

Evaluating the Effect of Global Channel Attention on EfficientNetB0 for Distracted Driver Classification under Diverse Lighting Conditions

Zufar Abdullah Rabbani^{*1}, Wahyu Syaifullah J. S², Alfian Rizaldy Pratama³

^{1,2,3}Data Science Departement, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
Email: ¹22083010080@student.upnjatim.ac.id, ²wahyu.s.j.saputra.if@upnjatim.ac.id,
³alfan.fasilkom@upnjatim.ac.id

Abstrak

Klasifikasi aktivitas distraksi pengemudi berbasis citra telah banyak diteliti sebagai upaya menekan risiko kecelakaan lalu lintas, namun sebagian besar penelitian menggunakan data dengan kondisi pencahayaan yang seragam. Padahal, variasi pencahayaan dapat menurunkan kemampuan arsitektur *Convolutional Neural Network* (CNN) dalam mengekstraksi fitur. *Channel attention* (CA) sebelumnya terbukti meningkatkan performa *EfficientNetB0*, tetapi *EfficientNetB0*+CA belum pernah diuji secara mendalam pada klasifikasi aktivitas distraksi pengemudi dengan kondisi pencahayaan beragam. Penelitian ini bertujuan mengevaluasi performa CA pada *EfficientNetB0* dalam klasifikasi aktivitas distraksi pengemudi pada kondisi pencahayaan beragam. Data yang digunakan adalah *State Farm Distracted Driver Dataset* dan data primer sebanyak 10.080 citra dari tiga subjek pada kondisi pencahayaan gelap, semi-terang, dan terang. Data sekunder dan dua subjek primer digunakan untuk pelatihan, sedangkan satu subjek primer yang belum pernah dilihat model digunakan untuk pengujian. Tahap *preprocessing* meliputi perubahan ukuran citra, normalisasi, dan augmentasi data. Model dilatih dengan *transfer learning* dua fase (*backbone* dibekukan, kemudian *fine-tuning*) dan dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Akurasi *EfficientNetB0* (*baseline*) dibandingkan *EfficientNetB0*+CA berturut-turut adalah 38,84% vs 60,36% (gelap), 61,07% vs 32,59% (semi-terang), 67,14% vs 84,73% (terang), dan 55,68% vs 59,23% (seluruh kondisi). CA meningkatkan akurasi pada kondisi gelap dan terang, tetapi menurunkannya pada kondisi semi-terang, sehingga peningkatan performa yang konsisten akibat penambahan CA tidak dapat disimpulkan.

Kata kunci: Channel Attention, EfficientNetB0, Klasifikasi Distraksi Pengemudi

Abstract

Image-based distracted driver classification has been widely studied to reduce the risk of traffic accidents, yet most studies use data captured under uniform lighting conditions. However, illumination variation can degrade feature extraction in Convolutional Neural Network (CNN) architectures. Channel attention (CA) has previously been shown to improve EfficientNetB0, but EfficientNetB0 with CA has not yet been extensively tested for distracted driver classification under diverse lighting conditions. This study aims to evaluate the performance of channel attention on EfficientNetB0 in classifying distracted driver activities under diverse lighting conditions. We used the State Farm Distracted Driver Dataset and a primary dataset of 10,080 images from three subjects captured under dark, semi-bright, and bright lighting conditions. The secondary data and two primary subjects were used for training, while one unseen primary subject was used for testing. Preprocessing included image resizing, normalization, and data augmentation. The models were trained using two-phase transfer learning (frozen backbone followed by fine-tuning) and evaluated using accuracy, precision, recall, and F1-score. The accuracies of baseline EfficientNetB0 versus EfficientNetB0+CA were 38.84% vs 60.36% (dark), 61.07% vs 32.59% (semi-bright), 67.14% vs 84.73% (bright), and 55.68% vs 59.23% (all conditions). CA improved accuracy in dark and bright conditions but reduced it in semi-bright conditions, so a consistent performance gain from adding CA cannot be concluded.

Keywords: Channel Attention, EfficeintNetB0, Distracted Driver Classification

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)



1. INTRODUCTION

Transportation is one thing that is very important for the smooth running of all human activities. The transportation that is frequently used and is increasing is private cars, because they are considered more practical for long trips. However, even though it is considered more practical, using a private cars has several risks, such as traffic accidents due to the driver being distracted by other activities, such as

making calls on a smartphone, drinking, or operating the radio[1], [2]. Accidents can be viewed from three factors, namely road environmental factors, vehicle factors and human factors. Among these three factors, the human factor is often determined to be the main factor causing accidents, around 90% of accidents are caused by human negligence[3], [4].

Advanced Driver Assistance Systems (ADAS), which is an in-vehicle system device interface to support driving safety, was created as an effort to prevent traffic accidents due to distracted driver focus. ADAS can be implemented well using cameras, these cameras are used to analyze activities in driver behavior such as using smartphones, drinking, or operating the radio[5], [6]. Convolutional Neural Network (CNN) is a type of neural network that is commonly used on image data, such as images captured by a camera[7]. In general, Convolutional Neural Network architectures, although they produce a high level of accuracy, require extensive computing resources, making them less suitable for mobile applications or devices with limited capabilities. Because in this context the scenario of real-time recognition of inattentive drivers is very important, and it must also be possible to do it on devices with limited computing resources, because not all mobile devices have extensive computing[8], [9].

In recent years, a number of scientific studies have focused on the development of efficient Convolutional Neural Network (CNN) architectures, one of which is for distracted driver classification, such as EfficientNetB0, which is able to achieve good accuracy even when constrained by limited computing resources[10], [11]. EfficientNetB0 achieves efficiency through a compound scaling methodology that simultaneously balances network depth, width, and input resolution, and it has been reported to extract multiscale features through inverted residual blocks with linear bottlenecks[12], [13], [14]. Each of these blocks is equipped with a Squeeze-and-Excitation (SE) module that recalibrates channel-wise feature responses[14], [15]. Furthermore, the incorporation of attention mechanisms such as Channel Attention (CA) into CNN and EfficientNetB0 architectures has seen a surge in popularity due to their ability to increase the model's sensitivity to important features with minimal parameter addition[10], [16]. In the context of distracted driver classification, EfficientNetB0 combined with CA has been reported to perform better than the backbone alone, and other attention modules have likewise been combined with EfficientNetB0 for this task[8], [10].

However, previous studies on image-based distracted driver classification, including those employing EfficientNetB0 with attention mechanisms, generally used publicly available datasets with relatively uniform lighting conditions and did not evaluate model performance separately under different lighting conditions[10], [17], [18], [19]. This is a limitation, because in real-world conditions people do not only drive under the same lighting conditions (for example, only during the day). It is also known from studies in other vision tasks that variations in lighting intensity, highlights and shadows, and noise due to low light conditions can significantly reduce the accuracy of CNN architectures in feature extraction[20], [21], [22]. Consequently, it remains unknown whether the performance improvement reported for CA on EfficientNetB0 is maintained under diverse lighting conditions. To the best of our knowledge, based on the literature reviewed, no previous study has evaluated EfficientNetB0 with additional CA for distracted driver classification under diverse lighting conditions, and this constitutes the research gap addressed in this study[8], [10], [17], [18], [19].

Therefore, this study uses EfficientNetB0 as the backbone architecture to address the problem of the need for a lightweight and efficient model. An additional CA module is then added for two reasons. First, CA has been reported to improve the performance of CNN architectures in general, including EfficientNetB0 for distracted driver classification on datasets with a single lighting condition[8], [10]. Second, attention mechanisms have been reported to help address illumination-related problems, such as uneven illumination and low-light conditions, in other image-related tasks[23], [24]. Based on these findings, CA is a plausible candidate for improving robustness to lighting variation in distracted driver classification, although its effect in this specific context has not yet been empirically verified.

The main objective of this study is to evaluate the performance of Channel Attention on EfficientNetB0 in classifying distracted driver activities under diverse lighting conditions. To achieve this objective, a baseline EfficientNetB0 and an EfficientNetB0+CA model are trained and compared using image data captured under dark, semi-bright, and bright lighting conditions. The contributions of this study are an evaluation of the effect of lighting conditions on the EfficientNetB0 architecture, and an evaluation of the effect of adding CA on model performance under each lighting condition, specifically for the context of distracted driver classification.

2. RESEARCH METHOD

This study uses the EfficientNetB0 architecture as the backbone model for image-based distracted driving activity classification. The following are the main steps designed to train the model: data collection, data preprocessing, EfficientNetB0 model training, and model evaluation. The research was conducted using the Python Programming Language and the Google Colab Platform.

