Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: ijeeemi.org; Vol. 7, No. 1, pp. 1-10, January 2025; e-ISSN: 2656-8624.
https://doi.org/10.35882/ijeeemi.v7i1.30

RESEARCH PAPER

OPEN ACCES

Thermal Image Classification of Autistic Children Using Res-Net Architecture

Ahmadiar Ahmadiar[1], Melinda Melinda[1], Zharifah Muthiah[1], Zulfan Zainal[2], Muharratul Mina Rizky[1]

[1] Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia
[2] Computer Engineering Departmen of Engineering Faculty, Universitas Serambi Mekkah, T. Imum Lueng Bata Street, Batoh , Banda Aceh City, Indonesia

ABSTRACT

The classification of thermal images has found significant applications across various fields, including the analysis of facial thermal images. This study introduces a novel approach to classifying thermal facial images of children with Autism Spectrum Disorder (ASD), a neurological condition affecting communication and social interaction. Current diagnostic methods for ASD are primarily reliant on human expertise and lack definitive biological markers. Early diagnosis of ASD is crucial as it can lead to more effective interventions, improving the quality of life for affected children. This research leverages deep learning, particularly Convolutional Neural Networks (CNNs) with transfer learning, to explore thermal imaging as a non-invasive diagnostic tool. Using the ResNet architecture (ResNet-18, ResNet-34, and ResNet-50), the study aims to develop a robust classification model for ASD detection. The methodology involves training models with identical parameters: 100 epochs, batch size of 2, Stochastic Gradient Descent (SGD) optimizer, cross-entropy loss, learning rate of 0.001, and momentum of 0.9. The experimental results demonstrate high classification accuracy: 97.22% for ResNet-18, 99.22% for ResNet-34, and 99.41% for ResNet-50, indicating that ResNet-50 achieves the best performance. These findings highlight the potential of thermal imaging combined with deep learning to support early ASD diagnosis, contributing to advancements in non-invasive medical imaging technologies.

PAPER HISTORY

Received Nov. 15, 2024
Revised Dec. 31, 2024
Accepted Jan. 05, 2025
Published Jan. 20, 2025

KEYWORDS

ASD;
classification;
thermal imaging;
ResNet;

CONTACT:

melinda@usk.ac.id

1. INTRODUCTION

Autism spectrum disorder (ASD) is a neurological disorder that severely interferes with communication skills necessary for daily life. People with ASD often have difficulty in social situations. Most of the neurophysiological symptoms of ASD are known to medical professionals, but there are no biological markers or pathological techniques that can definitively diagnose autism. [1]. Diagnostic methods that are often used are interviews, Autism Diagnostic Observation Schedule (ADOS), and Autism Diagnostic Interview-Revised (ADI-R) [2]. The diagnostic results of these methods will depend on the doctor and the accuracy of the information provided by the patient or parents. However, they are very reliable. Human error can reduce the accuracy of these procedures. Recent advances in Artificial Intelligence have led to advanced medical diagnosis systems. AI can improve the accuracy and efficiency of medical diagnoses by providing doctors with valuable information and insights that can help make decisions faster and more precise [3].

Diagnosis at a very early age can improve outcomes significantly, and scientific evidence shows that ASD children who receive medical treatment before the age of four have a higher average IQ (compared to the typical IQ values of ASD children) [4]. Accurate and early diagnosis of autism spectrum disorder (ASD) is essential to facilitate timely intervention and provide individualized care for affected individuals. Rapid advances in deep learning techniques in recent years have ushered in a new era of medical image analysis, particularly in ASD detection using facial images [4]. Early screening of ASD from facial images can provide significant benefits using convolutional neural network (CNN) models with transfer learning approaches [5-6].

Use thermal imaging as a passive medium to analyze physiological signals related to affective states unobtrusively. The study hypothesizes that the impact of changes in skin temperature is due to pulsating blood flow in the blood vessels in the front area of the face of autistic children [7]. Classification system comparing the facial skin temperature of autistic and non-autistic children using thermal imaging on various emotions using ReNet-50 with an accuracy of 90% [8]. CNN-based ASD face image classification with an accuracy of 94.6% [9]. The ResNet-50 architecture for the classification of thermal images of autistic children with accuracy results of 89.2% [10].

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

Previous research [11] developed a deep-learning model to classify the faces of autistic children using thermal images. The purpose of this study is to analyze and compare the performance of various deep learning architectures in image classification systems, especially in identifying patterns in certain data. This study also aims to provide a benchmark with consistent training parameters, so that it can be a reference for model development and future research.

This research contributes to the field by developing a classification system using ResNet-18, ResNet-34, and ResNet-50 architectures with standardized training parameters, such as 100 epochs, a batch size of 2, SGD optimizer, cross-entropy loss, a learning rate of 0.001, and a momentum of 0.9. Additionally, it provides a comparative analysis of the accuracy testing results among the three architectures to assess their performance in the classification task.

2. METHOD

This study focuses on classifying two classes: thermal facial images of normal and autistic children, aiming to support non-invasive ASD diagnosis through deep learning.

A. Autism Spectrum Disorder (ASD)

The term “autism” was initially introduced to describe behavioral symptoms in schizophrenia patients. Psychiatrist Dr. Leo Kanner and pediatrician Dr. Hans Asperger coined the term autism in the 1940s to introduce a syndrome in children with behavioral disorders, differences in social interaction and communication. Currently, the umbrella term “autism spectrum disorders” (ASD) is used to describe a clinically heterogeneous group of neurodevelopmental disorders that share common behavioral core features that affect social communication and include restrictive and repetitive stereotyped behavioral patterns and interests [12].

Early diagnosis of autism spectrum disorder (ASD) is essential and beneficial for early planning and treatment of special education, provision of family support, and providing appropriate medical care to the child promptly [13]. In general, the available modalities for diagnosing affective states for ASD are divided into three categories, namely behavioral responses, physical reactions, and physiological signals. Behavioral responses are usually measured through engagement or body language toward activities, while physical reactions are demonstrated through facial expressions and speech [7].

