The 2nd International Conference on Multidisciplinary Engineering
(2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00014 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Face Recognition System For Student Identification Using Vgg16 Convolutional Neural Network

Chrisnata Manihuruk[1] **Muhammad Fikry**[✉2] **Hafizh Al Kautsar Aidilof**[3]
[1] Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, chrisnata.210170180@mhs.unimal.ac.id
[2] Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, muh.fikry@unimal.ac.id
[3] Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, hafizh@unimal.ac.id
✉Corresponding Author: **muh.fikry@unimal.ac.id | Phone: +6282277870744**

Abstract
In this paper, we present a robust facial recognition system designed to identify students at Department of Informatics in Universitas Malikussaleh using a Convolutional Neural Network (CNN) algorithm, specifically the VGG16 architecture. The advancement of information technology and machine learning has significantly improved facial recognition capabilities, establishing it as a reliable alternative to traditional identification methods such as fingerprinting and iris scanning. Our approach leverages a diverse dataset captured from five different angles, enhancing the representation of facial features and improving model training. The system development comprises several critical stages, including image acquisition, preprocessing, model training with training and validation data, and performance evaluation. Experimental results indicate that the CNN model achieves an impressive accuracy of 99.09% on training data and 100% on both validation and testing datasets. These findings affirm the model's high classification accuracy across the tested classes, underscoring the effectiveness of the VGG16-based CNN in facial recognition applications. The implications of this study suggest that the developed system can significantly enhance digital attendance and security systems, catering to the growing demand for reliable AI-driven security technologies in contemporary society. We anticipate that with its promising outcomes, this system can be implemented on a larger scale, contributing to the ongoing advancement of AI-based security solutions.
Keywords: AI-driven security, Convolutional Neural Network, Facial Recognition, VGG16 Architecture.

Introduction

The rapid advancement of technology has significantly impacted various sectors, particularly in information technology. This evolution facilitates human tasks and enables meaningful changes across numerous domains [1]. One of the most notable developments is in the field of machine learning, which has found applications in diverse areas such as image processing, natural language processing, and data analysis. As machine learning continues to innovate through various algorithms, it adeptly processes increasingly complex data, including facial recognition.

Facial recognition technology has emerged as a key application of machine learning, profoundly influencing security measures and user experiences [1]. This technology enables the automatic identification of human faces and has been widely adopted in applications ranging from security systems to access control and digital attendance. Compared to conventional identification methods, such as fingerprinting or iris scanning, facial recognition provides a more practical solution that enhances both efficiency and accuracy. Given the growing demand for reliable security technologies in contemporary society, the implementation of facial recognition systems is becoming increasingly essential.

In the implementation of facial recognition, Convolutional Neural Network (CNN) algorithms have proven to be more accurate than traditional methods, such as Local Binary Patterns Histograms (LBPH). CNNs excel in learning complex and varied facial features [2], thereby increasing the effectiveness of facial recognition systems. Utilized the LBPH method, achieving an accuracy rate of 86% when recognizing faces with various expressions [3]. Additionally, developed an attendance system based on facial recognition using the Fisherface method, which recorded attendance automatically with an accuracy of 80%. These studies indicate that while LBPH and Fisherface methods show potential, CNNs offer superior accuracy and adaptability to a wider range of facial variations [4].

This study aims to explore the potential of employing CNN algorithms for facial recognition among students at the Universitas Malikussaleh's Department of Informatics. The research utilizes a dataset comprising five facial angles to ensure the system can recognize faces from various perspectives. By focusing on five student samples, the study aspires to create an accurate facial recognition system suitable for security and digital attendance purposes. The findings from this research could provide valuable insights into the application of advanced facial recognition technologies in academic settings.

The development of a robust facial recognition system using CNNs represents a significant advancement in the intersection of technology and education. This research not only contributes to the academic landscape but also addresses the pressing need for innovative solutions to enhance security measures in educational institutions. By leveraging the strengths of CNNs, this study endeavors to pave the way for more reliable and efficient facial recognition applications, ultimately supporting the growing demand for AI-driven security technologies in modern society.

Literature Review

Machine learning is a branch of computer science that focuses on developing systems to learn from data and experience without explicit programming. This method allows computers to recognize patterns in datasets, generate predictions, or make more informed decisions based on learned data. With these patterns, the computer can process the information further and store the algorithm as a machine learning model used for future decision making [5].