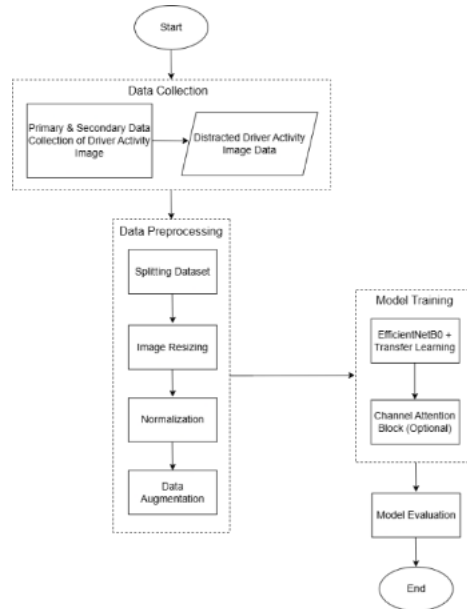


Figure 1. Research Flowchart

2.1. Data Collection

This stage is the initial step in the research, collecting image data of distracted drivers to be used as training and test data. This study used both primary and secondary data, so data collection was conducted using different methods. This study used the secondary dataset, the State Farm Distracted Driver Dataset, collected from the Kaggle.com platform [25]. This dataset has a total of 22,424 images, with 10 activity labels/classes, and each class has 1,900–2,500 images. Then for the primary data used in this study was collected manually using a webcam camera with a resolution of 1920x1080 pixels and 30 frames per second. Every 10 labels/classes have images that have 3 different lighting conditions to represent morning, afternoon, evening, and night conditions. The three different lighting conditions are measured using lux values, bright conditions have a value of 200-2000 lux, semi-bright conditions have a value of 70-200 lux, while dark conditions have a value below 70 lux. Each different lighting condition has a total of 336 images, which makes the total images per class is 1.008 images, and 10.080 total for primary data.

Table 1. Number of Secondary Dataset

State Farm Distracted Driver Dataset (Secondary Dataset)		
Source: https://www.kaggle.com/competitions/state-farm-distracted-driver-detection [25]		
Number	Class Activity	Number of Images
1	c0 (Safe driving)	2.489
2	c1 (Texting-right)	2.346
3	c2 (Calling on the phone-right)	2.326
4	c3 (Texting-left)	2.267
5	c4 (Calling on the phone-left)	2.317
6	c5 (Operating the radio)	2.312
7	c6 (Drinking)	2.325
8	c7 (Reaching behind)	2.002
9	c8 (Hair and makeup)	1.911

10	c9 (Talking to the passenger)	2.129
Total	10 Class Activity	22.424

Table 2. Number of Primary Dataset

Primary Dataset		
Source: Collected manually using a webcam camera with a resolution of 1920x1080 pixels and 30 frames per second.		
Number	Class Activity	Number of Images
1	c0 (Safe driving)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
2	c1 (Texting-right)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
3	c2 (Calling on the phone-right)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
4	c3 (Texting-left)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
5	c4 (Calling on the phone-left)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
6	c5 (Operating the radio)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
7	c6 (Drinking)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
8	c7 (Reaching behind)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
9	c8 (Hair and makeup)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
10	c9 (Talking to the passenger)	1.008 (3 subject x 112 bright, 3 x subject x 112 semi-bright, 3 subject x 112 dark condition)
Total	10 Class Activity	10.080

In the primary dataset, lighting conditions were categorized based on lux measurements obtained using a smartphone light meter application. Measurements were taken near the driver's face and directed toward the webcam to approximate the illumination received by the camera. However, the measurements were not obtained using a professionally calibrated lux meter, which may introduce measurement inaccuracies and variability across lighting conditions.

Primary data were collected from three voluntary participants using a webcam under various lighting conditions. Prior to data collection, all participants were informed about the objectives of the study and voluntarily agreed to the use of their image data for research purposes. The collected data were used exclusively for academic research and were not publicly distributed. Participant privacy and data confidentiality were considered during data collection, handling, and storage.

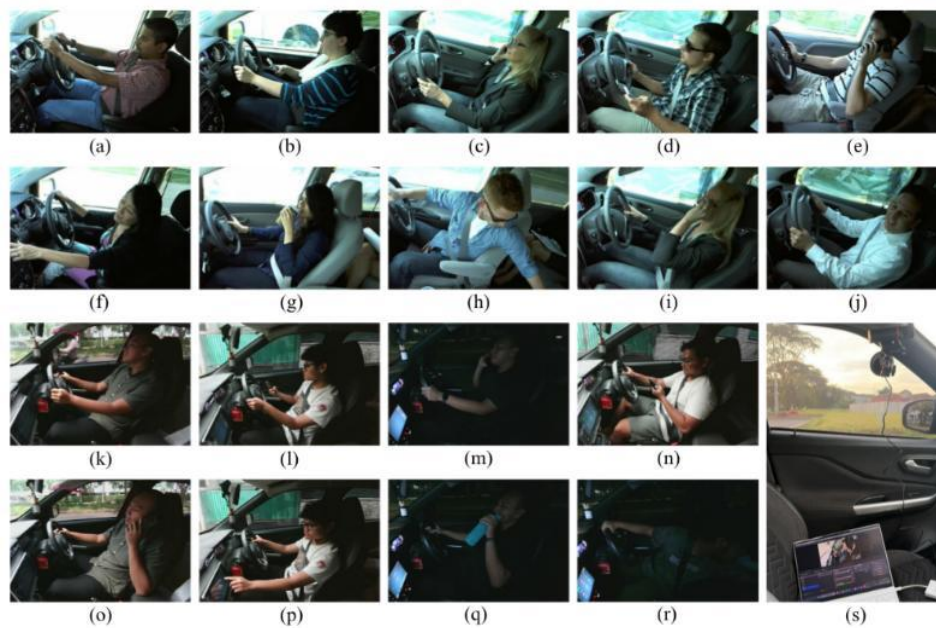


Figure 2. Distracted Driver Activity Image Sample. (a) Safe Driving (c0); (b) Texting-Right (c1); (c) Calling on the Phone-Right (c2); (d) Texting-Left (c3); (e) Calling on the Phone-Left (c4); (f) Operating the Radio (c5); (g) Drinking (c6); (h) Reaching Behind (c7); (i) Hair and Makeup (c8); (j) Talking to the Passenger (c9); (k) Safe Driving Bright (primary dataset); (l) Texting-Right Semi-bright (primary dataset); (m) Calling on the Phone-Right Dark (primary dataset); (n) Texting-Left Semi-bright (primary dataset); (o) Calling on the Phone-Left Bright (primary dataset); (p) Operating the Radio Semi-bright (primary dataset); (q) Drinking Dark (primary dataset); (r) Reaching Behind Dark (primary dataset); (s) Primary Data Collection Setup.

2.2. Data Preprocessing

This stage involves preparing the images for use by the EfficientNetB0 model. The preprocessing process carried out in this study is as follows:

- **Splitting Dataset**
Splitting the dataset is done to split the dataset into two main parts, namely the training set and the testing set. The training set is used to train the model to be able to learn visual patterns from various distracted driver activities, while the testing set is used to evaluate the model's performance with data that has never been seen before. In this study, the training set includes all secondary data and 67% primary data, where 50% comes from the same person subject, because in the primary data there are 3 people subjects used as shown in Figure 2. Therefore, if 2 subject is included in the training set plus secondary data, it can automatically be ensured that the model has never seen the third person subject from the primary data. This is very important because it can prevent data leakage that can cause the model evaluation process to become irrelevant.

Table 3. Dataset Splitting

Class	Train	Val	Test	Total
c0	2529	632	336	3497
c1	2414	604	336	3354
c2	2398	600	336	3334
c3	2351	588	336	3275
c4	2391	598	336	3325
c5	2387	597	336	3320

Class	Train	Val	Test	Total
c6	2398	599	336	3333
c7	2139	535	336	3010
c8	2067	516	336	2919
c9	2241	560	336	3137

- **Image Resizing**
Image resizing is performed to resize the collected images to a uniform size, for example, 224x224 pixels. This is necessary because the EfficientNetB0 model requires a fixed input size, making size equalization crucial before further processing [26].
- **Normalization**
Normalization is performed to align pixel values with the input values expected by the model being used. In this case, the EfficientNetB0 model expects input in the range [0, 255], therefore, normalization is performed to ensure the image data has an appropriate distribution [27].
- **Data Augmentation**
Data augmentation is performed with the aim of creating artificial variations from existing data, this is done to make the model more robust. In this study, the augmentation techniques used include rotation, horizontal flipping, width and height shift, brightness adjustment, zoom, and shear. This is done to make the model more robust against ambiguity [28].

2.3. Model Training

After the preprocessing stage, the next step is training the EfficientNetB0 model. The EfficientNetB0 backbone is trained using Transfer Learning pretrained on ImageNet to minimize the training time and minimize the need for a large dataset. Afterward, fine-tuning is performed to adapt the model to the existing dataset, namely image data of impaired driver activity.

EfficientNetB0 was used in this study because it offers a favorable balance between classification accuracy and computational efficiency, as elaborated below. In addition, this architecture utilizes the swish activation function and Mobile Inverted Bottleneck Convolution (MBConv) blocks combined with Squeeze-and-Excitation (SE) modules, which recalibrate channel-wise feature responses so that more informative channels are emphasized and less relevant ones are suppressed[14], [15]. Furthermore, the use of transfer learning techniques allows the use of pre-trained models that have been previously trained on a specific dataset with similar problems, then used as a starting point. This combination produces a relatively lightweight model, requires a shorter training time compared to other models, and is suitable for devices with limited computing resources in the classification of distracted driving activities based on images.

