An Identity-Aware Privacy-Preserving Deep Learning Framework for Culturally Sensitive Image Sharing

Mahmoud Obaid*, Hadeel Bkhaitan, Duha Maali, Saja Hammad, Thaer Thaher
Department of Computer Systems Engineering, Arab American University, Jenin, P.O. Box 240, Palestine

***Abstract*—The blistering development of digital image sharing raises privacy concerns, especially in cultural contexts where image exposure could be ethically and socially provocative. In Islamic societies, sharing images of women without hijab can be deeply sensitive. This study presents SITR, an identity-sensitive privacy-preserving deep learning system aimed at reducing unintended sharing of sensitive images involving female family members without hijab. SITR integrates three components in a unified deployment-ready pipeline: 1) face recognition with Multi-task Cascaded Convolutional Networks (MTCNN), 2) family-member authentication with FaceNet embeddings stored in a vector database, and 3) hijab detection with an optimized Densely Connected Convolutional Network (DenseNet). The hijab detection model was trained and evaluated on a cleaned dataset of 2,191 images with hijab and non-hijab cases with diverse visual conditions. DenseNet121 was benchmarked against ResNet50, MobileNetV2, and EfficientNet-B0, achieving the best overall performance. To further enhance its effectiveness, DenseNet121 was modified by integrating an Efficient Channel Attention (ECA) mechanism and applying hyperparameter tuning. The optimized selected model achieved 92.16% test accuracy, strong discrimination with precision of 91.63% , and 86.39% F1-score on a held-out test set. The model was deployed as a quantized RESTful API, reduced from 82 MB to 27 MB while maintaining predictive reliability. Results demonstrate the practicability of identity-conditioned, culturally-aware AI systems for privacy protection. This work highlights the role of context-sensitive computer vision beyond generic content moderation toward culturally-aware and ethically accountable applications.**

***Keywords*—Cultural sensitivity; deep learning; DenseNet121; Efficient Channel Attention; FaceNet; hijab detection; MTCNN**

I. Introduction

In this age of digital interactions, transfer of photos is quick and immediate. Individuals post pictures to social networks, transfer them through communication services and archive them on cloud-based systems [1]. However, this feature raises the potential of unintended privacy exposure, especially in settings where personal images carry ethical, social, or cultural sensitivity. When a picture is posted on the internet, the creator of the image might never know who might see, save, or repost it [2]. Recent studies on image privacy protection further confirm that digital images often contain sensitive personal information and that protecting such content remains an active research challenge [3]. This concern is special in the Islamic world, where people appreciate modesty and personal privacy [4], [5].

The hijab is a head covering worn by women, and it is a sign of modesty and devotion to religion. For many women, images taken without a hijab are considered private and may lead to discomfort, social harm, or ethical concerns if shared unintentionally [6]. This risk is not limited to social media platforms. It also extends to messaging applications, email, and cloud storage services, where private images may be accidentally opened or even accidentally sent [7]. Such images are likely to be affected by human error when such protection is done manually, it is likely to go unnoticed when sharing such images and can be time-consuming as it requires checking and modifying.

In recent years, Artificial Intelligence (AI) techniques have been widely used in image analysis, face recognition, and content moderation [8], [9]. Photo management systems, security applications, and social media monitoring tools incorporate AI-powered technologies that can detect faces, organize photo collections, and remove inappropriate content. However, most of these systems treat such tasks independently and are not designed for privacy enforcement in culturally sensitive scenarios. Consequently, current tools are generally insufficient for situations in which sharing decisions depend not only on what appears in the image, but also on who appears in it, particularly to guard hijab-wearing women against unintentional exposure in the online space. Therefore, privacy-aware face-related systems remain an open research area.

Despite the fact that the independent research on face detection, face recognition or hijab detection has been conducted in several studies, there has been an imperative gap in identity-aware and culturally sensitive AI systems, which combine the contextual identity verification with the visual attribute classification into a unified, deployable framework. Current solutions are either general content moderation systems, hijab-related image classification approaches, or face-recognition systems without privacy-enforcement logic. This limits their use in culturally sensitive contexts where the perception of privacy cannot be solely based on the visual information in the image, but also on the identity of the people in the picture. In addition, computer vision under partial facial covering or appearance changes remains challenging, as shown in recent surveys on masked and occluded face analysis [10].

To address this gap, this study proposes SITR, an AI-powered system application that integrates deep learning models to help prevent the unintended sharing of privacy-sensitive photos. The system combines three deep learning modules: 1) face detection using MTCNN, 2) family member verification using FaceNet, and 3) hijab detection using a fine-tuned DenseNet121 model. When a female face is detected and identified as a registered family member, the system checks for the presence of a hijab. If no hijab is found, the face and neck are automatically blurred before sharing the photo. This design reflects the actual system workflow, where face

*Corresponding author

detection, identity verification, and hijab classification operate as a single privacy-aware pipeline before image sharing.

This study presents the design, development, and evaluation of the SITR system. The main contributions of this work are:

- A novel identity-conditioned privacy-preserving framework that integrates face detection, family-member verification, and hijab classification into a unified mobile pipeline.
- A systematically optimized hijab-detection component using DenseNet121 architecture enhanced with Efficient Channel Attention (ECA) and validated through controlled ablation.
- A deployment-aware implementation including model compression and RESTful API integration for real-world mobile usage.
- An empirical evaluation demonstrating the trade-off between predictive performance and computational efficiency in culturally sensitive AI applications.

The remainder of this study is organized as follows: Section II reviews related work. The system architecture and methodology are described in Section III. Section IV discusses the experimental results and system limitations. Finally, Section V concludes the study and outlines future work.

II. Related Work and Technical Background

This section explores the technical background, systems, and research studies relevant to SITR. It also presents the main deep learning models and techniques that support the SITR pipeline. No existing systems were identified that fully match the goals and architecture of SITR. However, several applications and studies partially address specific components relevant to the proposed framework, offering valuable insights that helped shape the system design.

A. CNN-Based Deep Learning and DenseNet

A Convolutional Neural Network (CNN) is a deep learning model that is widely used for working with visual data especially images. It uses filters to scan the input image and learn patterns such as edges, textures, and shapes. Because of this ability, CNNs have been successfully used in image classification, object detection, and face recognition. Based on CNNs, many deep learning models have been developed to improve accuracy and overall performance. These models keep the main idea of CNNs for learning features, but they also add new architectural improvements. DenseNet, MobileNet, EfficientNet, and ResNet are among the most common CNN-based models that are widely used in image classification and differ in their depth and complexity [11].

Dense Convolutional Network (DenseNet) is a deep learning architecture for CNNs introduced by Huang et al. [12]. It revolutionized the field of computer vision by proposing a novel connectivity pattern within CNNs, addressing challenges such as feature reuse, vanishing gradients, and parameter efficiency. Unlike traditional CNNs, each layer passes its output to all the following layers. This helps the network reuse features more effectively and improves the flow of information. As a result, important information is less likely to be lost as it moves through the network. DenseNet-121 is a version of the DenseNet architecture that has 121 layers, including convolutional, pooling, and fully connected layers. Because of this design, DenseNet-121 can achieve strong performance while using fewer parameters. Its efficient structure makes it a good choice for tasks that need deep feature extraction with lower computational cost [13].

