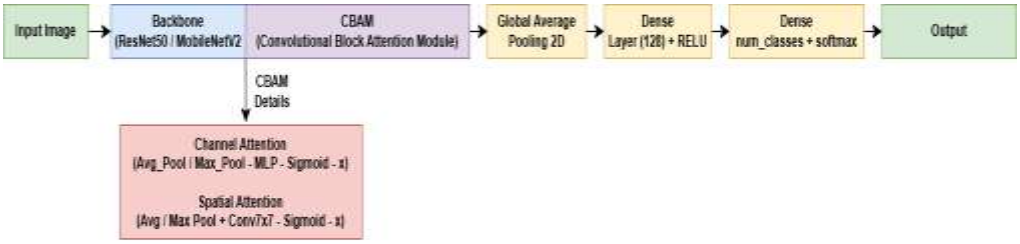


Fig. 4 CNN+CBAM Architecture

Hyperparameter

The hyperparameter configuration encompasses the learning rate schedule, optimizer settings, batch size, number of epochs, weight initialization strategy, and regularization techniques, each of which was selected to ensure stable convergence and reproducible training dynamics. A detailed summary of these settings is presented in Table 3 and Figure.

Table3. Hyperparameter Details

Components	Settings
Optimizer	Adam
Learning Rate	1e-4
Batch Size	32
Epochs	50
Scheduler	ReduceLROnPlateau
Loss Function	Categorical Cross-Entropy
Weight Initialization	He Normal
Regularization	Dropout / L2
Input Image Size	224 x 224
Augmentation	Rotation, Width Shift, Height Shift, Zoom, Brightness, Shear, Fill Mode
Preprocessing	K-Fold = 2

Train Procedure

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The training procedure was implemented using a two-fold cross-validation scheme to ensure a reliable assessment of model performance across different data partitions. For each fold, the dataset was divided into training and validation subsets, and all architectures, including ResNet50, MobileNetV2, and their CBAM-augmented variants (ResNet50+CBAM and MobileNetV2+CBAM), were trained independently. Each model was initialized with ImageNet pre-trained weights and processed through standardized preprocessing and augmentation pipelines. Training was performed on a GPU using the Adam optimizer with mini-batch execution, while early stopping and ReduceLROnPlateau were applied to control overfitting and stabilize convergence. Performance was monitored on the validation subset at every epoch, and the best-performing weights from each fold were preserved for comparative analysis.

Evaluation Metrics

The model evaluation phase aims to assess the performance and generalization capability of the proposed architectures using a two-fold cross-validation scheme. In this setup, the dataset is divided into two equally sized folds, and the training–validation process is repeated twice. In each iteration, one-fold functions as the validation subset while the other serves as the training subset. This approach enables a more reliable estimation of model performance, especially when working with a limited dataset, as every sample is used for both training and validation across different iterations. The evaluation focuses on the model's ability to correctly identify image patterns, maintain stability across folds, and avoid overfitting. Four primary metrics are used in this study: accuracy, precision, recall, and F1-score, all computed from predictions made on the validation fold of each cross-validation iteration.

Accuracy

the overall rate of correct predictions compared to the total amount of validation data.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Ctrue}}{\boldsymbol{Ctrue} + \boldsymbol{Cfalse}} \quad \textbf{(1)}$$

Precision

the extent to which the positive predictions generated by the model match the actual labels.

$$\boldsymbol{Precision} = \boldsymbol{Ratio}(\boldsymbol{TP}, \boldsymbol{TP} + \boldsymbol{FP}) \quad \textbf{(2)}$$

Recall

the model ability to recognize all data that should belong to the positive class.

$$\boldsymbol{Recall} = \boldsymbol{Ratio}(\boldsymbol{TP}, \boldsymbol{TP} + \boldsymbol{FN}) \quad \textbf{(3)}$$

F1-Score

provides a balance between precision and recall and is useful when data between classes is imbalanced.

$$\boldsymbol{F1} = \frac{2 \times \boldsymbol{Precision} \times \boldsymbol{Recall}}{\boldsymbol{Precision} + \boldsymbol{Recall}} \quad \textbf{(4)}$$

In addition to these metrics, a confusion matrix is generated for each fold to analyze classification outcomes for all classes. This matrix provides insight into which classes are correctly recognized, and which tend to produce misclassification errors. The metric values obtained from both cross-validation folds are averaged to obtain the final performance score for each model. These results are used to compare the four architectures: ResNet50, ResNet50+CBAM, MobileNetV2, and MobileNetV2+CBAM. The model achieving the highest average metrics and demonstrating consistent performance across folds is considered the most effective for the currency image classification task.

