

Face Mask Detection Using Deep Learning

Using Deep Learning Techniques

Ahood Faisal Al shahrani

Dr. Yasser Abdel Hamid

AI901 – Artificial Intelligence Thesis

Midocean University

Second Semester – 2026

Chapter 1

Introduction

The spread of the COVID-19 pandemic created a global health challenge and required many preventive measures to reduce infection rates. One of the most effective safety measures recommended by health organizations was wearing face masks in public places. Monitoring whether people are wearing masks in crowded environments such as airports, hospitals, schools, and shopping centers can be difficult when done manually.

With the development of artificial intelligence and computer vision technologies, automated systems can now help monitor safety measures. Deep learning techniques allow computers to analyze images and recognize patterns within visual data.

This project focuses on developing an intelligent system that can detect whether a person is wearing a face mask or not using deep learning. The system is trained using a dataset containing images of people wearing masks and people not wearing masks. The goal of the model is to classify images into two categories: with mask and without mask.

Chapter 2

Background and Related Work

Artificial Intelligence and Computer Vision

Artificial intelligence is a field of computer science that focuses on building systems capable of performing tasks that usually require human intelligence. Computer vision is a branch of artificial intelligence that enables machines to interpret and understand visual information such as images and videos.

Deep Learning

Deep learning is a subset of machine learning that uses neural networks with multiple layers to learn patterns from data. Deep learning has become widely used in image classification tasks due to its high accuracy and ability to automatically extract important features from images.

Convolutional Neural Networks

Convolutional Neural Networks (CNN) are one of the most effective deep learning architectures used for image recognition tasks. CNN models consist of several layers such as convolutional layers, pooling layers, and fully connected layers. These layers work together to extract features from images and perform classification.

Related Work

Many researchers have used deep learning models for face mask detection. Some studies used CNN models to classify images of masked and unmasked faces. Other research used real-time detection models such as YOLO to detect multiple faces in images and determine whether each person is wearing a mask.

Chapter 3

Proposed Approach

Dataset Description

The dataset used in this project consists of images of human faces divided into two categories: people wearing masks and people without masks. The dataset includes images captured under different lighting conditions, angles, and backgrounds.

Data Preprocessing

Before training the model, several preprocessing steps were performed. Images were resized to a fixed dimension to ensure consistent input to the neural network. Pixel values were normalized to improve training performance. The dataset was also divided into training and validation sets.

Model Architecture

The deep learning model used in this project is based on a Convolutional Neural Network. The model includes convolutional layers for feature extraction followed by pooling layers that reduce the spatial size of feature maps. After feature extraction, the features are passed to fully connected layers that perform the classification task.

Training the Model

The model was trained using the labeled dataset for several epochs. During training, the model adjusts its internal parameters to minimize classification errors. The Adam optimizer and binary cross-entropy loss function were used during

the training process.

Chapter 4

Experimental Results and Discussion

The trained model achieved good performance in classifying images of masked and unmasked faces. During training, the accuracy increased gradually while the loss decreased, indicating that the model successfully learned meaningful patterns from the dataset.

Training and validation graphs were used to evaluate the learning performance of the model. These graphs help analyze whether the model is improving during training and whether overfitting occurs.

Chapter 5

Conclusion and Future Work

This project presented a deep learning approach for detecting face masks using image classification. A convolutional neural network model was trained on a dataset containing masked and unmasked face images.

The results showed that the model can effectively classify images and detect whether individuals are wearing masks. This system can potentially be used in public monitoring systems to improve safety in crowded environments.

Future work may include using larger datasets, improving model accuracy, and implementing real-time face mask detection using live camera systems.

Face Mask Detection Using Deep Learning Techniques

Prepared by:

Ahood Faisal Al Shahrani

Supervised by:

Dr. Yasser Abdel Hamid

Course:

AI901 – Artificial Intelligence Thesis

Midocean University

Second Semester – 2026

Abstract

The COVID-19 pandemic has had a significant impact on global public health systems and daily human activities, highlighting the urgent need for effective preventive measures to reduce the spread of infectious diseases. Among these measures, wearing face masks in public places has been identified as one of the most effective strategies to limit the transmission of respiratory viruses. However, ensuring compliance with mask-wearing regulations in crowded environments such as hospitals, airports, shopping centers, and educational institutions remains a challenging task when performed manually. Manual monitoring requires continuous human effort, is time-consuming, and is often prone to errors.