B. Face Detection Using MTCNN

Face detection serves as a basic function of computer vision because it enables the detection and localization of human faces in digital images. Face-based applications usually start with this step, which supports later tasks such as facial recognition, emotion analysis, and identity verification. The Multi-task Cascaded Convolutional Network (MTCNN) developed by Zhang et al. functions as a deep learning system which detects multiple faces while finding their facial features. The system uses three cascading stages, including the Proposal Network (P-Net) and Refinement Network (R-Net) and Output Network (O-Net) [14]. MTCNN shows strong results on the WIDER FACE benchmark dataset because its cascaded design enables it to attain high accuracy and recall rates. MTCNN is used as the core method for detecting faces in user-uploaded images before further processing is applied.

C. Face Verification Using FaceNet Embeddings

Schroff et al. introduced FaceNet as a deep CNN model for face identification and verification, with the added capability of face clustering [15]. Rather than depending only on conventional classification, FaceNet projects face images into a compact Euclidean space where each face is represented by a discriminative embedding. These embeddings are learned using a triplet loss function, which encourages images of the same person to be placed closer together (anchor and positive samples) while forcing images of different people to remain farther apart (anchor and negative samples). Through training on large-scale datasets, FaceNet learns rich and robust facial features that allow it to perform effectively under variations in lighting, facial expression, and head pose. Another key strength of FaceNet is its ability to generate compact face embeddings, which makes storage and similarity comparison more efficient. This property makes the model well-suited for large-scale face recognition applications, particularly in mobile and cloud-based environments [16].

Within SITR, FaceNet is used to verify whether a detected face belongs to a known family member. By comparing the embedding of a new face to those stored in the database, the system can determine identity and proceed to hijab classification only for matched females.

D. Existing Related Studies and Systems

In this section, existing systems and related studies relevant to SITR are reviewed. No systems were identified that fully correspond to the vision and functionality of the proposed application; however, several platforms and research works address some of the targeted aspects and offer valuable insights that helped shape the proposed framework.

Ku and Dong [17] proposed a face recognition method based on the integration of MTCNN and a modified VGGNet architecture. Their model replaces max pooling layers with mean pooling and combines SoftMax Loss and Center Loss to enhance classification performance. The system achieves 98.53% accuracy on the Labeled Faces in the Wild (LFW) dataset. While this work shows strong performance in face recognition, it focuses on facial identification and does not include hijab detection or privacy-aware content protection.

In [5], a fully convolutional network (FCN) is used for automatic pixel-wise segmentation of faces, hijabs, and backgrounds. The model, trained on a set of 250 images, achieved 92.69% mean accuracy. This method simplifies hijab segmentation compared to manual approaches. However, the study focuses on segmentation and does not address identity verification, automated blurring, or mobile deployment. In another study, Khaliluzzaman et al. [18] introduced a visual feature-based hijab detection framework. Their method uses the Viola-Jones face detector and the YCbCr color model to identify hair and neck visibility. The approach achieved 96.47% accuracy on hijab images and 95% on non-hijab cases. It is important to note that this rule-based method operates under controlled visual conditions with explicitly visible neck and hair features, which may explain its higher reported accuracy compared to the deep learning baseline in this work. Deep learning approaches, however, offer greater generalizability and adaptability to unseen visual conditions, as demonstrated by the optimized model achieving 92.16% accuracy on a more heterogeneous dataset.

Other tools, such as HaramBlur, a browser extension, allows users to navigate the web with reduced browsing distractions by detecting and blurring "haram" content, including faces without hijab, using face detection and NSFWfiltering. However, it does not perform hijab classification or user verification. PimEyes [19], on the other hand, is a facial recognition search engine that allows users to locate images of a person online by uploading one photo. However, it lacks hijab detection and privacy-enforcement functionality. Table I presents a comparison of different face detection, authentication, and hijab detection systems alongside our proposed system, SITR, based on several key aspects.

E. Research Gap and Positioning of SITR

The reviewed systems and studies show meaningful progress in face detection, face recognition, and hijab-related image analysis. However, as shown in Table I, none of the reviewed systems provides a complete solution that combines face detection, face verification, and hijab detection, along with automated content blurring and mobile-friendly deployment. Most existing tools either focus on facial identification without considering cultural sensitivity or address hijab detection without verifying user identity.

This highlights the need for a privacy-focused tool designed for the specific needs of hijab-wearing users. SITR is positioned as an integrated mobile application that brings these components together in one pipeline. It is also worth noting that the proposed framework is specifically designed for the Islamic cultural context; its applicability to other cultural or religious contexts involving head coverings, such as medical caps, traditional headscarves in non-Islamic traditions, or sports equipment, is not evaluated in this work and represents a recognized scope limitation that should be addressed in future research.

III. System Overview and Methodology

This section describes the system design, architecture, and methodology of the proposed SITR application. The system integrates mobile and cloud components with deep learning models to enable privacy-aware image sharing. It also presents the operational workflow, hijab detection model, dataset preparation, training setup, and deployment process. Fig. 1 depicts an overview of the main processing pipeline used in the system. Overview of the SITR pipeline showing four sequential stages: 1) input image upload to the mobile application, 2) face detection localization using MTCNN, 3) identity verification via FaceNet embeddings and cosine similarity comparison against the ChromaDB database, and 4) hijab detection using the optimized DenseNet121 model. If a registered female family member is detected without a hijab, automatic face and neck blurring is applied before sharing. If a hijab is detected or the face does not match a registered member, it is safe to share.

A. System Architecture

SITR follows a modular architecture consisting of a mobile front-end, backend APIs, databases, and deep learning models. The architecture and data flow are illustrated in Fig. 2. As depicted in Fig. 2, the frontend layer communicates with the backend via HTTP REST calls. The backend orchestrates three sequential AI operations, including face detection (MTCNN), identity verification (FaceNet + ChromaDB), and hijab classification (DenseNet121), before returning a privacy enforcement decision to the mobile client.

The main system components are:

1) Frontend layer (mobile app): React Native and Expo are the used frameworks, and this layer will allow the interface to upload photos, manage family members, and see the results of detection. Support on user authentication and session management are provided to achieve secure access.

2) Backend/API layer: This layer is implemented using the Flask framework, which processes the image, along with the AI flow, by consuming RESTful APIs for face detection, face verification, and hijab classification.

3) Database layer: Convex stores structured user and family member information, while ChromaDB stores high-dimensional facial embeddings and enables real-time similarity queries using cosine similarity.

4) AI Engine layer: Includes three core models integrated in a sequential pipeline:

a) MTCNN: Detects faces and localizes facial landmarks with high accuracy under varied lighting and poses.

b) FaceNet: Generates 512-dimensional embeddings to verify detected faces against registered family members stored in ChromaDB.

c) DenseNet121-based hijab detection model: A fine-tuned DenseNet121 model used for hijab detection. ECA was later investigated as an enhancement during optimization.