Deep learning is a part of machine learning that mimics the structure of human neural networks. This neural network consists of several layers that process information independently without human intervention. Deep learning becomes more effective in processing complex data such as images or videos, and is known to be more accurate and reliable in object or feature recognition tasks from digital images. These algorithms perform automatic tuning and select the most optimal model for classification or prediction [6],[7].

Face recognition is a biometric technology capable of identifying individuals based on facial characteristics from photos or videos. This process involves comparing facial features with data already in the database. Face recognition is often used in security systems and requires devices such as cameras and sensors to capture facial images. This technology uses methods to determine whether the image is a human face and matches it with a corresponding profile [8].

Figure 1. Face Recognition

In the context of face recognition, a digital image is a two-dimensional image composed of pixels that carry information about colour and intensity. Every image used in a face recognition system must go through an image processing process, such as transformation to grayscale or separation of important features [9]. These processes help to improve the quality of the data used in model training, allowing the facial recognition system to identify faces more accurately and efficiently [10].

A dataset is a collection of data in a specific context that is used to train and test machine learning models. To train machine learning models, datasets with the same context are required, such as images of plant diseases, facial photographs, or stock price movements. The quality and variety of datasets greatly affect the accuracy of machine learning models. As done by [11], they improved the dataset quality by preprocessing to remove invalid, inconsistent, and duplicate data. In addition, [12] performs augmentation to increase the variety of the dataset, such as zooming in, zooming out, rotating, or changing the colour of photos based on existing data.

Materials & Methods

1. Data Collection

The training process in Face Recognition requires three main types of data: training data, testing data, and validation data. Training data is used to teach the program to recognise patterns from available faces. The validation data serves as a reference to assess how well the programme learnt the pattern from the training data. Meanwhile, the test data is used to measure the accuracy and reliability of the program in recognising new faces that have never been seen, based on the learning obtained from the training data.

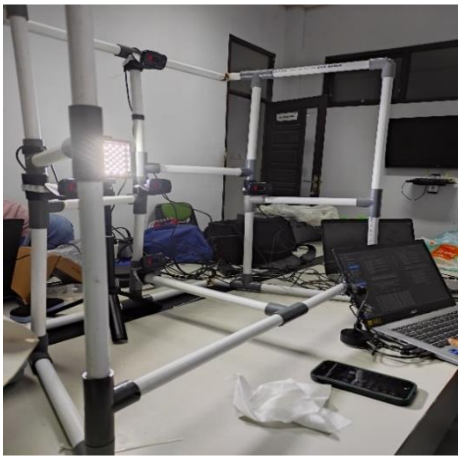

Figure 2. Modified Camera Frame

In this project, facial data from department of informatics students was used. For face data collection, a semi-circular camera frame was made from a pipe, allowing images to be taken from 5 different angles of the face. This method improves the training quality of the program, as more face angles are recognized, resulting in higher accuracy.

Figure 3. Example face image dataset

This data is organized into folders according to the type of dataset (dataset_train, dataset_test, dataset_evaluation). Each main folder also has separate subfolders for each student, organized by Name and NIM. Overall, there are a total of 2.125 images used in training, testing and evaluation, so the programme gets enough data to learn optimally.

2. Method

Convolutional Neural Network (CNN) is one of the deep learning methods designed to process visual images. CNN works by convolving the input image with filters, extracting important features, and distinguishing various objects or types of images. CNN has two main stages, namely feature learning and classification, where each layer performs feedforward and backpropagation processes to optimize prediction results. CNN became the main foundation in face recognition due to its ability to handle visual data effectively [10].