RESULT

Build Model

In the model-building stage, four convolutional neural network architectures were developed to facilitate comparative performance analysis. Both ResNet50 and MobileNetV2 were employed as backbone networks with pre-trained ImageNet weights, and their top classification layers were removed (include_top=False) to enable their use as feature extractors. A new classification head was then added in accordance with the number of target classes. Additionally, the Convolutional Block Attention Module (CBAM) was selectively integrated into architecture to enhance the refinement of spatial and channel-wise feature representations. This design process produced four model configurations: ResNet50, ResNet50 with CBAM, MobileNetV2, and MobileNetV2 with CBAM. These configurations were constructed to assess how the addition of attention mechanisms influences extraction quality, training stability, and validation accuracy. The structural differences between the baseline and CBAM-enhanced

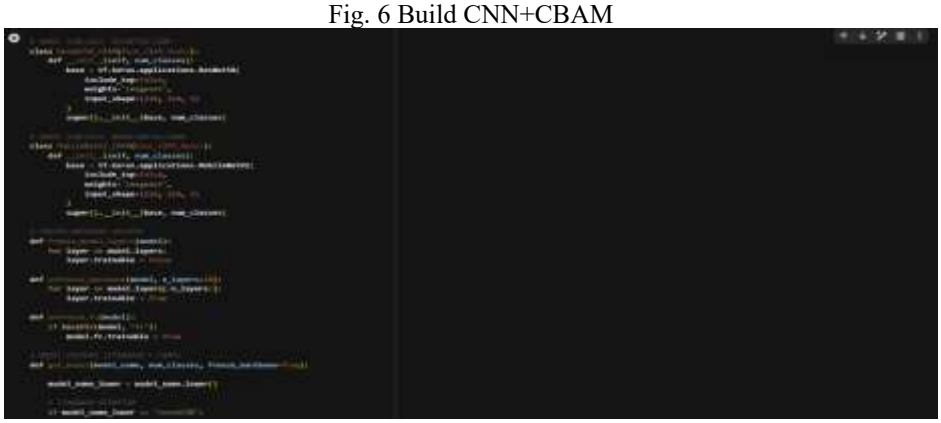
*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

models are depicted in Figures 5 and 6, which illustrate the modifications introduced within the feature extraction flow.

Fig. 5 Build CBAM

Fig. 6 Build CNN+CBAM



Train Cross Validation Performance

Figure 6 presents the training and validation accuracy as well as the loss progression across the four evaluated architectures: ResNet50, ResNet50+CBAM, MobileNetV2, and MobileNetV2+CBAM. The plots show a clear upward trend in accuracy and a steady reduction in loss for all models, indicating that the networks converged effectively during training.

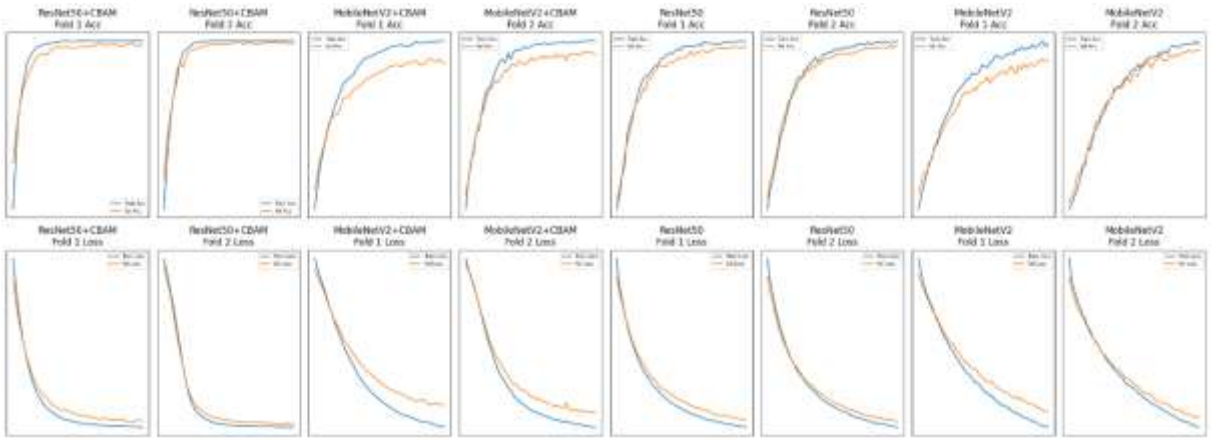


Fig. 7 Graph Cross Validation Accuracy and Loss

In the ResNet50 baseline, accuracy increased rapidly during the initial epoch and stabilized toward the end, although minor fluctuations in validation accuracy suggest the presence of mild overfitting. When CBAM was