TABLE I. COMPARISON OF RELATED SYSTEMS WITH SITR

System	Face Detection	Face Verification	Hijab Detection	Accuracy	Blurring Function	Remarks
MTCNN + Modified VGGNet [17]	Yes	Yes	No	98.53% (LFW)	No	Focused on face recognition using CNN with Center Loss.
FCN Segmentation Model [5]	No	No	Yes	92.69% (mean accuracy)	No	Pixel-wise segmentation into skin, hijab, and background.
Visual Feature Detection [18]	Yes	No	Yes	Hijab: 96.47%, Non-hijab: 95%	No	Rule-based detection using Viola-Jones and YCbCr color model.
HaramBlur Extension [20]	Yes	No	No	Not reported	Yes	Browser extension for NSFW content detection and blurring.
PimEyes Search Tool [19]	Yes	Yes	No	Not reported	No	Online facial search engine without cultural sensitivity.
SITR (Proposed)	Yes	Yes	Yes	92.16% (test set)	Yes	Integrates detection, verification, hijab classification, and blurring in a mobile application.

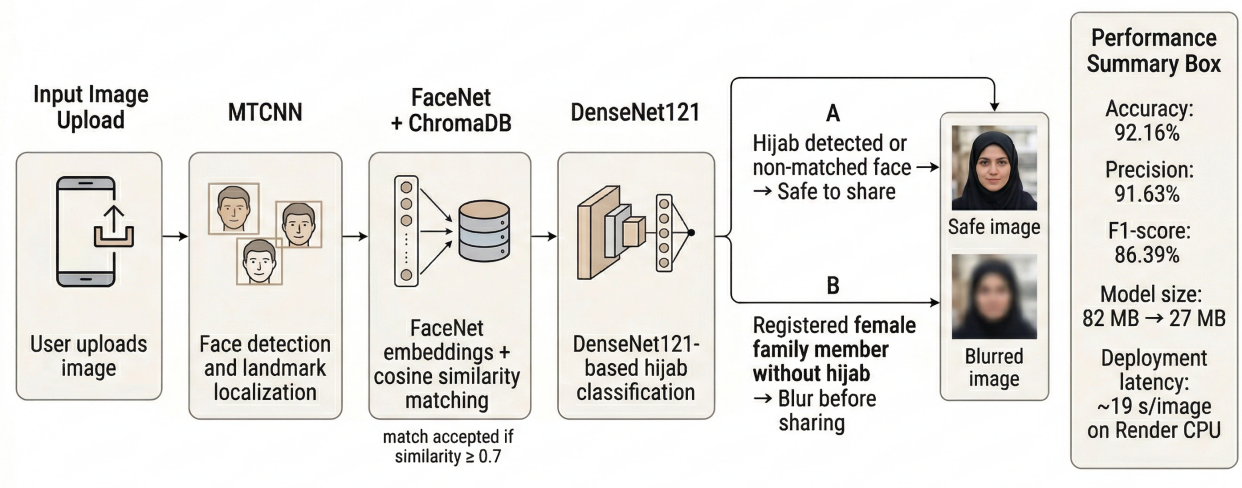


Fig. 1. Overview of the SITR pipeline: identity verification, hijab classification, privacy-enforcement decisions, and final deployment outcomes.

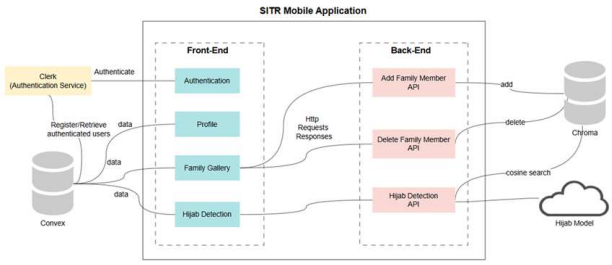


Fig. 2. System architecture with component interaction and data flow.

B. Operational Workflow

The operational workflow of SITR consists of two main stages, as illustrated in Fig. 3.

1) Enrollment stage: The user loads family member images into the application; the images are resized and passed through MTCNN for face detection, then through FaceNet to generate 512-dimensional facial embeddings, which are stored in ChromaDB for future matching.

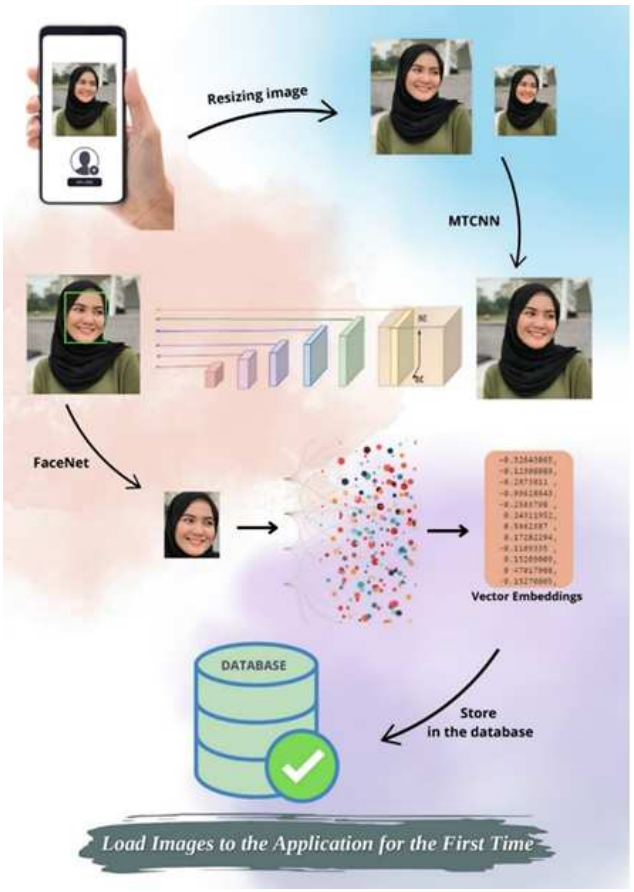
2) Detection stage: An uploaded image is resized, and MTCNN detects all faces. For each detected face, FaceNet generates an embedding that is compared against stored embeddings using cosine similarity (threshold $>= 0.7$). If a female family member is matched, the DenseNet121 hijab detection model classifies the region. If no hijab is detected, the face and neck region are automatically blurred before the image can be shared. Users are notified of the outcome and may choose to share or cancel.

