Detection of Autism Spectrum Disorder (ASD) Using Lightweight Ensemble CNN Based on Facial Images for Improved Diagnostic Accuracy

Andi Kurniawan Nugroho[1], Jajang Edi Priyanto[2], D. S. P. Vinski[3]
Electrical Department, Universitas Semarang, Semarang, Indonesia[1]
Public Health Department, Institut Kesehatan Indonesia, Jakarta, Indonesia[2]
Celtech Stem Cell Centre Laboratory and Banking Jakarta, Indonesia[3]

***Abstract*—Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects how people talk to each other and act. The fact that ASD is becoming more common and that diagnosing it can be difficult means that early detection is important for improving treatment outcomes. This study's goal is to use lightweight ensemble Convolutional Neural Networks (CNN) to make it easier to classify ASD from facial photos. The study looks at different CNN architectures, like MobileNetV2 and EfficientNet variations, to find the best model for diagnosing ASD quickly and accurately. The method involves training and testing five lightweight CNN models on a set of facial photos. We use pre-processing methods like scaling and data augmentation to help the model learn better. The study tests how well ensemble CNN models work by combining predictions from different architectures using averaging and voting methods. We use important performance metrics like accuracy, precision, recall, and F1-score to see how well each model works. The results show that the best balance between accuracy and computational efficiency is achieved by combining MobileNetV2 and EfficientNetB0. This combination achieves an accuracy of 0.8299, a precision of 0.8514, a recall of 0.8182, and an F1_score of 0.8344. Other models, like ResNet50 combined with EfficientNetB0, have higher precision but lower recall, making them less useful for finding all ASD cases. This study was also compared with other researchers, and the proposed study was found to have greater accuracy than other researchers. The results show that ensemble CNN models can significantly improve the accuracy of classifying ASD compared to single CNNs. This study shows that lightweight ensemble CNN models are good at finding ASD in pictures of people's faces. The method is fast and can be used on devices with limited processing power, making it a good way to find ASD early in both clinical and real-world settings.**

***Keywords*—Component; autism spectrum disorder (ASD); early detection; ensemble convolutional neural network (CNN); facial images; classification; accuracy**

I. Introduction

Autism Spectrum Disorder (ASD) is a developmental disorder that affects how a person talks, interacts with others, and thinks and acts. People with Autism Spectrum Disorder (ASD) often have trouble interacting with others, have trouble learning new words, and have rigid habits that they do over and over. Many conditions that used to be called classic autism, Asperger's syndrome, and undefined pervasive developmental disability are now part of ASD [1]. This spectrum shows how different people can have very different symptoms and how bad they can be. It is thought that both genetic and environmental factors play a role in the causes of ASD, but these factors are not fully understood. Recent studies have confirmed the idea that a combination of genetic and prenatal factors has a big effect on the development of ASD. The number of people with ASD around the world has gone up a lot in the last ten years.

Center for Disease Control and Prevention (CDC) in the United States says that the number of children with Autism Spectrum Disorder (ASD) went from 1 in 150 in the early 2000s to 1 in 36 in 2023 [2]. There are many reasons for this increase, including changes in how autism is diagnosed, more people and healthcare professionals being aware of it, and better detection of autism across a wider range of ages. There aren't many statistics available in Indonesia, but reports from the Ministry of Health of the Republic of Indonesia and a few local studies show that the number of people with ASD is rising at a similar rate. Studies show that 2 to 5 out of every 10,000 children have Autism Spectrum Disorder (ASD). This trend is growing because more people in Indonesia know about ASD, and diagnostic tools are getting better [3]. Sadly, it is still hard to get a diagnosis and early help for ASD in Indonesia, especially in remote areas. This makes it harder to treat and support people with ASD.

A multidisciplinary approach is often used in ASD screening and diagnosis. This includes clinical evaluations based on DSM-5 criteria and a range of screening tools designed to find ASD symptoms in both children and adults. The Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) are the two most common tests used to diagnose Autism Spectrum Disorder (ASD). The ADOS is a series of exercises that are meant to test a child's social skills, communication skills, and behavior. The ADI-R is a semi-structured interview with parents or caregivers to find out how a child is doing with their early developmental milestones. In addition to these observational assessment methods, questionnaire-based screening tools like the Modified Checklist for Autism in Toddlers (M-CHAT) are used to find ASD in young children. These steps make it easier to find problems early, but they require specialized knowledge and experience to give an accurate and appropriate assessment.

As technology has improved, research on using artificial intelligence (AI) methods to classify ASD has made a lot of progress. Artificial intelligence has a lot of potential to help with diagnosing ASD by looking at different types of data, such as neuroimaging, audio recordings, behavioral patterns, and genetic information. Support Vector Machines (SVM), Random Forest, Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN) are some of the algorithms that have been used to make models for classifying ASD that are more accurate and useful. Here are eleven papers that came before this one that used AI methods to classify ASD.

K. Vakadkar et al. (2021) used a data-driven method with Decision Tree and Naive Bayes algorithms to sort children with Autism Spectrum Disorder. The study showed that AI models can be very accurate at finding people with ASD [4]. Duda et al. (2016) made an AI-based diagnostic system that uses motion sensor data to find signs of ASD in kids. The SVM algorithm that was used worked very well, with an accuracy of almost 90% [5]. Heinsfeld et al. (2018) used deep learning and transfer learning to sort Autism Spectrum Disorder using fMRI data. This study shows that AI can tell the difference between different brain connection patterns in people with ASD [6].