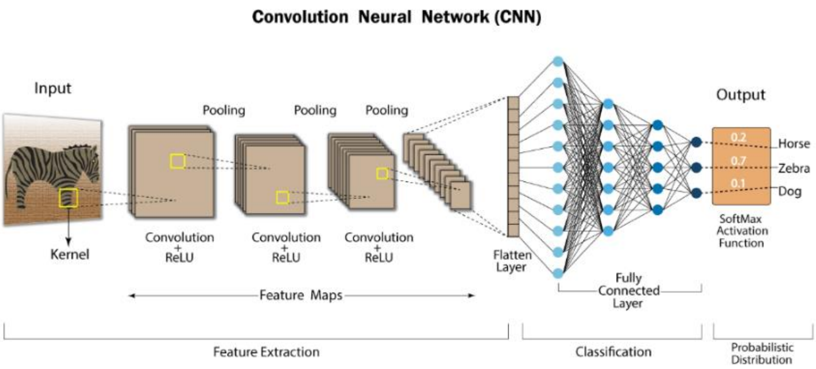

Figure 4. Illustration of Convolutional Neural Network

Conceptually, a Convolutional Neural Network (CNN) is an algorithm that operates hierarchically, where the output of each layer is used as input for the next layer [11]. The process in CNN is explained through an illustration as shown in Figure 1, which depicts the structure of a CNN model with various types of layers such as convolution layer, max pooling, flatten layer, and fully connected layer. The illustration also shows the interconnection between each output of these layers, which becomes the input for the next layers in the CNN architecture.

VGG16 is a Convolutional Neural Network (CNN) architecture known for its simplicity and effectiveness in image classification tasks. Developed by the Visual Geometry Group at the University of Oxford, VGG16 consists of 16 layers with learnable parameters: 13 convolution layers with small 3x3 filters, 5 max-pooling layers to lower the spatial dimension, and 3 fully connected layers at the end. VGG16 uses ReLU activation in each layer to introduce non-linearity, followed by a softmax layer for multi-class classification. Despite having a high number of parameters (approximately 138 million), making it computationally heavy, VGG16 is widely used for transfer learning and feature extraction due to its robust and adaptable performance [1].

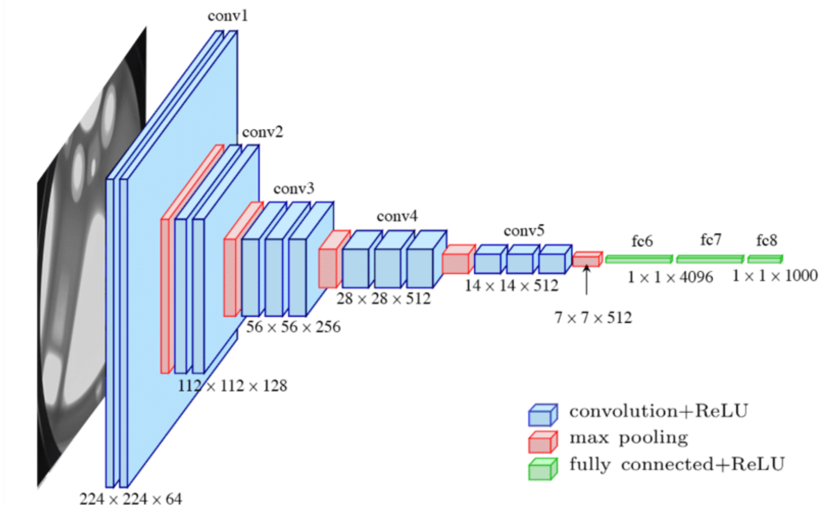

Figure 5. VGG16 Architecture

VGG16 is a CNN architecture with the input used in the form of an RGB image measuring 224 × 224 pixels. There are 2 types of convolutional layers used in this architecture, namely convolutional layers with 3x3 filter size (conv3) and 1x1 filter size (conv1). The size of the convolutional layer used varies, namely 64x64, 128x128, 256x256, and 512x512. This model is the best in localization and classification [13].

3. Evaluation

The process of creating a Face Recognition system involves several stages, starting from image capture by the camera to the identification of the captured image.

An image is a two-dimensional representation of a three-dimensional physical object. Image acquisition can be in the form of black-and-white images, photographs, or moving colour images. The Face Recognition process starts with the camera capturing the face image, and often multiple images are captured in a single process, which is called the multi-Face Recognition technique. The captured images are stored in a dataset.

Preprocessing is the initial processing of image data before going through the Convolutional Neural Network (CNN) algorithm. Some of the stages in preprocessing include image cropping. This stage aims to reduce or eliminate noise, clarify important features in the image, resize the image as needed, convert the original image to fit the required format.

A total of 500 images were used in the training dataset, which included various facial expressions to enrich the available data. The expressions include smiles, frowns, bulging eyes, squinting eyes, as well as facial expressions while talking. All images were taken without the use of accessories such as glasses or masks. This training dataset serves as the main data in the programme training process.