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

integrated into ResNet50, both training and validation accuracy curves became smoother, and the loss decreased more consistently. This pattern indicates that the attention mechanism helped the model focus on more informative features, leading to improved training stability and better generalization. The MobileNetV2 model, designed as a lightweight architecture, exhibited a slower but stable improvement in accuracy. Its validation curve remained relatively smooth, showing that the model maintained consistent learning despite having fewer parameters. Surprisingly, the MobileNetV2+CBAM variant did not consistently surpass the baseline; while its training accuracy improved steadily, the validation accuracy in some folds trailed the original MobileNetV2. This behavior suggests a diminishing return effect, where the added complexity introduced by CBAM may not align with MobileNetV2's compact structure, leading to suboptimal benefit from the attention mechanism.

Overall, the training curves demonstrate that CBAM integration was effective for ResNet50 by improving learning stability and refining feature extraction. However, its impact on MobileNetV2 was mixed, highlighting that attention modules must be matched carefully with the underlying architecture to achieve optimal gains.

Table4. Results Train Model

Models	K–Fold Train	Best Train Accuracy	Min Train Loss	Epoch
ResNet50	1	0.9937	0.1737	50
ResNet50	2	0.9891	0.2221	50
ResNet50+CBAM	**1**	**1.0000**	**0.0141**	**50**
ResNet50+CBAM	**2**	**1.0000**	**0.0069**	**50**
MobileNetV2	1	0.9344	0.5110	50
MobileNetV2	2	0.9033	0.6127	50
MobileNetV2+CBAM	**1**	**0.9922**	**0.1552**	**50**
MobileNetV2+CBAM	**2**	**1.0000**	**0.0702**	**50**

Table5. Results Mean Train Model

Models	K–Fold Train	Mean Train Acc	Mean Train Loss
ResNet50	1&2	0.9914	0.1979
ResNet50+CBAM	1&2	**1.0000**	**0.0105**
MobileNetV2	1&2	0.9188	0.5619
MobileNetV2+CBAM	1&2	**0.9961**	**0.1127**

Cross Validation Performance

The two-fold cross-validation results show that each model responds differently to variations in data partitioning, yet a consistent pattern emerges regarding the impact of the attention mechanism. Both ResNet50 and ResNet50+CBAM exhibit highly stable performance across the two folds, indicating strong generalization capability and robust feature extraction despite changes in training–validation composition. Notably, the integration of CBAM further enhances this stability by enabling the model to selectively emphasize informative channel and spatial features, which results in substantially higher validation accuracy and markedly lower validation loss compared to the baseline ResNet50. MobileNetV2 also demonstrates improvements across folds, although with slightly higher variability, a characteristic consistent with its lightweight architecture and reduced representational capacity. The incorporation of CBAM into MobileNetV2, however, significantly mitigates this limitation. MobileNetV2+CBAM achieves higher accuracy and a smoother reduction in validation loss, indicating that the attention mechanism helps the model focus on discriminative patterns that might otherwise be overlooked. These results collectively suggest that CBAM benefits both deep and lightweight architectures, with the most pronounced performance gains observed in MobileNetV2+CBAM and ResNet50+CBAM. The enhancement is attributed to CBAM's ability to refine feature importance dynamically, leading to more discriminative learning and improved generalization across folds.

Table6. Results Cross Validation

Models	Best Val Acc	Best Val Loss	Epoch
ResNet50	0.9672	0.2711	50
ResNet50+CBAM	**0.9945**	**0.0488**	**50**
MobileNetV2	0.8571	0.6774	50
MobileNetV2+CBAM	**0.9282**	**0.3098**	**50**

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Error Classification

The error classification analysis revealed that most misclassifications occurred among denominations with similar visual characteristics, such as color tone, pattern, or illumination. These similarities often caused the model to confuse adjacent nominal values, particularly in lower denominations, with subtle textual or texture details. Additionally, environmental factors in image acquisition, including lighting variations and background clutter, contributed to inconsistent predictions. Despite these limitations, the overall error rate remained low, indicating that the model's feature extraction and attention mechanisms performed effectively under most conditions.