Bone et al. (2017) made a Machine Learning model that uses voice recordings and speech patterns to sort people with Autism Spectrum Disorder (ASD). This study shows that AI has a lot of potential for analyzing non-verbal data [7]. R. Thapa et al. (2023) looked at how likely a child was to develop ASD at an early stage, which made the diagnosis process easier. To predict ASD traits, it uses several machine learning models, such as Logistic Regression (LR), Naive Bayes (NB), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest Classifier (RFC). The study found that Logistic Regression was the best model for the dataset. However, machine learning models could do a better job of finding ASD if they were trained on bigger datasets and used deep learning methods like Convolutional Neural Networks (CNNs). The authors stressed how important it is to find problems early and how machine learning can help speed up the diagnostic process [8].

Using gene interaction network data, D. Bandara [9] made a Graph Neural Networks (GNN) model to figure out how likely it is that a gene is linked to Autism Spectrum Disorder (ASD). The author used different types of GNNs as starting points for three classification tasks: binary, multi-class, and syndromic risk classification. These types of GNNs included Graph Sage, Graph Convolutional Networks (GCN), and Multi-Layer Perceptron (MLP). The experimental results show that the Sage Graph model consistently outperforms the other models with the highest accuracy for all three classification tasks: 85.80% accuracy for binary risk classification, 81.68% accuracy for multi-class risk classification, 90.22% accuracy for syndromic gene classification.

J. W. Song et.al concluded that neuroimaging studies in neurodevelopmental disorders, such as autism spectrum disorder (ASD) and Attention-Deficit or Hyperactivity Disorder (ADHD), have uncovered various biomarkers. However, limitations in comprehensively analyzing such biomarkers limit clinically significant results. The use of deep learning (DL) offers a new approach to improve diagnosis accuracy and assist clinicians in determining more appropriate treatment. DL enables earlier and more accurate detection of biomarkers by analyzing complex patterns of neuroimaging data [10].

A. Ashraf et al. used deep learning methods, specifically an optimized Convolutional Neural Network (CNN), providing better results in the classification of individuals with Autism Spectrum Disorder (ASD) than previous conventional methods. Transfer learning demonstrated efficacy, achieving an accuracy of 82.31% on the ABIDE-I and ABIDE-II datasets. This study looks at how fMRI data can help find Autism Spectrum Disorder (ASD) early on so that kids with autism can have a better environment. The results showed better performance, but the authors stressed how important data augmentation is for improving accuracy even more [11].

R. A. Bahathiq et al. R. A. Bahathiq et al. have used machine learning (ML) and deep learning (DL) to diagnose autism spectrum disorder (ASD) using structural magnetic resonance imaging (sMRI) data. This is a promising approach, but it faces a lot of challenges. There are still a number of problems with ML/DL research that make it hard to find biomarkers for ASD and improve diagnostic accuracy. These include small sample sizes, differences in brain biomarkers, and problems with generalizing results. This study stresses how important it is to have bigger datasets and more rigorous methods to deal with these problems [12].

Y. Lin and others have shown that machine learning methods can accurately predict risk genes linked to Autism Spectrum Disorder (ASD). Using spatiotemporal gene expression patterns in the human brain, gene-level constraint metrics, and other factors of gene variation, researchers found new genes that increase the risk of ASD and confirmed genes that were already known to do so. The study showed that the genes that were expected to be found are linked to important biological processes like neural signaling, neurogenesis, chromatin remodeling, and protein ubiquitination.. Although the results are encouraging, the study acknowledges limitations, including the necessity for additional confirmation via independent replication studies [13].

AI has a lot of promise for helping to diagnose and classify ASD, but there are still other issues and shortcomings that need to be fixed. A primary difficulty is the quality and diversity of data utilized for training AI algorithms. Inadequate or unrepresentative datasets might result in biased models that lack generalizability across diverse populations. Furthermore, the acquisition of high-quality data, particularly pertaining to brain imaging and genetic information, is frequently expensive.

The restricted interpretability of AI models is a significant concern in medical contexts, particularly Autism Spectrum Disorder (ASD). Most Deep Learning algorithms operate as "black boxes", rendering their internal mechanics challenging for medical professionals and researchers to comprehend. This results in challenges in evaluating outcomes and making clinical decisions based on AI model outputs. Transfer learning presents a viable option to address some limitations of conventional AI methodologies. The Lightweight Ensemble CNN for ASD classification using facial photos combines fast processing with strong predictive power. The use of lightweight CNN architectures like MobileNetV2 and EfficientNetB0 makes it possible for the system to work well on devices with limited computing power while still being very accurate, thanks to the ensemble method. Things that helped with this study are:

1) The strategy improves the overall accuracy of the classification by using an ensemble approach that combines predictions from several lightweight CNN models, such as MobileNetV2 and EfficientNetB0, instead of just one model. This is important for classifying ASD because the details of facial features can be very different and subtle.

2) Using a lightweight CNN architecture makes sure that the models can do their jobs quickly. This makes training and inference faster, which means that these models can be used on devices with limited processing power, like mobile phones or edge devices used in clinical settings.

