StealthFace:Transfer Learning-Based Ensemble Model for Disguised Face Recognition using Skin-Segmented Images

Padmashree G, Karunakar A Kotegar

Abstract—**Disguised face identification is challenging since people cover their identities by wearing masks, hats, sunglasses, or other disguises. These disguises dramatically modify face features, making identifying individuals a difficult task. In this study, we introduce StealthFace, an ensemble model that utilizes transfer learning and skin-segmented images to improve the recognition of disguised faces. To improve the accuracy and resilience of disguised face recognition, StealthFace leverages the power of deep learning algorithms, transfer learning, and ensemble methods. The framework employs pre-trained convolutional neural network models, such as ResNet50 and DenseNet121 for feature extraction and to learn discriminative features that are important in identifying individuals despite the disguises. The ensemble approach combines models' predictions, utilizing their collective expertise and capturing multiple perspectives on disguised faces. Our ensemble model outperforms other state-of-the-art methods with an overall accuracy of 99.24% and contributes to enhancing security measures and advancing the development of face recognition systems.**

Index Terms—**Deep Learning, Disguised Faces, Ensemble Approach, Face Recognition, Skin Segmentation.**

I. Introduction

ONE of the most investigated subjects in computer vision and machine learning is face recognition. Face recognition has made great progress since the introduction of deep learning-based techniques [1], [2], [3]. However, when faces are disguised, the performance of face recognition algorithms might suffer significantly. Simple disguises such as spectacles or hats can be worn, as can more complicated ones such as makeup, facial hair, or masks. Face recognition algorithms are significantly hindered by the variety and unpredictable nature of disguises [4], [5].

Disguised Facial Recognition (DFR) is the technique of recognizing a person who is wearing a disguise or changing their physical appearance. DFR is an essential activity with numerous applications such as surveillance, law enforcement, security, and access control. Traditional face recognition algorithms often rely on the availability of high-quality face photos for training and testing. However, the lack of disguised face datasets and difficulty in capturing high-quality photos in real-world circumstances make DFR difficult. DFR has attracted a great deal of attention from researchers recently, who are working to create reliable algorithms that function well under a variety of forms. Due to their capacity to learn discriminative features from significant volumes of data, deep learning-based techniques have demonstrated promising outcomes in DFR.

Some of the key issues with DFR are: (i) **Alterations to facial features:** Disguises can change the size, color, and texture of a person's face, making it challenging for conventional facial recognition algorithms to correctly identify them. (ii) **Inadequate training data:** Developing good algorithms for disguised face recognition necessitates a substantial amount of training data. However, it might be challenging to find high-quality pictures of people who are wearing disguises, which results in a scarcity of relevant training data. (iii) **Variability in disguises:** Disguises can take several forms, such as masks, hats, sunglasses, makeup, or facial hair. Because different types of disguise can modify a person's look in different ways, it is challenging for algorithms to generalize and properly identify individuals in various types of disguise. (iv) **Intraclass variability:** Even within the same sort of disguise, there can be significant differences in how it is worn and applied, making it difficult to precisely identify individuals. (v) **Inter-class similarity:** Disguises can make people with diverse faces look alike, confusing recognition systems. Some of the sample disguise images from the Sejong face dataset are shown in Figure 1.

Outlined below are the main contributions of this research:

- A complete review of the research contributions involved in disguised facial recognition is presented.
- StealthFace, an ensemble framework, is presented using the transfer learning approach to reduce training convergence time and optimize a large number of parameters for the identification of disguised faces.
- In-depth ablation study that explores multiple factors influencing StealthFace's performance, including the impact of batch size, the comparative analysis and selection of pre-trained models, the advantages of ensemble learning, and the evaluation of various loss functions, all aimed at optimizing performance for disguised face recognition
- To avoid overfitting, various strategies such as dropout, batch normalization, global average pooling, and early stopping methods are utilized.

The paper is organized as follows. Section II reviews related works in the field, providing context and identifying gaps addressed by this research. Section III outlines the proposed approach, detailing the methodology and innovations introduced. Section IV presents the experiments and

Manuscript received May 13, 2024; revised January 23, 2025.

Padmashree G is an assistant professor in the Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India e-mail: g.padmashree@manipal.edu.

Karunakar A Kotegar is a professor in the Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India e-mail: karunakar.ak@manipal.edu.