EfficientNetB0 was selected over other lightweight architectures, such as MobileNetV3 and ShuffleNetV2, for two main reasons. First, a previous study on driver distraction detection[10] compared EfficientNetB0 with several CNN backbones, namely VGG16, ResNet50, Xception, and InceptionV3, together with their channel attention variants, and reported that EfficientNetB0 achieved the best performance for this task. Second, EfficientNetB0 offers a favorable accuracy–efficiency trade-off for deployment on edge devices. As reported in the original studies, EfficientNetB0 has approximately 5.3 million parameters and 0.39 billion FLOPs, and reaches 77.1% top-1 accuracy on ImageNet[14]. MobileNetV3-Large has a similar model size (approximately 5.4 million parameters, 219 million multiply–adds) but a lower top-1 accuracy of 75.2%[29], whereas ShuffleNetV2 1× requires fewer parameters (approximately 2.3 million) and computations (146 MFLOPs), but its top-1 accuracy is lower at 69.4%[30]. For a task that must distinguish visually similar driver activities under difficult lighting conditions, this higher accuracy at a model size still suitable for edge devices was preferred. Nevertheless, a direct comparison with MobileNetV3 and ShuffleNetV2 on distracted driver classification was not performed in this study and is left for future work.

Although the EfficientNetB0 model already implements an channel attention, SE, in each block, this study added a global channel attention to examine whether it improves model performance on datasets with varying lighting conditions, particularly dark conditions. Therefore, there are EfficientNetB0 models trained with the addition of a global channel attention and also those without the

addition of a global channel attention, only the EfficientNetB0 backbone. This was done to determine the effect of the channel attention on model performance.

Table 4. Training Hyperparameters

Hyperparameter	Phase 1 (Frozen)	Phase 2 (Fine-tuning)
Epoch phase	1 – 20	21 – 50
Learning rate	0.001	0.00001
Optimizer	Adam	Adam
Backbone trainable	False (all layer frozen)	Partial (30 last layer)
Total epoch maximum	50	50
Batch size	16	16

Table 4 presents the hyperparameters used in the two-phase training strategy for all model variants. Phase 1 (epochs 1–20) freezes the EfficientNetB0 backbone and trains only the classification head using Adam with a learning rate of 0.001. Phase 2 (epochs 21–50) applies partial fine-tuning by unfreezing the last 30 backbone layers and reducing the learning rate to 0.00001 to minimize excessive changes to pretrained weights. Training can stop earlier if no improvement in accuracy occurs for 10 consecutive epochs.

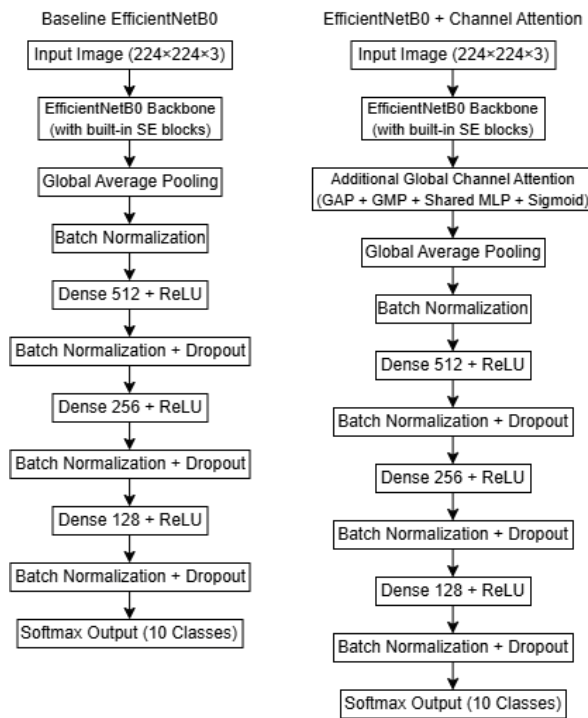


Figure 3. Architecture Block Diagram

The proposed attention module does not replace the original Squeeze-and-Excitation (SE) blocks embedded within EfficientNetB0. Instead, an additional global channel attention module is appended after the EfficientNetB0 backbone output to further refine high-level feature representations prior to classification. The attention module was positioned after the backbone to enhance global semantic feature refinement while preserving the original EfficientNetB0 architecture and pretrained weights. Previous studies reported that EfficientNetB0 combined with additional attention mechanisms achieved promising performance on distracted driver classification using publicly available datasets with relatively consistent illumination conditions. Therefore, this study extends the investigation by evaluating the effectiveness of the approach under more diverse real-world lighting conditions collected from primary data [10].

2.4. Model Evaluation

The evaluation method used in this study aims to measure the performance of several trained models in classifying distracted driving activities based on images. The evaluation was conducted using metrics such as accuracy, precision, recall, f1-score, and confusion matrix. The evaluation was conducted on the entire testing set and under each lighting condition to determine the model's performance under specific lighting conditions.

Accuracy is used to determine the proportion of correct predictions across the entire testing set. However, since accuracy alone is not always sufficient, precision and recall are also used. Precision measures correct positive predictions (the classification accuracy of each class), while recall indicates the actual number of samples successfully recognized by the model. The F1-score is also calculated to provide a more comprehensive evaluation, as the F1-score is the average of precision and recall, making it very useful for assessing the balance of model performance across all classes [31].

In addition to numerical metrics, a confusion matrix is also used to observe patterns of misclassification between classes. The confusion matrix visually depicts the number of correct and incorrect predictions for each class, allowing us to identify which classes the model frequently mispredicts compared to the others [31]. With this combination of metrics, model evaluation becomes more comprehensive and can be used to assess the effectiveness of the model in accurately recognizing various distracted driver activities.

Furthermore, in this study, evaluations were conducted for 2 models, namely EfficientNetB0 (without additional global channel attention) and EfficientNetB0+CA (adding global channel attention). The two models will be evaluated and compared in performance, this is done to determine the effect of additional global channel attention on the EfficientNetB0 backbone. So that later we will get complete evaluation and comparison results for each model in each lighting condition in the context of image-based distracted driver activity classification.

3. RESULTS AND DISCUSSION

The following are the results and discussion of research using the distracted driver classification modeling methodology using EfficientNetB0 with additional Channel Attention.

3.1. Data Preprocessing Results

Before the image classification process is carried out, the data image goes through a series of preprocessing stages first. These steps include dataset splitting, image resizing & normalization, and data augmentation. All of these processes are designed to prepare the data for further processing, namely model training.

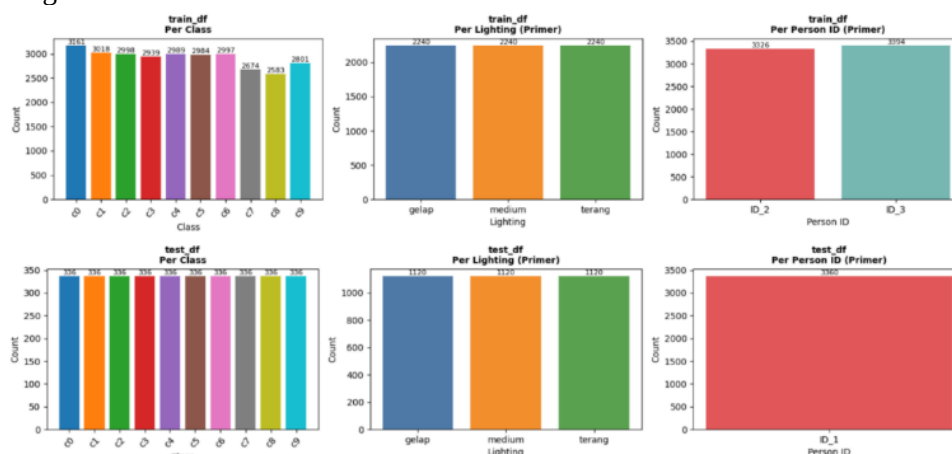


Figure 4. Results of splitting the dataset

As previously stated, the dataset is splitted into two main parts: the training set and the testing set. In this study, the training set (train_df) consists of all secondary data plus 2 human subjects in the primary data, and it also includes 3 different lighting conditions for each human subject. Then, the testing set (test_df) contains the remainder of the primary data, namely 1 different human subject that is

not included in the training set. Because, as previously explained, the primary data is from 3 different human subjects, of the three human subjects, 2 human subjects are used for the training set along with the secondary data, and 1 human subject for the testing set. The diagram in the upper left corner of Figure 3 shows the number of images in each class after combining the secondary data and primary data for the training set. The number of images per class for the primary data added to the training set is 672 images, which means the number of secondary data per class is the number in the diagram minus 672 (e.g., secondary c0 = $3161 - 672 = 2489$). That's why the diagram in the bottom left corner has 336 images for all classes, because for the primary data, each person has 336 images for each class. Then, the middle diagram shows that each person in the primary data has the same 3 lighting conditions with the same number, namely 1120 for each lighting condition for each person. Then, the rightmost diagram shows that the training set added people with IDs 2 and 3, while the testing set used person with ID 1. Below are samples of the image data used in both datasets.