B. Thermal Images

Thermal imaging has been at the forefront of affective computing research. Image-based techniques appear to be influenced by lighting, the subject's skin color, brightness, etc., but thermal image-based processes address all this [14]. Compared with conventional visual imaging systems or cameras, thermal imaging systems can detect and measure the temperature of objects produced by thermal images more accurately. Therefore, monitoring human body temperature identifies heat sources effectively to deliver consistent image quality [11]. The data for this study was sourced from previous research [11], where ethical considerations were addressed through the use of a Consent Form, ensuring compliance with ethical guidelines for research involving human subject

C. Convolutional Neural Network

CNN is one of the most well-known deep learning algorithms widely used to process data. Many layers are trained in an excellent way [15], [16], [17]. The CNN layers are convolutional and pooling, followed by fully connected layers. The basic structure of CNN is like Fig. 2.

Based on Fig. 2, the input image classification process will go through several stages, namely the convolution, pooling, and fully connected layers [18]. At each stage, it has a different approach. Convolution is the first layer that takes the input image, pooling to reduce the dimensions of the input image, a fully connected layer that will perform the classification [19].

a) Convolution Operation

The convolution operation is the core of CNN. In the convolutional layer (Eq. (1)),, features are extracted by sliding a filter kernel over the input feature map. The convolution operation can be expressed as:

$$h_j^{(n)} = \sum_{k=1}^{K} h_k^{(n-1)} \otimes w_{kj}^{(n)} + b_{kj}^{(n)} \quad (1)$$

The output of the *j-th* feature map in the n -th hidden layer, denoted as $h_j^{(n)}$, is obtained by applying a 2D convulation ($\otimes$) operation between the input feature map $h_k^{(n-1)}$ from the $(n-1)$*-th* layer and the weights $w_{kj}^{(n)}$ of the k*-th* channel in the j*-th* feature map of the n *-th* layer. This operation is followed by adding the corresponding bias term $b_{kj}^{(n)}$

b) Pooling Layer

The pooling layer (Eq. (2)) is used to reduce the dimensionality of data, preserve significant features, and prevent overfitting. The pooling operation can be described as:

$$h_j^{(pool)} = Pooling(h_j^{(n)}) \quad (2)$$

The pooling operation, denoted as $h_j^{(pool)}$ is applied to the output of the j*-th* feature map in the n -th hidden layer, $h_j^{(n)}$, as represented in Equation (2). This operation, referred to as pooling, serves to reduce the spatial dimensions of the feature map while retaining the most important information. Depending on the specific application or design choice, the pooling method can be either max-pooling, which selects the maximum value

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

within a defined region, or average-pooling, which calculates the average value within the same region. Both approaches aim to achieve dimensionality reduction and improve computational efficiency.

c) Activation Function

After the convolution or pooling operation, an activation function (Eq. (3)), such as ReLU $f(x) = \max(0, x)$, is applied to introduce non-linearity:

$$a_j^{(n)} = f\,(h_j^{(n)}) \tag{3}$$

The activation of the *j-th* feature map in the n -th hidden layer, denoted as $a_j^{(n)}$, is obtained by applying an activation function f to the corresponding output $h_j^{(n)}$ of the same layer. The activation function f introduces non-linearity into the network, enabling it to learn and model complex patterns and relationships within the data. Common activation functions include ReLU (Rectified Linear Unit), sigmoid, and tanh, each serving specific purposes depending on the architecture and requirements of the neural network. This process ensures that the neural network can generalize and perform effectively on tasks such as classification, regression, or feature extraction.

d) Loss Function

To train CNNs, the primary goal is to minimize the loss. One commonly used loss function is the cross-entropy loss:

$$E = -\frac{1}{m}\sum_{i=1}^{m}\sum_{c=1}^{C} y_i^{(c)} \log(\hat{y}_i^{(c)}) \tag{4}$$

The loss function E, as expressed in Equation (4), represents the cross-entropy loss used to evaluate the performance of a classification model. It is calculated as the negative average over all training samples, denoted by m, and classes, represented by C. For each data sample i and class c, the true label is denoted by $y_i^{(c)}$, $\hat{y}_i^{(c)}$ represents the predicted probability of the model assigning the sample to class c. The logarithmic term $\log(\hat{y}_i^{(c)})$ penalizes incorrect predictions more heavily, ensuring that the model prioritizes learning accurate probabilities. This loss function is widely used in multi-class classification problems due to its effectiveness in training models to output meaningful probability distributions.

e) Softmax for Class Probabilities

At the output layer, the softmax function (Eq. (5)) is used to generate probabilities for each class:

$$P(y = c|x) = \frac{\exp\,(z_c)}{\sum_j^C \exp\,(z_j)} \tag{5}$$

The probability of a sample belonging to class c, given its features x, denoted as $P(y = c|x)$, is calculated using the softmax function, as shown in Equation (5). This is computed by taking the exponential of the logit score for class c, z_c, and ividing it by the sum of the exponentials of the logit scores for all classes j from 1 to C, where C represents the total number of classes. The softmax function ensures that the output probabilities for all classes sum to 1, making it a suitable choice for multi-class classification tasks.

D. Residual Network

Residual Network architecture or abbreviated as ResNet is a type of pre-trained neural network learning that is efficient and accurate. ResNet-18 is a model consisting of 18 convolutional layers and several other layers, including a batch normalization layer and a ReLU activation layer. On the other hand, ResNet-34 has 34 convolutional layers with a deeper structure. This makes ResNet-34 more complex than ResNet-18 and perhaps better able to handle more complex tasks. [20].