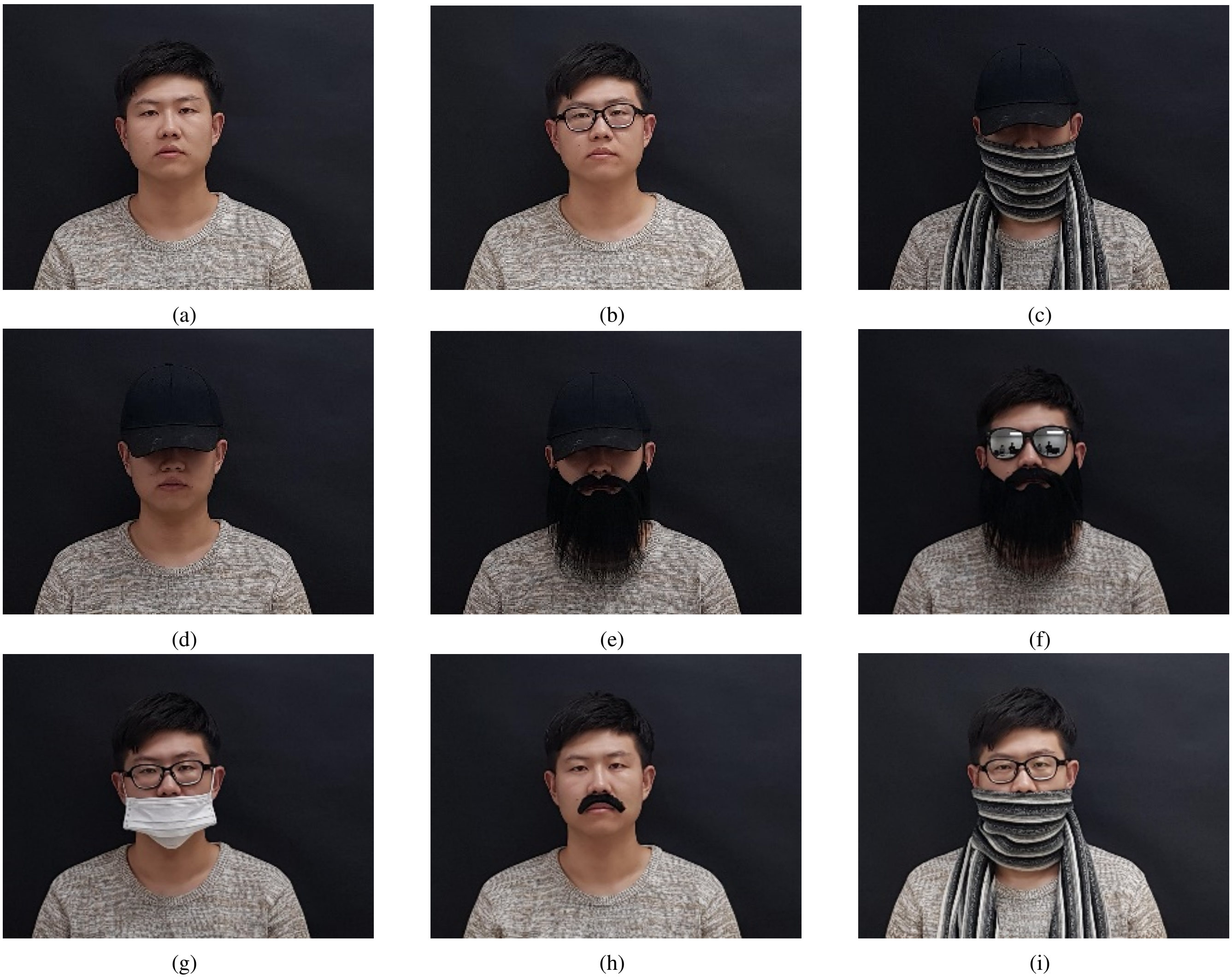


Fig. 1: Sample Disguised Images From Sejong Face Dataset (a) Normal (b) Glasses (c) Scarf and Cap (d) Cap (e) Fakebeard and Cap (f) Glasses and fakebeard (g) Glasses and mask (h) Fake mustache (i) Glasses and scarf

results, demonstrating the effectiveness and performance of the proposed method. Section V includes an ablation study, which assesses the impact of various components of the approach. Section VIII benchmarks the StealthFace against state-of-the-art techniques, offering a comparative analysis. Finally, Section IX concludes the paper, summarizing the key findings and implications of the research.

II. Related Works

In recent decades, significant progress has been made in face recognition research, leading to remarkable advancements. Many state-of-the-art face recognition frameworks have achieved exceptional accuracy, particularly for unconstrained face datasets originating from controlled environments. However, despite continuous improvements in accuracy, these frameworks often face significant challenges when it comes to recognizing faces under disguise. Disguise poses a difficult obstacle that many face recognition systems still struggle to overcome, resulting in performance limitations and potential failures.

As the research in disguised face recognition continues to evolve, it remains a complex task with various challenges. In this context, [6] proposed a model aimed at recognizing disguised faces. Their approach involved a two-stage training process. In the first stage, they employed two Deep Convolutional Neural Networks (DCNNs) to extract identity features from both aligned and unaligned images. These extracted features were later fused to enhance the representation. Moving to the second stage, they computed the Principal Component Analysis (PCA) transformation matrix to facilitate the recognition of disguised faces. By applying this approach, they achieved an impressive accuracy of approximately 79%, positioning their model among the top performers in the Disguise Faces in the Wild (DFW) competition phase-1. This contribution demonstrates the efficacy of their model in effectively recognizing disguised faces, and the attained accuracy showcases its competitiveness in the field.

