Suez Canal Engineering
Energy and Environmental
Science Journal
Faculty of Engineering – Suez Canal University

2025, Volume 3, NO. 2, pages 65– 75

Advancing Smart Infrastructure Monitoring Systems Through Adaptive COVID-19 Responses and 6G Network Integration

Nourhan osama[1]*, Ahmed M. Mohamed[1], Basem E. Elnaghi[1], Khalid Abd EL Salam[1,2]

[1]Electrical Engineering Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt,
[2]Department of Computer Science, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt.

**Corresponding author, Email address: nourhanu4@gmail.com.*
DOI: 10.21608/sceee.2025.352340.1060

Article Info:

Article History:
Received: 12\01\2025
Accepted: 18\03\2025
Published: 30\04\2025

DOI: 10.21608/sceee.2025.352340.1060

Abstract

The following research work is intended to enhance the efficiency of the facemask detection system, which is important in limiting airborne disease transmission, especially in places where the infection rate is most likely. Two approaches are proposed in this paper for enhancing surveillance: the first model is a custom system model using convolution neural network (CNN), which gave high sensitivity and 96.4% accuracy and The second approach is a hybrid model system that uses CNNs for feature extraction along with a pre-trained classifier algorithm Darknet. This hybrid method leverages the strengths of both CNNs and pre-trained algorithms improved accuracy, stability, and reduced loss. These results indicate that the best by reaching an accuracy of 98%, This model is sensitive to delay and thus highly adaptable across different datasets as it is trained on a huge dataset with a variety of images hence, we suggest using it on a 6G network at an estimated data rate of one terabit per second and taking advantage of 6G technologies, These will provide valuable inputs not only for research in the future but also hold immense promise for greatly improving practical applications in image classification tasks.

Keywords: *Airborne Diseases, Convolution Neural Network, Face Mask Detection, 6G network*

Suez Canal Engineering, Energy and Environmental Science Journal (2025), *Vol.3*, NO.2.

1. Introduction

Airborne diseases are a major health problem worldwide, affecting millions yearly. These infectious diseases spread through respiratory droplets or airborne particles, making them particularly challenging to control in populated areas. Airborne pathogens can remain suspended in the air for hours, potentially infecting others who enter these spaces. The impact of airborne diseases on global health systems is huge. Respiratory infections constitute about 20% of total hospital admissions in healthcare facilities at any one time during peak seasons around the world and airborne diseases are among the top 10 causes of death cases according to the World Health Organization (WHO)(World Health Organization: WHO, 2024). The economic burden is equally great, with estimates suggesting that these airborne diseases cost global economies billions of dollars annually through healthcare expenses and lost productivity. The modern approaches incorporate the traditional methods of hand hygiene and mask-wearing with advanced technologies involving high-efficiency particulate air filter (HEPA) filtration systems and ultraviolet light (UV) air purification. Public health experts emphasize that the best way to prevent the disease is multi-layered, involving personal protective measures and environmental controls. Face masks have emerged as a critical tool in preventing the transmission of airborne diseases. They work as physical barriers to partially block respiratory droplets and particles from entering or leaving the respiratory system of the wearer. Many studies have been conducted and confirmed that well-worn masks can reduce infectious agent transmission considerably in healthcare and community settings.

The following are the important benefits associated with the wearing of face masks:

- Source Control: Masks help prevent respiratory droplets from being emitted by infected people, including asymptomatic individuals.
- Personal Protection: These offer a barrier to incoming particles, hence reducing the chances of infection for a wearer.
- Community Protection: Widespread mask-wearing creates a collective barrier that can help reduce community transmission rates.

Artificial intelligence (AI) and convolutional neural networks (CNNs) have brought in a revolution for face mask detection surveillance systems through an efficient solution to mask compliance monitoring within public spaces, which has become very vital in the present context of preventing airborne disease. CNNs are special kinds of deep learning algorithms and work exceptionally well for image processing and recognition, and deep CNNs extend that by adding additional layers of processes to allow higher states of feature abstraction and analysis. This leads to the advantages of the implementation of AI-powered mask detection systems, which include scalability, in which they can monitor big crowds and multiple locations simultaneously; consistency, as they provide 24/7 monitoring without human fatigue or bias; accuracy, as the new systems boast of very high detection rates with a lot less false positives; and cost-effectiveness, as they reduce the need for manual monitoring while improving coverage. These systems have come in quite useful, especially for high-traffic areas like airports, shopping malls, and medical facilities, where the keeping-on of a mask is most required to avoid the contraction of airborne diseases. So, the use of AI and deep learning in face mask identification is a huge step forward for public health surveillance since it combines cutting-edge technology with useful health safety precautions. Given the ongoing difficulty in preventing airborne illnesses, these intelligent monitoring systems will be a crucial instrument in implementing and upholding preventative measures.

The CNN architecture is purposed for visual data processing and analysis, just like that of the human brain, with high performance in visual information perception to identify patterns or features in an image. Therefore, in detecting face masks, CNN checks on images obtained from surveillance cameras to identify the faces and the fact that whether an individual has the face mask or not.