Error Classification ResNet50+CBAM K-Fold 1 Error Classification ResNet50+CBAM K-Fold 2

Fig. 8 Error Classification of ResNet50+CBAM

Error Classification MobileNetV2+CBAM K-Fold 1 Error Classification MobileNetV2+CBAM K-Fold 2

Fig. 9 Error Classification MobileNetV2+CBAM Model

Error Classification ResNet50 K-Fold 1 Error Classification ResNet50 K-Fold 2

Fig. 10 Error Classification ResNet50 Model

Error Classification MobileNetV2 K-Fold 1 Error Classification MobileNetV2 K-Fold 2

Fig. 11 Error Classification MobileNetV2 Model

Accurate Classification

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The models exhibited strong performance across all currency denominations. ResNet50 and its CBAM-augmented variant delivered the most consistent accuracy, particularly when distinguishing denominations with subtle visual similarities. MobileNetV2 also achieved competitive results, and its performance improved further with the addition of CBAM. Overall, incorporating CBAM enhanced the stability and generalization capability of every evaluated architecture.

Accurate Classification ResNet50+CBAM K-Fold 1 Accurate Classification ResNet50+CBAM K-Fold 2

Fig. 12 Accurate Classification of ResNet50+CBAM Model

Accurate Classification MobileNetV2+CBAM K-Fold 1 Accurate Classification MobileNetV2+CBAM K-Fold 2

Fig. 13 Accurate Classification of MobileNetV2+CBAM Model

Accurate Classification ResNet50 K-Fold 1 Accurate Classification ResNet50 K-Fold 2

Fig. 14 Accurate Classification of ResNet50 Model

Accurate Classification MobileNetV2 K-Fold 1 Accurate Classification MobileNetV2 K-Fold 2

Fig. 15 Accurate Classification of MobileNetV2 Model

Evaluate Metrics

The evaluation results demonstrate distinct performance patterns across the four architectures. ResNet50 shows stable and high classification capability, achieving accuracies of 0.9610 and 0.9578 across the two folds, with consistent precision, recall, and F1-scores near 0.96. The number of misclassified samples remains relatively low

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(25 and 27), indicating strong generalization and balanced performance across classes. ResNet50+CBAM exhibits the highest accuracy among all models, reaching 0.9906 in both folds. Precision, recall, and F1-scores are consistently above 0.989, and the number of misclassifications is reduced to only six samples per fold. These results indicate that the attention mechanism effectively enhances feature discrimination, enabling the model to capture subtle variations present in the banknote images. MobileNetV2 presents lower performance compared to the ResNet-based models, with accuracies of 0.8393 and 0.8328. Precision, recall, and F1-scores follow the same trend, and the misclassified samples reach 103 and 107. This outcome is expected given the lightweight nature of MobileNetV2, which limits its capacity for fine-grained feature extraction. MobileNetV2+CBAM demonstrates noticeable improvements over the baseline MobileNetV2. Accuracy increases to 0.8783 and 0.9406, accompanied by higher precision, recall, and F1-scores. The reduction in misclassified samples—from over 100 to 78 and 38—indicates that the addition of CBAM substantially enhances the model's ability to identify discriminative visual features despite its compact architecture. Overall, the metrics confirm that incorporating CBAM consistently improves classification performance across both ResNet50 and MobileNetV2 architectures, with the most significant gains observed in fine-grained feature recognition required for banknote classification.