To address this challenge, this study presents a deep learning-based system for automatic face mask detection using image classification techniques. The primary objective of this study is to develop an accurate and robust intelligent model capable of classifying facial images into two categories: individuals wearing face masks and individuals not wearing face masks.

The proposed system is based on a Convolutional Neural Network (CNN), which is widely recognized for its effectiveness in image recognition tasks. The model was trained using a labeled dataset consisting of approximately 10,000 facial images, equally divided between masked and unmasked classes. The dataset includes variations in lighting conditions, facial orientations, background complexity, and image quality, enhancing the model's ability to generalize to real-world scenarios.

To prepare the data for training, several preprocessing steps were applied, including resizing images to a fixed dimension, normalizing pixel values, and splitting the dataset into training and validation sets. In addition, data

augmentation techniques such as rotation, flipping, and zooming were used to increase dataset diversity and reduce the risk of overfitting.

The model was trained for multiple epochs using the Adam optimizer and binary cross-entropy loss function. During training, both accuracy and loss were monitored to evaluate model performance and ensure effective learning behavior.

Experimental results demonstrated that the model achieved high classification performance in distinguishing between masked and unmasked faces. The training and validation curves showed consistent improvement, indicating effective learning. The model achieved an accuracy of approximately 95%, along with strong precision, recall, and F1-score values, confirming its reliability and generalization capability.

These findings clearly highlight the effectiveness of deep learning techniques in developing automated monitoring systems for public safety. The proposed system can be integrated into surveillance systems to support health regulations and reduce the spread of infectious diseases. Future work may focus on improving model accuracy, increasing dataset size and diversity, and implementing real-time detection systems using live video streams.

Keywords: Face Mask Detection, Deep Learning, CNN, Computer Vision, Image Classification, COVID-19

Table of Contents

Cover Page	i
Abstract	ii
Table of Contents	iii
List of Abbreviations	iv
List of Tables	v
List of Figures	vi
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Significance of the Study	3
1.4 Scope of the Study	4
1.5 Objectives	5
Chapter 2: Background and Related Work	6
2.1 Artificial Intelligence	6
2.2 Computer Vision	7
2.3 Deep Learning	8
2.4 Convolutional Neural Networks	9
2.5 Related Work	10
2.6 Research Gap	11
2.7 Link to This Study	12
Chapter 3: Proposed Approach	13
3.1 Dataset Description	13
3.2 Dataset Statistics	14

3.3 Data Preprocessing	15
3.4 Model Architecture	16
3.5 Training Process	17
3.6 Evaluation Methods	18
Chapter 4: Experimental Results and Discussion ...	19
4.1 Training Performance (Accuracy and Loss)	19
4.2 Confusion Matrix	20
4.3 Performance Metrics	21
4.4 Discussion	22
Chapter 5: Conclusion and Future Work	23
5.1 Conclusion	23
5.2 Future Work	24
References	25

List of Abbreviations

AI – Artificial Intelligence

ML – Machine Learning

DL – Deep Learning

CNN – Convolutional Neural Network

CV – Computer Vision

COVID-19 – Coronavirus Disease 2019

TP – True Positive

TN – True Negative

FP – False Positive

FN – False Negative

F1-Score – Harmonic Mean of Precision and Recall

List of Tables

Table 3.1: Dataset Distribution

Table 4.1: Model Performance Metrics

List of Figures

Figure 3.1: Sample Images from the Dataset

Figure 4.1: Training and Validation Performance

Figure 4.2: Confusion Matrix of the Model Predictions

Chapter 1: Introduction

1.1 Introduction

In recent years, global health crises have emphasized the importance of intelligent technological solutions in managing public safety.

The rapid spread of the COVID-19 pandemic has significantly impacted global health systems and daily human activities. As a result, governments and health organizations introduced various preventive measures to reduce the transmission of the virus. One of the most important and effective measures was the use of face masks in public places.

Wearing face masks has proven to be an essential method for limiting the spread of respiratory infections. However, monitoring compliance with mask-wearing regulations in crowded environments such as hospitals, airports, shopping centers, and educational institutions remains a challenging task when performed manually. Human monitoring requires significant effort, time, and resources, and it is often prone to human error.