The Key Components of CNNs in Face Mask Detection:

- Convolutional Layers:

The convolutional layers apply filters called kernels on the input images to extract some crucial features, including edges, textures, and patterns. In mask detection, some convolutional layers will identify facial landmarks, contours, and positioning of a mask.

- Pooling Layers:

It reduces the spatial dimensions of feature maps while preserving important information. This step improves computation efficiency and avoids overfitting; hence, it is crucial for real-time mask detection in a crowded environment.

- Fully Connected Layers:

Take as input the features extracted through high-level feature representations given by convolution and pooling, and map those for specific outputs: "mask worn correctly", "mask worn incorrectly", and "no mask."

- Activation Functions:

Non-linear, but easy functions, allow the model to learn complicated patterns within the data so that it would effectively classify mask-wearing.

- SoftMax Layer:

The SoftMax layer in the case of a multi-class face mask detection system will provide the probabilities for every class and indicate the likelihood of a given input image.

In addition to this, custom-designed CNN models; the pre-trained algorithms further enhance detection systems for face masks. Leveraging pre-existing knowledge from large, diverse datasets. These models are particularly useful in scenarios where limited computational resources or insufficient labeled data hinder the development of custom models from scratch. Popular pre-trained algorithms include models like ResNet, Squeeznet, YOLO (You Only Look Once), and Alexnet. Tand will enhance their deployment in broader contexts, from detecting other public health risks to ensuring compliance with safety protocols, solidifying their role as essential tools in global health surveillance systems. In the context of an advanced face mask detection smart system, the integration of 6G features will take this system to a new level of scalability, efficiency, and effectiveness. 6G wireless technology is expected to expand and improve upon 5G's capabilities. Development for the sixth generation of wireless technology is on the go; it's going to be a revolution in the landscape of telecommunication. Artificial intelligence significantly impacts the handling and control mechanisms of complex, dynamic, and heterogeneous 6G networks. In 6G networks, data rates are targeted as high as 1 Tbps, which shall enable new and innovative applications such as real-time 3D holograms, Ultra-HD video streaming, and massive IoT deployment. Beyond data transmission, 6G will also feature sensing and interpretation of the environment, thus opening potentially new applications: health, disaster management, and smart cities (Haque et al., 2024).

2. 6G development and its vision

2.1 Features and Key Technologies for 6G

1. Ultra-high data rates and Low Latency

THz Bandwidth: Enables ultra-high-speed data transmission, The Data rate of 6G is projected to 1 terabit per second and reduced latency to microseconds (Singh et al., 2023).

2. Greater support for machine-to-machine (M2M) connections and Enhanced Connectivity

Support for ubiquitous connectivity: Everything connected, anytime, anywhere by anyone and anything besides the Integration of satellites, UAVs, and terrestrial networks for global coverage, especially in rural and remote areas.

3. Higher network reliability

Higher network reliability6G will be building on and expanding the 5G-powered URLLC service further. Some key methods of ensuring higher reliability correspond to device-to-device communication, multiple wireless hops, simultaneous transmission, and AI/ML, among others. Thereby, as the use for network penetration is concerned, 6G would, therefore, considerably outshine generation 5. Besides, against a backdrop of comparative studies from past generations, in M2M interactions, the error rates get reduced tenfold, with the enhancement of over one hundred times improvements in network reliability.

4. The use of AI and ML for optimal connectivity

5G will enable the full potential of machine learning (ML) and artificial intelligence (AI) technologies. AI/ML will eventually be applied to different network services, network layers, and network components. AI/ML will help achieve greater efficiency at lower computing complexity, from improving beamforming in the radio tier to planning at the cell site with self-optimizing networks, Intelligent Mobility and Handover Management in 6G Networks, Intelligent Spectrum Management, Self-Optimizing Networks (SON), AI routing, etc.

5. Non-RF (VLC and optical)

Visible Light Communication (VLC) acts as a complementary and alternative technology for applications with requirements for minimal electromagnetic interference and high data security especially in hospitals to limit the use of electromagnetic waves which have bad and dangerous side effects on human health. Although the need for a Line of Sight (LoS) path limits VLC's applicability to geometries, it also enhances security because eavesdroppers are equally subject to these limitations. VLC provides a wider spectral range without RF interference when compared to the radio frequency (RF) spectrum. This has benefits, particularly for indoor or industrial settings like workplaces, hospitals, and aircraft. However, in combination with VLC, RISs - consisting of mirror arrays, metasurfaces, or liquid crystals (LCs) - are a new approach, that tackles the remaining challenges like range and limited field of view (FOV) of receivers by using reflectors and refractive elements, the RISs can perform beam steering, splitting, scattering, light polarization, and photon absorption (What Is a 6G Network? Working and Benefits, 2025), but it still has a lot of limitation because of a Line of Sight (LoS) path restricts the usability of VLC to specific geometries but also adds benefits for security as these restrictions apply to eavesdroppers as well. Compared to the radio frequency (RF) spectrum, VLC offers an extended spectrum range without interference with RF. This offers advantages, especially in indoor or industrial applications such as in hospitals, airplanes, or work environments. So, immerging between VLC and RF will be a powerful solution, as the person inside the place where the communication will be based on VLC but if goes outside away from the place that is prepared for VLC it changes into RF communication directly. This is to limit the use of RF. This method of communication is not only low-cost, requiring just a Light light-emitting diode (LED) and a photodiode, but also dual-purpose, with LEDs serving for both illumination and data transmission. VLC's immunity to electromagnetic interference enhances its suitability for environments where traditional RF communication poses risks (Cano et al., 2022).