II. Methodology

A. Design CNN

The dataset of face images for ASD was obtained from Kaggle and data from an autism center in Semarang, with 1.468 labeled as ASD and 1.468 labeled as normal. The dataset was divided into training and testing data with an 80:20 ratio. The block diagram and flowchart of the proposed research are shown in Fig. 1 and Fig. 2. The dataset has pictures of people's faces that are either classified as having Autism Spectrum Disorder (ASD) or not. This information could come from a variety of places, such as clinics, past research, or relevant public statistics. Before training, the facial images will go through a number of steps to prepare them, such as resizing, normalizing, and adding more data through techniques like rotation, flipping, and adjusting the contrast.

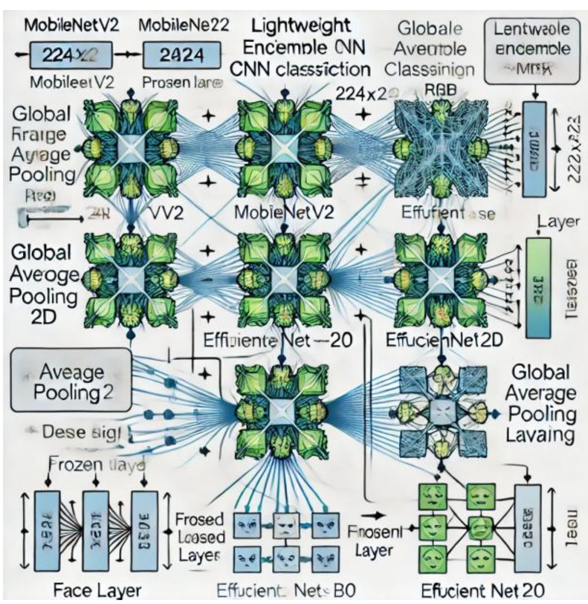


Fig. 1. Block diagram of proposed research.

The goal of these steps is to improve the quality of the dataset and make the image selection more varied, which will help the model learn better. We chose lightweight CNN designs like MobileNetV2 and EfficientNetB0 because they are fast and have small models. They use depthwise separable convolutions and compound scaling methods to make inferences quickly while keeping accuracy high. When you put a picture of a face into the model, the CNN network will extract features. This method uses a number of convolutional layers to find important patterns in the image, such as the shape of the face, the expression on it, and the texture of the skin. The result of this extraction is a number that shows what the face looks like.

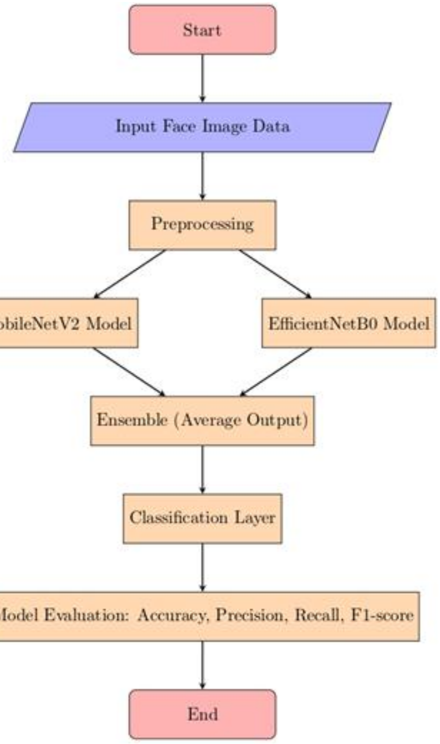


Fig. 2. Research flowchart.

The ensemble method combines the results of the two CNN models after they have independently classified the data. To get the final prediction, this method might use the average of the probability from both models. Using a voting system in which the most common result from both models is chosen as the final category. The ensemble method makes the system more reliable and accurate by lowering the chance of errors from a single model and making it better at generalizing new data. The model will give a final output that shows the chance that the person has Autism Spectrum Disorder (ASD) after the ensemble phase is over. A softmax layer is usually used to turn the model's output into a probability distribution that shows how sure the model is of its categorization. It is possible to make a final decision about whether someone has ASD based on the resulting probabilities. A specific threshold is set to tell the difference between the positive (indicating ASD) and negative (indicating absence of ASD) classifications. We use a number of metrics to judge how well this architecture works, such as accuracy, precision, recall, F1-score, and AUC. This review is very important for figuring out how well the model works at finding ASD and figuring out what the system does well and what it doesn't. This pseudo-code shows the main steps for making an ASD classification model using an ensemble method that includes two small CNN models. As shown in Fig. 3, this method combines the best features of two CNN architectures to create a model that is more accurate and robust for classifying facial images related to ASD.

```
Algorithm 1 ASD Classification using Lightweight Ensemble CNN
Input: Dataset of face images (JPEG format)
Output: ASD classification labels (ASD or non-ASD)
Initialize two pre-trained lightweight CNN models: MobileNetV2 and EfficientNetB0
procedure PREPROCESS DATA
    Load the face image dataset in JPG format
    Resize images to a fixed size (e.g., 224 × 224 pixels)
    Normalize pixel values (scaling between 0 and 1)
    Split dataset into train, validation, and test sets
end procedure
procedure MODEL DEFINITION
    Load MobileNetV2 with pre-trained weights from ImageNet
    Load EfficientNetB0 with pre-trained weights from ImageNet
    Remove top layers (head) from both MobileNetV2 and EfficientNetB0
    Add GlobalAveragePooling2D to each model output
end procedure
procedure ENSEMBLE STRATEGY
    Concatenate the outputs of MobileNetV2 and EfficientNetB0
    Add a fully connected layer (Dense layer) with 128 units and ReLU activation
    Add Dropout layer with a dropout rate of 0.5
    Add output layer with 1 unit and Sigmoid activation (binary classification)
end procedure
procedure COMPILE MODEL
    Compile the ensemble model using Binary Crossentropy loss
    Use Adam optimizer
    Define metrics for evaluation: Accuracy, Precision, Recall, F1-score
end procedure
procedure TRAIN MODEL
    Train the ensemble model on the training data
    Validate on the validation dataset
    Train for a set number of epochs (e.g., 50 epochs)
    Use EarlyStopping callback to prevent overfitting
end procedure
procedure EVALUATE MODEL
    Evaluate the trained model on the test dataset
    Calculate accuracy, precision, recall, F1-score, and confusion matrix
    Plot accuracy and loss curves across epochs
end procedure
procedure PREDICTION
    Input new face images
    Preprocess the image in the same way as the training dataset
    Predict ASD or non-ASD using the trained ensemble model
end procedure
```