With the advancement of artificial intelligence, particularly in the field of computer vision, automated systems have become capable of analyzing images and detecting patterns with high accuracy (Russell & Norvig, 2021). Deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown remarkable success in image classification and object detection tasks (Krizhevsky et al., 2012).

This research focuses on developing a deep learning-based system that can automatically detect whether individuals are wearing face masks or not. The system uses image classification techniques to distinguish between two categories: masked faces and unmasked faces. By leveraging CNN

models, the system aims to provide accurate and efficient detection that can be applied in real-world environments.

The importance of this research lies in its practical application in enhancing public safety. Automated face mask detection systems can be integrated into surveillance systems to ensure compliance with health regulations and reduce the spread of infectious diseases.

In addition, this research contributes to the growing field of artificial intelligence by demonstrating the effectiveness of deep learning techniques in solving real-world problems. The results of this study can support future developments in intelligent monitoring systems and smart city applications.

1.2 Problem Statement

Despite the importance of wearing face masks, ensuring compliance in public spaces remains difficult. Manual monitoring systems are inefficient and cannot handle large crowds effectively. They also require continuous human supervision, which increases cost and reduces reliability over time.

Therefore, there is a need for an automated system that can accurately detect whether individuals are wearing face masks using image processing and deep learning techniques. Such a system can support public safety efforts by providing fast, accurate, and scalable monitoring.

This limitation highlights the need for intelligent automated systems capable of handling large-scale monitoring efficiently.

This challenge necessitates the development of intelligent automated systems capable of ensuring efficient and scalable monitoring.

1.3 Significance of the Study

This research is important for several reasons.

First, it provides an automated solution for monitoring compliance with public safety measures. Instead of relying entirely on human observation, institutions can use intelligent systems to improve efficiency and reduce error.

Second, the study demonstrates the value of artificial intelligence in addressing real-world healthcare and public safety challenges. During health crises such as COVID-19, intelligent monitoring systems can play an important role in enforcing preventive measures.

Third, the research contributes academically by applying deep learning and computer vision techniques to a practical classification problem. This makes the study useful not only as a technical project but also as an example of how artificial intelligence can be used in real-world environments.

Finally, the study may serve as a foundation for future work in real-time surveillance, smart city systems, and health monitoring applications.

Furthermore, this study supports the advancement of smart surveillance systems in modern cities.

1.4 Scope of the Study

This study focuses on developing an image classification model using deep learning techniques to detect face masks. The system is trained and tested using a dataset of facial images categorized into masked and unmasked classes.

The research emphasizes model development, training, and evaluation. It includes dataset preparation, preprocessing, CNN model design, and performance analysis using standard classification metrics.

The study is limited to image-based classification and does not include object detection or real-time deployment. Instead, it focuses on building and evaluating the classification model under controlled experimental conditions.

1.5 Objectives

The main objectives of this research are:

1. To design and develop a deep learning-based model for face mask detection.
2. To apply Convolutional Neural Networks (CNNs) for image classification.
3. To train the model using a labeled dataset of masked and unmasked faces.
4. To evaluate the model performance using accuracy, precision, recall, and F1-score.
5. To analyze the model's performance under different conditions.
6. To suggest improvements for future real-time applications.

Chapter 2: Background and Related Work

2.1 Artificial Intelligence

Artificial Intelligence (AI) is a rapidly evolving field of computer science that focuses on developing systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, problem-solving, and decision-making. AI systems are designed to analyze data, recognize patterns, and make predictions or decisions with minimal human intervention (Russell & Norvig, 2021).

In recent years, AI has become one of the most influential technologies in many sectors, including healthcare, security, transportation, and education. During the COVID-19 pandemic, AI-based systems were widely explored for disease prediction, health monitoring, and public safety applications.

In the context of this research, AI serves as the broader framework for developing an intelligent system capable of automatically detecting whether a person is wearing a face mask. This demonstrates how AI can be applied to practical challenges that affect public health and safety.

2.2 Computer Vision

Computer vision is a branch of artificial intelligence that enables machines to interpret and understand visual data such as images and videos. The main goal of computer vision is to allow computers to extract useful information from visual content in a way that resembles human vision (Szeliski, 2010).

Computer vision has many important applications, including facial recognition, autonomous driving, object detection, image classification, and medical image analysis. Traditional computer vision methods often relied on manually engineered features, which limited their ability to perform well in complex environments.