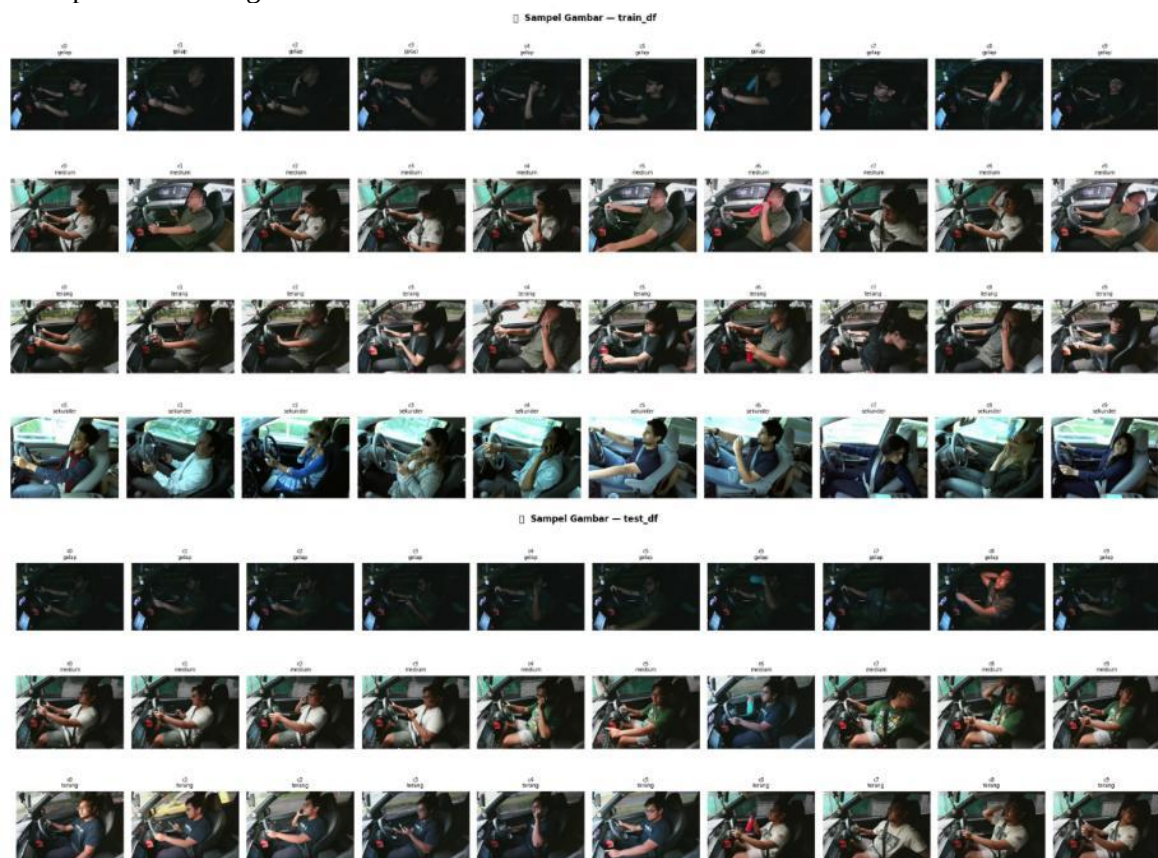


Figure 5. Sample Data Used in the Training Set and Testing Set

After dividing the dataset and viewing data samples, the next result produced by this preprocessing process is data augmentation. The augmentation techniques used in this research include:

- Rotating 15 degrees, rotate the image randomly to the left or right up to a maximum of 15 degrees. This technique helps the model still recognize objects even though their position is tilted.
- Horizontal Flip, flips the image horizontally (like looking in a mirror). This is very commonly used because objects (such as faces or objects) usually retain the same meaning even if they are facing in opposite directions.
- Width & Height Shift 0.1, shifts the image horizontally or vertically by 10% of the total width/height of the image. This trains the model not to only focus on objects that are right in the center of the frame (translation invariance).

- Brightness Adjustment 0.8, 1.2, changes the light intensity of the image randomly. A range of 0.8 means the image can be 20% darker, and 1.2 means it can be 20% brighter. This helps the model deal with lighting variations in the real world.
- Zoom 0.15, zoom in or zoom out randomly by 15%. This simulates different camera distances to the object.
- Shear 0.1, shifts the corners of the image so that the image looks like it was pulled diagonally (corner distortion of 10%). This helps the model recognize objects from different points of view (perspectives).
- Fill Mode 'nearest', When the image is rotated or panned, there will be an empty area at the edge of the image. 'nearest' will fill the empty area with the closest pixel color so that the image does not have jet black areas that could interfere with the training process.

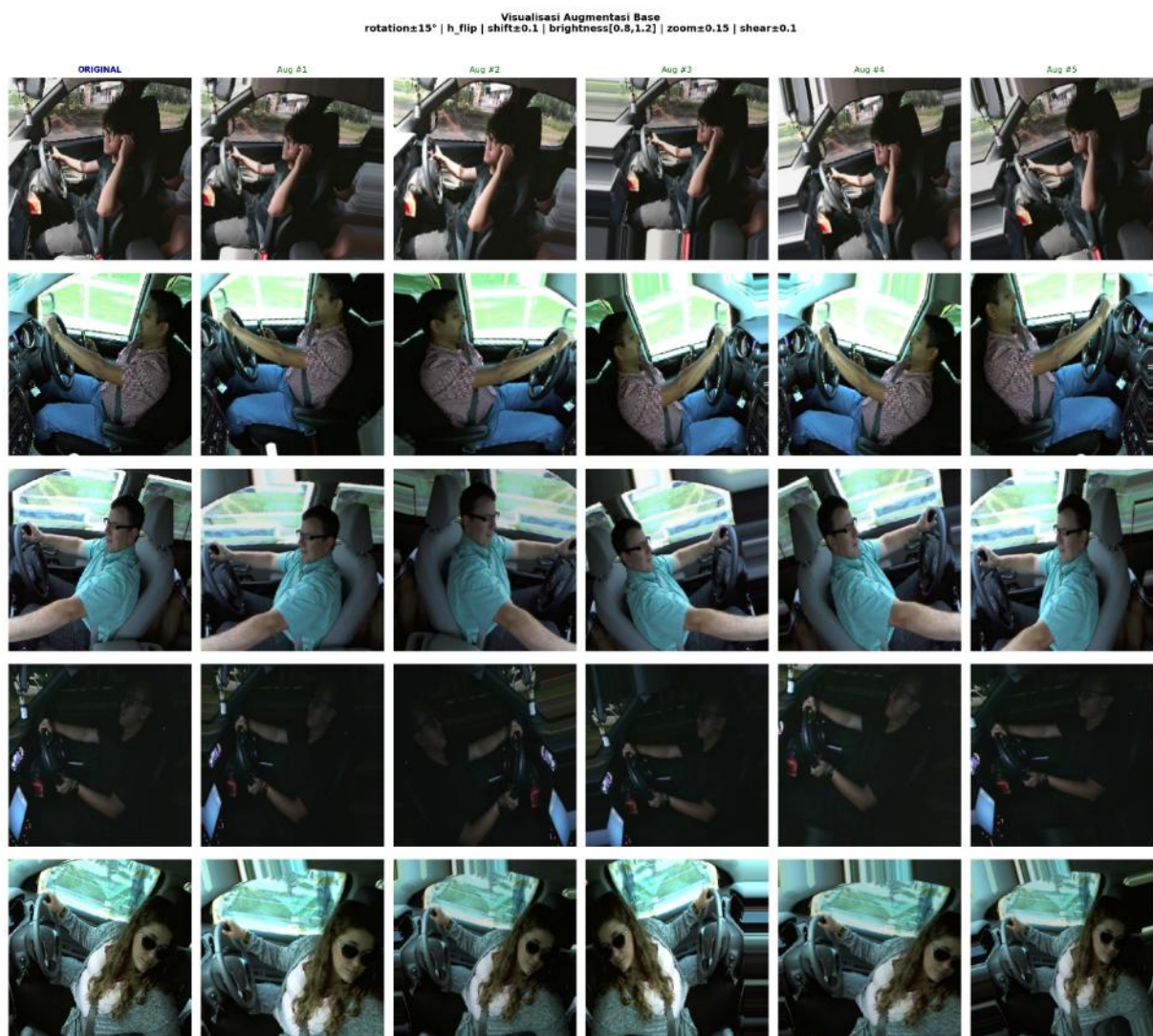


Figure 6. Data Augmentation Visualization Results

3.2. Model Training

This study implements the EfficientNetB0 architecture as a backbone with several configuration variants. There are two variants for the trained model, namely the model with Channel Attention (CA) and without CA, so that overall there are 2 model variants trained and tested. The complete model architecture consists of several layers arranged sequentially. After the EfficientNetB0 backbone layer (pretrained ImageNet), the output features pass through the GlobalAveragePooling2D layer, followed by three consecutive Dense blocks with neuron sizes of 512, 256, and 128. Each Dense block is equipped

with BatchNormalization and Dropout with different regularization levels (0.5, 0.4, and 0.3 respectively) to prevent overfitting. L2 regularization with a value of 0.001 is also applied to each Dense layer. The output layer uses the softmax activation function for multiclass classification.