2.2 A Review of Hybrid VLC/RF Networks

Systems based on visible light communication (VLC) and radio frequency (RF) will be integrated to create hybrid systems that can address and largely meet these needs. A hybrid network architecture minimizes non-coverage zones, boosts system data throughput, and enhances load balancing by permitting complementary cooperation between the two technologies without interfering with one another. For Internet of Things (IoT) applications such as location-based services, smart lighting, home automation, smart healthcare, and industrial IoT, VLC/RF hybrid networks can offer reliable and efficient communication solutions. Considering that Wi-Fi and LTE cover a far wider region and are more resilient to link quality deterioration brought on by movement than the VLC, therefore, RF communication can maintain a stable wireless connection in the case that the VLC is not available and can conveniently provide an uplink data and control channel. As a result, hybrid VLC/RF networks are essential components of communication systems of the future.

In (Li et al., 2016), a load-balancing technique between the VLC and LTE networks is suggested. In (X. Li et al., 2015), the cooperative load balancing problem is resolved and several VLC cell generation and frequency reuse patterns are examined. Wang and Haas (Dynamic Load Balancing with Handover in Hybrid Li-Fi and Wi-Fi Networks, 2014) solved a utility maximization problem to provide a load balancing strategy between Wi-Fi and Li-Fi.

A resource allocation issue for the hybrid VLC and RF femtocell networks with delay restrictions is examined in (Jin et al., 2014). The problem of maximizing energy efficiency in hybrid RF and VLC networks is resolved by (Kashef et al., 2016).

In (Protocol Design and Capacity Analysis in Hybrid Network of Visible Light Communication and OFDMA Systems, 2014), a changeover protocol is suggested for the VLC and orthogonal frequency-division multiple access (OFDMA) hybrid network.

The power allocation issue for hybrid power-line communication (PLC), VLC, and RF communication systems has been examined by Kashef et al. (Alvarez et al., 2023). To formulate and solve a tractable resource allocation problem, these previous publications on the subject oversimplify hybrid RF and VLC communication systems. In (Pratama & Choi, 2018) The paper introduces a hybrid Li-Fi/Wi-Fi network that incorporates a dynamic handover routine to address system load balancing to balance the system load effectively while minimizing the handover overhead associated with user mobility.

6G is a leap in communication technology, ultra-speeds integrated with intelligence in the networks, and connectivity all over the world. It will reduce the digital divide, enable disruptive applications, and reshape industries. However, only addressing technical, economic, and environmental challenges can allow the full potential of this technology to be realized. So, for our smart system face mask detection it will improve it in several ways:

- Real-Time Analysis: Ultra-low latency combined with high data rates will facilitate seamless real-time detection and reporting of mask compliance in crowded areas.
- Edge AI Integration: With 6G-enabled edge computing, detection systems can process data locally, reducing reliance on centralized servers and further improving efficiency.
- Scalability and Global Reach: The integration of satellites for connectivity will further extend the deployment in remote and underserved areas, hence making the systems universally accessible.
- Improved Accuracy: Smarter, context-aware mask detection systems, powered by advanced AI capabilities coupled with high-precision sensing, can make a difference between masks and other facial obstructions. Therefore, 6G will render smart face mask detection a necessity to be taken for public health, while response times would be shorter with more effective management of the pandemic.
- Li-Fi Integration: Li-Fi or Light Fidelity adoption in 6G networks will enable mask detection systems to use VLC for ultra-fast data transmission in those environments where radio signals are constrained, like hospitals, airplanes, and underground spaces. Li-Fi can ensure secure and interference-free communication for such systems.

3. Related work

Many studies that have been published about face mask detection are summarized in this section. It focuses on current research that has used DL for comparable objectives. Regardless of how they align, the model described in this research can recognize faces in images and provide accurate segmentation masks for them.

The study introduces the RILFD (Real Image-based Labeled Face Mask Dataset) (Wang et al., 2023), This study presents the face mask detection system, ResNet Hybrid-Dilation-Convolution Face-Mask-Detector (RHF), which uses hybrid dilation convolutional networks. The authors created the Light Masked Face Dataset (LMFD) and the Masked Face Dataset (MFD) to train and evaluate the model. By improving the perception of the convolutional kernel, the hybrid dilation convolution network resolves issues with image discontinuity. The RHF model achieves better identification results with a mean Average Precision (mAP) of 93.45% while requiring a significantly shorter amount of training time than ResNet50. This study highlights the system's practicality by highlighting its effectiveness in identifying masks in a variety of scenarios.