With the integration of deep learning, computer vision systems have become far more powerful and accurate. Instead of manually defining image features, modern systems can automatically learn them from data. In this research, computer vision is used to analyze facial images and determine whether a face belongs to a masked or unmasked category.

2.3 Deep Learning

Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers to learn complex patterns from data. It has become one of the most successful approaches in artificial intelligence, particularly in tasks involving large and complex datasets (LeCun et al., 2015).

One of the main strengths of deep learning is automatic feature extraction. Unlike traditional approaches, deep learning models can learn directly from raw input data without requiring extensive manual feature design. This makes deep learning particularly useful in fields such as image classification, speech recognition, natural language processing, and medical diagnosis.

For visual tasks, deep learning models can identify low-level and high-level image features through hierarchical learning. In the context of this research, deep learning is used to train a model to recognize the visual differences between masked and unmasked faces.

2.4 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialized type of deep learning model designed specifically for image-related tasks. CNNs are highly effective because they can capture spatial patterns in images, such as edges, textures, and shapes (Krizhevsky et al., 2012).

A typical CNN architecture consists of convolutional layers, pooling layers, activation functions, and fully connected layers. The convolutional layers extract features from the image, while pooling layers reduce dimensionality and help the model focus on the most important information. The fully connected layers then use these extracted features to make a classification decision.

CNNs are widely used in image recognition, object detection, facial analysis, and many other computer vision applications. In this research, CNN is selected because it is well suited for classifying facial images into two categories: with mask and without mask.

2.5 Related Work

Several studies have explored the use of deep learning techniques for face mask detection, especially during the COVID-19 pandemic. Many researchers have developed CNN-based systems to classify facial images into masked and unmasked categories, reporting promising results.

Some studies used transfer learning techniques with pre-trained models such as MobileNet (Howard et al., 2017), ResNet, and VGGNet to improve performance and reduce training time. These approaches are especially useful when the dataset is limited, as they make use of previously learned features from large image datasets.

Other studies focused on real-time mask detection using object detection algorithms such as YOLO (Redmon et al., 2016). These systems are capable of detecting multiple faces in a single image or video frame and classifying each one based on mask usage.

Although previous studies achieved strong results, several challenges remain. Some models perform well only under controlled conditions and may struggle in real-world scenarios with poor lighting, occlusions,

varying face angles, and complex backgrounds. In addition, some studies rely on limited or unbalanced datasets, which can reduce model reliability and generalization performance.

These studies demonstrate strong performance; however, many of them rely on specific datasets and may not generalize well to real-world conditions.

Compared with standard CNN models, transfer learning approaches such as MobileNet can reduce training time, while YOLO-based methods are more suitable for real-time detection tasks.

2.6 Research Gap

Based on the reviewed studies, it is clear that face mask detection systems have made significant progress, but certain limitations still need to be addressed. Many existing models are sensitive to changes in environmental conditions such as lighting variations, image quality, occluded faces, or unusual face orientations.

Another limitation is that some research focuses more on model accuracy than on data diversity and generalization. A model may perform well on a specific dataset but show reduced performance when applied in real-world settings.

Therefore, this research aims to develop a CNN-based model trained on a balanced and diverse dataset, with the goal of improving generalization and achieving reliable classification performance.

This gap highlights the need for a more generalized and robust model capable of handling diverse real-world scenarios.

2.7 Link to This Study

The concepts discussed in this chapter form the theoretical foundation of this research. Artificial intelligence provides the broad framework, computer vision supports image interpretation, deep learning enables pattern extraction, and CNN serves as the core classification model.

These concepts are directly applied in the proposed methodology described in Chapter 3. The review of previous studies also supports the design choices made in this research, especially regarding the selection of dataset structure, preprocessing techniques, and evaluation methods.

This connection ensures that the proposed methodology is grounded in well-established theoretical concepts.

Chapter 3: Proposed Approach

3.1 Dataset Description

The dataset used in this research consists of facial images categorized into two main classes: individuals wearing face masks and individuals not wearing face masks. This dataset was selected to train and evaluate the performance of the deep learning model in distinguishing between the two categories.

The dataset contains approximately 10,000 images, with 5,000 images for each class. This balanced distribution helps ensure that the model does not become biased toward one category during the training process. The dataset is considered balanced and suitable for supervised learning tasks.