ResNet-50 ResNet-50 is one of the largest models, with 50 convolutional layers. This model incorporates deeper and more complex residual blocks, making it easier to extract complex features. ResNet-101 is a larger model, with 101 layers, making it deeper and more powerful in handling very complex recognition tasks. Although it has more parameters [21]. The residual architecture in ResNet is represented as:

$$y = F\,(x) + x \tag{6}$$

In Equation (6), the output y is calculated by adding the input x to the transformation $F\,(x)$. Here, x represents the input to the operation, while $F\,(x)$ is a transformation that can involve operations such as convolution, batch normalization, or the ReLU activation function. This approach, often referred to as a residual connection or shortcut connection, combines the transformed input with the original input. It helps preserve information from earlier layers and improves the flow of gradients during backpropagation, making it particularly effective in deep neural networks.

ResNet has different parameter values, ResNet-18 usually contains around 11-12 million parameters, and ResNet-34 around 21-22 million parameters. ResNet-50 has 25-26 million parameters and ResNet-101 has around 44-45 million parameters. The number of different parameters will result in longer training and greater computing resources [21].

The initial difference that can be seen is that ResNet-18 and ResNet-34 have the first and second blocks of two consecutive convolution layers with a kernel size of 3x3 and an output of 64 channels, compared to ResNet-50 and ResNet-101. The first block starts with 1x1 convolution which produces 64 channels, then followed by 3x3 convolution with 64 channels, and finally followed again by 1x1 convolution which produces 256 channels. The second and subsequent blocks in ResNet-50 also follow this pattern, namely using 1x1 convolution to

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

reduce the number of channels, then 3x3 convolution with 64 channels, and ending with 1x1 convolution which increases the number of channels [8], [20], [22], [23], [24], [25], [26],[27], [28], [29].

E. Evaluation Of Metrics

Evaluation of metric values is obtained from the confusion matrix values. In the confusion matrix, there is a True Positive (TP) value which refers to the amount of positive data and positive predictions, and False Positive (FP) refers to the amount of positive data but negative predictions. True Negative (TN) refers to the amount of negative data and negative predictions, and False Negative (FN) refers to the amount of negative data, but positive predictions [30]. The equation for each ratio measure can be calculated mathematically as follows [31]. Accuracy (Eq. (7)) represents the proportion of correctly predicted instances (true positives and true negatives) out of the total number of instances.

a) *Accuracy*: $\frac{TP+TN}{TP+TN+FP+FN}$ (7)

Accuracy is a metric used to evaluate the performance of a classification model. It is calculated by dividing the total number of correct predictions (true positives TP and true negatives TN) by the total number of predictions, which includes true positives (TP), and true negatives (TN), false positives (FP), and false negatives (FN). In simple terms, accuracy measures the proportion of correctly classified instances out of all instances and is commonly expressed as a percentage.

Precision (Eq. (8)) measures the proportion of correctly predicted positive instances (true positives) out of all instances predicted as positive.

b) *Precision*: $\frac{TP}{TP+FP}$ (8)

Precision measures the accuracy of positive predictions made by a classification model. It is calculated as the ratio of true positives (TP) to the total predicted positives, which includes both true positives (TP) and false positives (FP). In other words, precision indicates how many of the predicted positive results are actually correct.

Sensitivity (Recall) (Eq. (9)) indicates the proportion of correctly identified positive instances (true positives) out of the total actual positives.

c) *Sensitivity*: $\frac{TP}{TP+FN}$ (9)

Sensitivity, also known as recall or true positive rate, measures the model's ability to correctly identify positive cases. It is calculated as the ratio of true positives (TP) to the total actual positives, which includes both true positives (TP) and false negatives (FN). Sensitivity indicates how well the model detects positive instances. Specificity (Eq. (10)) quantifies the proportion of correctly identified negative instances (true negatives) out of the total actual negatives.

d) *Specificity:* $\frac{TN}{TN+FP}$ (10)

Specificity measures a model's ability to correctly identify negative cases. It is calculated as the ratio of true negatives (TN) to the total actual negatives, which includes both true negatives (TN) and false positives (FP). Specificity indicates how well the model avoids incorrectly labeling negative instances as positive.

3. RESULTS

A. Training Model

In previous research [11] the data only used a conventional CNN architecture; therefore, this research will carry out model training using CNN architectures ResNet-18, ResNet-34 and ResNet-50. The following is the framework that will be implemented in this study. Several vital parameters have been determined for the training process in the implemented ResNet model. The training was conducted over 100 epochs, with each epoch covering an iteration through the entire training data set. The batch size is set at 2, indicating that the model updates its weights every two data samples, allowing for more responsive and adaptive updates to data variations. As a loss function, Cross-entropy is used, which is generally effective in classification tasks. The optimizer applied is SGD (Stochastic Gradient Descent), with a learning rate of 0.001 and momentum of 0.9. Momentum aims to increase convergence speed and help overcome fluctuations in learning rate. A softmax activation function is also used, which is essential in generating the desired class output probability. With this parameter configuration, the ResNet model is expected to be able to learn data patterns well and produce accurate class predictions.

At this stage, model training is carried out from previously designed models (transfer learning). The training process is learning from data, therefore the better the training data, the better the model performance will be. In this research, 100 epochs are used, each epoch will perform an iteration according to the total train data and batch size value. The total train data is 2,040 images and the batch size value is 2, then the iteration value will be 1,020, meaning that in one epoch the weights will be updated 1,020 times.

In evaluating the model's performance for autism identification, the confusion matrix provides a detailed description of the classification results between the two main classes, namely positive (autistic) and negative (normal). True Positive (TP) reflects how well the model recognizes autistic cases, while True Negative (TN) measures the model's ability to identify normal instances correctly. False Positive (FP) reflects an error in classifying normal cases as autistic, while False Negative (FN) demonstrates an error in classifying autistic cases as normal. With a deep understanding of this confusion matrix, we can evaluate the extent to which the model can differentiate between the two classes and identify areas where misclassification may occur. Thus, the confusion matrix helps inform the development and improvement of models to increase the accuracy of autism case identification.