Recognizing faces under various challenges such as image resolution variations, illumination changes, age differences, and pose variations become even more complex when individuals are disguised. To address this issue, [7] proposed an approach for identifying disguised faces and distinguishing them from impersonators. Their approach involved training two distinct DCNNs, namely Inception ResNet-v2 and ResNet-101 [8], during the training phase. To enhance the Softmax loss, they utilized the L2-softmax loss. The features extracted from the two networks were fused by computing the average of the scores. These fused features were then embedded into the discriminative subspace of a metric learning framework. During testing, the features of the given pair of faces were extracted using the DCNNs and embedded into the subspace. Finally, the similarity score between the embedded features was calculated. The authors achieved a promising accuracy of approximately 72.9%, which indicates the potential of their approach and paves the way for future research to further improve the understanding and performance of disguised face recognition. Overall, their approach demonstrates the effectiveness of using DCNNs and metric learning techniques for recognizing disguised faces and highlights areas for future exploration and advancements in the field.

To authenticate disguised faces, researchers [9] devised a transfer learning approach based on deep learning that utilizes the Residual Inception network framework coupled with center loss. Their approach, known as ”Deep Disguise Recognizer (DDR)”, involves a two-phase training process. In the first phase, deep inception ResNet networks are trained on a large-scale face database to learn face representations. These networks serve as the basis for extracting features. In the second phase, the pre-trained model is transferred to the DDR, which encodes the face representations of facial disguises. The authors noted that the accuracy of verification varied across different databases and different DCNN models. Moreover, their analysis revealed that the DDR-MSCeleb framework exhibited better performance for genuine male subjects compared to genuine female subjects. The proposed approach demonstrates the effectiveness of using transfer learning and the Residual Inception network framework for verifying disguised faces. However, the study highlights the importance of considering database variations and the influence of DCNN models on verification accuracy. Furthermore, the authors observed gender-related performance differences, emphasizing the need for further investigation and improvements in recognizing disguised faces, particularly for female subjects.

[10] presented a method for identifying individuals from disguised and impostor photographs. Their approach utilizes a VGG-face architecture combined with a contrastive loss based on a cosine distance metric, employing the Disguised Faces in the Wild (DFW) dataset. Compared to the DFW baseline, their proposed network achieved a significant accuracy improvement of 27.13%. This enhancement was achieved through the augmentation of data and resulted in improved generalization capabilities of the network. The findings highlight the efficacy of employing the VGG-face design and the contrastive loss with a cosine distance metric for robust identification of individuals in the presence of disguises. The substantial accuracy improvement achieved by the proposed approach demonstrates its potential for advancing the field of disguised face recognition and improving the performance of identification systems.

The DFW dataset, introduced in [3], comprises more than 11,000 photos capturing 1,000 identities, each with variations in different types of disguise accessories. This dataset includes both genuine and impostor obfuscated face photos, offering a comprehensive resource for studying the challenges of disguised face recognition. The dataset is further segmented into easy, medium, and hard difficulty levels to demonstrate the problem's complexity, enabling a complete analysis. This paper gives thorough descriptions of the DFW dataset, including baseline results, evaluation methodology, performance analyses of entries submitted to the First International Workshop and Competition on DFW, and insights into the three difficulty levels of the DFW challenge dataset. The availability of the DFW dataset, along with the comprehensive analysis provided in this research, serves as a valuable reference for understanding and addressing the intricate nature of disguised face recognition. The diverse range of images and difficulty levels enable researchers to evaluate and develop robust algorithms for this challenging task.

The research [11] introduces A2-LINK, an active learning system, to handle the challenge of face recognition in the presence of disguise. A2-LINK starts with a face recognition machine learning model, intelligently chooses training samples from the target domain, and then uses hybrid noises like adversarial noise to fine-tune a network that performs well in both the presence and absence of disguise. Experimental findings on the DFW and DFW2019 datasets with cutting-edge deep-learning feature models like DenseNet and ArcFace show the efficacy and generalizability of the proposed approach.

[12] proposed a novel encoder-decoder network called DED-Net, which focuses on learning both local and global features of disguised and non-disguised images. The network utilizes cosine and mahalanobis distance metrics to capture the variations in class characteristics. The entire framework is named Disguise Resilient (D-Res). The research also takes into account low-resolution images from benchmark datasets DFW2018 and DFW2019, specifically considering resolutions of 32×32, 24×24, and 16×16. By employing the D-Res framework, an impressive accuracy of 96.3% is achieved, surpassing state-of-the-art techniques by an improvement of 3%. The DED-Net network, combined with the use of distance metrics and the D-Res framework, demonstrates its effectiveness in capturing both local and global features of disguised and non-disguised images. The incorporation of low-resolution images and the notable increase in accuracy highlight the potential of this approach in advancing the field of disguised face recognition.