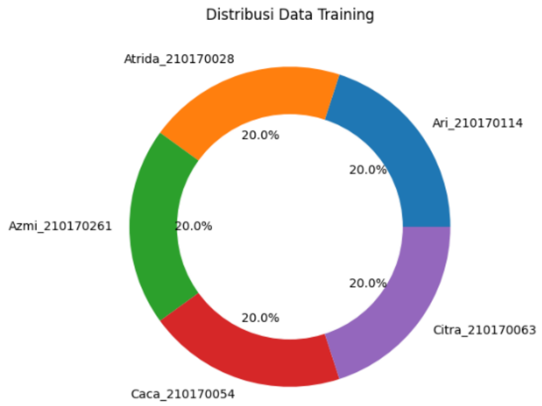

Figure 6. Training data distribution diagram.

The diagram above illustrates the distribution of training data from the dataset used in this study. Each class has an equal number of 500 images, so each segment represents 20% of the total training dataset. This balanced distribution is very important in the model training process, as it ensures that the model can learn well from each class. Thus, the model will be better able to recognize and classify the faces accurately when faced with new data.

Validation data is data that serves as a comparison with train data. The goal is to measure the accuracy of the model that has been trained with the training data. With a variety of expressions and taken from various points of view as seen in **Figure.6.** These images help the programme test the accuracy of the model on patterns that have never been seen before.

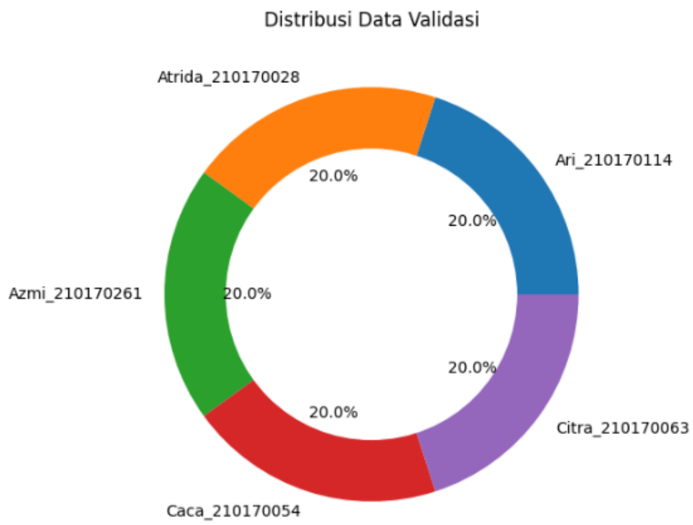

Figure 7. Validation data distribution diagram

The diagram above shows the distribution of validation data from the dataset used in this study. Each segment in the diagram represents the proportion of each student face class, which consists of five individuals: 'Ari_210170114', 'Atrida_210170028', 'Azmi_210170261', 'Caca_210170054', and 'Citra_210170063'. Each class has an equal number of 90 images, so each segment on the diagram represents 20% of the total dataset. With this even distribution, the research can ensure that the trained model has an equal opportunity to learn from each class, which is important for improving accuracy in the face classification process.

The test data is used to compare the identification results of the models that have been trained using the training data and validation data, which are then stored in the weight_train_validation.h5 file. This test dataset contains 85 unique images that are different from the training data and validation data. The images in the test dataset include a variety of facial expressions with five different viewing angles. After comparison with the test data, the model will produce the final identification results

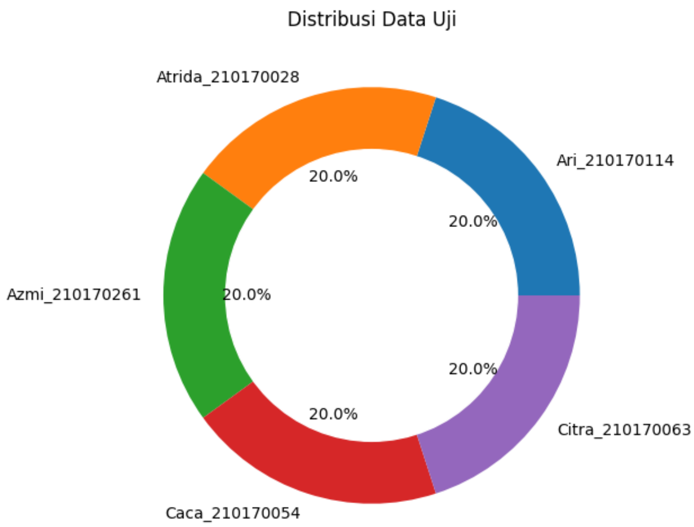

Figure 8. Test data distribution diagram

Each class consists of 85 images, so the total number of images for all classes is 425. A balanced proportion, where each class gets 20% of the total test data, is very important in model evaluation. This even distribution ensures that the model is fairly tested against all classes, allowing for a more precise assessment of accuracy.