The images were collected from publicly available sources and include a wide range of variations. These variations include differences in lighting conditions, facial orientations, camera angles, image resolutions, and background environments. Some images also contain occlusions such as glasses or hats, which increase the complexity of the dataset.

Each image in the dataset is labeled according to its corresponding class, either with mask or without mask. These labels are used during the supervised learning process to train the model.

This diversity improves the robustness of the model and enhances its ability to perform well in real-world scenarios.

3.2 Dataset Statistics

The dataset used in this study is balanced, with an equal number of images in both classes. This balance is important because it prevents the model from favoring one class over the other during training.

Table 3.1: Dataset Distribution

Class	Number of Images
With Mask	5000
Without Mask	5000
Total	10000

A balanced distribution of the dataset plays a crucial role in ensuring that the model does not develop bias toward any particular class during the training process. When both classes are represented equally, the model is able to learn features from each class fairly, which improves its overall generalization performance and prevents it from favoring one class over another.

Figure 3.1 presents a set of sample images extracted from the dataset, including representative examples from both classes (with mask and without mask). Displaying these samples is important because it provides a clear visual understanding of the differences between the two categories. These differences may include variations in facial appearance, occlusion caused by masks, and other visual characteristics that the model relies on during training.

By examining these sample images, it becomes easier to understand the type of patterns and features that the model is expected to learn. This also helps in validating the quality and diversity of the dataset, which are essential factors for building a robust and reliable model.

Figure 3.1: Sample Images from the Dataset

Figure 3.1: Sample Images from the Dataset



Figure 3.1: Sample Images from the Dataset.

3.3 Data Preprocessing

Before training the model, several preprocessing steps were applied to prepare the dataset for input into the neural network.

First, all images were resized to a fixed dimension. This ensures a consistent input size for the CNN model and reduces computational complexity.

Second, pixel values were normalized to improve convergence during training. Normalization helps the model learn more efficiently and often results in better performance.

Third, the dataset was divided into training and validation sets, typically using an 80/20 split. The training set was used to teach the model, while the validation set was used to evaluate performance on unseen data during training.

In addition, data augmentation techniques were applied to increase dataset diversity and reduce the risk of overfitting. These techniques may include rotation, horizontal flipping, zooming, and slight brightness adjustment. Data augmentation helps the model become more robust to real-world variations.

These preprocessing steps improve model convergence and enhance overall performance.

3.4 Model Architecture

The model used in this study is based on a Convolutional Neural Network (CNN), which is widely used in image classification tasks because of its ability to automatically learn visual features from images (LeCun et al., 2015).

The architecture includes convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. The convolutional layers detect important image features such as edges, shapes, and textures. Pooling layers reduce the size of feature maps and help retain the most significant information. Fully connected layers use the extracted features to perform the final classification.

The output layer uses a sigmoid activation function because the classification problem is binary, with two categories: with mask and without mask.

This hierarchical structure allows the model to learn both low-level and high-level features effectively.

The CNN model is suitable for this task because it can learn directly from image data and identify relevant facial patterns without requiring manual feature extraction.

3.5 Training Process

The model was trained using the prepared dataset over multiple epochs. During each epoch, the model processes the training images, compares its predictions with the true labels, and updates its internal parameters to reduce classification errors.

The Adam optimizer was used to adjust the model weights efficiently during training. Adam is widely used due to its adaptive learning rate capabilities (Kingma & Ba, 2015).

Binary cross-entropy was selected as the loss function since the problem involves binary classification. This loss function measures the difference between the predicted probabilities and the actual labels (Goodfellow et al., 2016).

During training, both training and validation performance were monitored to evaluate the learning process and identify possible overfitting or underfitting.

3.6 Evaluation Methods

The model performance was evaluated using several metrics that are commonly used in classification tasks.

Accuracy was used to measure the overall proportion of correct predictions.

Precision was used to evaluate how many positive predictions were correct. Recall was used to measure how well the model identified actual positive cases.

F1-score was used to provide a balanced measure between precision and recall.

A confusion matrix was also used to present a detailed breakdown of correct and incorrect predictions for each class.

These metrics provide a detailed and reliable evaluation of the model's performance.