Fig. 3. Pseudo-code of ASD classification algorithm with weighted ensemble CNN methodology.

B. Lightweight Ensemble CNN in ASD Classification

Lightweight Group Convolutional Neural Networks (CNN) use face images to group people with Autism Spectrum Disorder (ASD) into groups. This involves mathematically defining the network architecture and the ensemble method [14]. This is a breakdown of the parts that make up the final equation:

1) Individual model outputs: Let us label two lightweight CNN models as M1 and M2. Each model receives an input image X and produces a probability distribution across the classes (e.g., ASD or non-ASD).

The result produced by model M1 is:

$$P_1 = M_1(X) \text{ (probabilities from model 1)} \tag{1}$$

The output of model M_2 is:

$$P_2 = M_2(X) \text{ (probabilities from model 1)} \tag{2}$$

2) Ensemble method: The ensemble may be constructed by several techniques, including averaging or voting. We will examine the average method, which integrates the outputs of both models.

The ultimate ensemble output can be expressed as:

$$P_{ensemble} = \frac{P_1+P_2}{2} \tag{3}$$

3) Activation function: The final output $P_{ensemble}$ typically needs to be passed through an activation function (like softmax) to ensure it produces a valid probability distribution. The softmax function is defined as:

$$P_{final}(c) = \frac{e^{P_{ensemble}(c)}}{\sum_j e^{P_{ensemble}(j)}} \tag{4}$$

Here, c represents a specific class (e.g., ASD or non-ASD), and the sum in the denominator runs over all classes j.

4) Final equation: Combining these components, the complete mathematical representation of the Lightweight Ensemble CNN for ASD classification can be summarized as:

$$P_{final}(c) = softmax\left(\frac{M_1(X)+M_2(X)}{2}\right) \tag{5}$$

5) Training the models: During training, the loss function (e.g., categorical cross-entropy) is applied to the final predictions compared to the ground truth labels. The loss function can be expressed as:

$$L = -\sum_i y_i \log(P_{final}(c_i)) \tag{6}$$

where,

y_i is the true label for the i-th instance,

c_i is the class corresponding to the i-th instance.

III. RESULT AND ANALYSIS

The performance comparison of the five lightweight ensemble CNN models presented in Table I highlights the differences in accuracy, precision, recall, and F1-score for each model. Here's a detailed breakdown of the comparison: MobileNetV2 + EfficientNetB0. This model exhibits the most balanced performance across all metrics. It provides strong accuracy and maintains a good trade-off between precision and recall, which results in a high F1-score. This indicates that the model can consistently detect ASD cases while keeping a low false-positive rate. MobileNetV2 + EfficientNetB5, this model achieves very high precision but poor recall, indicating it identifies positive cases with high accuracy but misses a significant number of actual ASD cases (low sensitivity). The imbalance between precision and recall leads to a relatively low F1-score, showing it is less effective in capturing true positive cases.

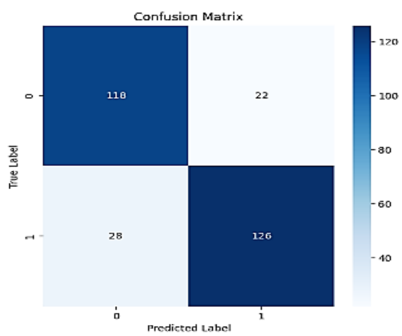
The MobileNetV2 and EfficientNetB7 work together to give very high accuracy, similar to the B5 combination, while greatly improving recall. The higher recall makes it easier to find real ASD patients, which gives it a better F1-score than the B5 variation. This means that it is better at finding both positive and negative cases overall. The ResNet50 and EfficientNetB0 models together have the best accuracy of all the models, but their recall is very low. It correctly finds positive cases, but it misses a lot of real ASD cases, as shown by the low recall. The big difference between precision and recall gives the group the lowest F1-score, which means that the model is not good enough for situations that require high sensitivity.