Accuracy is the most common metric used to assess the effectiveness of an algorithm by estimating the correct probability of a class, and is calculated by dividing the number of correct predictions by the total predicted data. Accuracy calculations in machine learning often involve the confusion matrix, which provides information on the comparison between the classification results produced by the model and the actual classification. In this context, some frequently used performance metrics include accuracy, precision, and recall, each of which can be calculated using a specific formula [10].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Precission = \frac{TP}{TP+FP} \quad (2)$$

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

Explanation:

1. True Positive (TP), the value at which the model correctly predicts the positive class.
2. True Negative (TN), the value at which the model correctly predicts the negative class.
3. False Positive (FP), the model result incorrectly predicts the positive class.
4. False Negative (FN), the model result incorrectly predicts the negative class.

Results and Discussion

1. Coding Stage

This stage includes creating program code using a predetermined programming language and algorithm. In this study, the code was created with the Visual Studio Code application, and Using Jupyter Notebook, specifically. The algorithm used is Convolutional Neural Network (CNN), with a dataset that has been designed as needed.

2. Enrollment

This section serves to capture and store facial images as datasets. This stage is the first step in writing code, where the camera is used to take pictures that will later be stored as part of the dataset.

3. Model

This section covers model building and training using TensorFlow, a machine learning framework from Google that supports the development of artificial neural networks, including CNNs. The model used in this research is VGG16, a CNN model that helps identify images more accurately at larger scales.