C. Cosine Similarity for Face Matching

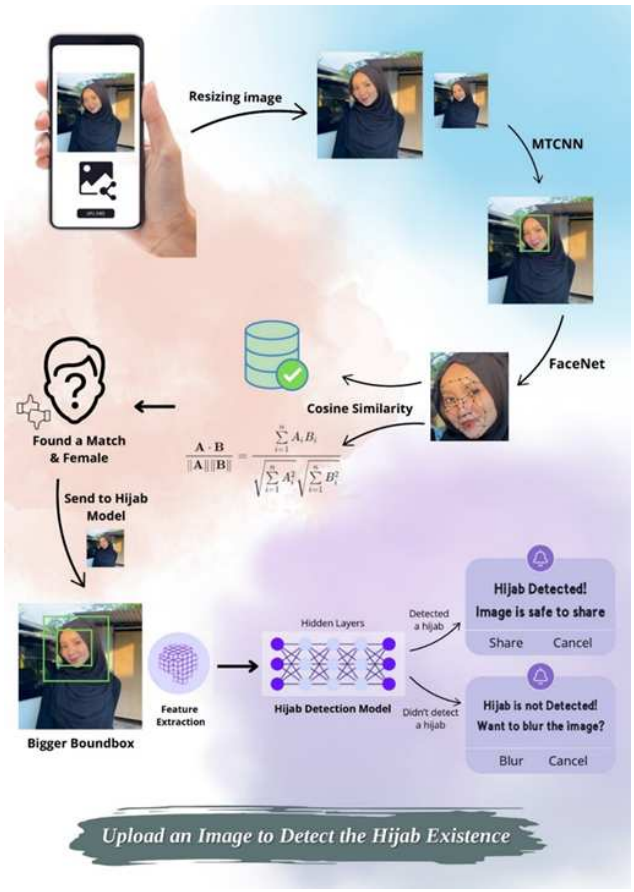
Once facial embeddings are generated using FaceNet, the next step in face verification involves comparing these embeddings to determine similarity. One of the most commonly used techniques for this comparison is cosine similarity. In SITR, this metric is used to verify whether a detected face belongs to a registered family member by comparing the embedding of the detected image with the stored face embeddings in the database. The cosine similarity between two embeddings is computed as follows:

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} \tag{1}$$

where, A and B are two face embeddings, $A \cdot B$ is the dot product, and $\|A\|$ and $\|B\|$ are the magnitudes of the vectors. A cosine similarity threshold of 0.7 was selected in SITR following a series of empirical trials in which thresholds



(a) Initial enrollment with embedding storage.



(b) hijab detection and privacy enforcement.

Fig. 3. SITR operational workflow including: (a) Family member enrollment and (b) Privacy-aware hijab detection.

ranging from 0.45 to 0.9 were evaluated on enrollment and test image pairs. The threshold of 0.7 yielded the best balance between true acceptance rate and false acceptance rate on the available face verification samples. It is acknowledged, however, that a formal ROC curve analysis and threshold sweep using a dedicated held-out verification set were not conducted; this represents a limitation of the current study and is recommended for future validation. Scores greater than or equal to 0.7 are accepted as a match, while scores below 0.7 are rejected.

D. Hijab Detection Model Based on DenseNet121

The hijab detection module of the SITR system was developed using DenseNet121 [12], a deep CNN pre-trained on the ImageNet dataset. This model was selected after evaluating multiple CNN architectures due to its robust feature extraction capabilities and proven effectiveness in image classification tasks. DenseNet121 is well known for its strong performance because it captures fine-grained features through dense blocks, where each layer is connected to every other layer in a feed-forward fashion [21]. This design promotes feature reuse and effective combination [22], which makes it suitable for detecting small or subtle details, such as whether a female subject is wearing a hijab. Since it is pre-trained on a large and diverse dataset, DenseNet121 provides a strong foundation for this classification task and adapts well to varied lighting conditions and viewing angles, allowing high accuracy even with a relatively small dataset [13].

To further improve its effectiveness, additional optimization steps were investigated:

1) Efficient Channel Attention (ECA): ECA [23], a light and effective channel-attention module based on local cross-channel interaction via 1D convolution, was added after each dense block to help the model concentrate on the most informative channel features.

2) Hyperparameter tuning: A grid search [24] was conducted to determine optimal values for learning rate, dropout, batch size, and weight decay.

The customized DenseNet121 architecture used for hijab detection is illustrated in Fig. 4. It shows the complete modified DenseNet121 architecture. The input image (224×224×3 px) passes through an initial 7×7 convolution and 3×3 max pooling layer before entering four dense blocks. After each dense block, an ECA attention layer is inserted for lightweight channel-wise feature recalibration using a 1D convolution (kernel size k=3). Transition layers (1×1 convolution + 2×2 average pooling) reduce spatial dimensions between dense blocks. After the fourth Dense Block, a 7×7 global average pooling layer is applied, followed by a dropout layer (rate=0.3)

and a fully connected linear layer that produces the binary classification output (0 = no hijab; 1 = hijab). The effect of these optimization steps is evaluated later in the Results and Discussion section.

E. Dataset and Preprocessing

A dataset of 3,082 annotated images collected from Roboflow was used to train the hijab detection model. Following data cleaning, 891 samples were removed due to inconsistent annotation or poor image quality. The final valid dataset contained 2,191 images (including 1,248 training images, 63 validation images, and 880 testing images). There was no bias due to class imbalance as there was a balance between the hijab and non-hijab classes (50.8% vs. 49.2, respectively). All the splits were done using a fixed seed randomly to guarantee reproducibility. The test set was not used in the hyperparameter tuning and was strictly held out. It is acknowledged that this dataset size is relatively small for training a robust deep learning classifier. In particular, the validation split of only 63 images may limit the reliability of hyperparameter selection decisions. These constraints stem from the limited availability of publicly annotated, culturally appropriate data and represent a recognized limitation of this study. Future work should aim to collect larger and more diverse datasets incorporating varied geographic regions, hijab styles, and demographic backgrounds to improve model generalizability. Cross-dataset validation is also recommended to assess robustness beyond the current data distribution.

F. Training and Evaluation Setup

The hijab detection model was trained and tested on Google Colab with GPU support, which ensured faster and more efficient experiments. All CNN models were trained under consistent conditions for fairness, as presented in Table II. The Colab environment was equipped with 2 vCPUs, 12 GB RAM, and an NVIDIA T4 GPU with 16 GB VRAM. Moreover, default PyTorch parameters were used unless otherwise noted. To improve reliability, four independent runs were conducted for each model, and results were averaged.

TABLE II. TRAINING SETUP AND PARAMETERS USED FOR ALL CNN MODELS

Factor	Value
Image Input Size	224×224 px
Batch Size	4
Learning Rate	1e-5
Weight Decay	1e-6
Dropout	0.3
Epochs	20
Optimizer / Loss	Adam with binary cross-entropy loss

G. Evaluation Metrics

To evaluate the effectiveness and reliability of the SITR hijab detection model, several commonly used metrics in classification tasks were adopted, as summarized in Table III. All metrics were calculated from multiple independent runs to reduce the impact of random variations in model performance and provide a more trustworthy assessment of the model's effectiveness. All the reported results are the averages over four independent runs.

TABLE III. SUMMARY OF EVALUATION METRICS USED TO ASSESS SITR MODEL.