A Mask R-CNN model (He et al., 2017) has been proposed to detect whether people wear masks or not and whether the mask is used incorrectly in public and crowded environments. The proposed Mask R-CNN model with a ResNet101 backbone achieved a mAP of 83%, and an F1-score of 86%.

The study introduces the Rapid Real-Time Face Mask Detection System (RRFMDS), an automated system made to use CCTV camera video feeds to continuously check for face mask compliance. The system successfully recognizes faces with or without a mask by using a Single-Shot MultiBox Detector (SSD) for face detection and a refined MobileNetV2 model for mask classification trained using 14,535 photos from a custom dataset (Sheikh & Zafar, 2023).

In the context of COVID-19, AI-Based Monitoring of Various Risk Levels Melo et al. (Melo et al., 2021) developed a CNNs model to monitor COVID-19 risk levels by detecting the presence of a face mask and measuring body temperature. By preparing input photos using synthetic data generation techniques and a large dataset, the model trained by YOLOv5 and a ResNet-50-based key point detector for object detection is guaranteed to contain images in all and diverse scenarios. The system detects masks, spectacles, and caruncles using thermal imaging with 96.65% and 78.7% accuracy, respectively. Its average accuracy at identifying masks in RGB images is 82.4%.

Loey et al. (Loey et al., 2020) used YOLOv2 with ResNet-50 to detect medical face masks, they used resnet 50 for extract features and yolo v2 for detection, and they merged Two datasets, Medical Masks Dataset (MMD) and Face Mask Dataset (FMD) for training the model and split it into 90% for validation and 10% for test. The model's performance was optimized using the Adam optimizer, achieving an average precision (AP) of 81% for medical face mask detection.

4. Material and Methods

We used the Mendeley Dataset (Melo et al., 2021b) dataset in this investigation. Melo, César, et al. created this dataset. They generated a dataset since the quantity of data used is another crucial factor in producing reliable and strong models. The development of a technology capable of producing artificial visuals has become essential. This program was developed to enable the application of a wide variety of masks on publicly available datasets, The samples came from pre-existing datasets such as Group Images (Gallagher & Chen, 2009), IMM (Nordstrøm et al., 2004), Wider (Yang et al., 2015), Helen (Le et al., 2012), Celeba (Liu et al., 2014), and Coco (Lin et al., 2014). The Mendeley face mask dataset is an extended and diversified image dataset that is well-organized into classes like "with mask" and "without mask." It will be a very useful dataset, considering the number of images in the dataset that could help in training a robust machine learning model as it contains 37469 images of people with masks and 19957 images of people without masks, so the total of the images was 57426 images after filtering them. This large volume guarantees a representation of real-world scenarios, including variations in pose, lighting conditions, facial orientation, and quality. This diversity enhances the generalization of the model on unseen data and minimizes biases for reliable and accurate predictions in diverse environments. The size of the dataset will go a long way in enhancing the training process and enabling deep learning models to learn intricate features with high generalization, avoiding overfitting. Similarly, it contains images with groups of people and images with individuals so its applicability to real-life challenges like crowd monitoring and identification of individuals. Thus, highly reliable face mask detection systems can be developed with this dataset, which is of importance in public health monitoring, safety enforcement, and observing compliance with mask-wearing regulations.

(a) Samples of Images of a group of people

(b) Samples of Images of Individual person

Figure 1. Samples from Mendeley Dataset (a) samples of images of a group of people (b) samples of images of a person

4.1 Preprocessing

In this work, we started by dividing the dataset into two groups of images: those that contain groups of people and those that contain a single individual. This was a very important initial step in handling the dataset, whereby the whole processing pipeline would be tuned to cope with each of these types effectively. Since face detection is the basic step for most of the face-related technologies, such as face recognition and verification, it constituted the basis for our methodology. To detect faces in images, we used the algorithm by Viola-Jones, first proposed in 2001. The algorithm has gained much popularity since then because of its performance for real-time face detection applications. It operates on grayscale images by using Haar-like features in a cascading classifier to meet the demands of both speed and accuracy. The cascade structure allows the algorithm to scan an image efficiently. This approach ensures reliable performance in detecting faces in images. Once the faces were detected, each face was extracted from the images. This step was important for processing each detected face separately to ascertain if every single one of them had a mask or not. Images extracted from there were then pre-processed to make the dimensions uniform, as the original dataset contained images with low resolution to high resolution. All the detected faces were rescaled to an identical size of 500×500 pixels for unifying the dataset, which would contribute to better performance in model training. Moreover, the images were cropped to appropriate dimensions so that only the regions of interest were considered, excluding irrelevant background portions from the analysis. This pipeline of preprocessing significantly optimized the dataset for subsequent face mask classification tasks. All prior preprocessing procedures are shown in Figure 2.