B. ResNet-18 Training

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

ResNet-18 has a total of 18 main layers. Suitable for common tasks in computer vision with lower computational requirements. Consists of 18 layers, including convolution, batch normalization, Relu, max pooling, and classifier layers. Uses a basic residual block consisting of 2 3x3 convolution layers in one block. Lighter and faster in terms of computing. Complete training process in 110m 17s, training graph in Fig. 1 (a) ResNet-18. Fig. 1 (a) ResNet-18 displays the training graph using ResNet-18. It can be seen that the train graph value was very good, reaching 1, however, during data validation, the average value was 0.8, so it was stated that this training was still not good. In the context of autism identification, positive results (TP = 0.9804) indicate the accuracy of the system in recognizing autistic conditions, providing a positive impact on early detection efforts. Meanwhile, the negative result (TN = 0.9647) in the normal group shows the model's ability to identify individuals without autism with a high level of accuracy. However, there were also false positives (FP = 0.0196) and false negatives (FN = 0.0353), which indicated errors in classifying several cases. Despite this, the overall accuracy of the model reached 97.25%, reflecting a significant success rate in the recognition of autism and normals by minimizing classification errors.

C. ResNet-34 Training

ResNet-34 is an expansion of ResNet-18 by adding more residual blocks. Deeper residue block structure with multiple residue blocks. Complete training results in 170m 58s, training graph in Fig. 1 (b) ResNet-34.

Fig. 1 (b) ResNet-34 displays the training graph using ResNet-34, it can be seen that the train graph value is better than ResNet-18, the data validation value has reached 0.9. In the ResNet34 implementation, positive results (TP = 0.9922) and negative results (TN = 0.9922) show excellent ability to classify both classes accurately. The low false positive (FP = 0.0078) and false negative (FN = 0.0078) rates indicate minimal errors in recognizing positive and negative classes. With an overall accuracy of 99.22%, the ResNet34 model can be considered a very reliable and effective model in a given classification task, providing high certainty in determining classification results.

D. ResNet-50 Training

ResNet-50 Uses more complex residue blocks. with a bottleneck residue block structure, which allows to learn richer and deeper feature-representations. Training completes in 227m 28s, training graph in Fig. 1 (c) ResNet-50. Fig. 1 (c) ResNet-50 displays the training graph using ResNet-50, it can be seen that the train graph value is better than ResNet-18 and ResNet-34, and the average validation value reaches 0.9. In the case of using ResNet50, the positive results (TP = 0.9882) and negative results (TN = 1.000) show a very high level of accuracy in recognizing positive and negative classes. This model can avoid classification errors with minimal false positives (FP = 0.0118) and false negatives (FN = 0.0). The overall accuracy reached 0.9941, reflecting the reliability and effectiveness of the ResNet50 model in the given classification task. This high level of accuracy provides strong confidence in the use of this model for identification and classification purposes.

Table 1. summarizing the results from the three confusion matrices for ResNet-18, ResNet-34, and ResNet-50

Model	TP (Autism)	N (Autism)	TN (Non-Autism)	FP (Non-Autism)
ResNet-18	98.04%	1.96%	96.47%	3.53%
ResNet-34	99.22%	0.78%	99.22%	0.78%
ResNet-50	99.82%	1.18%	100%	0.00%

Table 1 presents The confusion matrices shown above illustrate the classification performance for ResNet-18, ResNet-34, and ResNet-50 models in distinguishing between autistic and non-autistic children.

From the three confusion matrix values, metric evaluation can be calculated, in this study using the Sensitivity and Specificity values, which can be seen in Table 2.

Table 2. Performance Comparison Of Several Method

Parameters	ResNet-18	ResNet-34	ResNet-50
Accuracy	97.25%	99.22%	99.41%
Precision	98.04%	99.22%	98.82%
Sensitivity	96.52%	99.22%	100%
Specificity	98.80%	99.22%	98.88%

Table 2 presents the results of the evaluation metrics, with information on the comparison values of accuracy, sensitivity, and specificity of the three architectures, namely ResNet-18, ResNet-34, and ResNet-50. In ResNet-18 the highest value is specificity 98.80%. ResNet-34 obtained 99.22% in all comparisons of accuracy, sensitivity, and specificity, and ResNet-50 had a value of 100% in sensitivity.

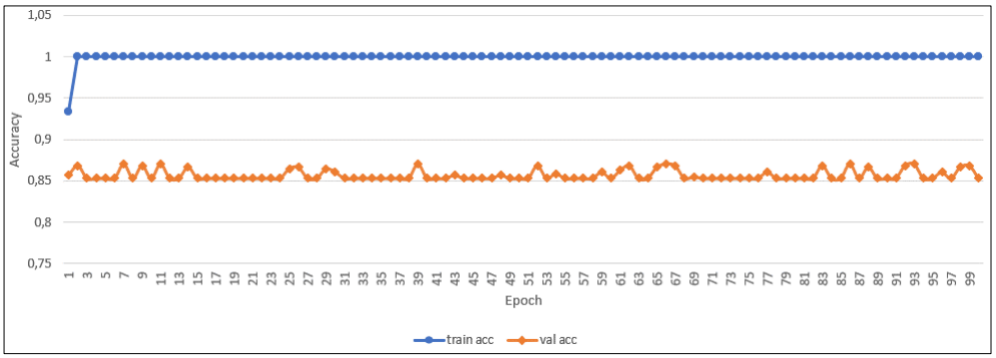
4. DISCUSSION

This graph of model training results provides valuable insight into model performance during the training process. The blue line representing train accuracy shows the extent to which the model could correctly predict the training data across various epochs. If the blue line continues to increase, the model will effectively learn from the training data. Meanwhile, the orange line representing validation accuracy shows how much the model can generalize and predict data that has never been seen before. Ideally, we would like to see an increase in both lines over time, indicating that the model can understand general patterns and not just memorize the training data. However, if train accuracy continues to increase while validation accuracy slows, it could indicate overfitting, where the model is too specific for the training data. Therefore, monitoring changes in both lines along the epoch axis helps in assessing the performance and generalization ability of the model, as well as providing information on when the training process should be