Metric	Description / Purpose
Accuracy	Percentage of correctly classified images (hijab / non-hijab).
Precision	Fraction of predicted "hijab" images that are actually correct.
Recall	Fraction of actual hijab images correctly detected by the model.
F1-score	Harmonic mean of precision and recall for balanced evaluation.
Computational Time	Training and inference time, reflecting model efficiency.
Loss Values	Indicates training stability and convergence during evaluation.

H. Deployment Setup

To support practical use of the optimized DenseNet121 model, it was deployed as a REST API using the Flask framework on the Render cloud platform. Dynamic quantization reduced the model size to 27 MB from the original 82 MB, showing the memory efficiency and speed benefits. The SITR mobile application uses the backend API to communicate with Convex for handling user data and with ChromaDB to support similarity search during family-member verification.

IV. RESULTS AND DISCUSSION

This section presents quantitative and qualitative results for evaluating the SITR application and its core hijab detection model. The experiment was performed through three different phases: initially by benchmarking CNN models, followed by optimization of the DenseNet121 model with Efficient Channel Attention (ECA), and finally by tuning hyperparameters for maximum performance.

A. Quantitative Results

1) Benchmarking advanced CNN models: To identify the baseline model for hijab detection, four pre-trained CNN architectures, including ResNet50 [25], MobileNetV2 [26], EfficientNet-B0 [27], and DenseNet121) were investigated on the curated dataset. These models are selected because they are widely used for image classification and represent different levels of depth and complexity [13].

Table IV summarizes the average performance metrics across four separate runs. Fig. 5 presents grouped bar charts comparing accuracy, precision, recall, and F1-score across the four architectures, with each bar representing the mean value across four independent training runs to allow assessment of both peak performance and training consistency. The results show that DenseNet121 achieved the best overall performance. It recorded an average accuracy of 79.49%, precision of 73.48%, recall of 62.69%, and an F1-score of 65.67%. These findings represent a clear improvement over other model. Its accuracy is 5.3% higher than MobileNetV2, 3.9% higher than ResNet50, and 0.8% higher than EfficientNet-B0. In terms of F1-score, DenseNet121 also outperformed its closest competitor, EfficientNet-B0, by 1.1%. It should be noted that the baseline DenseNet121 accuracy of 79.49% is lower than the 96.47% reported for the rule-based visual feature detection method proposed in [18]. This discrepancy is

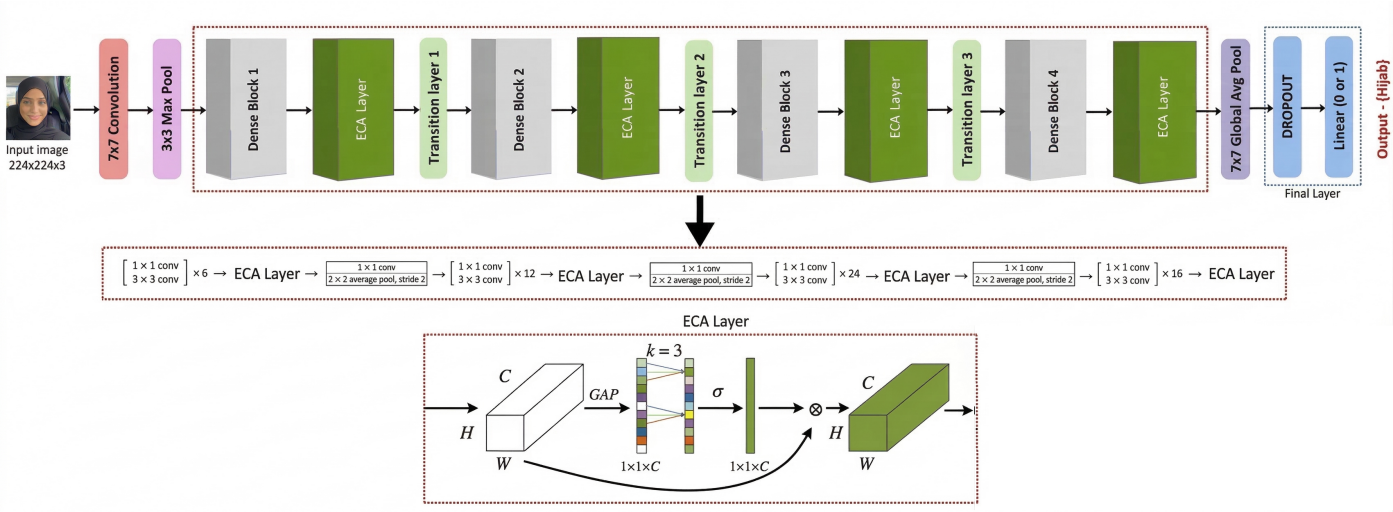


Fig. 4. The customized DenseNet121 architecture used for hijab detection model.

primarily attributable to differences in evaluation conditions: the rule-based approach was tested under controlled settings with clearly visible neck and hair features, whereas the deep learning models in this study were evaluated on a more heterogeneous dataset encompassing diverse lighting conditions, poses, partial occlusions, and backgrounds. Deep learning approaches such as DenseNet121 are expected to generalize better to unseen real-world conditions, as evidenced by the improved accuracy of 92.16% achieved after optimization, while rule-based methods may degrade significantly under uncontrolled visual conditions.

TABLE IV. AVERAGE PERFORMANCE METRICS ACROSS FOUR RUNS ON THE TEST SET IN TERMS OF CLASSIFICATION METRICS AND LOSS.

Model	Classification Metrics (%)				Loss
	Accuracy	Precision	Recall	F1-Score	
ResNet50	75.63	60.91	77.89	66.52	0.4512
MobileNetV2	73.18	55.87	60.36	57.84	0.4541
EfficientNet-B0	78.69	67.63	60.73	64.95	0.481
DenseNet121	79.49	73.48	62.69	65.67	0.486

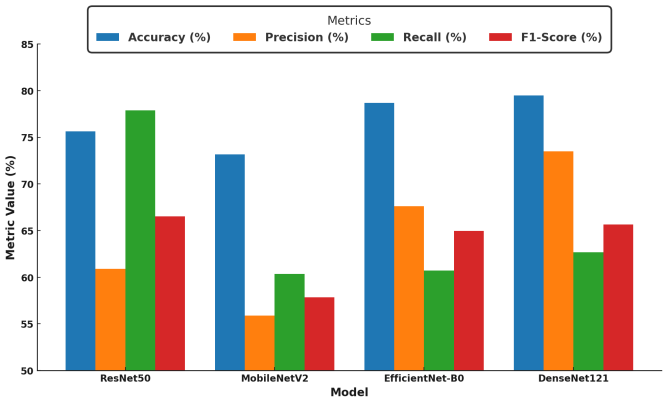


Fig. 5. Test classification results for the four models (averaged over four runs).

Precision is a critical metric in this context because it reflects the model's ability to reduce false positives. DenseNet121 achieved the highest precision at 73.48%, which indicates greater reliability in distinguishing hijab from non-hijab images. While ResNet50 captured slightly more true hijab samples, its lower precision reduced its overall consistency. In contrast, DenseNet121's architecture uses dense connectivity for better feature reuse. This allowed it to capture fine-grained details of hijab texture and coverage, leading to more stable and accurate classification.