The implemented Channel Attention module adopts a modified Squeeze-and-Excitation mechanism. The CA implementation works by combining two parallel processing paths, namely Global Average Pooling and Global Max Pooling, to summarize spatial information from all feature maps. The results of both paths are then processed through a Shared MLP consisting of three Dense layers, where each Dense layer in the hidden part is equipped with BatchNormalization and ReLU activation. The outputs of both paths are summed (element-wise addition) and activated with a sigmoid function to produce attention weights per channel. These weights are then multiplied (scaled) again with the input feature map so that the model can selectively highlight the most informative channels.

The training process adopts a two-phase training strategy commonly used in transfer learning approaches:

- Phase 1: Frozen Backbone (Epochs 1–20). In this phase, all EfficientNetB0 backbone weights are frozen, allowing only the head layers to be trained. The learning rate used is 0.001. This strategy aims to allow the randomly initialized head weights to reach a stable state without disrupting the pretrained feature representations of the backbone.

- Phase 2: Fine-tuning (Epochs 21–50). After the 20th epoch, the backbone is partially unwrapped. Only the last 30 layers of EfficientNetB0 are activated for training, while the initial layers remain frozen. The learning rate is drastically reduced to 0.00001 to avoid catastrophic forgetting of the low-level feature representations learned from ImageNet.

Both variations of the model were developed utilizing the Adam optimization algorithm in conjunction with a categorical cross-entropy loss function, which is deemed appropriate for multiclass classification endeavors employing one-hot encoding of labels. Evaluation metrics during training included accuracy, precision, and recall, calculated as a weighted average to accommodate potential imbalance in class distribution. The training process is controlled by callback mechanisms such as EarlyStopping and ReduceLROnPlateau. EarlyStopping monitors validation accuracy (the ratio between training set and validation set is 80:20%) with a patience of 10 epochs. In the absence of any enhancement over a span of 10 successive epochs, the training process is halted automatically, and the optimal weights are reinstated. The ReduceLROnPlateau function observes the validation loss and diminishes the learning rate by a factor of 0.5 if there is no reduction in loss over a duration of 5 consecutive epochs, with a stipulated minimum learning rate threshold of 1×10^{-6} . This mechanism allows the model to adaptively exit the optimization plateau without manual intervention.

3.3. Model Evaluation Results

After the model training process is complete, the model is evaluated using primary data that the model has never seen before, as explained previously. The following table shows the evaluation results of the two model variations using the evaluation metrics of accuracy, precision, recall, and f1-score. The two models are also evaluated under different lighting conditions (bright, semi-bright, dark), as well as across the entire testing set, meaning all lighting conditions.

Table 5. Model Evaluation Results

		EfficientNetB0 (No CA)	EfficientNetB0+CA (With CA)
Dark	Accuracy	0.3884	0.6036
	Precision	0.4405	0.6748
	Recall	0.3884	0.6036
	F1-Score	0.3289	0.5702
Semi-bright	Accuracy	0.6107	0.3259
	Precision	0.7051	0.5805
	Recall	0.6107	0.3259
	F1-Score	0.5885	0.2836

Bright	Accuracy	0.6714	0.8473
	Precision	0.6522	0.8621
	Recall	0.6714	0.8473
	F1-Score	0.6342	0.8454
All Conditions	Accuracy	0.5568	0.5923
	Precision	0.5781	0.7080
	Recall	0.5568	0.5923
	F1-Score	0.5398	0.6010

Tables 5 show the evaluation results of the two trained models. The results indicate that the EfficientNetB0+CA model performs slightly better overall than the baseline EfficientNetB0 model without additional channel attention. The proposed model achieves higher accuracy under both bright and dark lighting conditions, contributing to improved overall performance. However, a notable decrease in accuracy is observed under semi-bright conditions, limiting its consistency across all lighting environments. One possible explanation is that EfficientNetB0 already incorporates built-in Squeeze-and-Excitation (SE) attention blocks, and the addition of another channel attention module may introduce redundant feature recalibration. This additional attention mechanism may excessively suppress certain feature channels that remain important in semi-bright conditions, where illumination characteristics are more ambiguous and transitional compared to clearly dark or bright environments. This behavior suggests that the additional attention mechanism may improve feature selectivity under extreme illumination conditions while reducing generalization capability under intermediate lighting conditions. The following confusion matrix analysis is presented to further examine the classification errors between classes.

Each figure presents the raw counts (left) and the row-normalized proportions (right), where the diagonal represents the per-class recall. Every class contains 112 test images in each lighting condition and 336 images when all conditions are combined, so the off-diagonal cells show which classes absorb the errors of a given class. Figures 7–10 present the baseline EfficientNetB0 and Figures 11–14 present EfficientNetB0+CA, each in the order of dark, semi-bright, bright, and all lighting conditions.



Figure 7. Confusion Matrix Model EfficientNetB0 In Dark Condition

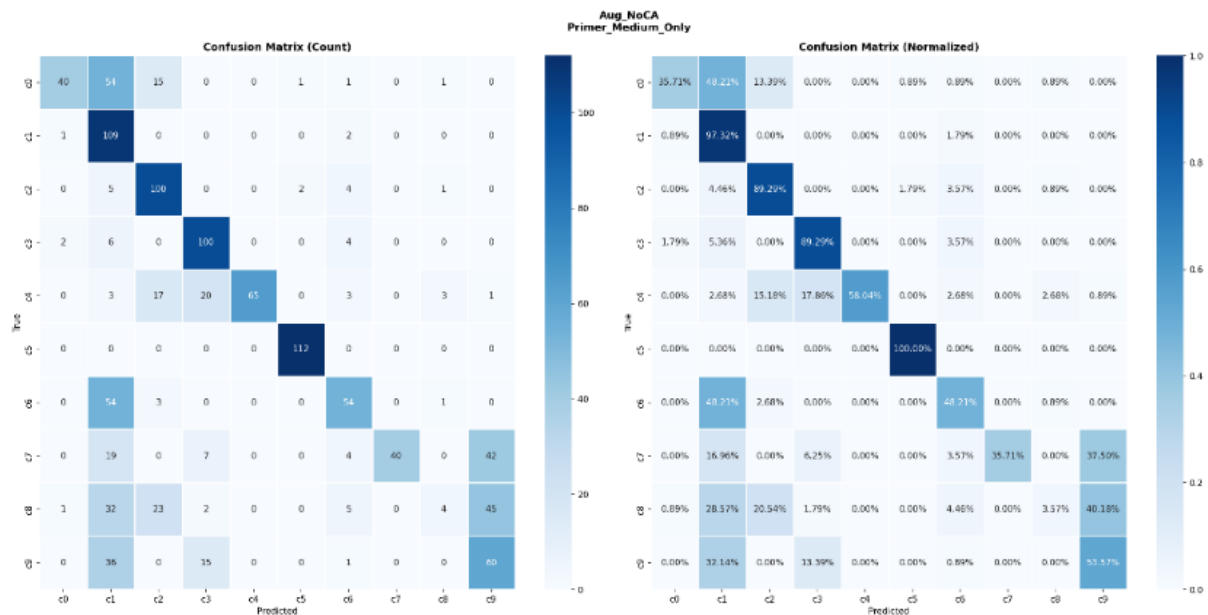


Figure 8. Confusion Matrix Model EfficientNetB0 In Semi-bright Condition



Figure 9. Confusion Matrix Model EfficientNetB0 In Bright Condition

Figure 7 shows that the baseline model collapses in dark conditions, none of the 112 safe driving (c0) images is classified correctly, and about one third of all predictions (360 of 1,120) are assigned to c7 (reaching behind). The texting-left class (c3) is also almost entirely lost, with 92 of 112 images predicted as c8 (hair and makeup). Figure 8 shows that under semi-bright conditions the baseline still recognizes the phone and radio classes (c1, c2, c3, and c5) well, while errors are mainly found in c0, c6, c7, and c8, which are confused with c1 or c9, c8 is the weakest, with only 4 correct images. Figure 9 shows that in bright conditions most classes are recognized well, and the most notable error is between c9 (talking to the passenger) and c0 (safe driving), which is 53 of the 112 c9 images are predicted as c0 and only 4 are classified correctly. The left-hand phone class (c4) is also confused with c6 and c8 in this condition. Figure 10 summarizes all conditions, which is the baseline recognizes c1 (97.6% recall) and c5 best, whereas c0, c8, and c9 each remain at around 25% recall, indicating that these classes are the main source of the baseline's overall accuracy of 55.68%.