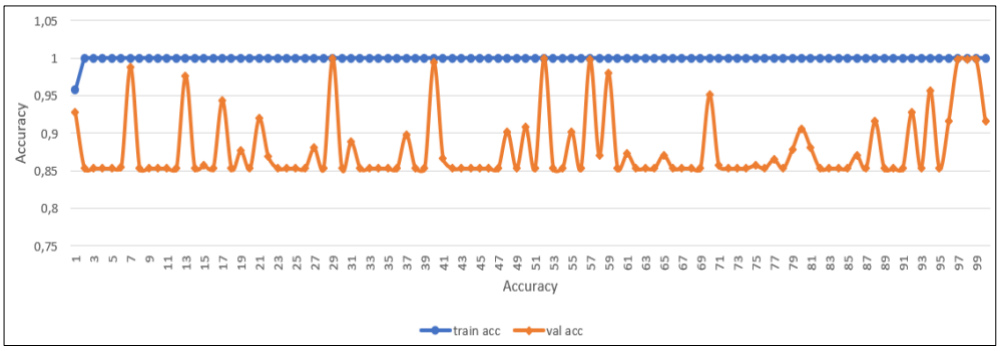
Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

stopped to avoid overfitting.

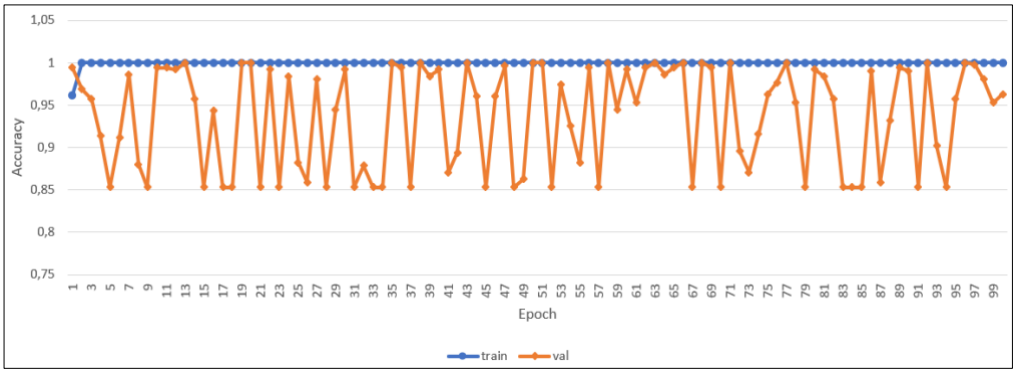
difference between the two accuracies, this could indicate



(a)



(b)



(c)

Fig. 1. Train Accuracy and Validation Accuracy Result Curves, (a) ResNet-18, (b) ResNet-34, (c) ResNet-50

Fig. 1 (a) ResNet-18 depicts the model's performance during the training process, which has several unique characteristics. A stable blue line at 1 indicates that the model achieved complete accuracy on the training data. This suggests that the model can perfectly predict and fit the training set, but it is essential to note whether this is a sign of overfitting to the training data. Meanwhile, the orange line, which is above 0.85 but does not reach 0.9, indicates that the model can provide relatively accurate predictions on validation data but only partially reaches the desired level of accuracy (0.9 or more). Model evaluation needs to pay attention to the difference between train accuracy and validation accuracy. If these two lines develop simultaneously, the model properly understands the data patterns and can generalize well to never-before-seen data. However, if there is a significant overfitting.

Fig. 1 (b) ResNet-34 depicts a model with specific characteristics that must be considered. The stability of the blue line at a value of one indicates that the model can achieve complete accuracy on the training data, which means that the model can learn and adapt well to the data. However, it is worth noting that absolute accuracy on the training data may show signs of overfitting, where the model over-understands specific training data patterns and may need to be more capable of generalizing to new data. The orange line, which varies at values 0.85, 0.9, 0.95, and even touches the number 1, shows fluctuations in validation accuracy. Although these fluctuations may reflect natural variations in model performance, the frequent presence of values above 0.85 at certain epochs indicates more consistent variation. This

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

can mean a point where the model may have difficulty generalizing or face specific challenges in validation data. Fig. 1 (c) ResNet-50 depicts the model with the blue line stable at a value of one, indicating that the model can obtain complete accuracy on the training data. However, the difference in the behavior of the orange line (validation accuracy) compared to ResNet-34, with variations in values of 0.085, 0.9, 0.95, and even touching the number 1, is interesting to note.

This comparison may reflect the characteristic differences between ResNet-50 and ResNet-34 in the generalization ability and adaptation of the model to validation data. If the orange line on the ResNet-50 graph is often above the value 0.95, this could indicate that ResNet-50 may be more easily affected by variations in the validation set or has a higher level of sensitivity to changes in the validation data. To understand these differences further, it can be helpful to perform further analysis on the specifications, such as the model configuration, the number of parameters, or the way the training and validation data are processed. Evaluation of these differences can help better understand each model's characteristics and guide corrective or enhancement measures that may be necessary, depending on the specific goals and needs of the classification task at hand.

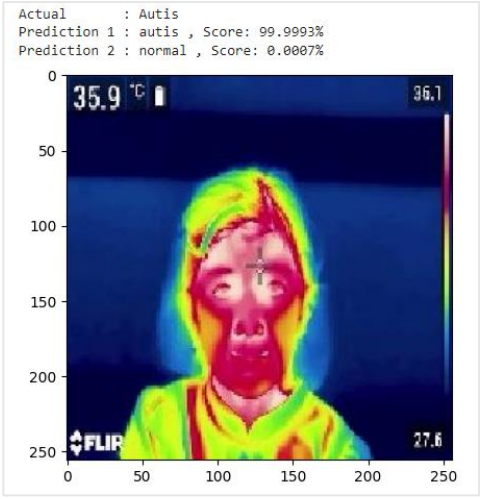
Of the three architectures used, ResNet-50 has high accuracy results, one of the reasons being that ResNet-50 has a deeper network than the previous ResNet. Figure 2 displays a representative comparison of thermal images for autism and non-autism. In Fig. 2 (a), with actual autistic data input, the model produces predictions with high confidence. The first prediction shows the autistic label with a probability of 99.9993%, while the second prediction shows the normal label with a very low likelihood, only 0.0001%. A similar thing can be seen in Figure 2 (b) with actual normal data input, where the first prediction shows a normal label with a confidence level of 99.9979%, and the second prediction shows an autistic label with a very low probability, only 0.021%. Notably, these percentages come from the softmax function, which maps the model output into a probability distribution for each class. Therefore, the highest probability value is taken as the class prediction produced by the model. These results reflect the confidence level of the ResNet-50 model in classifying test data, and the high probability of the central prediction shows the model's ability to provide very confident predictions.