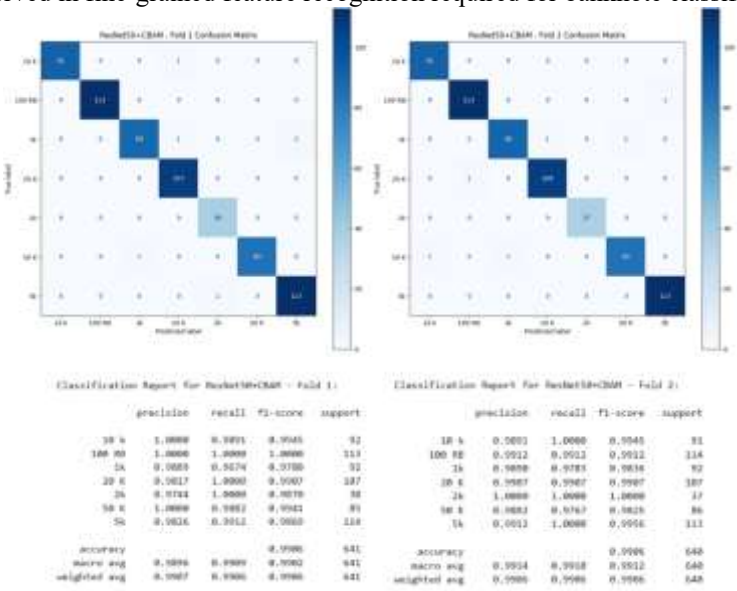


Fig. 16 Confusion Metrics and Classification Report ResNet50+CBAM

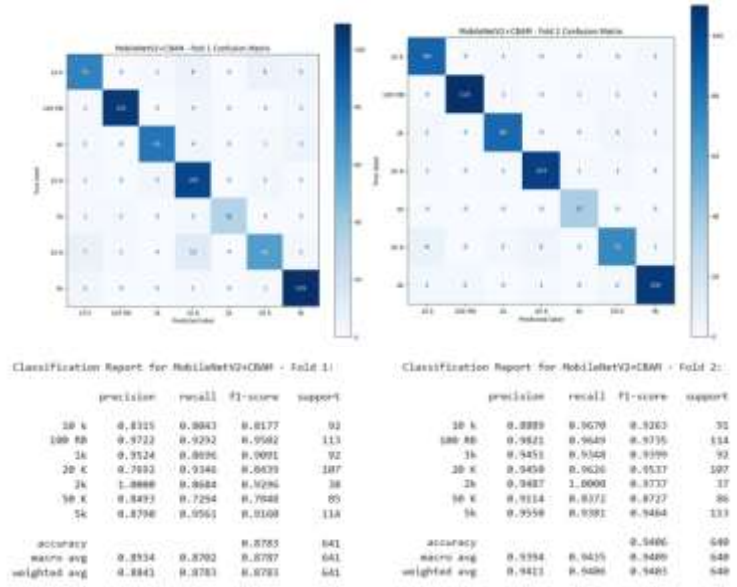


Fig. 17 Confusion Metrics and Classification Report MobileNetV2+CBAM

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15558
e-ISSN : 2541-2019
p-ISSN : 2541-044X