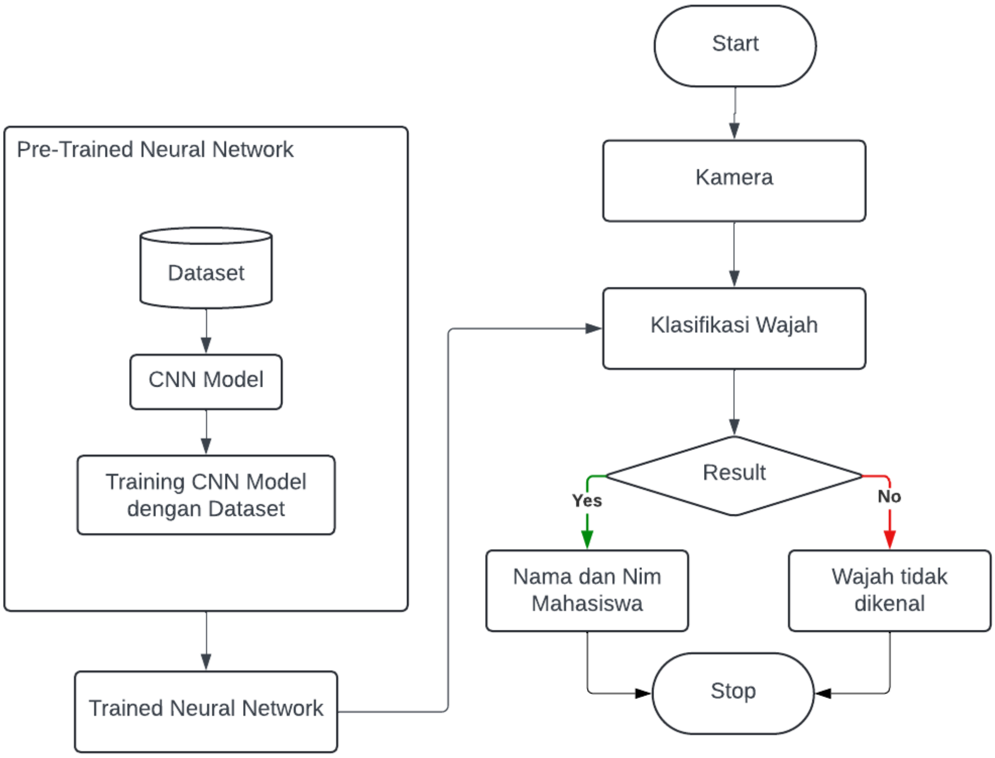

Figure 9. Design System

The scheme of this system follows several stages, starting with the division of datasets that have been collected as needed (train, validation, and test datasets) manually. The split datasets will undergo preprocessing, including image size adjustment. Next, the datasets will go through data augmentation involving filtering and geometry transformation processes to evaluate their impact on the neural network. The scheme includes several key components: Pre-Trained Neural Network, which makes use of the labelled dataset to train a CNN model and then test the accuracy of the model using validation data; a camera to capture facial images; and a face classification process that uses the trained model to identify the captured faces. If the classification result shows 'Yes' (recognised face), then the system displays the student's name and NIM. If the result is 'No' (unknown face), then the system indicates that the face is unknown.

4. Training

The training process involves matching data from various datasets, such as training data and validation data. Images that have been categorized in different datasets are compared and then merged into a single data file. At this stage, training data and validation data are used to train the model to recognize facial images more accurately. **Figure 12** shows the flowchart of the training process. The train dataset will be trained and validated using validation data that is different from the train dataset, then saved in a single .h5 format file with the name *weight_train.h5*. The data in this *model_train.h5* file will be used in the next stage to identify the captured image or images.

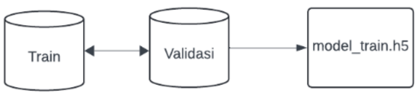

Figure 10. Flowchart Of the training process

5. Indentification

The identification stage is the main process that performs the functions for face identification. This process includes several functions that are responsible for data organization, logic, and display of output results. **Figure 13** shows the diagram of the matching process between the test data and the *weight_train.h5* file. The process starts by calling the required dataset, namely *weights_train.h5*, along with the test data from the directory where the file is stored. The result of this process is the confusion matrix, which is used to measure the accuracy of the trained model.

Figure 11. Flowchart Of the identification process

6. Classification Results

In this research, the Convolutional Neural Network (CNN) method was used with the VGG16 architecture model. The resulting model shows the ability to classify images into five different classes of student faces, namely 'Ari_210170114', 'Atrida_210170028', 'Azmi_210170261', 'Caca_210170054', and 'Citra_210170063'.

In Figure 12 and 13, the process of training the model to recognize and learn the patterns contained in the images is shown. This training takes place over a number of epochs, with this study using 5 epochs. The performance of the model is evaluated using training data and validation data. This process is performed with a dataset consisting of training data and validation data. Once trained, the model will be used to predict the class labels of the images taken from the validation dataset.

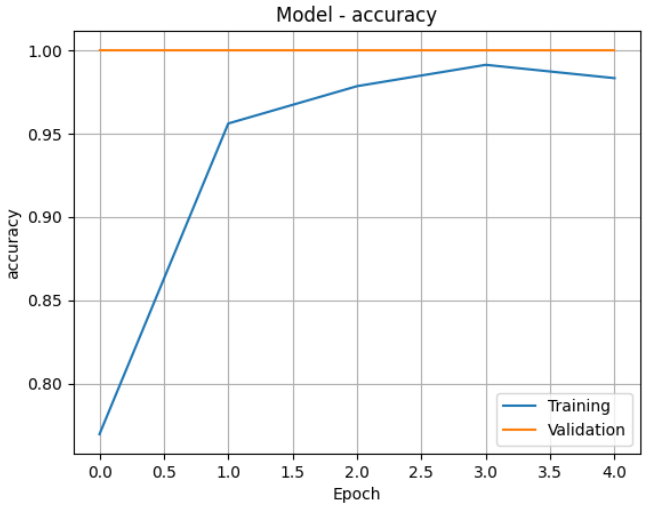

Figure 12. VGG16 Accuracy Training Chart

In Fig. 14, the performance of the model during the training process on the dataset used. At the beginning of the first epoch, the model shows an accuracy of 58.63% with a loss of 1.0849 for the training data, and a val accuracy of 100% with a loss of 0.0206. After completing the second epoch, the model accuracy increased to 93.98% and continued to rise until it reached 97.64% in the third epoch, with val accuracy remaining 100%. In the fourth epoch, the model accuracy reached 99.09% with a much lower loss, while in the fifth epoch, the accuracy slightly decreased to 98.57%, but still showed excellent performance with val accuracy remaining 100%.