In this section, we present the classification results of thermal facial images using three ResNet architectures. Table 1 presents The confusion matrices ResNet-18, ResNet-34, and ResNet-50. The evaluation metrics, including accuracy and misclassification rate, were derived from the corresponding confusion matrices for each model.

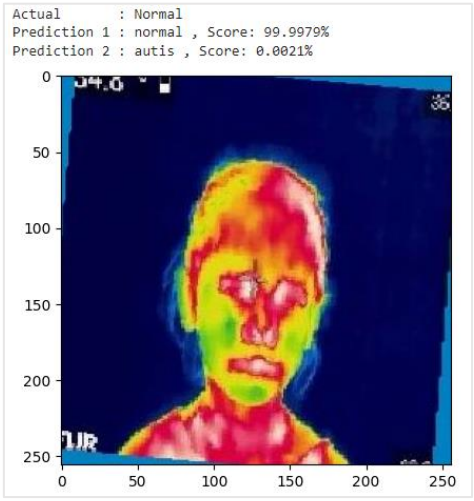
The confusion matrix for ResNet-18 shows that the model achieves an accuracy of 97.25%, with a misclassification rate of 2.75%. It correctly classifies 98.04% of autistic children and 96.47% of non-autistic children, but misclassifies 3.53% of non-autistic children as autistic and 1.96% of autistic children as non-autistic.

For ResNet-34, the model demonstrates improved performance, with an accuracy of 99.22% and a lower misclassification rate of 0.78%. The true positive rates are 99.22% for autistic children and 99.22% for non-autistic children, with very minimal misclassification between the two classes.

The confusion matrix for ResNet-50 highlights its superior classification capability, achieving an accuracy of 99.41% and the lowest misclassification rate of 0.59%. It perfectly classifies all non-autistic children (100%), while it slightly misclassifies 1.18% of autistic children as non-autistic.



(a)



(b)

Figure 2. Comparison of testing accuracy of thermal images of children's faces, (a) autism, (b) non-autism

In this section, we present the classification results of thermal facial images using three ResNet architectures. Table 1 presents The confusion matrices ResNet-18, ResNet-34, and ResNet-50. The evaluation metrics, including accuracy and misclassification rate, were

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

derived from the corresponding confusion matrices for each model.

The confusion matrix for ResNet-18 shows that the model achieves an accuracy of 97.25%, with a misclassification rate of 2.75%. It correctly classifies 98.04% of autistic children and 96.47% of non-autistic children, but misclassifies 3.53% of non-autistic children as autistic and 1.96% of autistic children as non-autistic.

For ResNet-34, the model demonstrates improved performance, with an accuracy of 99.22% and a lower misclassification rate of 0.78%. The true positive rates are 99.22% for autistic children and 99.22% for non-autistic children, with very minimal misclassification between the two classes.

The confusion matrix for ResNet-50 highlights its superior classification capability, achieving an accuracy of 99.41% and the lowest misclassification rate of 0.59%. It perfectly classifies all non-autistic children (100%), while it slightly misclassifies 1.18% of autistic children as non-autistic.

Table 3. Performance Comparison Of Several Method

Ref	Method	Accuracy
[7]	Customized Neural Network	96%
[9]	Customized Neural Network	95%
[10]	Convolutional Neural Network	98%
This Research	ResNet-50	99.41%

Table 3 presents the comparative values of accuracy testing parameters from other similar studies using thermal images of the faces of children with autism and non-autism. Research [8] using a Customized Neural Network architecture obtained an accuracy value of 96%, while research [11] using a CNN architecture obtained an accuracy value of 998%. This research obtained a superior score of 99.41% using the ResNet-50 architecture.

However, this study has some limitations. The dataset used is relatively small, which could affect the generalizability of the model. A larger, more diverse dataset would help the model perform better in real-world applications. Moreover, overfitting may still be a concern, particularly for deeper models like ResNet-50, even with regularization techniques such as dropout. Another limitation is the absence of behavioral or clinical data that could provide a more comprehensive view of ASD characteristics, which might further enhance classification performance.

The implications of this study are significant for non-invasive autism detection. The high accuracy of ResNet-50 indicates that thermal facial imaging could be a powerful tool for early autism diagnosis, especially in settings where traditional diagnostic methods (e.g., behavioral analysis) are not readily available or feasible. Future research should focus on increasing the dataset size and diversity, incorporating additional clinical and behavioral data, and exploring other deep learning models to improve the robustness of the system. These steps will ensure that the model can be effectively used in clinical practice, aiding in the early detection of autism and potentially reducing diagnostic delays.

5. CONCLUSION

This research uses the dataset [11] to carry out data training using the ResNet-18, ResNet-34, and ResNet-50 architectures using the same training parameters, namely epoch 100, batch size 2, SGD, Cross-entropy, learning rate 0.001, momentum 0.9. The length of training time varies, namely ResNet-18 110m 17s, ResNet-34 170m 58s, ResNet-50 227m 28s with final comparison values of testing accuracy 97.22%, 99.22%, and 99.41%.
Further research of this study is an opportunity to compare with other facial thermal datasets of autism and normal children, the application of various CNN architectures, and changing the hyperparameter value to get better accuracy. It is also hoped that further research can apply to a mobile application so that it can be one of the doctor's aids in deciding the results of autism and typical diagnoses.