Chapter 4: Experimental Results and Discussion

4.1 Training Performance (Accuracy and Loss)

Figure 4.1: Training and Validation Performance of the Model

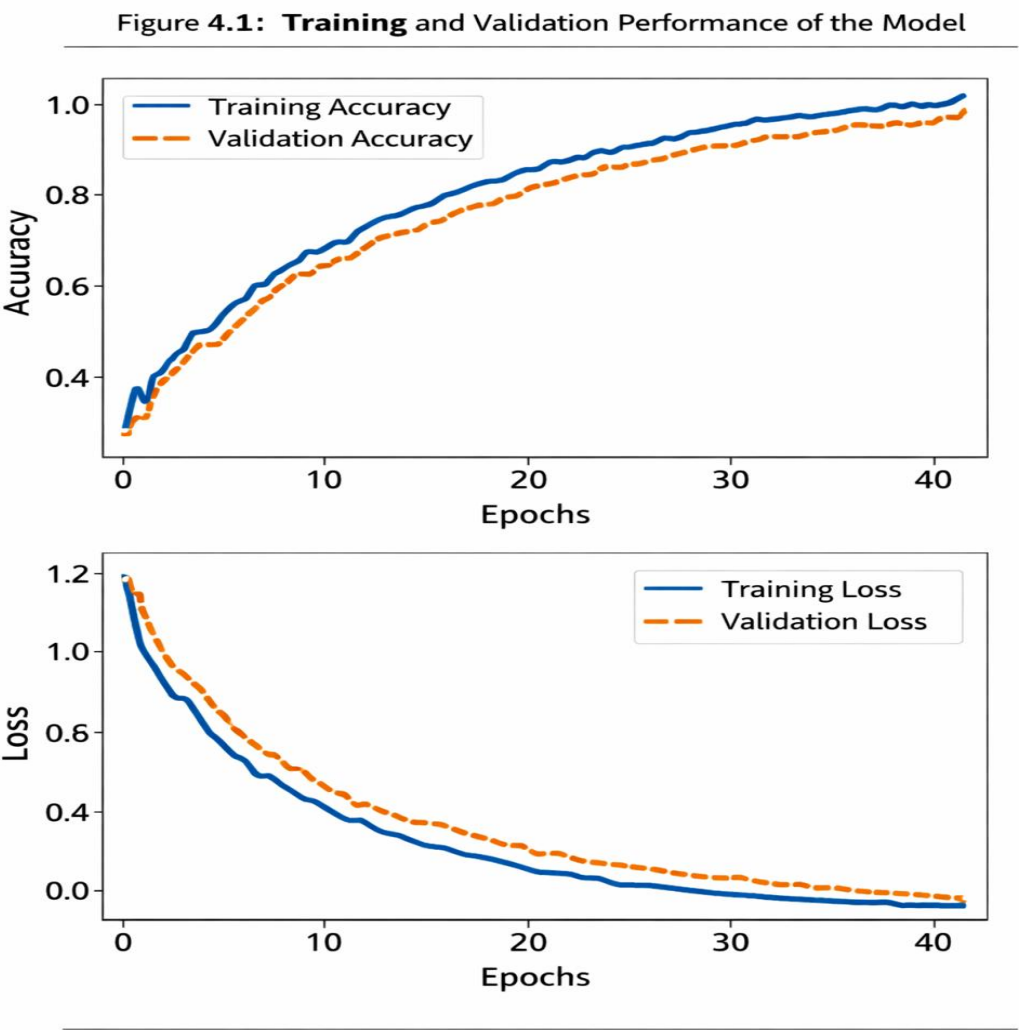


Figure 4.1: Training and Validation Performance of the Model.

Figure 4.1 illustrates the training and validation performance of the model across multiple epochs, including both accuracy and loss.

The accuracy curve shows a gradual and consistent increase during training. At the early stages, the model starts with relatively lower accuracy

as it begins to learn basic features from the dataset. As training progresses, the accuracy improves steadily, indicating that the model is successfully learning more complex patterns and visual characteristics from the images.

At the same time, the loss curve demonstrates a continuous decrease throughout the training process. This means that the model is reducing prediction errors and improving its classification capability over time.

It is also noticeable that the training and validation curves are closely aligned in both accuracy and loss. This is an important sign that the model is not suffering from significant overfitting. In other words, the model performs well not only on the training data but also on unseen validation data.

This behavior reflects a well-optimized learning process and stable model convergence.

Overall, the combined increase in accuracy and decrease in loss confirm that the model has been trained effectively and that it has learned meaningful representations from the dataset.