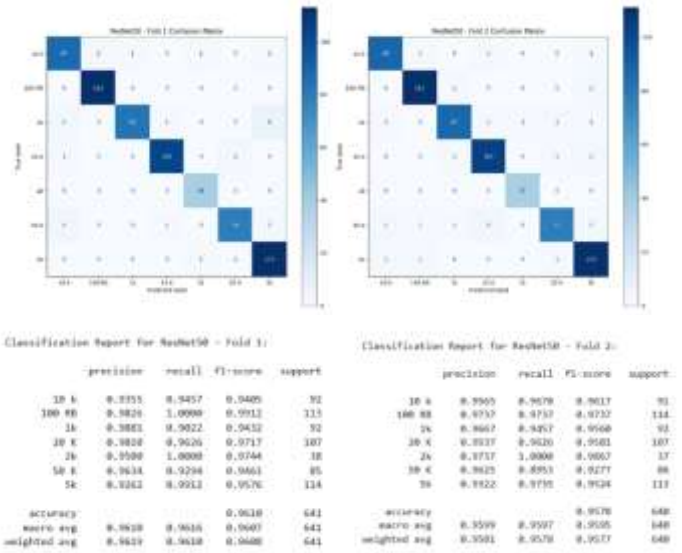


Fig. 18 Confusion Metrics and Classification Report ResNet50

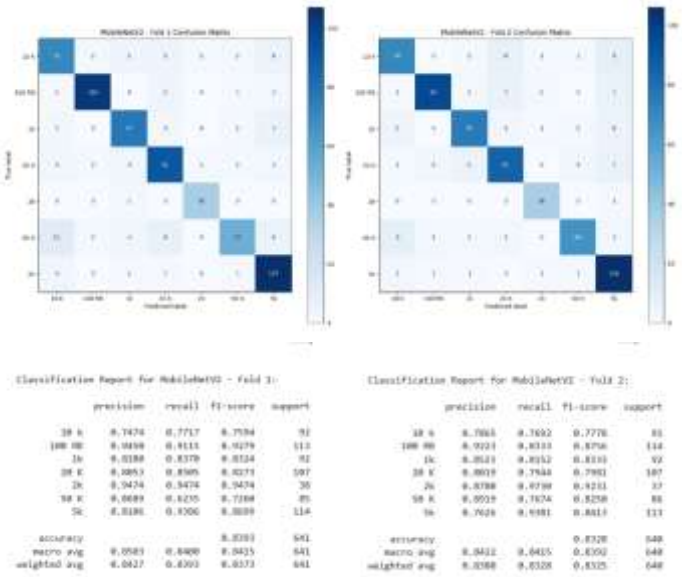


Fig. 19 Confusion Metrics and Classification Report MobileNetV2

Table7. Results Summary Evaluate Metrics

Models	Fold	Acc	Precision	Recall	F1-Score	Misclassified	Total Samples
ResNet50	1	0.9610	0.9610	0.9616	0.9607	25	641
ResNet50	2	0.9578	0.9599	0.9597	0.9595	27	640
ResNet50+CBAM	**1**	**0.9906**	**0.9896**	**0.9909**	**0.9902**	**6**	**641**
ResNet50+CBAM	**2**	**0.9906**	**0.9914**	**0.9910**	**0.9912**	**6**	**640**
MobileNetV2	1	0.8393	0.8503	0.8400	0.8415	103	641
MobileNetV2	2	0.8328	0.8422	0.8415	0.8392	107	640
MobileNetV2+CBAM	**1**	**0.8783**	**0.8934**	**0.8702**	**0.8787**	**78**	**641**
MobileNetV2+CBAM	**2**	**0.9406**	**0.9394**	**0.9435**	**0.9409**	**38**	**640**

Overall Models Comparison

Overall, the comparative evaluation showed that both architectures benefited from the integration of CBAM. ResNet50-based models demonstrated superior training stability and classification accuracy, with the CBAM-

*name of corresponding author

enhanced variant achieving further gains in feature discrimination. MobileNetV2 also exhibited notable improvements when augmented with CBAM, resulting in more reliable recognition of visually similar denominations while retaining its computational efficiency. Collectively, these findings indicate that CBAM consistently strengthens feature extraction and supports better generalization across both ResNet50 and MobileNetV2 families for currency classification tasks.

Table8. Results Overall Models CNN + CBAM

Models	Train Accuracy	Val Accuracy	Train Loss	Val Loss	Total Samples	Rank Model
ResNet50	99.14%	96.72%	0.1979	0.2711	640	3
ResNet50+CBAM	**100%**	**99.45%**	**0.0105**	**0.0488**	**640**	**1**
MobileNetV2	91.88%	85.71%	0.5619	0.6774	640	4
MobileNetV2+CBAM	**99.61%**	**92.82%**	**0.1127**	**0.3090**	**640**	**2**

DISCUSSIONS

In this section, the discussion interprets experimental results and evaluates model performance across the different architectures. The results show that ResNet50 and ResNet50+CBAM consistently surpass the MobileNetV2 variants. Deeper CNN architecture demonstrates a stronger ability to capture fine-grained visual patterns required for currency discrimination. CBAM-enhanced models achieve additional improvement through spatial and channel attention, which direct the network toward subtle local features, including micro-textures, shading variations, and small graphical elements characteristic of Rupiah banknotes. Studies in attention-based CNNs also indicate that CBAM strengthens a model's focus on discriminative regions in visually similar objects.

The integration of CBAM supports faster convergence and improves generalization by refining the feature selection process, particularly for denominations with similar color tones or design structures. MobileNetV2 maintains strength in computational efficiency and remains suitable for resource-limited or embedded environments, although its lightweight architecture restricts its ability to capture complex fine-grained details. Both ResNet50 and MobileNetV2 demonstrate measurable performance gains when augmented with CBAM. Attention mechanisms enhance recognition accuracy and robustness in banknote classification systems that require detailed feature extraction.