ACKNOWLEDGEMENT

I would like to acknowledgement to Universitas Syiah Kuala for the Lektor grant with No: 248/UN11.2.1/PT.01.03/PNBP/2023 Tanggal 3 Mei 2023.

REFERENCES

[1] W. Jones et al., “Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism,” JAMA Netw. open, vol. 6, no. 9, p. e2330145, 2023, doi: 10.1001/jamanetworkopen.2023.30145.
[2] M. G. Perinelli and M. Cloherty, “Identification of autism in cognitively able adults with epilepsy: A narrative review and discussion of available screening and diagnostic tools,” Seizure, vol. 104, no. June 2022, pp. 6–11, 2023, doi: 10.1016/j.seizure.2022.11.004.
[3] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review,” Healthc., vol. 10, no. 3, pp. 1–30, 2022, doi: 10.3390/healthcare10030541.
[4] M. S. Alam et al., “technologies Efficient Deep Learning-Based Data-Centric Approach for Autism Spectrum Disorder Diagnosis from Facial Images Using,” pp. 1–27, 2023.
[5] T. Ghosh et al., “Artificial intelligence and internet of things in screening and management of autism spectrum disorder,” Sustain. Cities Soc., vol. 74, p. 103189, 2021, doi: 10.1016/j.scs.2021.103189.
[6] ZB Akhtar and AT Rawol, "Biomedical Engineering (BME) Perspectives Exploration for Investigation Analysis: Artificial Intelligence (AI) and Extended Reality (XR)" Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics August 2024, Vol. 6, No. 3, pp. 132-146, e-ISSN:

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

2656-8624, https://ijeeemi.ijahst.org https://doi.org/10.35882/jeeemi.v6i3.4.
[7] N. Rusli, S. N. Sidek, H. M. Yusof, N. I. Ishak, M. Khalid, and A. A. A. Dzulkarnain, "Implementation of Wavelet Analysis on Thermal Images for Affective States Recognition of Children with Autism Spectrum Disorder," IEEE Access, vol. 8, pp. 120818–120834, 2020, doi: 10.1109/ACCESS.2020.3006004.
[8] K. Ganesh, S. Umapathy, and P. Thanaraj Krishnan, "Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging," Proc. Inst. Mech. Eng. Part H J. Eng. Med., vol. 235, no. 10, pp. 1113–1127, 2021, doi: 10.1177/09544119211024778.
[9] M. Beary, A. Hadsell, R. Messersmith, and M.-P. Hosseini, "Diagnosis of Autism in Children using Facial Analysis and Deep Learning," 2020, [Online]. Available: http://arxiv.org/abs/2008.02890
[10] F. C. Tamilarasi and J. Shanmugam, "Convolutional neural network based autism classification," Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020, no. Icces, pp. 1208–1212, 2020, doi: 10.1109/ICCES48766.2020.09137905.
[11] M. Melinda, Ahmadiar, Maulisa Oktiana, "A Novel Autism Spectrum Disorder Children Dataset Based on Thermal Imaging," 2023, doi: http://dx.doi.org/10.1109/ICCCE58854.2023.10246072.
[12] A. K. Sauer, J. E. Stanton, S. Hans, and A. M. Grabrucker, "Autism Spectrum Disorders: Etiology and Pathology," Autism Spectr. Disord., pp. 1–16, 2021, doi: 10.36255/exonpublications.autismspectrumdisorders.2021.etiology.
[13] B. R. G. Elshoky, E. M. G. Younis, A. A. Ali, and O. A. S. Ibrahim, "Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images," ETRI J., vol. 44, no. 4, pp. 613–623, 2022, doi: 10.4218/etrij.2021-0097.
[14] I. Engineering and I. Engineering, "Relevance of Thermal Imaging and Respiration Signals in Recognizing Human Emotions," vol. 7, no. 1, pp. 5166–5175, 2022.
[15] L. Mohammadpour, T. C. Ling, C. S. Liew, and A. Aryanfar, "A Survey of CNN-Based Network Intrusion Detection," Appl. Sci., vol. 12, no. 16, 2022, doi: 10.3390/app12168162.
[16] S. Bellamkonda, N. P. Gopalan, C. Mala, and L. Settipalli, "Facial expression recognition on partially occluded faces using component based ensemble stacked CNN," Cogn. Neurodyn., vol. 17, no. 4, pp. 985–1008, 2023, doi: 10.1007/s11571-022-09879-y.
[17] M. N. Kartheek, M. V. N. K. Prasad, and R. Bhukya, "Texture based feature extraction using symbol patterns for facial expression recognition," Cogn. Neurodyn., vol. 0123456789, 2022, doi: 10.1007/s11571-022-09824-z.
[18] H. A. Razak, N. K. Zakaria, N. F. M. Zamri, N. M. Tahir, and A. A. Almisreb, "Detection of Criminal Behavior at the Residential Unit based on Deep Convolutional Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 2, pp. 804–813, 2022, doi: 10.14569/IJACSA.2022.0130293.
[19] M. M. Adnan et al., "An Improved Automatic Image Annotation Approach Using Convolutional Neural Network-Slantlet Transform," IEEE Access, vol. 10, pp. 7520–7532, 2022, doi: 10.1109/ACCESS.2022.3140861.
[20] M. Gao, P. Song, F. Wang, J. Liu, A. Mandelis, and D. Qi, "A Novel Deep Convolutional Neural Network Based on ResNet-18 and Transfer Learning for Detection of Wood Knot Defects," J. Sensors, vol. 2021, 2021, doi: 10.1155/2021/4428964.
[21] D. Su, H. Zhang, H. Chen, J. Yi, P. Y. Chen, and Y. Gao, "Is robustness the cost of accuracy? – A comprehensive study on the robustness of 18 deep image classification models," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11216 LNCS, pp. 644–661, 2018, doi: 10.1007/978-3-030-01258-8_39.
[22] Feni Isdaryani, Noor Cholis Basjaruddin, and Aldi Lugina, "Masked Face Recognition and Temperature Monitoring Systems for Airplane Passenger Using Sensor Fusion," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 11, no. 2, pp. 140–147, 2022, doi: 10.22146/jnteti.v11i2.3835.
[23] Syukron Abu Ishaq Alfarozi and Azkario Rizky Pratama, "CNN-Based Model for Copy Detection Pattern Estimation and Authentication," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 12, no. 1, pp. 44–49, 2023, doi: 10.22146/jnteti.v12i1.6205.
[24] H. Wang, K. Li, and C. Xu, "A New Generation of ResNet Model Based on Artificial Intelligence and Few Data Driven and Its Construction in Image Recognition Model," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/5976155.
[25] I. A. Ahmed et al., "Eye Tracking-Based Diagnosis and Early Detection of Autism Spectrum Disorder Using Machine Learning and Deep Learning Techniques," Electron., vol. 11, no. 4, 2022, doi: 10.3390/electronics11040530.
[26] N. A. Selamat, S. H. M. Ali, K. N. Minhad, and J. Sampe, "Feature Selection Analysis of Chewing Activity Based on Contactless Food Intake Detection," Int. J. Integr. Eng., vol. 13, no. 5, pp. 38–48, 2021, doi: 10.30880/ijie.2021.13.05.006.
[27] S. K. Jarraya, M. Masmoudi, and M. Hammami, "A comparative study of Autistic Children Emotion recognition based on Spatio-Temporal and Deep analysis of facial expressions features during a Meltdown Crisis," Multimed. Tools Appl., vol. 80, no. 1, pp. 83–125, 2021, doi: 10.1007/s11042-020-09451-y.
[28] T. Thanarajan, Y. Alotaibi, S. Rajendran, and K. Nagappan, "Eye-Tracking Based Autism Spectrum Disorder Diagnosis Using Chaotic Butterfly Optimization with Deep Learning Model," Comput.