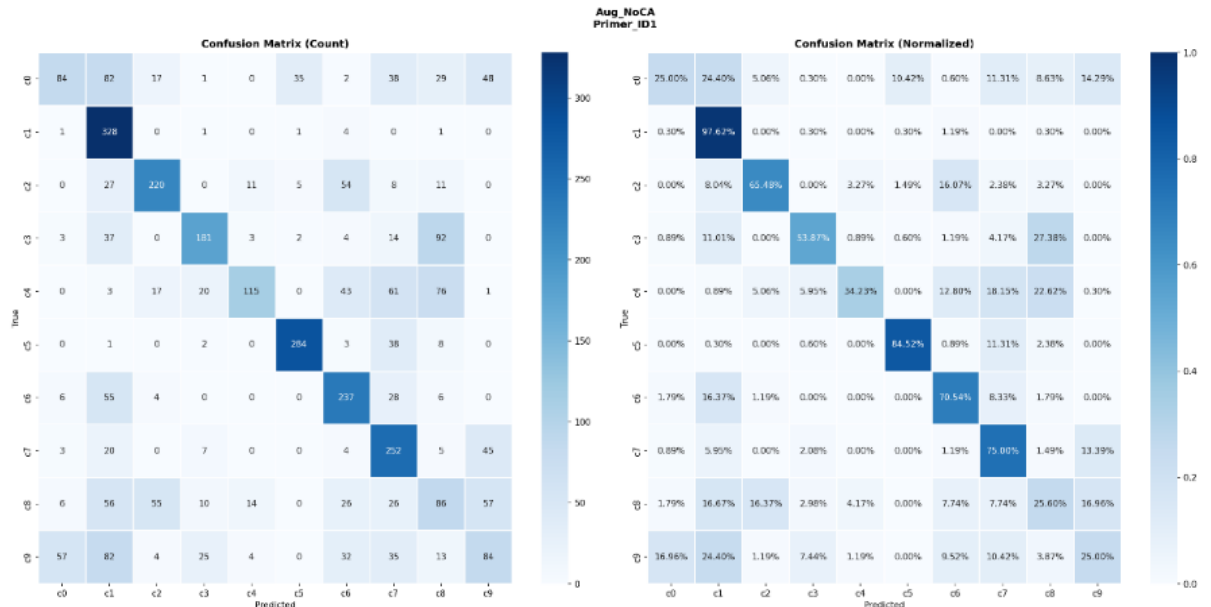


Figure 10. Confusion Matrix Model EfficientNetB0 In All Lighting Conditions

Note on the number of images. The primary dataset used in this study consists of 10.080 images collected from three subjects, which equals 1.008 images per class. Each subject contributed 336 images per class, namely 112 images for each of the dark, semi-bright, and bright conditions, as summarized in the primary dataset in Table 2. Two subjects (IDs 2 and 3), totaling 6.720 images (672 per class), were included in the training set together with the secondary data, while one subject (ID 1), totaling 3.360 images (336 per class), was used as the testing set. Therefore, the figure of 6.720 images refers only to the primary images in the training set, not to the entire primary dataset. All confusion matrices in Figures 7–14 are computed on the testing set only, so each class contains 112 images in each lighting condition (1.120 images per condition) and 336 images when all lighting conditions are combined (3.360 images in total). Since the same testing set was used for all evaluations, the number of images in the confusion matrices is not affected by the clarification of the primary data count, which concerns the description of the total and training data only.

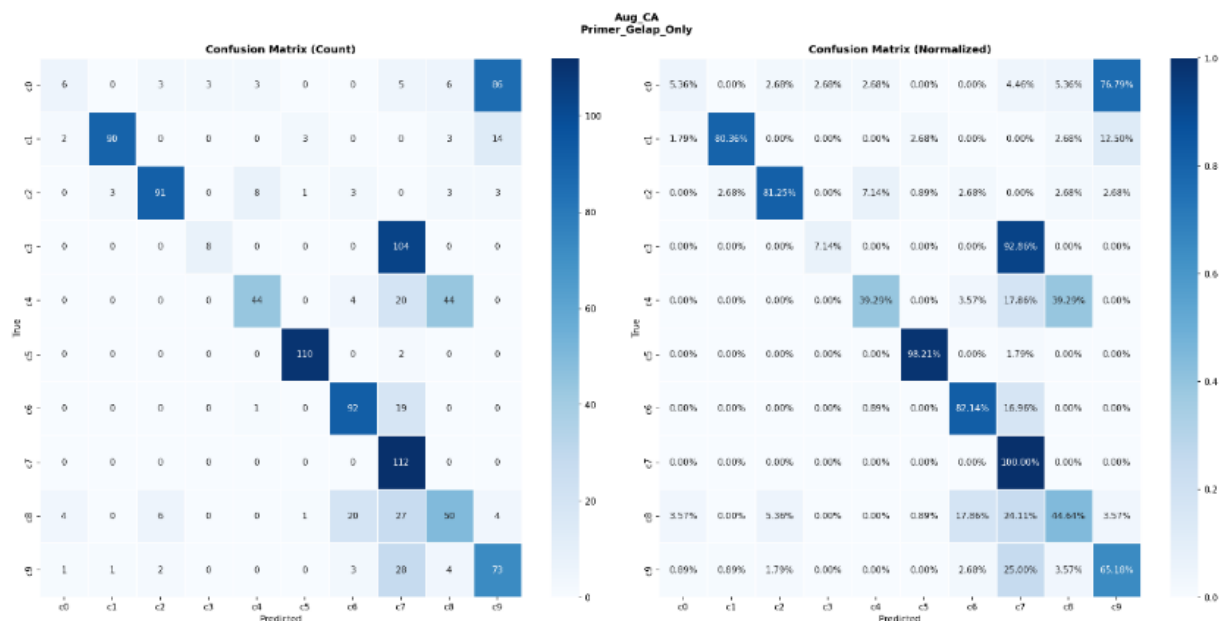


Figure 11. Confusion Matrix Model EfficientNetB0+CA In Dark Condition



Figure 12. Confusion Matrix Model EfficientNetB0+CA In Semi-bright Condition

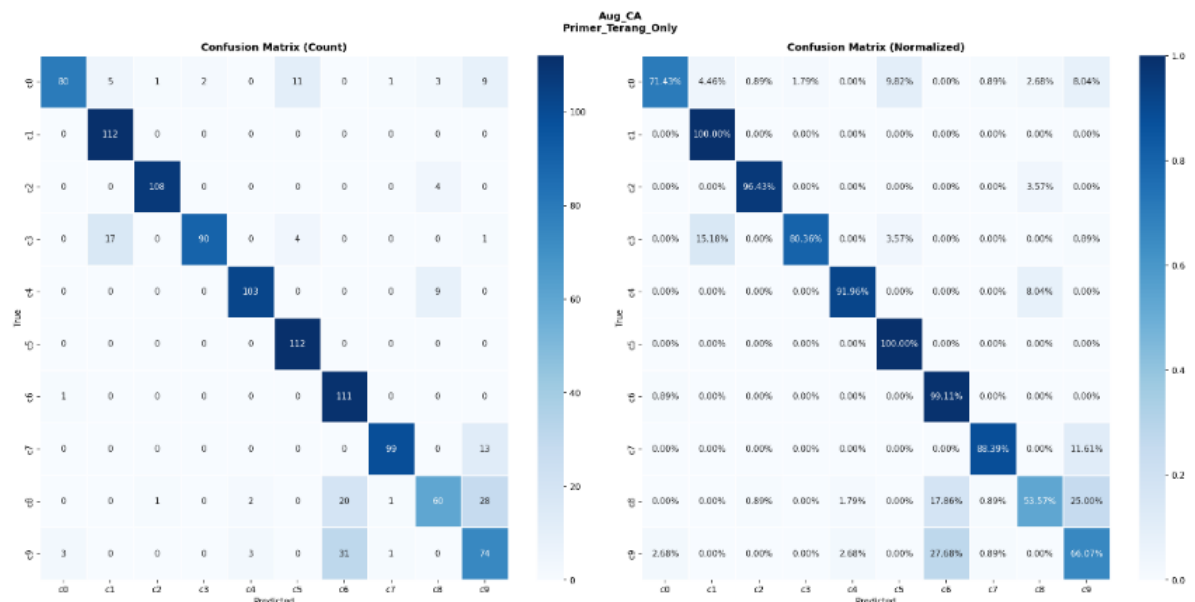


Figure 13. Confusion Matrix Model EfficientNetB0+CA In Bright Condition

Figure 11 shows that, in dark conditions, adding CA reduces the collapse toward c7 seen in Figure 7 and substantially improves several classes, notably c2 (from 14 to 91 correct images), c9 (from 20 to 73), c5 (from 63 to 110), and c4 (from 9 to 44), although c1 decreases slightly (from 109 to 90). However, the model still correctly classifies only 6 of the 112 safe driving (c0) images, most of which are predicted as c9 (86 images), and c3 is still largely predicted as c7 (104 images). The higher dark-condition accuracy therefore comes from the distracted classes rather than from safe driving. Figure 12 shows the opposite behavior in semi-bright conditions, where predictions collapse onto c6 (drinking) and c9, which together receive almost three quarters of all predictions (498 and 325 of 1,120). Class c6 alone absorbs 87, 100, and 103 images of c0, c1, and c2, respectively, while c8 and c4 are almost never recognized (0 and 3 correct images). Figure 13 shows that in bright conditions six classes (c1, c2, c4, c5, c6, and c7) are recognized with at least 99 of 112 correct images, and the remaining errors are mainly found in c8 (28 images predicted as c9 and 20 as c6), c9 (31 images predicted as c6), and, to a lesser degree, c0 and c3. Figure 14 shows that across all conditions the gain from CA mainly originates from c9 (from 84 to 250 correct images), with smaller gains for c4, c8, c3, and c6, while c1 decreases from 328 to 214 correct images because 100 of its images are misclassified as c6. Safe driving (c0) remains

the weakest class, with 92 of 336 images (27.4%) classified correctly, compared with 84 (25.0%) for the baseline.