TABLE I. ENSEMBLE CNN LIGHTWEIGHT MODEL PERFORMANCE TEST

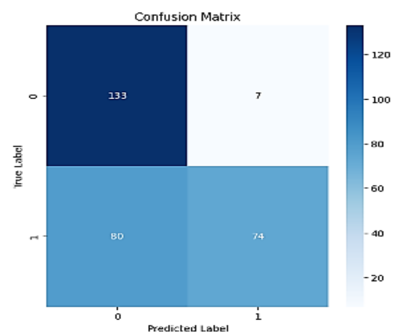
Model LightWeight + Ensemble CNN	Model Performance			
	Accuracy	**Precision**	**Recall**	**F1_score**
MobilNetV2 + EfficientB0	0.8299	0.8514	0.8182	0.8344
MobilNetV2 + EfficientB5	0.7109	0.9059	0.5000	0.6444
MobilNetV2 + EfficientB7	0.8027	0.9138	0.6883	0.7851
ResNet50 + EfficientB0	0.6667	0.9515	0.3831	0.5364
VGG16 + EfficientB0	0.7687	0.8644	0.6623	0.7500

The VGG16 + EfficientNetB0 model performs about the same as the MobileNetV2 + EfficientNetB0 combination, but it is generally worse. It has a moderate recall, which means it can find a lot of cases of ASD, but its overall accuracy and F1-score are lower, which means it is less reliable. The MobileNetV2 + EfficientNetB0 combination has the best overall performance, with balanced metrics and a high F1-score, making it the strongest model in the comparison. Most Accurate: The ResNet50 + EfficientNetB0 model is very accurate, but it doesn't do as well on recall, which makes its performance uneven. A trade-off between precision and recall.

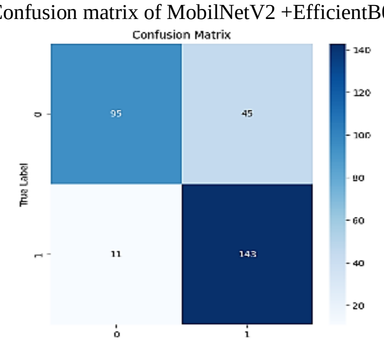
The MobileNetV2, EfficientNetB5, and EfficientNetB7 models exhibit considerable discrepancies in precision and recall, with B5 prioritizing precision and B7 providing a more favourable balance by enhancing recall. The MobileNetV2 + EfficientNetB0 model is the most effective, providing a balanced and high-performing solution for ASD classification. Other models like ResNet50 + EfficientNetB0 or MobileNetV2 + EfficientNetB5 may be valuable, where precision is the priority, but their low recall limits their effectiveness in real-world scenarios where sensitivity is equally important.



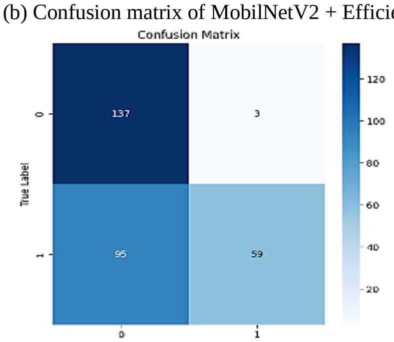
(a) Confusion matrix of MobilNetV2 +EfficientB0.



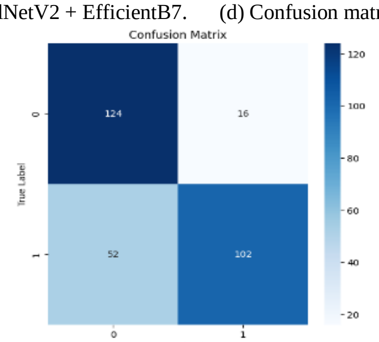
(b) Confusion matrix of MobilNetV2 + EfficientB5.



(c) Confusion matrix of MobilNetV2 + EfficientB7.



(d) Confusion matrix of ResNet50 + EfficientB0.



(e) Confusion Matrix of VGG16+ EfficientB0.

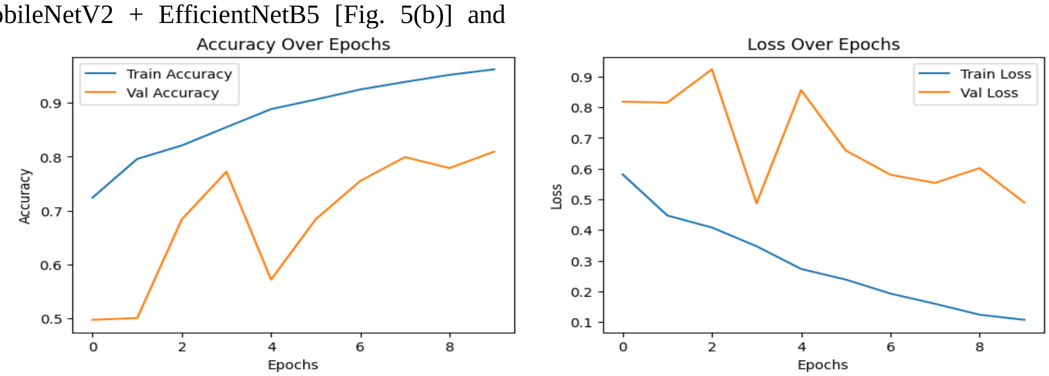
Fig. 4. Confusion matrix of ensemble CNN lightweight model.