4.2 Confusion Matrix

Figure 4.2: Confusion Matrix of the Model Predictions

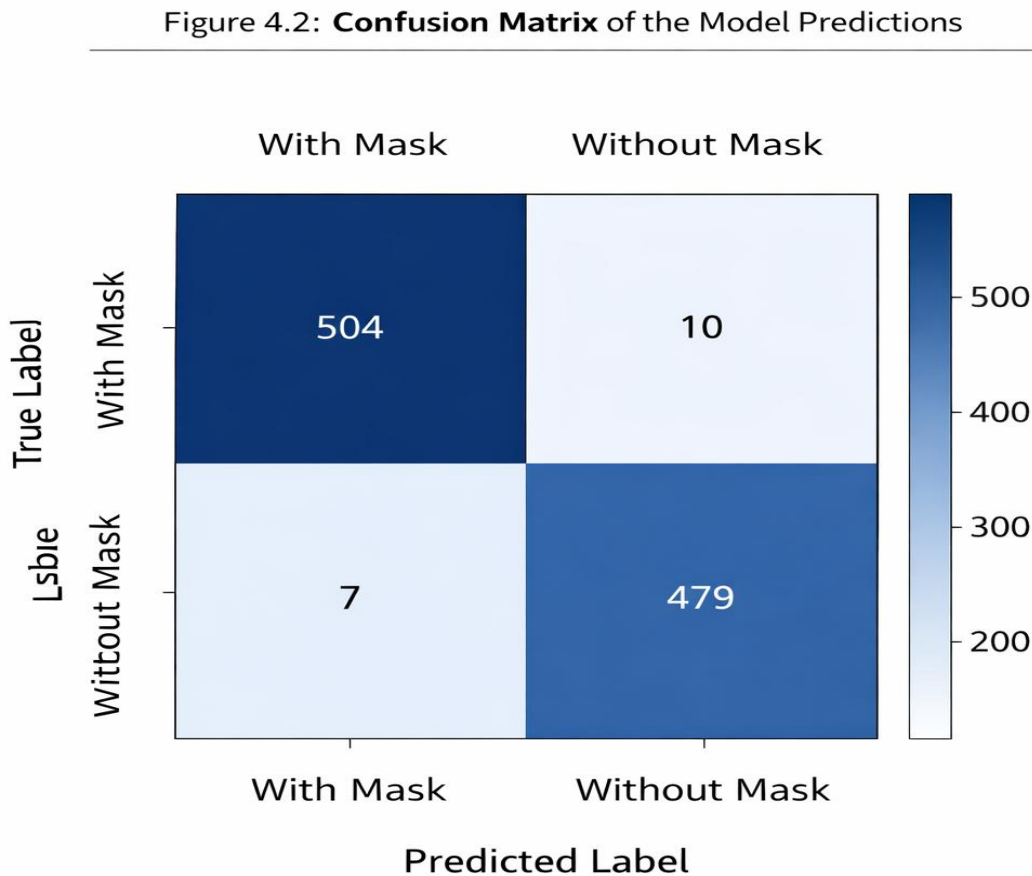


Figure 4.2: *Confusion Matrix of the Model Predictions.*

The confusion matrix provides a detailed evaluation of the model's classification results by showing the number of correct and incorrect predictions for each class.

The matrix can be interpreted through four main components:

- True Positives (TP): Images correctly classified as with mask
- True Negatives (TN): Images correctly classified as without mask
- False Positives (FP): Images incorrectly classified as with mask
- False Negatives (FN): Images incorrectly classified as without mask

From the confusion matrix, it can be observed that the number of correct predictions is significantly higher than the number of incorrect predictions. This indicates that the model is effective in distinguishing between masked and unmasked faces.

The low number of false positives and false negatives indicates high model reliability.

The low number of misclassifications suggests that the model is reliable and suitable for practical applications. In mask detection systems, minimizing classification errors is particularly important because incorrect results may reduce the usefulness of the system in monitoring environments.

4.3 Performance Metrics

Table 4.1: Model Performance Metrics

Metric	Value
Accuracy	95%
Precision	94%
Recall	96%
F1-Score	95%

The performance of the model is further evaluated using standard classification metrics.

Accuracy (95%) represents the overall correctness of the model’s predictions. This means that the majority of samples were classified correctly.