CONCLUSION

The experimental results show that ResNet50+CBAM achieves the highest performance, with a validation accuracy of 99.45%, demonstrating a consistent improvement over the baseline ResNet50 validation accuracy 96.72%. This outcome aligns with previous studies indicating that attention mechanisms enhance fine-grained feature extraction in image classification tasks. MobileNetV2 exhibits lower validation accuracy 85.71%, which is consistent with literature highlighting the limitations of lightweight CNN architectures when dealing with visually detailed datasets. Nevertheless, the integration of CBAM significantly improves the performance and fold-to-fold stability of MobileNetV2 validation accuracy 92.82%, supporting earlier findings on the effectiveness of attention modules for compact models. This study has several limitations, particularly the relatively small dataset size and the absence of an independent test set, resulting in performance evaluation that relies entirely on validation data. These constraints may introduce bias in estimating the true generalization capability of the models. Alignment with Previous Research. The results reflect established insights that deeper architectures generally perform better on fine-grained classification and that spatial–channel attention typically contributes 1–3% accuracy gains by focusing on salient regions. These findings support the initial hypotheses: ResNet50 outperforms MobileNetV2, MobileNetV2 remains more computationally efficient but less accurate, and CBAM consistently enhances accuracy and robustness across architectures. Study limitations include a relatively small dataset and absence of a dedicated test set, which may bias generalization estimates. Future work should expand the dataset, including an independent test set, investigate more advanced architectures, and integrate Explainable AI, Grad-CAM to improve interpretability for real-world banknote classification systems.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to the Department of Informatics Engineering for providing computational facilities and research support throughout this study. This work represents a small but significant step toward advancing intelligent vision systems for currency recognition in the era of digital transformation. The authors also acknowledge all collaborators who contributed ideas and technical assistance, enabling the integration of deep learning and attention mechanisms for future-oriented financial technology applications.

*name of corresponding author

REFERENCES

F. Amaludin, M. I. Zulfa, & H. Siswantoro (2025). Pengaruh Hyperparameter Tuning Pada Kinerja Mobilenetv2 Dengan Transfer Learning Untuk Deteksi Penyakit Kulit. *Jurnal SINTA: Sistem Informasi Dan Teknologi Komputasi*, *2*(2). https://doi.org/10.61124/sinta.v2i2.43

D. Aprillia, T. Rohana, T. Al. Mudzakir, & D. Wahiddin. (2024). Deteksi Nominal Mata Uang Rupiah Menggunakan Metode Convolutional Neural Network dan Feedforward Neural Network. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(4), 2068–2077. https://doi.org/10.30865/klik.v4i4.1711

M. Z. Hanif, W. A. Saputra, Y. H. Choo, & A. P. Yunus. (2024). Rupiah Banknotes Detection Comparison of The Faster R-CNN Algorithm and YOLOv5. *JURNAL INFOTEL*, *16*(3), 502–517. https://doi.org/10.20895/infotel.v16i3.1189

A. R. Hermanto, A. Aziz, & S. Sudianto. (2024). Perbandingan Arsitektur MobileNetV2 dan RestNet50 untuk Klasifikasi Jenis Buah Kurma. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, *12*(4), 630–637. https://doi.org/10.26418/justin.v12i4.80358

M. M. Ibrahim, R. Rahmadewi, & L. Nurpulaela. (2023). Pendeteksian Nominal Uang Pada Gambar Menggunakan Convolutional Neural Network: Integrasi Metode Pra-Pemrosesan Citra Dan Klasifikasi Berbasis CNN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(2), 1394–1400. https://doi.org/10.36040/jati.v7i2.6863

J. Ismail, L. Tanti, & W. Wanayumini. (2025). Development of Skin Cancer Pigment Image Classification Using a Combination of Mobilenetv2 and CBAM. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *10*(4), 770–780. https://doi.org/10.33480/jitk.v10i4.6541

K. L. Kohsasih, M. Zarlis, & B. H. Hayadi. (2022). Comparison of CNN Architecture for White Blood Cells Image Classification. *ICOSNIKOM 2022 - 2022 IEEE International Conference of Computer Science and Information Technology: Boundary Free: Preparing Indonesia for Metaverse Society*. https://doi.org/10.1109/icosnikom56551.2022.10034875

K. Maulana Azhar, I. Santoso, & Y. A Adi Soetrisno. (2021). Implementasi Deep Learning Menggunakan Metode Convolutional Neural Network Dan Algoritma Yolo Dalam Sistem Pendeteksi Uang Kertas Rupiah Bagi Penyandang Low Vision. *Transient: Jurnal Ilmiah Teknik Elektro*, *10*(3), 502–509. https://doi.org/10.14710/transient.v10i3.502-509