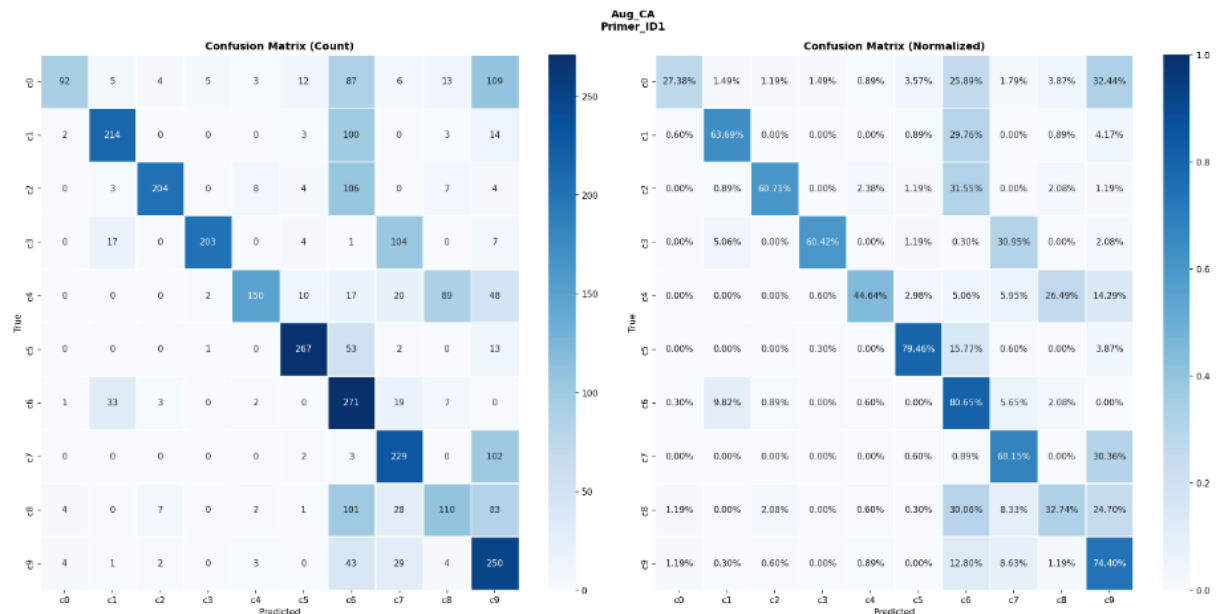


Figure 14. Confusion Matrix Model EfficientNetB0+CA In All Lighting Conditions

3.4. Discussion

This section discusses why the results in Section 3.3 were obtained and how they compare with previous studies. Overall, the addition of CA to EfficientNetB0 increased accuracy by 21.52 percentage points in dark conditions and by 17.59 percentage points in bright conditions, but decreased it by 28.48 percentage points in semi-bright conditions, resulting in a modest overall improvement of 3.55 percentage points across all lighting conditions. This indicates that the effect of CA in this study depends on the lighting condition rather than being uniformly beneficial.

The improvement under dark and bright conditions may be explained by the more consistent illumination characteristics of these two conditions. In dark images, many feature channels are likely to respond to noise and low-contrast regions, and CA may help the model to give more weight to channels that carry activity-related cues. This is consistent with the comparison of Figures 7 and 11, where the strong collapse of predictions toward class c7 in the baseline model is reduced, and several classes are recognized far better, such as c2 (from 14 to 91 correct images) and c9 (from 20 to 73 correct images). In bright conditions, objects and body parts such as the phone, bottle, and hands are clearly visible, and the channel recalibration may further sharpen the separation between classes, as shown by the increase of the F1-score from 0.6342 to 0.8454 and by six classes being recognized with at least 99 out of 112 correct images in Figure 13. These explanations should nevertheless be interpreted with caution, because the improvement is not uniform across classes, as discussed below.

However, the higher accuracy of EfficientNetB0+CA in dark conditions does not necessarily mean that the model is good overall. This model only managed to correctly classify 6 of the 112 c0 (safe driving) images (Figure 11), which means that even when the driver is driving safely in dark conditions (e.g. at night), the model is more likely to identify the driver as distracted. This misclassification of class c0 occurs not only in the EfficientNetB0+CA model, but also in the baseline model, which did not classify any c0 image correctly in the same condition (Figure 7). This indicates that the main challenge is not only caused by the additional attention mechanism, but is also closely related to the visual characteristics of class c0 itself under low lighting conditions. Unlike classes related to distractions that contain distinctive movements or objects, the safe driving class relies mostly on subtle visual cues such as the placement of both hands on the steering wheel, and this is similar to class c9 (talking to the passenger), which also places both hands on the steering wheel but with the head looking at the rear passenger. This is supported by Figures 9 and 11, where 53 of the 112 c9 images are predicted as c0 by the baseline model in bright conditions, and 86 of the 112 c0 images are predicted as c9 by

EfficientNetB0+CA in dark conditions. The absence of strong movement patterns in this class, combined with dark lighting, makes the important visual cues less distinguishable, causing the model to rely on ambiguous high-level features.

The decrease in performance under semi-bright conditions may be related to the characteristics of this lighting condition in the primary data. Semi-bright illumination is transitional, and it may involve oblique light directions and light reflections (for example, in the late afternoon) that create uneven highlights and shadows on the driver and the cabin. Such patterns can disturb the feature extraction of the model, so that the features of some activities become less representative. Since CA is designed to emphasize feature channels that appear important, it is possible that it also amplified channels responding to these illumination artifacts instead of activity-related cues, making the problem more severe than in the baseline model. This interpretation is consistent with Figure 12, where the predictions of EfficientNetB0+CA concentrate on only two classes, c6 and c9, which receive 498 and 325 of the 1,120 predictions, respectively, while class c6 alone absorbs 87, 100, and 103 images of classes c0, c1, and c2. In addition, since EfficientNetB0 already contains built-in Squeeze-and-Excitation (SE) blocks, the additional channel attention may introduce redundant feature recalibration that suppresses subtle contextual information needed to distinguish activities under these lighting conditions. It should be noted that these explanations are based on the observed results and the characteristics of the primary data. They have not been verified through controlled experiments, for example with a fixed light direction, future research could conduct this test for further development.

Compared with the previous study[10], which reported performance improvements of EfficientNetB0 with CA on public datasets, the benefit of CA in this study is not consistent across lighting conditions. Because the backbone, the CA design, and the number of training epochs in this study are the same as in[10], the difference is unlikely to be caused by the architecture or the training configuration. A more plausible explanation is the data used. The previous study used open-source datasets with relatively uniform lighting conditions, whereas this study added primary data with dark, semi-bright, and bright conditions and evaluated the models on a subject that was not included in the training set. Under these conditions, the improvement obtained by CA under uniform lighting does not necessarily carry over to all lighting conditions, particularly the semi-bright condition. This study cannot separate the effect of lighting variation from other characteristics of the primary data, such as camera position, cabin environment, and the subjects themselves, since these factors vary together. In addition, the test set consists of only one subject, so some of the differences may also be affected by subject-specific appearance. Therefore, these findings should be understood as showing that the effectiveness of CA depends on the conditions of the data, rather than contradicting the results of[10]. They also suggest that improvements reported under uniform lighting should be verified under the lighting conditions expected in real use, and that per-class metrics, particularly for the safe driving class, should be reported in addition to overall accuracy.

4. CONCLUSION

This study aimed to evaluate the performance of channel attention on EfficientNetB0 in classifying distracted driver activities under diverse lighting conditions. This objective was achieved, as both the baseline EfficientNetB0 and the EfficientNetB0+CA models were trained and evaluated under dark, semi-bright, and bright conditions, as well as across all conditions combined. However, this evaluation did not result in a model that performs well under all lighting conditions.

The evaluation shows that the effect of CA depends on the lighting condition. CA increased accuracy in dark conditions (from 38.84% to 60.36%) and in bright conditions (from 67.14% to 84.73%), but decreased it in semi-bright conditions (from 61.07% to 32.59%). Across all conditions combined, accuracy increased only slightly, from 55.68% to 59.23%. The best result in each condition was achieved by EfficientNetB0+CA in dark (60.36%) and bright (84.73%) conditions, and by the baseline EfficientNetB0 in semi-bright conditions (61.07%).