Fig. 4 in the study presents the confusion matrices for the five lightweight ensemble CNN models discussed in Table I. These matrices provide a detailed breakdown of how each model performs in terms of correctly and incorrectly classifying instances of Autism Spectrum Disorder (ASD) and non-ASD cases. The confusion matrix is a useful tool for understanding the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each model. Fig. 4(a) has the best overall performance, with low false positives and false negatives, indicating strong accuracy and a balanced ability to correctly identify both ASD and non-ASD cases.

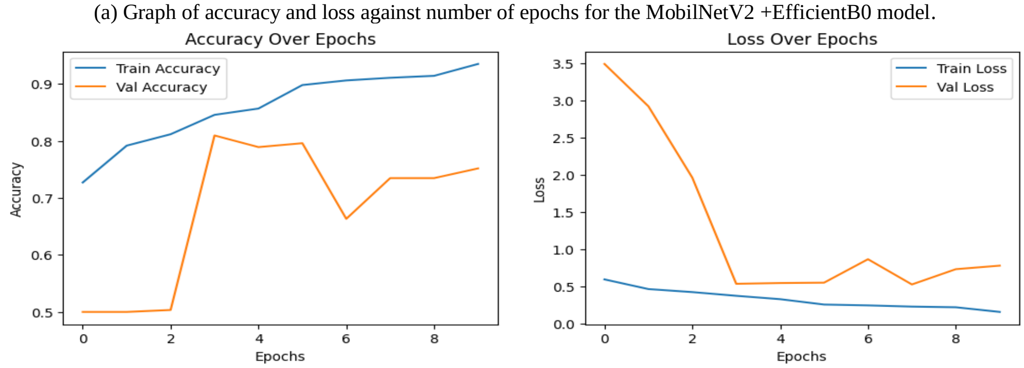
Fig. 4(b) and Fig. 4(d) are highly precise, but their confusion matrices reveal that they suffer from high false negatives, meaning they miss many ASD cases and have poor recall. Fig. 4(c) improves recall significantly compared to the B5 variant, with fewer false negatives, offering a better balance between precision and recall. Fig. 4(e) provides a moderate performance across all metrics, as seen in the relatively balanced false positives and false negatives, making it a decent model for general ASD classification.

Fig. 5 of the study presents the graphs of accuracy and loss plotted against the number of epochs for the five lightweight ensemble CNN models. These graphs provide insights into how each model learns over time and how their performance improves or stabilizes during training. MobileNetV2 + EfficientNetB0 [Fig. 5(a)] has the smoothest and most effective learning curve, with early stabilization of both accuracy and loss. This indicates a well-optimized model with minimal overfitting. MobileNetV2 + EfficientNetB5 [Fig. 5(b)] and MobileNetV2 + EfficientNetB7 [Fig. 5(c)] show slower learning with more fluctuations, indicating less stable models that might struggle with generalization. ResNet50 combined with EfficientNetB0 [Fig. 5(d)] exhibits a more gradual enhancement and a greater final loss, rendering it less successful than the other models. VGG16 combined with EfficientNetB0 [Fig. 5(e)] demonstrates moderate performance, exhibiting relatively rapid initial learning but ultimately stabilizing at a lower accuracy than the highest-performing models. The MobileNetV2 + EfficientNetB0 model demonstrates optimal efficiency and stability, attaining elevated accuracy and little loss alongside a seamless learning trajectory. Alternative models such as MobileNetV2 combined with EfficientNetB5 and ResNet50 paired with EfficientNetB0 exhibit slower learning rates and diminished performance, characterized by elevated loss and reduced final accuracy, indicating potential difficulties with complicated data or a need for more tuning to enhance outcomes.

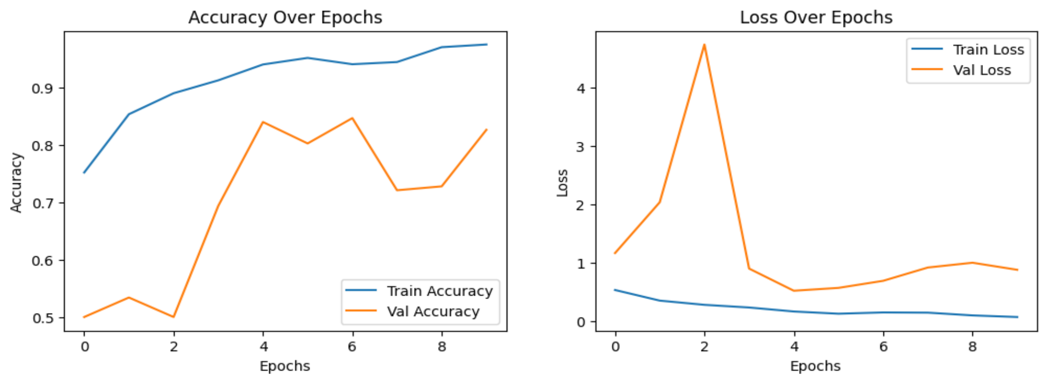
As shown in Table II, this study was also compared with previous studies. Cardoso et al. (2021) conducted a study for ASD diagnosis. Eye tracking data was used. The method used was an ensemble of random forests. The accuracy results were 0.75 and 0.82 for the F1-score [15]. Kollias et al. (2022) used Decision Trees, Logistic Regression, and Transfer Learning methods. The accuracy achieved was 0.805 [16]. The proposed study achieved an accuracy of 0.8299, which is higher than previous studies, using only a combination of the MobilNetV2 and EfficientB0 models with a facial images dataset.