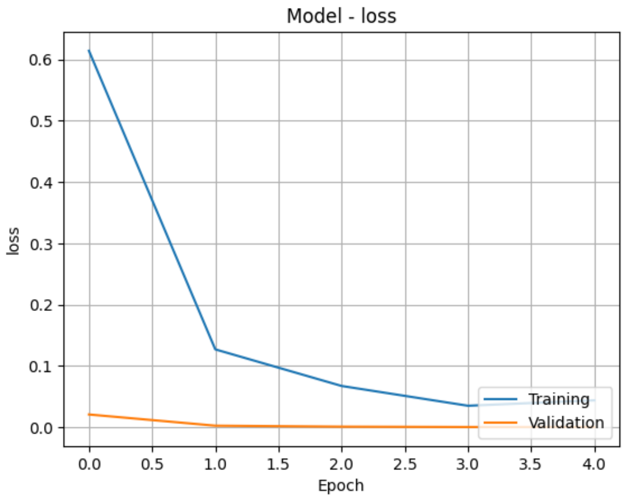

Figure 13. VGG16 Loss Training Chart

In Fig. 15, shows the performance of the model during the training process on the dataset used. At the beginning of the first epoch, the model recorded a loss of 1.0849 for training data and 0.0206 for validation data. After completing the second epoch, the model showed significant improvement with an accuracy of 93.98% and a lower loss of 0.1603. Furthermore, in the third epoch, the accuracy increased again to 97.64% with a loss of 0.0773, and in the fourth epoch, the model reached an accuracy of 99.09% with a very small loss of 0.0346. In the fifth epoch, the model accuracy slightly decreased to 98.57%, but the loss remained low at 0.0401. Performance on the validation data remains excellent with 100% accuracy and near zero loss.

7. Classification results of all classes

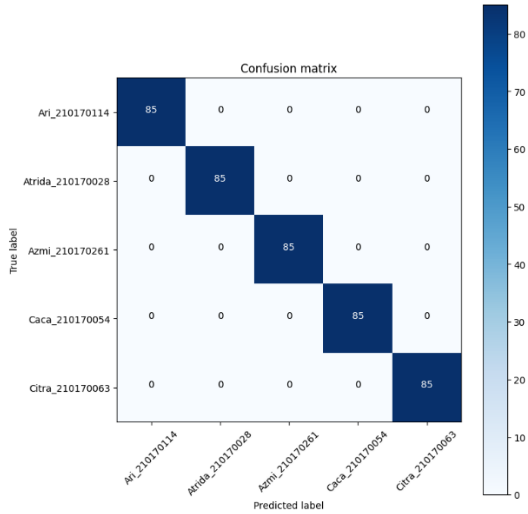

Figure 14. Confusion Matrix on Test Data

Table 1 shows the prediction results obtained using the CNN model with VGG16 architecture in the testing process using dataset_uji. This dataset consists of a total of 85 samples for each class, which includes five different classes. The model showed accurate classification results for each class, resulting in high accuracy across all tested datasets. The prediction results are shown in the attached confusion matrix, with details as follows:

1. For the class "Ari_210170114," which consists of 85 images, the model successfully classified all images correctly.
2. For the class "Atrida_210170028," with 85 images, the model was able to correctly identify all images.
3. For the class "Azmi_210170261," which also has 85 images, the model classified all images accurately.
4. In the class "Caca_210170054," out of a total of 85 images, the model was able to classify all images correctly.
5. In the class "Image_210170063," the model again showed full accuracy with 85 images correctly classified.

This confusion matrix shows that the CNN model with VGG16 architecture achieved high accuracy in face classification, with all images in the test dataset correctly classified without any errors, reflecting optimal performance in identifying each class in the test dataset.

Table 1. Calculate Matrix Report

	Precision	Recall	F1-Score	Support
Ari_210170114	1.00	1.00	1.00	85
Atrida_210170028	1.00	1.00	1.00	85
Azmi_210170261	1.00	1.00	1.00	85
Caca_210170054	1.00	1.00	1.00	85
Citra_210170063	1.00	1.00	1.00	85
Accuracy			1.00	425
Macro Avg	1.00	1.00	1.00	425
Weighted Avg	1.00	1.00	1.00	425

In Figure 11 above, the level of precision that indicates the accuracy of the model for each class is as follows: "Ari_210170114," 'Atrida_210170028,' 'Azmi_210170261,' 'Caca_210170054,' and 'Citra_210170063' all achieve a precision of 100%. This shows that the VGG16 model successfully classified each instance of each class accurately.