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.

Mater. Contin., vol. 76, no. 2, pp. 1995–2013, 2023, doi: 10.32604/cmc.2023.039644.
[29] H. Sewani and R. Kashef, “An autoencoder-based deep learning classifier for efficient diagnosis of autism,” Children, vol. 7, no. 10, 2020, doi: 10.3390/children7100182.
[30] M. S. Al Reshan et al., “Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model,” Healthc., vol. 11, no. 11, 2023, doi: 10.3390/healthcare11111561.
[31] M. Badawy, A. M. Almars, H. M. Balaha, M. Shehata, M. Qaraad, and M. Elhosseini, “A two-stage renal disease classification based on transfer learning with hyperparameters optimization,” Front. Med., vol. 10, 2023, doi: 10.3389/fmed.2023.1106717.

AUTHOR BIOGRAPHY

Ahmadiar His Bachelor of Electrical Engineering was obtained from Electrical Engineering Department of Syiah Kuala University in 2004. He then continuing his degree in Master of Science in Information Technology at the School of Computer Science, Universiti Sains Malaysia (USM), Penang, Malaysia in 2008. He Served as a Lecturer at the Department of Electrical Engineering, Faculty of Engineering Unsyiah from 2008 until now. His research focus on the Information Systems, and Disaster Information and Management System. He also teach some subject like Engineering Economic, Technopreuneurship, Disaster Telecommunication and Information Technology and Information System Analysis and Practices.

Melinda was born in Bireuen, Aceh, on June 10, 1979. She received a B.Eng degree from the Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh in 2002. She completed her master’s degree at the Faculty of Electrical Department, University of Southampton, United Kingdom, with a concentration in field study of Radio Frequency Communication Systems in 2009. She has already completed her Doctoral degree at the Department of Electrical Engineering, Engineering Faculty of Universitas Indonesia in February 2018. She has been with the Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University since 2002. She is also a member of IEEE. Her research interests include multimedia signal processing and fluctuation processing. She can be contacted at email: melinda@usk.ac.id.

Zharifah Muthi’ah was born on March 17, 2001, in Sigli, Aceh. She earned her bachelor's degree in Computer Engineering from Syiah Kuala University in 2023, where she focused on multimedia technology. During her undergraduate studies, she gained extensive knowledge and skills in areas such as multimedia systems, digital content development, and related technologies. Motivated to deepen her expertise, she pursued a master's degree in Electrical Engineering at Syiah Kuala University, specializing in the biomedical field. She successfully completed her master's degree in 2024. Through her academic journey, Zharifah developed a strong foundation in both multimedia and biomedical engineering, equipping her with a multidisciplinary approach to tackle modern technological challenges.

Zulfan Zainal is a lecturer at Universitas Serambi Mekkah, an educational institution categorized as a Colleges and Universities company with an estimated workforce of 43 employees. His research interests span various advanced technological fields, including digital image processing, multimedia content production, biometric systems, and computer vision. As a dedicated academic, Zulfan actively contributes to the development of innovative solutions in these areas, fostering advancements that bridge the gap between technology and practical applications. Currently based in Indonesia, he plays a significant role in promoting education and research within his institution, inspiring students and colleagues alike through his expertise and commitment to technological excellence.

Muharratul Mina Rizky was born on August 10, 1990, in Langsa, Aceh. He earned his bachelor's degree in Informatics Engineering from Ubudiyah Indonesia University in 2014, where he gained a strong foundation in computer systems, programming, and software development. Driven by a passion for advancing his knowledge, he pursued a master's degree in Information Technology at the Department of Electrical Engineering and Computer, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, and graduated in 2021. In 2024, he joined the Department of Electrical Engineering and Computer, Faculty of Engineering, Universitas Syiah Kuala, where he contributes to the academic and research community. His journey reflects a continuous commitment to professional growth and education in the field of information technology and engineering.

Corresponding author: Melinda, melinda@usk.ac.id, Department of Electrical Engineering and Computer, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh 23111, Indonesia.