Figure 2. Preprocessing Steps of the Dataset Images

4.2 The Model Architecture

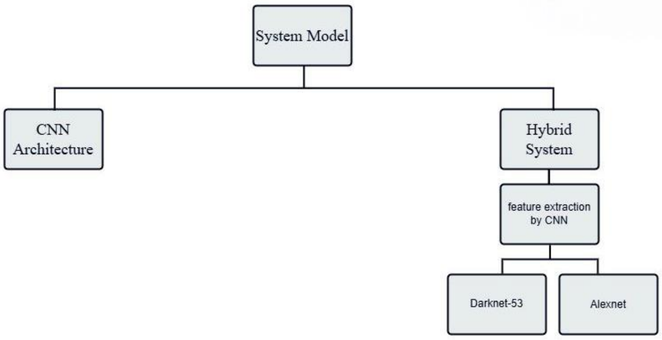

Figure 3. The model architecture

We have tried two proposed model systems to get the most accurate and the most efficient. So, we apply two proposed systems, the first one is by using customized CNN architecture, and the second one is the hybrid model system by using customized CNN architecture to extract features and labels then we use pre-trained algorithms as classifiers learned from the extracted features and network.

4.2.1 First proposed model system

Our first proposed model is a customized deep convolution neural network system model, CNN is an effective method for acquiring and analyzing the data by combining basic characteristics like edges and curves to create a complex features map, Convolutional, nonlinear pooling, and fully connected layers are some of CNN's hidden layers. First, we divided the dataset into 80% for training and 20% for testing the model. The input images are rescaled to 224x224 pixels and three RGB colour channels that the CNN architecture can understand. It consists of five convolutional layers using 3×3 filters to extract the spatial information from an input layer with a pixel size of 128×128×3. Batch Normalization after each convolutional layer regularizes the training while a ReLU active layer adds nonlinearity to our model. Features with reduced spatial dimensionality are prepared by max-pooling layers, ensuring model efficiency and increasing robustness when there is a change in size in the inputs. Next comes the fully connected layer, which from the extracted feature takes classification over two classes or outputs depending upon or without a facemask existence. The other layer is SoftMax, which takes the output of the network and converts the output into probabilities for each class. Finally, the classification layer: Assigns the final label, based on the highest probability.

- Training the CNN

It trains using the SGDM optimizer with a total of 10 epochs and a mini-batch size of 32. Each epoch shuffles the data for better generalization. Every 10 iterations, the trained network will be validated; that way, the progress during the training process of the network can be followed by the plots displayed online.3. Training of CNN.
The training process would run for a total of 20 max epochs, using the SGDM optimizer with a mini-batch size of 128. At each epoch, the shuffling of data is done to enhance generalization. Every 10 iterations, there is a need for validation regarding how the network has performed in this training, hence training progress can be visualized by live plots.

4.2.2 Second proposed model system

After discussing the proposed customized CNN systems, a hybrid system is proposed that will combine the good features of convolutional neural networks and pre-trained algorithms to yield better accuracy and stability for the models. We combined a CNN feature extractor with a pre-trained classifier Darknet.
The CNN model as a feature extractor was designed on a 2D convolution layer with 256 filters of size 3×3. The ReLU method was used for this layer's activation. The rest of the two max pooling layers used down sampling for the architecture of the fully connected layer and SoftMax layer at the output for classification purposes. The Adam optimizer was chosen due to its potential for self-adaptive learning of the learning rates of different parameters while improving convergence rates; therefore, problems like vanishing or exploding gradient can be avoided. That's why this work tried for different maximum epochs: 5, 7, and 10, out of which 7 yielded the best results. For this experiment, the mini-batch size selected was 64 for efficient training. Data augmentation was considered at first but eventually resulted in lower accuracy; therefore, it was not included here. Among the experimented values of the learning rate, 0.01, 0.001, and 0.5 optimal accuracies, along with the stability while training for this model could be achieved only for a learning rate of 0. 001.

The dataset was further divided into 80% for training and 20% for testing. The features obtained with the CNN feature extractor were given, along with training and testing labels, to the pre-trained Darknet. This hybrid mode constituted the best system that combined the strengths of Darknet on feature classification and CNN feature extraction.

5. Results and Discussion

5.1 Implementation of the system

The following hardware and software configurations have been used to test and evaluate the implemented frameworks:

Processor Intel(R) Core (TM) i7-6820HQ CPU @ 2.70GHz, 2701 MHz, 4 Core(s), 8 Logical Processor(s).

- Operating system: Windows 10 Pro.
- Installed RAM: 16.0 GB (15.9 GB usable).
- System type: 64-bit operating system, x64-based processor.
- Compiler: MATLAB R2020b.

5.2 Evaluation metrics

The study uses several measures to evaluate how well a classification model performs when it comes to machine learning-based automatic picture classification. If there is an imbalance in the distribution of class labels, the conventional accuracy metric might not be adequate. When working with severely imbalanced classes, precision-recall metrics are very helpful. Precision gauges the model's capacity to locate pertinent data points, while recall gauges its capacity to locate all pertinent cases. Additionally, in the situation of imbalanced classes, the confusion matrix is an appropriate way to summarize a classification algorithm's performance (Publications, 2024).