F. A. Nugroho, & N. Wiliani. (2025). Perbandingan Kinerja ANN dan CNN dalam Tugas Klasifikasi Citra Berbasis Pembelajaran Mesin. *Teknomatika: Jurnal Informatika Dan Komputer*, *18*(1), 22–27. https://doi.org/10.30989/teknomatika.v18i1.1561

A. M. Nur Hidayat, Antamil, & I. Zakiyah M. (2023). Identifikasi Nominal Mata Uang Rupiah Bagi Penyandang Tunanetra Dengan Algoritma Convolutional Neural Network Berbasis Android. *Journal Software, Hardware and Information Technology*, *3*(2), 60–65. https://doi.org/10.24252/shift.v3i2.102

A. Pratap, & N. Sardana. (2022). Machine learning-based image processing in materials science and engineering: A review. *Materials Today: Proceedings*, *62*(P14), 7341–7347. https://doi.org/10.1016/j.matpr.2022.01.200

V. Raja R, Jimson L, Gnanaprakasam C, Jerrin Simla A, Sharmila J. L. B, Lincy Jemina S. (2024). Enhanced Brain Tumor Analysis: Integrating ResNet50 with Convolutional Block Attention Modules for Advanced Insights. *Journal of Electrical Systems*, *20*(6s), 2601–2612. https://doi.org/10.52783/jes.3272

M. Rakha, M. D. Sulistiyo, D. Nasien, & M. Ridha. (2024). A Combined MobileNetV2 and CBAM Model to Improve Classifying the Breast Cancer Ultrasound Images. *Journal of Applied Engineering and Technological Science (JAETS)*, *6*(1), 561–578. https://doi.org/10.37385/jaets.v6i1.4836

N. Ratnasri, & T. Sharmilan. (2021). Vending Machine Technologies: A Review Article. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *58*(2), 160–166. https://www.gssrr.org/journalofbasicandapplied/article/view/12579

A. E. Rewina, S. Sulistyowati, M. Kurniawan, M. Dinarta N, & S. F. Yunanda. (2024). Penerapan Metode CNN (Convolutional Neural Network) dalam Mengklasifikasi Uang Kertas dan Uang Logam. *TIN: Terapan Informatika Nusantara*, *4*(12), 778–785. https://doi.org/10.47065/tin.v4i12.5128

Riski Rahmadan, Nurliani, Efendi Rahayu, Saudah, Ayu Puspita Sari Sinaga, & Enda Ribka Meganta P. (2025). Evaluasi Penerapan Convolutional Neural Network (CNN) untuk Klasifikasi Penyakit Daun Jagung Menggunakan Pendekatan Systematic Literature Review. *RJOCS (Riau Journal of Computer Science)*, *11*(1), 19–27. https://doi.org/10.30606/rjocs.v11i1.3068

Rissa Ilmia Agustin, Jamaludin Indra, Sutan Faisal, Ahmad Fauzi, & Rija Nur Hijriyya. (2024). Klasifikasi Pecahan Uang Kertas Rupiah Menggunakan Transfer Learning Dengan Model Mobilenetv2. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, *9*(2), 242–250. https://doi.org/10.24252/INSTEK.V9I2.49123

R. Robet, J. T. K. P. Angin, & T. H. Siregar. (2025). Attention Augmented Deep Learning Model for Enhanced Feature Extraction in Cacao Disease Recognition. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *9*(4), 1965–1977. https://doi.org/10.33395/sinkron.v9i4.15249

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

T. Turahman, E. Hasmin, & K. Aryasa. (2024). Analisis Perbandingan Metode Convolutional Neural Network (CNN) dan MobileNet dalam Klasifikasi Penyakit Daun Padi. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *9*(1), 368–377. https://doi.org/10.35870/jtik.v9i1.3218

M. Zakariah, & A. Alnuaim. (2024). Recognizing human activities with the use of Convolutional Block Attention Module. *Egyptian Informatics Journal*, *27*. https://doi.org/10.1016/j.eij.2024.100536

Y. Zhang, W. Feng, Z. Wu, W. Li, L. Tao, X. Liu, F. Zhang, Y. Gao, J. Huang, & X. Guo. (2023). Deep-Learning Model of ResNet Combined with CBAM for Malignant–Benign Pulmonary Nodules Classification on Computed Tomography Images. *Medicina (Lithuania)*, *59*(6). https://doi.org/10.3390/medicina59061088

*name of corresponding author