Regarding recall, which measures the success rate of the model in identifying positive examples, the results also show a value of 100% for all tested classes. This means that the model was able to recognize every positive example from all classes without error.

Each class showed perfect precision, recall, f1-score, and support values, resulting in an overall accuracy of 100%. These results show that the VGG16 model performed optimally in classifying the test dataset, with highly accurate performance across all tested classes.

Conclusions

Based on the results of research on *Face Recognition* systems using Convolutional Neural Network (CNN) models with

VGG16 architecture, it can be concluded that the use of diverse facial data for training, validation, and testing plays an important role in improving the accuracy of the model. In this study, the facial data of five Informatics Engineering students was taken with a unique method, using a semicircular camera frame that allows taking pictures from five different viewpoints. This method proved to be able to improve model accuracy by providing a wider variety of angles and facial expressions.

The VGG16 CNN model trained with an even distribution of data in each class showed optimal performance during training and testing. The training accuracy of the model increased significantly from 58.63% to reach 99.09% after four *epochs*, and the accuracy of the validation data was consistent at 100%. During testing, the model was able to classify all test images in each class without error, resulting in an overall accuracy of 100%. The high level of precision in all classes shows that the VGG16 architecture is able to recognize faces accurately and consistently.

This research also shows that CNN-based *Face Recognition* technology can be applied on a larger scale and wider security scope. This model has the potential to be used in various sectors that require artificial intelligence-based face recognition systems to improve efficiency and security. The results of this research can be used as a reference for further development in the field of face recognition technology for implementation in various environments.

References

[1] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, *35*(4), 399-458.

[2] Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association.

[3] Ranjini, M. M. D., Jeyaraj, M. P., Kumar, M. S., Prasath, T. A., & Prabhakar, G. (2023, July). Haar Cascade Classifier-based Real-Time Face Recognition and Face Detection. In *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 990-995). IEEE.

[4] Budiman, A., Yaputera, R. A., Achmad, S., & Kurniawan, A. (2023). Student attendance with face recognition (LBPH or CNN): Systematic literature review. *Procedia Computer Science*, *216*, 31-38.

[5] Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, *10*(07), 3897-3904.

[6] Liu, F., Chen, D., Wang, F., Li, Z., & Xu, F. (2023). Deep learning based single sample face recognition: a survey. *Artificial Intelligence Review*, *56*(3), 2723-2748.

[7] Lin, Y., Xie, S., Ghose, D., Liu, X., You, J., Korhonen, J., ... & Dash, S. P. (2024). FishIR: Identifying Pufferfish Individual Based on Deep Learning and Face Recognition. *IEEE Access*.

[8] Autade, A., Adhav, P., BabarPatil, A., Dhumal, A., Vispute, S., Rajeswari, K., ... & Rathi, S. (2023, August). Automated Multi Face Recognition and Identification using Facenet and VGG-16 on Real-World Dataset for Attendance Monitoring System. In *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)* (pp. 1-5). IEEE.

[9] Muhammad Fikry. "Performance Analysis of Smart Technology With Face Detection Using YOLOv3 and InsightFace for Student Attendance Monitoring". International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 4, June 2024, pp. 3490 -, https://ijisae.org/index.php/IJISAE/article/view/6865.

[10] Boutros, F., Struc, V., Fierrez, J., & Damer, N. (2023). Synthetic data for face recognition: Current state and future prospects. *Image and Vision Computing*, *135*, 104688.

[11] Krichen, M. (2023). Convolutional neural networks: A survey. *Computers*, *12*(8), 151.

[12] Yamsani, N., Jabar, M. B., Adnan, M. M., Hussein, A. H. A., & Chakraborty, S. (2023, December). Facial Emotional Recognition Using Faster Regional Convolutional Neural Network with VGG16 Feature Extraction Model. In *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1-6). IEEE.

[13] Riehl, K., Neunteufel, M., & Hemberg, M. (2023). Hierarchical confusion matrix for classification performance evaluation. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(5), 1394-1412.