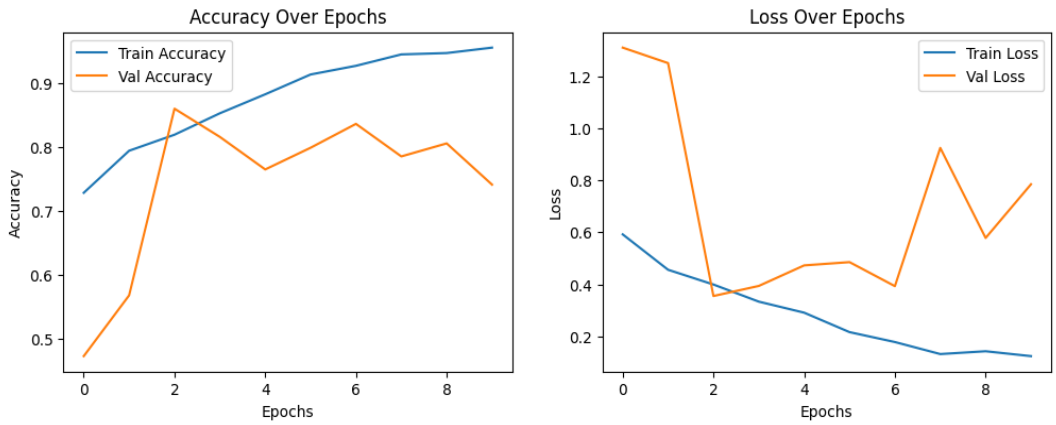
(a) Graph of accuracy and loss against number of epochs for the MobilNetV2 +EfficientB0 model.



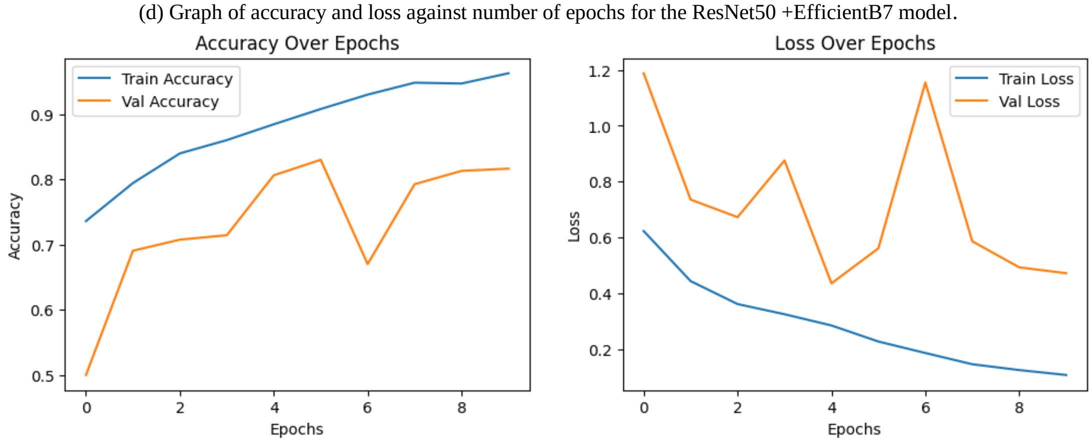
(b) Graph of accuracy and loss against number of epochs for the MobilNetV2 +EfficientB5 model.



(c) Graph of accuracy and loss against number of epochs for the MobilNetV2 +EfficientB7 model.



(d) Graph of accuracy and loss against number of epochs for the ResNet50 +EfficientB7 model.



(e) Graph of accuracy and loss against number of epochs for the VGG16 +EfficientB7 model.

Fig. 5. Graph of accuracy and loss against number of epochs for ensemble CNN lightweight model.

TABLE II. PRESENTATION OF THE RESULTS OF PREVIOUS STUDIES DISCOVERING ASD

Study	Method	Accuracy
Cardoso et al [15]	RF Classifier + ET signals	0.7500
Kollias et al [16]	Transfer learning +DT, logistic refression	0.8050
Proposed Method	MobilNetV2 + EfficientB0 +Facial Images	0.8299

IV. CONCLUSION

ResNet50 combined with EfficientNetB0 [Fig. 5(d)] exhibits a more gradual enhancement and a greater final loss, rendering it less successful than the other models. VGG16 combined with EfficientNetB0 [Fig. 5(e)] demonstrates moderate performance, exhibiting relatively rapid initial learning but ultimately stabilizing at a lower accuracy than the highest-performing models. The MobileNetV2 + EfficientNetB0 model demonstrates optimal efficiency and stability, attaining elevated accuracy and little loss alongside a seamless learning trajectory. Alternative models such as MobileNetV2 combined with