The subsequent equations are commonly employed metrics in the field of machine learning. In these metrics:

True Positive (TP): a result in which the model accurately forecasts positive values.

False Positive (FP): When the model forecasts positive results inaccurately.

False Negative (FN): When the model forecasts negative numbers inaccurately.

True Negative (TN): this is the result of the model accurately predicting negative values.

Accuracy: this metric measures how the classifier predicts correctly all the classes to the total number of instances.

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

Recall (sensitivity): it is also called true positive Rate It calculates the percentage of real positives in a classification problem that are genuine positives. This measure shows how consistently the model labels positive units in the dataset.

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \quad (2)$$

Precision: determines the percentage of true positives in a classification issue relative to all positive predictions. This score shows how reliable the model is at classifying a person as positive.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \quad (3)$$

Specificity (or true negative rate, TNR): This metric measures the ratio of correctly predicted negative cases to all actual negative cases.

$$\boldsymbol{Specificity = \frac{TN}{TN + FP}} \quad (4)$$

F1-Score: combines the precision and recall scores of a model. F1-score is especially useful when dealing with imbalanced datasets where one class has significantly more samples than the other.

$$\boldsymbol{F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}} \quad (5)$$

Confusion Matrix: The confusion matrix, shown in Figure 4, is the most thorough performance metric for a classification model and offers a detailed examination of the model's behavior. The confusion matrix for a binary classifier shows a matrix that aids in assessing the overall performance of the model. The model's predictions are represented by the rows of the matrix, while the actual labels of the data samples are represented by the columns. Understanding the model's

advantages and disadvantages through the analysis of the confusion matrix will help us fine-tune future training to make the model better.
(Draelos, 2019).

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 4. Confusion Matrix (Draelos, 2019).

5.3 Training and Results

The results and settings we used in the two suggested procedures are shown and discussed in this section. In the first strategy, we used a customized CNN model and in the second, we tested a hybrid system that combined a classifier for final classification with CNN for feature extraction. To get greater accuracy, stability, and reduced loss, the hybrid system made use of the benefits of both CNNs and pre-trained algorithms. Following the extraction of features and labels from the trained network, we used these features to train classifiers (Darknet) and assess how well they performed on a test set. The hybrid system showed how pre-trained algorithms can be efficiently used to achieve improved accuracy and stability when a CNN is used for feature extraction and a classifier is used for final classification. The model's performance was evaluated using confusion matrices to calculate metrics like accuracy, sensitivity (recall), and specificity after extensive experimentation revealed the most successful parameters. All things considered, this method speeds up model development, increases accuracy, and boosts flexibility across various domains and datasets, which makes it appropriate for tasks demanding accurate and trustworthy categorization. By providing insightful information and opportunities for further research, the approaches suggested in this study can significantly aid in real-world applications in face mask recognition and other picture classification problems.

5.3.1 First Proposed System

We proposed a CNN model to classify the images into two classes, first, we uploaded the dataset and split it into 80% for training and 20% for testing which is equal to 11485, then the images resized to 224 x 224 pixels with 3 RGB Channels, the model system composed of Six convolutional layers with filters of sizes 32, 64, 128, 256, 512, and 1024, each with a 3x3 kernel size and 'same' padding and each convolution layer followed by batch normalization layer and ReLU activation function, max pooling layer to reduce the spatial dimensions of the input volume and A fully connected layer with 1024 neurons followed by a ReLU activation Then the output layer was a fully connected layer with 2 neurons, followed by a SoftMax layer to convert the outputs to probabilities. The network was trained using Stochastic Gradient Descent with Momentum (SGDM), we got an accuracy of 96.4% and a sensitivity of 92.68%.

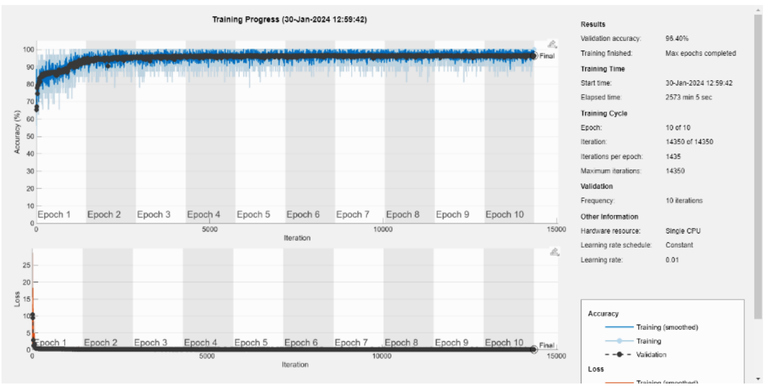

Figure 5. Accuracy and Loss During the training process by the proposed CNN

Table 1. Performance of the first Proposed Model Systems

System model	Accuracy	Precision	Sensitivity	Specificity	F1-score
Customized CNN Model System	96.4%	96.83%	92.68%	98%	93.9%

5.3.2 The Second Proposed System

In this paper, we will propose a new hybrid system that takes full advantage of the strengths of both models by using a CNN as a feature extractor, combining pre-trained algorithms as classifiers (Darknet) to improve the accuracy and stability of the model. The CNN Model is designed with a convolutional layer with a 3x3 kernel size and 256 filters, ReLU activation, followed by two max-pooling layers, and a fully connected layer, also SoftMax function at the output layer to convert it to probabilities. This setup optimizes the performance through adaptive tuning of learning rates for each parameter, thus giving in faster convergence and thereby overcoming problems of vanishing or exploding gradients effectively. Although we tried different data augmentation techniques, the accuracy dropped with those. The trained network along with the respective test and train labels was therefore extracted and used as input for the pre-trained algorithm Darknet.