Fig. 6 depicts the training and validation loss for each model. It is clear that DenseNet121 shows the smoothest and most consistent convergence. Its training and validation losses decreased steadily and remained close to each other. This pattern indicates strong generalization and reliable optimization. EfficientNet-B0 also achieves low final losses. However, its validation curve fluctuated during the middle epochs, which suggests it adapted more slowly to variations in the data. Meanwhile, validation loss for MobileNetV2 declines less uniformly This observation is consistent with its lower accuracy and recall scores in Table IV. ResNet50 reaches a reasonable loss value early in training, but its validation loss stabilizes at a higher level compared to DenseNet121. Overall, DenseNet121 converged faster and reached the lowest combined losses, supporting its superior quantitative results.

2) Model optimization with ECA: After selecting DenseNet121 as the best-performing baseline, it was further enhanced by integrating Efficient Channel Attention (ECA) after each dense block to prioritize informative channels. As shown in Table V, the ECA-enhanced version achieves an accuracy of 86.62% and an F1-score of 77.32%. This marks a significant improvement over the standard DenseNet121, which records 79.49% accuracy and a 65.67% F1-score. Overall, accuracy increases by 7.13% and the F1-score by 11.65%.

Other metrics improve as well. Precision rises from 73.48% to 79.91%, and recall increases from 62.69% to 74.91%. These findings confirm that the attention mechanism creates a better balance between identifying true hijab samples

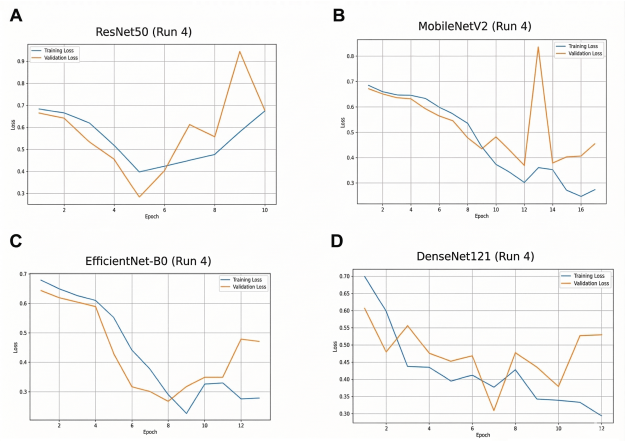


Fig. 6. Training and validation loss curves of four tested models over multiple epochs (Run 4).

and reducing false detections. The lower loss value, 0.369 compared to 0.486, also indicates more stable convergence. Together, these results demonstrate that incorporating ECA helps DenseNet121 learn richer feature representations and improves its overall robustness in hijab detection. However, this comparison reflects improvement over the untuned standard DenseNet121 only.

TABLE V. COMPARISON OF STANDARD AND ECA-ENHANCED DENSENET121 MODELS (AVERAGE OVER 4 RUNS).

Variant	Classification Metrics (%)				Loss
	Accuracy	Precision	Recall	F1-Score	
Standard DenseNet121	79.49	73.48	62.69	65.67	0.486
ECA-Enhanced	86.62	79.91	74.91	77.32	0.369

3) Hyperparameter tuning: A grid search was performed across various learning rates, dropout rates, batch sizes, and weight decays. The hyperparameter and the optimal configuration are given in Table VI. Due to GPU limitations, 19 different combinations were evaluated.

TABLE VI. GRID SEARCH HYPERPARAMETERS AND OPTIMAL VALUES

Parameter	Search Space	Optimal
Learning Rate	[0.001, 0.0001, 0.00001]	0.00001
Dropout	[0.3, 0.45, 0.7]	0.3
Batch Size	[4, 8, 16]	4
Weight Decay	[0.0001, 0.00001, 0.000001]	0.000001

Although the ECA-enhanced variant performed exceptionally well, it did not surpass the regular DenseNet121 after it was properly tuned. As such, we chose the tuned DenseNet121 as our final model on the basis of its superior validation performance in terms of accuracy, F1-score, and overall generalization capability.

The final deployment model was selected based on validation performance after hyperparameter tuning and then evaluated on the held-out test set. It achieved a test accuracy of 92.16%, an F1-score of 86.39%, and a test loss of 0.2174. It also achieved a precision of 91.63%. This high-precision value is particularly significant for the privacy-enforcement context of SITR. A false positive — where a non-hijab image is misclassified as containing a hijab — results in the system incorrectly declaring a privacy-sensitive image safe to share, representing the most critical failure mode. The achieved precision of 91.63%, therefore, reflects a low false-positive rate, directly aligned with the system's primary objective of protecting user privacy before image sharing.

B. Qualitative Results

To verify practical system behavior, the SITR mobile application was tested for image upload, family member verification, hijab detection, and privacy enforcement. During the enrollment stage, family members were successfully registered using FaceNet embeddings, which supported reliable verification in later stages. In the photo-sharing stage, uploaded images were processed and the system automatically verified hijab status and applied blurring to sensitive faces when necessary. These examples illustrate the intended behavior of the SITR pipeline in real image-sharing scenarios.

As illustrated in Fig. 7, the system correctly identified hijabi family members and allowed their images to be safely shared. The left screenshot shows the application in the analysis phase, where MTCNN detects the face and FaceNet confirms the identity as a registered female family member with a cosine similarity above the 0.7 threshold. The center screenshot displays the result: the DenseNet121 model detects a hijab and clears the image for sharing, displaying a "You can share the photo safely" confirmation message. The right screenshot shows the upload interface after analysis, indicating that no blurring was applied. This scenario corresponds to Path A in the SITR pipeline. In contrast, Fig. 8 shows how the system automatically blurred the head and neck region when a registered female family member was detected without a hijab, thereby enforcing privacy protection before the image could be shared. These examples highlight SITR's real-world functionality in supporting privacy-sensitive image sharing.

C. Discussion on Model Efficiency and Limitations

The strong performance of DenseNet121 can be attributed to its dense connectivity mechanism, which enables effective feature reuse and preserves low-level texture information across layers—characteristics particularly well-suited to hijab detection, which requires recognition of subtle visual patterns including fabric coverage, hair visibility, and neck exposure. The introduction of ECA improves the channel-wise feature recalibration and enables the network to focus on semantically valuable channels, and also reduces redundant responses. The mechanism is especially useful in visually complicated backgrounds where noise in the context may disrupt classification.