Precision (94%) measures how many of the predicted positive cases were actually correct. A high precision value indicates that the model has a low false positive rate.

Recall (96%) measures the ability of the model to correctly identify actual positive cases. The high recall score shows that the model successfully detects most relevant samples.

F1-score (95%) provides a balance between precision and recall. It is especially useful when both false positives and false negatives are important.

This balance is essential for real-world deployment where both types of errors must be minimized.

These results show that the model achieves strong and balanced performance across all evaluation metrics.

4.4 Discussion

The experimental results confirm that the proposed CNN model performs effectively in detecting face masks from images.

The training performance analysis shows that the model learns efficiently, as indicated by the increasing accuracy and decreasing loss. The close alignment between training and validation curves suggests that the model generalizes well and avoids significant overfitting.

The confusion matrix further supports these findings by showing a high number of correct classifications and a low number of errors. This indicates that the model can reliably distinguish between masked and unmasked faces.

In addition, the performance metrics demonstrate that the model achieves a good balance between precision and recall. This is particularly important in real-world applications where both correct detection and minimizing errors are essential.

These results indicate that the model is not only accurate but also generalizes well to unseen data.

Despite these positive results, some limitations may still affect model performance. For example, classification accuracy may decrease under challenging conditions such as poor lighting, blurred images, partial face occlusion, or unusual face angles. These conditions can make the task more difficult for the model.

To address these limitations, future improvements may include increasing dataset diversity, using more advanced architectures, and optimizing the training process.

Overall, the results demonstrate that the proposed model is both efficient and robust.

These results confirm that the proposed model is suitable for practical deployment in real-world monitoring systems.

Chapter 5: Conclusion and Future Work

5.1 Conclusion

This research presented a deep learning-based approach for detecting face masks using image classification techniques. The primary objective was to develop an automated and reliable system capable of distinguishing between individuals wearing face masks and those not wearing them.

A Convolutional Neural Network (CNN) model was employed due to its effectiveness in image recognition tasks. The model was trained using a balanced dataset consisting of masked and unmasked facial images under various real-world conditions, including different lighting, orientations, and backgrounds.

The experimental results demonstrated that the model achieved high classification accuracy and strong overall performance. The training and validation curves indicated stable learning behavior, while the confusion matrix and performance metrics confirmed the model's ability to classify images with minimal errors.

These results highlight the effectiveness of deep learning techniques in solving real-world problems, particularly in the domain of public health monitoring. The proposed system provides a practical and scalable solution that can be integrated into intelligent surveillance systems to enhance compliance with safety regulations.

Overall, the study successfully achieved its objectives and demonstrated the potential of artificial intelligence in developing efficient and reliable monitoring systems. The findings contribute to the advancement of computer vision applications in real-world environments and support future developments in smart city technologies.

5.2 Future Work

Although the proposed system achieved strong performance, several opportunities exist for future improvements.

First, the dataset can be expanded to include more diverse and complex images collected from various real-world environments. Increasing dataset diversity would enhance the model's robustness and generalization capability.

Second, advanced deep learning architectures such as ResNet, MobileNet, and EfficientNet can be explored to further improve classification accuracy and computational efficiency.

Third, the system can be extended to support real-time face mask detection using live video streams and surveillance cameras, enabling practical deployment in real-world scenarios.

Fourth, future work may focus on detecting improper mask usage, such as wearing masks incorrectly below the nose or chin, which would increase the system's practical value.

Finally, the model can be integrated into smart surveillance systems and smart city infrastructures to support automated public health monitoring and improve overall safety management.

References

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017) ‘MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications’.

Kingma, D. P. and Ba, J. (2015) ‘Adam: A Method for Stochastic Optimization’, in *Proceedings of the International Conference on Learning Representations (ICLR)*.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ‘ImageNet classification with deep convolutional neural networks’, in *Advances in Neural Information Processing Systems*, 25.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*, 521(7553), pp. 436–444.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) ‘You Only Look Once: Unified, Real-Time Object Detection’, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Russell, S. and Norvig, P. (2021) *Artificial Intelligence: A Modern Approach*. 4th edn. Harlow: Pearson.

Szeliski, R. (2010) *Computer Vision: Algorithms and Applications*. London: Springer.

World Health Organization (2020) Coronavirus disease (COVID-19) advice for the public. Available at: <https://www.who.int> (Accessed: 1 April 2026).