The integration with Darknet is done by feeding the extracted features from the CNN into the architecture of Darknet as input. Darknet is a high-performance neural network framework that, after processing these features with its robust layers and weights, classifies the data accordingly. For compatibility, the features are reformatted in a form to match the input dimensions and specifications required by Darknet. The classifier we use is the pre-trained weights of Darknet fine-tuned on larger datasets for improved stability and higher accuracy. The hybrid system proposed will combine the feature extraction capability of CNN with the efficiency of Darknet classification. This not only enhances classification but also saves computational overhead due to the reusability of pre-trained weights, hence making the system efficient for practical applications.

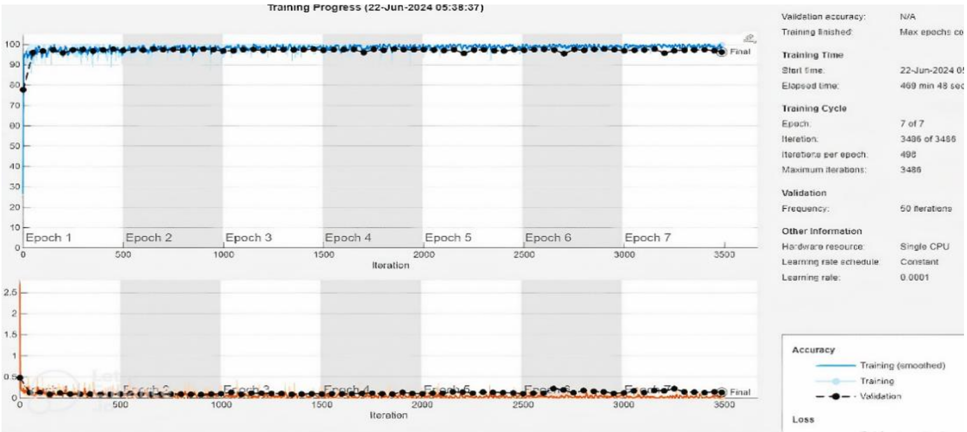

Figure 6. Accuracy and loss during the training process by CNN-Darknet

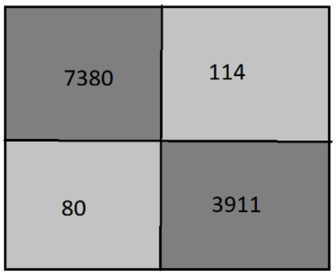

Figure 7. Confusion matrix for CNN-Darknet

Table 2. Performance of the second Proposed Model Systems

System model	Accuracy	Precision	Sensitivity	Specificity	F1-score
CNN-Darknet53	98.00%	98.42%	98.33%	97.07%	98.38%

6. Conclusions

In this study, we must improve the facemask detection system's efficacy because it is crucial in the age of airborne illness transmission and the rise in disease-related fatalities. We investigated the efficiency of facemask detection using two suggested techniques to facilitate surveillance, particularly in congested areas like hospitals. The first technique used

CNNs to customize a system model, which has a high sensitivity and 96.4% accuracy. By integrating a CNN for feature extraction with a pre-trained algorithm as the classifier for final predictions, we created a second hybrid system model. The advantages of CNNs and pre-trained algorithms are used in this hybrid technique to improve accuracy and stability and reduce loss. We use classifiers like Darknet after removing features and labels from the trained network. Finding the highest accuracy and stability of all, the results demonstrated that employing a CNN for feature extraction followed by a pre-trained classifier significantly improves performance. Overall, because it expedites model creation, enhances accuracy, and boosts adaptability across multiple domains and datasets, this approach is suitable for applications requiring reliable and accurate categorization. In addition to offering useful information and directions for future research, the methods proposed in this study have the potential to significantly improve practical applications in image classification tasks, such as face mask identification. And since it's extremely useful in real-time applications due to its sensitivity to delays, we advise utilizing this model system over a 6G network One terabit per second is the estimated data rate, in addition to enabling data transmission by light (VLC), which is safer for people's health, particularly in hospitals.

References

World Health Organization: WHO. (2024, August 7). The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Haque, M. A., Ahmad, S., Abboud, A. J., Hossain, M. A., Kumar, K., Haque, S., Sonal, D., Rahman, M., & Marisennayya, S. (2024). 6G wireless communication networks. International Journal of Business Data Communications and Networking, 19(1), 1–27. https://doi.org/10.4018/ijbdcn.339889.