The optimized model was successfully deployed using Render after applying dynamic quantization. This technique reduces the model size from 82 MB to 27 MB. This compression ratio demonstrates that the proposed system is well-suited for resource-constrained deployment environments, including mobile and edge devices. The findings suggest that identity-aware, culturally sensitive AI systems of this nature are practically feasible beyond the laboratory setting and can be integrated as lightweight API services into existing mobile applications. An

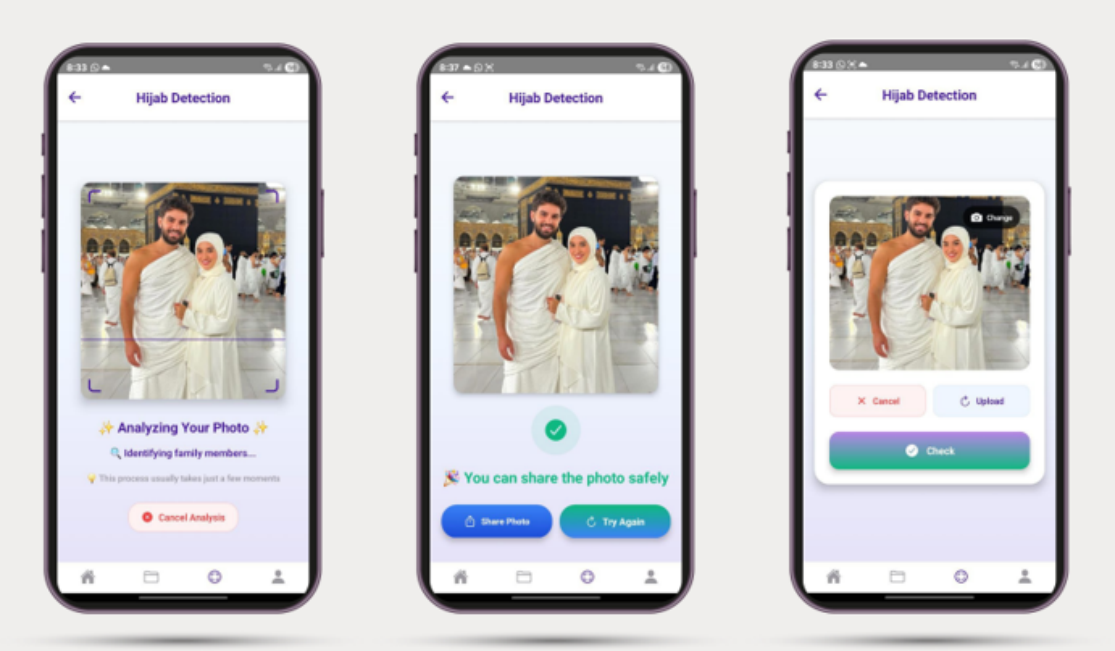


Fig. 7. Example of a hijabi family member correctly identified as safe for sharing.

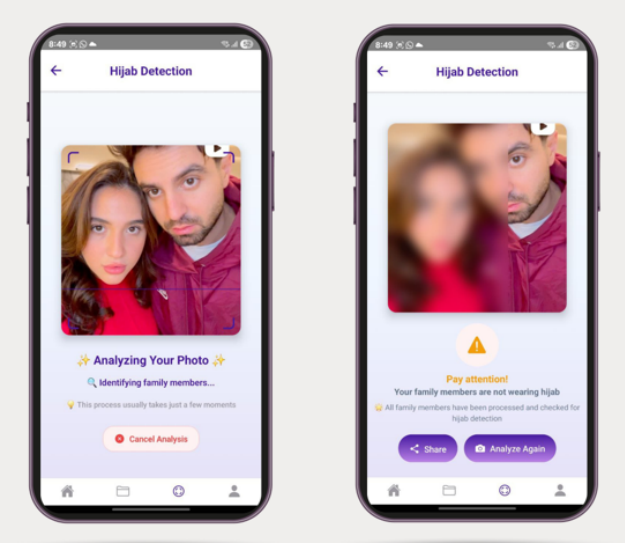


Fig. 8. Example of automatic blurring applied to a family member not wearing a hijab.

average inference time of approximately 19 seconds per image was recorded on a CPU-only cloud instance. This latency is acceptable for privacy-sensitive manual image-sharing workflows but is not yet suitable for real-time streaming applications.

From a broader research perspective, SITR demonstrates the viability of combining identity verification with visual attribute classification in a culturally grounded AI pipeline. This design pattern may be extended to other culturally sensitive contexts - such as age-appropriate content filtering or faith-based social media moderation—although such extensions would require separate dataset collection, model training, and ethical evaluation.

Regarding the absence of an end-to-end system evaluation, the current work evaluates each pipeline component (MTCNN, FaceNet, DenseNet121) individually. No integrated pipeline-level evaluation was conducted to measure the compounded effect of errors across all three stages on the final privacy decision output. A missed face detection or an incorrect identity match may propagate errors that are not captured by the individual component metrics reported in this study. This is a recognized limitation, and future work should include end-to-end evaluation using realistic multi-face test scenarios to quantify the true system-level accuracy.

Despite strong results, the system has several notable limitations:

1) Small and limited dataset: The hijab detection model was trained on only 2,191 images after removing 891 samples for quality issues. The validation split of only 63 images further limits the reliability of hyperparameter selection. This small and geographically constrained dataset may hinder generalization to diverse populations, hijab styles, and demographic backgrounds.

2) No end-to-end pipeline evaluation: Each of the three pipeline components — face detection, identity verification, and hijab classification — was evaluated independently. The compounded error across the full pipeline in realistic multi-face image sharing scenarios remains unmeasured.

3) Restricted cultural scope: The framework is explicitly designed for the Islamic cultural context. It does not account for women who wear head coverings for non-religious reasons (e.g., medical head coverings, traditional or cultural dress outside the Islamic context), which could produce misclassifications. The system is not validated for other cultural or religious head-covering traditions.

V. Conclusion

This study proposed the SITR framework, a culturally-aware and identity-sensitive deep learning system designed to prevent unintended sharing of privacy-sensitive images, specifically targeting the Islamic cultural context of hijab wearing. In contrast to traditional content moderation systems, SITR bases decisions on privacy enforcement on confirmed identity as well as classification in visual attributes in a single pipeline. The combination of face detection, identity verification through facial embeddings, and optimized hijab recognition shows how AI can be implemented in context-sensitive ethical use. The final tuned DenseNet121 model achieved an accuracy of 92.16% and an F1-score of 86.39%. Experimental results showed that ECA improved the untuned baseline significantly, but the tuned standard DenseNet121 achieved the best final performance and was selected for deployment. Dynamic quantization reduced the model size from 82 MB to 27 MB, making it lightweight and suitable for REST API deployment. The system demonstrated smooth end-to-end communication between the mobile frontend and backend, supporting practical integration in privacy-sensitive image-sharing scenarios.

Despite these promising results, two primary limitations must be explicitly acknowledged. First, the hijab detection model was trained on a relatively small dataset (2,191 images) with a minimal validation split (63 images), which limits confidence in the generalizability of the learned representations to diverse cultural, demographic, and environmental contexts. Second, no end-to-end system evaluation was conducted to quantify the compounded error across the three-stage pipeline (face detection $\rightarrow$ identity verification $\rightarrow$ hijab classification), meaning the true system-level accuracy in realistic multi-face image-sharing scenarios remains unmeasured. Additionally, the cosine similarity threshold of 0.7 used for face verification was selected empirically without a formal ROC-based analysis, which represents a further methodological gap.