Because neither model outperformed the other across all lighting conditions and each performs poorly in at least one condition, neither can be considered sufficiently robust for distracted driver classification, where unreliable predictions may endanger the driver and passengers. Therefore, it cannot be concluded that adding CA consistently improves the performance of EfficientNetB0 in this context.

For further research, it is recommended to explore the influence of attention mechanisms further on the EfficientNetB0 model or other CNN architectures for the context of distracted driver classification under varying lighting conditions. Because in this study the addition of attention mechanisms is only limited to using Channel Attention only, so a combination of further models is needed to determine more clearly and definitively the influence of attention mechanisms on CNN architectures for distracted driver classification under varying lighting conditions. Evaluation of the use of other lightweight CNN architectures along with further data augmentation techniques can also be carried out to improve the performance of the classification model. Future work may also employ professionally calibrated lux meters and more controlled lighting environments to obtain more objective and standardized illumination measurements. Future work should also repeat the experiments with more subjects and apply statistical tests so that differences between models can be assessed more rigorously.

REFERENCES

- [1] D. Amalia and F. Utaminigrum, "Deteksi Kantuk pada Pengemudi melalui Jumlah Kedipan Mata Menggunakan Facial Landmark berbasis Intel NUC," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 12, pp. 5529–5535, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] H. Suseno, "Kecelakaan Menghambat Tumbuh dan Berkembang Suatu Bangsa," *Jurnal Indonesia Sosial Sains*, vol. 3, no. 10, Oct. 2022, doi: <https://doi.org/10.59141/jiss.v3i10.716>.
- [3] S. Firdaus and K. D. Artika, "DETEKSI KELELAHAN PENGEMUDI MOBIL MENGGUNAKAN CITRA WAJAH," *ELEMEN : JURNAL TEKNIK MESIN*, vol. 8, no. 1, pp. 16–21, Jun. 2021, doi: 10.34128/je.v8i1.154.
- [4] Armizoprades, A. Fadhli, A. Azhar, and T. Fauzan, "ANALYSIS OF FACTORS CAUSING TRAFFIC ACCIDENTS IN THE KM 45 SICINCIN AREA, PADANG PARIAMAN REGENCY," *Journal of Scientech Research and Development*, vol. 7, no. 1, Jun. 2025, doi: <https://doi.org/10.56670/jsrd.v7i1.1161>.
- [5] G. Hermawan, "Klasifikasi Pengemudi Terganggu Berdasarkan Citra Menggunakan ResNet-50," *Jurnal Informatika Komputer, Bisnis dan Manajemen*, vol. 22, no. 3, 2024, doi: <https://doi.org/10.61805/fahma.v22i3.137>.
- [6] N. Darapaneni *et al.*, "Distracted Driver Monitoring System Using AI," in *2022 Interdisciplinary Research in Technology and Management (IRTM)*, IEEE, Feb. 2022, pp. 1–8. doi: 10.1109/IRTM54583.2022.9791545.
- [7] M. Faizal Nazili, A. B. Firmansyah, and R. Purbaningtyas, "Klasifikasi Keparahan Demensia Alzheimer Menggunakan Metode Convolutional Neural Network pada Citra MRI Otak," *Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 1–7, 2023.
- [8] X. Shi, F. Li, G. Lu, Y. Xie, F. Dong, and K. Chen, "A Recognition Algorithm for Distracted Driving Behavior Based on CBAM-EfficientNetB0," in *The 4th International Conference on Frontiers of Electronics, Information and Computation Technologies*, 2025, pp. 127–138. doi: 10.1007/978-981-96-5318-8_12.
- [9] H. Cai *et al.*, "Enable Deep Learning on Mobile Devices: Methods, Systems, and Applications," *ACM Transact. Des. Autom. Electron. Syst.*, vol. 27, no. 3, pp. 1–50, May 2022, doi: 10.1145/3486618.
- [10] T. Khan, G. Choi, and S. Lee, "EFFNet-CA: An Efficient Driver Distraction Detection Based on Multiscale Features Extractions and Channel Attention Mechanism," *Sensors*, vol. 23, no. 8, Apr. 2023, doi: 10.3390/s23083835.
- [11] Z. Zhao, "A comparative study of large-scale and lightweight convolutional neural networks for ImageNet classification," *Applied and Computational Engineering*, vol. 47, no. 1, pp. 101–110, Mar. 2024, doi: 10.54254/2755-2721/47/20241236.
- [12] D. Akhiyar, H. Herasmus, and Nofriadiman, "ADAPTASI SPESIFIK EFFICIENTNETB0 DENGAN LAPISAN KUSTOM UNTUK IDENTIFIKASI BUAH TROPIS," *Jurnal Sains Informatika Terapan*, vol. 4, no. 2, pp. 308–313, 2025, Accessed: Sep. 14, 2025. [Online]. Available: <https://rcf-indonesia.org/home/>
- [13] Sanjay Balwani and Narendra Bawane, "A Multiscale Fusion Network Integrating ConvNeXtSmall and EfficientNetB0 with Enhanced Image Preprocessing for Robust

- Classification,” *Metallurgical and Materials Engineering*, vol. 31, no. 3, pp. 382–391, Mar. 2025, doi: 10.63278/1388.
- [14] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 2020. [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [15] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE/CVF, 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.
- [16] S. Bala et al., “Improving Kidney Stone Detection with YOLOV10 and Channel Attention Mechanisms in Medical Imaging,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 3, pp. 951–963, Jul. 2025, doi: 10.35882/jeeemi.v7i3.868.
- [17] M. Uzzol Hossain, M. Ataur Rahman, M. Manowarul Islam, A. Akhter, M. Ashraf Uddin, and B. Kumar Paul, “Automatic driver distraction detection using deep convolutional neural networks,” *Intelligent Systems with Applications*, vol. 14, p. 75, 2022, doi: 10.1016/j.iswa.2022.20.
- [18] J. Wang and Z. Wu, “Model Lightweighting for Real-time Distraction Detection on Resource-Limited Devices,” *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, Dec. 2022, doi: 10.1155/2022/7360170.
- [19] Y. Lin, D. Cao, Z. Fu, Y. Huang, and Y. Song, “A Lightweight Attention-Based Network towards Distracted Driving Behavior Recognition,” *Applied Sciences (Switzerland)*, vol. 12, no. 9, May 2022, doi: 10.3390/app12094191.
- [20] L. Pasqualetto Cassinis, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, “On-ground validation of a CNN-based monocular pose estimation system for uncooperative spacecraft: Bridging domain shift in rendezvous scenarios,” *Acta Astronaut.*, vol. 196, pp. 123–138, Jul. 2022, doi: 10.1016/j.actaastro.2022.04.002.
- [21] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, “Vision-based human action recognition: An overview and real world challenges,” *Forensic Science International: Digital Investigation*, vol. 32, p. 200901, Mar. 2020, doi: 10.1016/j.fsidi.2019.200901.
- [22] J. Yang, H. Zou, Y. Zhou, and L. Xie, “Robust adversarial discriminative domain adaptation for real-world cross-domain visual recognition,” *Neurocomputing*, vol. 433, pp. 28–36, Apr. 2021, doi: 10.1016/j.neucom.2020.12.046.
- [23] K. Chao, W. Song, S. Shao, D. Liu, X. Liu, and X. B. Zhao, “CUI-Net: a correcting uneven illumination net for low-light image enhancement,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-39524-5.
- [24] X. Niu and G. Zheng, “Research on Low-Light Image Enhancement Algorithm Based on Multi-Dimensional Attention,” *Academic Journal of Engineering and Technology Science*, vol. 8, no. 1, 2025, doi: 10.25236/ajets.2025.080109.
- [25] A. Montoya, D. Holman, T. Smith, and W. Kan, “State Farm Distracted Driver Detection,” <https://kaggle.com/competitions/state-farm-distracted-driver-detection>.
- [26] H. Talebi and P. Milanfar, “Learning to Resize Images for Computer Vision Tasks,” 2021.
- [27] N. Fei, Y. Gao, Z. Lu, and T. Xiang, “Z-Score Normalization, Hubness, and Few-Shot Learning,” 2021.
- [28] M. Kang, M. Kang, S. W. Lee, and S. Kim, “Mixup Mask Adaptation: Bridging the gap between input saliency and representations via attention mechanism in feature mixup,” *Image Vis. Comput.*, vol. 146, Jun. 2024, doi: 10.1016/j.imavis.2024.105013.
- [29] A. Howard et al., “Searching for MobileNetV3,” in *International Conference on Computer Vision (ICCV)*, IEEE/CVF, 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.
- [30] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, “ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design,” 2018, pp. 122–138. doi: 10.1007/978-3-030-01264-9_8.
- [31] Darmatasia and M. H. Hasrul, “Model Efisien Berbasis Mobile Vision Transformer (MobileViT) untuk Klasifikasi Jenis Tanah dari Citra,” *JURNAL INSTEK*, vol. 10, no. 2, pp. 518–533, Oct. 2025, [Online]. Available: <https://journal.uin-alauddin.ac.id/index.php/instek>