EfficientNetB5 and ResNet50 paired with EfficientNetB0 exhibit slower learning rates and diminished performance, characterized by elevated loss and reduced final accuracy, indicating potential difficulties with complicated data or a need for more tuning to enhance outcomes. While other models, such as ResNet50 + EfficientNetB0, showed high precision, they lacked the recall necessary to detect all ASD cases effectively, making them less suitable for comprehensive diagnosis. This study was also compared with other researchers, and the proposed study was found to have greater accuracy than other researchers. The findings emphasize the potential of deep learning and AI in augmenting ASD screening processes, particularly in areas where access to specialized healthcare professionals is limited. However, challenges remain, including the need for larger, more diverse datasets to improve generalization across different populations and further development of explainable AI techniques to enhance model interpretability for clinicians. Future research should focus on refining the model's ability to generalize across different facial types and exploring its integration with other diagnostic tools for a multi-modal approach to ASD diagnosis.

Acknowledgment

The author would like to thank the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for its support for this research through the Regular Fundamental Grant scheme with Contract No. 016/LL6/PL/AL.04/2025.

References

[1] Diagnostic and statistical manual of mental disorders: DSM-5TM, 5th ed.," Diagnostic and statistical manual of mental disorders: DSM-5TM, 5th ed. American Psychiatric Publishing, Inc., Arlington, VA, US, pp. xliv, 947–xliv, 947, 2013. doi: 10.1176/appi.books.9780890425596.

[2] O. I. Talantseva et al., "The global prevalence of autism spectrum disorder: A three-level meta-analysis.," Frontiers in psychiatry, vol. 14. Switzerland, p. 1071181, 2023. doi: 10.3389/fpsyt.2023.1071181.

[3] A. D. Hernawan, A. Diningrum, S. N. Jati, and M. Nasip, "Risk Factors of Autism Spectrum Disorder (ASD)," Unnes J. Public Heal., vol. 7, no. 2, pp. 104–112, 2018, doi: 10.15294/ujph.v7i2.20565.

[4] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques," SN Comput. Sci., vol. 2, no. 5, pp. 1–9, 2021, doi: 10.1007/s42979-021-00776-5.

[5] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," Transl. Psychiatry, vol. 6, no. 2, pp. 1–5, 2016, doi: 10.1038/tp.2015.221.

[6] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," NeuroImage Clin., vol. 17, no. November 2016, pp. 16–23, 2018, doi: 10.1016/j.nicl.2017.08.017.

[7] D. Bone, S. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi- instrument fusion," J Child Psychol Psychiatry, vol. 57, no. 8, pp. 927–937, 2017, doi: 10.1111/jcpp.12559.Use.

[8] R. Thapa et al., "Machine Learning Differentiation of Autism Spectrum Sub-Classifications," J. Autism Dev. Disord., 2023, doi: 10.1007/s10803-023-06121-4.

[9] D. Bandara and K. Riccardi, "Graph Node Classification to Predict Autism Risk in Genes," Genes (Basel)., vol. 15, no. 4, 2024, doi: 10.3390/genes15040447.

[10] J. W. Song, N. R. Yoon, S. M. Jang, G. Y. Lee, and B. N. Kim, "Neuroimaging-based deep learning in autism spectrum disorder and attention-deficit/hyperactivity disorder," J. Korean Acad. Child Adolesc. Psychiatry, vol. 31, no. 3, pp. 97–104, 2020, doi: 10.5765/jkacap.200021.

[11] A. Ashraf, Z. Qingjie, W. H. K. Bangyal, and M. Iqbal, "Analysis of Brain Imaging Data for the Detection of Early Age Autism Spectrum Disorder Using Transfer Learning Approaches for Internet of Things," IEEE Trans. Consum. Electron., vol. 70, no. 1, pp. 4478–4489, 2024, doi: 10.1109/TCE.2023.3328479.

[12] R. A. Bahathiq, H. Banjar, A. K. Bamaga, and S. K. Jarraya, "Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging," Front. Neuroinform., vol. 16, no. September, 2022, doi: 10.3389/fninf.2022.949926.

[13] Y. Lin, S. Afshar, A. M. Rajadhyaksha, J. B. Potash, and S. Han, "A Machine Learning Approach to Predicting Autism Risk Genes: Validation of Known Genes and Discovery of New Candidates," Front. Genet., vol. 11, no. September, pp. 1–11, 2020, doi: 10.3389/fgene.2020.500064.

[14] Y. Li, W. C. Huang, and P. H. Song, "A face image classification method of autistic children based on the two-phase transfer learning," Front. Psychol., vol. 14, no. August, pp. 1–14, 2023, doi: 10.3389/fpsyg.2023.1226470.

[15] T. V. Cardoso et al., "Autism spectrum disorder diagnosis based on trajectories of eye tracking data," in Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2021, pp. 50–55. doi: 10.1109/CBMS52027.2021.00016.

[16] K. F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "Autism detection in High-Functioning Adults with the application of Eye-Tracking technology and Machine Learning," in 2022 11th International Conference on Modern Circuits and Systems Technologies, MOCAST 2022, IEEE, 2022, pp. 1–4. doi: 10.1109/MOCAST54814.2022.9837653.