Future work should address these gaps through the following directions: 1) collecting larger, more culturally and demographically diverse hijab detection datasets and applying cross-dataset validation; 2) conducting a full end-to-end pipeline evaluation with realistic test scenarios to measure true system-level precision and error propagation; 3) performing formal ROC curve analysis for threshold selection in the face verification component; and 4) exploring GPU-optimized and lighter model architectures to reduce inference latency below the current approximately 19 seconds per image, enabling real-time or near-real-time privacy enforcement in dynamic image-sharing contexts.

Declaration on Generative AI

During the preparation of this manuscript, the authors used a generative AI tool (ChatGPT) only to assist with language refinement, paraphrasing, and text organization. After using this tool, all generated content was carefully reviewed and edited by the authors as needed. The authors remain fully responsible for the scientific content, accuracy, originality, and integrity of this manuscript.

Acknowledgment

The authors would like to thank the Department of Computer Systems Engineering at the Arab American University for supporting the senior project that formed the basis of this work. The authors also acknowledge all contributors who supported the implementation, testing, and refinement of the SITR system.

References

[1] K. Ghazinour and J. Ponchak, "Hidden privacy risks in sharing pictures on social media," *Procedia Computer Science*, vol. 113, pp. 267–272, 2017.

[2] B. Debatin, J. P. Lovejoy, A.-K. Horn, and B. N. Hughes, "Facebook and online privacy: Attitudes, behaviors, and unintended consequences," *Journal of Computer-Mediated Communication*, vol. 15, pp. 83–108, 2009.

[3] L. Rakhmawati, Wirawan, and Suwadi, "Image privacy protection techniques: A survey," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 0076–0080.

[4] T. Afnan, Y. Zou, M. Mustafa, M. Naseem, and F. Schaub, "Aunties, strangers, and the FBI: Online privacy concerns and experiences of Muslim-American women," Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022). USENIX Association, 2022, pp. 387–406.

[5] A. Madani and M. W. F. D. M. Madkour, "Automatic face and hijab segmentation using convolutional network," *International Journal of Integrated Engineering*, vol. 11, pp. 60–66, 2019.

[6] T. Afnan, Y. Zou, M. Mustafa, M. Naseem, and F. Schaub, "Aunties, strangers, and the fbi: online privacy concerns and experiences of muslim-american women," in *Proceedings of the Eighteenth USENIX Conference on Usable Privacy and Security*, ser. SOUPS'22. USA: USENIX Association, 2022.

[7] M. Franco, S. A. Falioun, K. E. Fisher, O. Gaggi, Y. Ghamri-Doudane, A. J. Nashwan, C. E. Palazzi, and M. Shwamra, "A technology exploration towards trustable and safe use of social media for vulnerable women based on islam and arab culture," in *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, ser. GoodIT '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 138–145. [Online]. Available: https://doi.org/10.1145/3524458.3547259

[8] T. Gillespie, "Content moderation, ai, and the question of scale," *Big Data & Society*, vol. 7, 2020.

[9] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: a complete review," *Multimedia Tools Appl.*, vol. 80, no. 3, p. 4825–4880, Jan. 2021. [Online]. Available: https://doi.org/10.1007/s11042-020-09850-1

[10] K. M. Hosny, N. AbdElFattah Ibrahim, E. R. Mohamed, and H. M. Hamza, "Artificial intelligence-based masked face detection: A survey," *Intelligent Systems with Applications*, vol. 22, p. 200391, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667305324000668

[11] S. Ram, S. Vinoth, R. N. Gopalakrishnan, A. A. Balakumar, L. Kalinathan, and T. A. J. Velankanni, "Leveraging diverse CNN architectures for medical image captioning: DenseNet-121, MobileNetV2, and ResNet-50 in ImageCLEF 2024," in *CLEF 2024 Working Notes, Proceedings of the 15th International Conference of the CLEF Association*, ser. CEUR Workshop Proceedings, vol. 3740. Grenoble, France: CEUR-WS.org, 2024, pp. paper–160. [Online]. Available: https://ceur-ws.org/Vol-3740/paper-160.pdf

[12] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[13] T. Thaher, M. Saffarini, M. Mafarja, A. Alashbi, A. H. Mohamed, and A. A. El-Saleh, "A hybrid approach for heavily occluded face detection using histogram of oriented gradients and deep learning models," *Computer Modeling in Engineering & Sciences*, vol. 144, no. 2, pp. 2359–2394, 2025. [Online]. Available: http://www.techscience.com/CMES/v144n2/63698

[14] Z. L. Z. Zhang and Y. Q. K. Zhang, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, pp. 1499–1503, 2016.

[15] D. Kalenichenko and J. P. F. Schroff, "Facenet: A unified embedding for face recognition and clustering." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.

[16] A. Zhalgas, B. Amirgaliyev, and A. Sovet, "Robust face recognition under challenging conditions: A comprehensive review of deep learning methods and challenges," *Applied Sciences*, vol. 15, no. 17, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/17/9390

[17] W. K. H., & Dong, ""face recognition based on mtcnn and convolutional neural network."," *Frontiers in Signal Processing, vol. 4, no. 1*, pp. 37–42, 2020.

[18] A. Begum, S. K. Fatama, and M. M. I. M. Khaliluzzaman, "Detection and analysis of hijab based on visual feature of neck and hair," *in 2017 20th International Conference of Computer and Information Technology (ICCIT)*, pp. 1–6, Dec. 2017.

[19] Pimeyes - advanced face recognition search engine. [Online]. Available: Available: https://pimeyes.com/en

[20] A. Aljanzory. (2024) Haramblur - blur haram, nsfw images & videos. [Online]. Available: Available: https://chromewebstore.google.com/detail/haramblur-blur-haram-nsfw/pbcoegikffnadpahojjhgdladmmddeji?hl=en

[21] P. Podder, F. B. Alam, M. R. H. Mondal, M. J. Hasan, A. Rohan, and S. Bharati, "Rethinking densely connected convolutional networks for diagnosing infectious diseases," *Computers*, vol. 12, no. 5, 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/5/95

[22] F. A. Azis, H. Suhaimi, and E. Abas, "Real-time face mask classification with convolutional neural network for proper and improper face mask wearing," *International Journal of Computing*, vol. 22, no. 2, pp. 184–190, Jul. 2023. [Online]. Available: https://computingonline.net/computing/article/view/3087

[23] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 531–11 539.

[24] J. Ilemobayo, O. Durodola, O. Alade, O. Awotunde, T. Adewumi, O. Falana, A. Ogungbire, A. Osinuga, D. Ogunbiyi, I. Odezuligbo, O. Edu, and A. Ifeanyi, "Hyperparameter tuning in machine learning: A comprehensive review," *Journal of Engineering Research and Reports*, vol. 26, pp. 388–395, 06 2024.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[27] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds. PMLR, 2019, pp. 6105–6114. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html