Singh, P. R., Singh, V. K., Yadav, R., & Chaurasia, S. N. (2023). 6G networks for artificial intelligence-enabled smart cities applications: A scoping review. Telematics and Informatics Reports, 9, 100044. https://doi.org/10.1016/j.teler.2023.100044.

What is a 6G network? Working and benefits. (2025, March 10). Spiceworks Inc. https://www.spiceworks.com/tech/networking/articles/what-is-6g/.

Cano, H. P. Z., Zaharis, Z. D., Yioultsis, T. V., Kantartzis, N. V., & Lazaridis, P. I. (2022). Pattern Reconfigurable Antennas at Millimeter-Wave Frequencies: A Comprehensive survey. IEEE Access, 10, 83029–83042. https://doi.org/10.1109/access.2022.3196456.

Li, L., Zhang, Y., Fan, B., & Tian, H. (2016). Mobility-Aware load balancing scheme in hybrid VLC-LTE networks. IEEE Communications Letters, 20(11), 2276–2279. https://doi.org/10.1109/lcomm.2016.2598559.

Li, X., Zhang, R., & Hanzo, L. (2015). Cooperative load balancing in hybrid visible light communications and WiFi. IEEE Transactions on Communications, 63(4), 1319–1329. https://doi.org/10.1109/tcomm.2015.2409172.

Dynamic load balancing with handover in hybrid Li-Fi and Wi-Fi networks. (2014, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/7136231.

Jin, F., Zhang, R., & Hanzo, L. (2014). Resource allocation under Delay-Guarantee constraints for heterogeneous Visible-Light and RF Femtocell. IEEE Transactions on Wireless Communications, 14(2), 1020–1034. https://doi.org/10.1109/twc.2014.2363451.

Kashef, M., Ismail, M., Abdallah, M., Qaraqe, K. A., & Serpedin, E. (2016). Energy-efficient resource allocation for mixed RF/VLC heterogeneous wireless networks. IEEE Journal on Selected Areas in Communications, 34(4), 883–893. https://doi.org/10.1109/jsac.2016.2544618.

Protocol design and capacity analysis in hybrid network of visible light communication and OFDMA systems. (2014, May 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/6637084.

Alvarez, L. B., Montejo-Sánchez, S., Rodríguez-López, L., Azurdia-Meza, C., & Saavedra, G. (2023). A review of hybrid VLC/RF networks: features, applications, and future directions. Sensors, 23(17), 7545. https://doi.org/10.3390/s23177545.

Pratama, Y. S. M., & Choi, K. W. (2018). Bandwidth Aggregation Protocol and Throughput-Optimal Scheduler for hybrid RF and visible light communication systems. IEEE Access, 6, 32173–32187. https://doi.org/10.1109/access.2018.2844874.

Wang, S., Wang, X., & Guo, X. (2023). Advanced face mask detection model using hybrid dilation convolution-based method. Journal of Software Engineering and Applications, 16(01), 1–19. https://doi.org/10.4236/jsea.2023.16100.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1703.06870.

Sheikh, B. U. H., & Zafar, A. (2023). RRFMDS: Rapid Real-Time Face Mask Detection System for effective COVID-19 monitoring. SN Computer Science, 4(3). https://doi.org/10.1007/s42979-023-01738-9.

Melo, C., Dixe, S., Fonseca, J. C., Moreira, A. H. J., & Borges, J. (2021). AI-based monitoring of different risk levels in the COVID-19 context. Sensors, 22(1), 298. https://doi.org/10.3390/s22010298.

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 65, 102600. https://doi.org/10.1016/j.scs.2020.102600.

Melo, C., Dixe, S., Fonseca, J. C., Moreira, A. H. J., & Borges, J. (2021b). AI-based monitoring of different risk levels in the COVID-19 context. Sensors, 22(1), 298. https://doi.org/10.3390/s22010298.

Gallagher, A. C., & Chen, T. (2009). Understanding images of groups of people. 2009 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2009.5206828.

Nordstrøm, M. M., Larsen, M., Sierakowski, J., & Stegmann, M. B. (2004). The IMM Face database is an annotated dataset of 240 face images. https://www.semanticscholar.org/paper/The-IMM-Face-Database%2C-An-Annotated-Dataset-of-240-Nordstr%C3%B8m-Larsen/67539b4c16f53f3ff4abefd93615c9d3ce32c82e.

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2015). WIDER FACE: a face detection benchmark. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1511.06523.

Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive Facial Feature Localization. In Lecture notes in computer science (pp. 679–692). https://doi.org/10.1007/978-3-642-33712-3_49.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2014). Deep learning face attributes in the wild. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1411.7766.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In Lecture notes in computer science (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48.

Publications, P. B. (2024). Journal of Machine Learning Technologies. Journal of Machine Learning Technologies. https://doi.org/10.9735/2229-3981.

Draelos, V. a. P. B. R., MD PhD. (2019, May 18). Measuring Performance: the Confusion Matrix. Glass Box. https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/.