Addressing the challenge of limited database availability in the research domain, researchers [13] introduced a multimodal disguised face dataset. This dataset encompassed 100 participants, each with 15 diverse disguised photos and 8 distinct face add-ons. The dataset captured these samples under various modalities, including infrared, visible, thermal spectra, and visible plus infrared.

[14] proposed a framework called Disguise Invariant Face Recognition (DIFR) for effectively recognizing disguised faces. The framework utilizes the Viola-Jones face detector to detect faces, which is further enhanced by a noise-based augmentation technique to improve detection accuracy. To learn discriminative features for disguised face recognition, a fine-tuned pre-trained Convolutional Neural Network (CNN) is employed. Four different pre-trained models are utilized in the framework to identify disguised faces, with ResNet-18 [8] achieving the highest accuracy of 98.19%. The DIFR framework provides an effective technique for dealing with

disguise modifications in face recognition. It achieves outstanding performance in successfully recognizing disguised faces by employing a combination of face detection algorithms and fine-tuned CNN models. ResNet-18's improved accuracy illustrates its use in capturing discriminative traits and detecting disguised identities.

[15] presented a unique system for occlusion face recognition that included joint segmentation and feature learning. The framework is divided into three parts: the occlusion prediction (OP) module, the channel refinement (CR) network, and the feature purification (FP) module. The OP module predicts an occlusion mask, which is then transformed into a channel-wise mask matrix. This matrix is used to attenuate the occlusion characteristics while highlighting more discriminative visible features in both spatial and channel dimensions. To enhance the viability of candidate embeddings, the FP module is specifically designed to refine the non-occluded feature maps. Rather than directly embedding the non-occlusion feature maps, this module enhances the combined original and occlusion-free feature maps. Furthermore, the researchers introduced an upgraded occlusion face dataset called Webface-OCC+ for evaluating the proposed framework's generalization capabilities. The combined framework, consisting of the OP module, CR network, and FP module, demonstrates the promising potential for occlusion face recognition by jointly addressing segmentation and identification challenges. The evaluation of the Webface-OCC+ dataset further validates the framework's effectiveness in achieving improved generalization performance.

[16] put forth a novel DCNN that combines gait and facial biometric qualities for individual recognition. The proposed framework merges the feature vectors extracted from gait energy images and facial images, which are then input into the CNN model for further feature extraction. The experimental results demonstrated remarkable performance, achieving an accuracy of 97.5% on benchmark datasets including ORL Face, FEI Face, and CASIA Gait. This integration of gait and facial information within a deep learning framework showcases the effectiveness of the proposed approach in achieving highly accurate individual recognition.

In response to the scarcity of annotated datasets featuring low-resolution images with disguises, [16] introduced the D-LORD dataset. This collection consists of low-resolution surveillance films and high-resolution mugshot photos. It contains about 1.2 million frames from $14,098$ recordings and $2,100$ people. Under varied lighting situations, the captured subjects wear various disguise artifacts such as hats, monkey caps, wigs, sunglasses, and face masks. The D-LORD dataset is a great resource for advancing study in this sector since it addresses the difficult challenge of low-resolution face identification with disguise variations.

Researchers have paid close attention to the topic of disguised face recognition in recent years due to its broad potential applications in a variety of domains. The studies discussed in this part give light on the fundamental obstacles of disguised face identification, such as image quality, lighting conditions, and occlusions induced by disguises. To address these issues, a variety of ways have been proposed, including the use of deep learning-based techniques. However, there is still a significant performance gap between the latest innovations and traditional face recognition systems, highlighting the need for additional study and improvements in this sector. Overall, the findings from the reviewed studies provide significant views and guidance for future efforts aiming at improving the accuracy and robustness of disguised face recognition.

III. Proposed Approach

We provide a strong and effective model for recognizing disguised faces in this work. With limited computation resources, a deep ensemble neural network with transfer learning is described to minimize false positives and false negatives. Figure 2 depicts the general workflow of the disguised face recognition process. The initial step involves utilizing a YOLO face detector for the detection of faces within the images. Subsequently, a region-based watershed algorithm is employed to extract the skin region from the detected faces. Once the skin-segmented images are obtained, features are extracted from them. These features play a crucial role in the subsequent recognition of disguised faces. Finally, the extracted features are fed into the classification block, which performs the task of recognizing and classifying disguised faces. By following this series of steps, the proposed framework effectively detects faces, isolates the skin regions, extracts meaningful features, and performs the classification process for disguised face recognition. The utilization of the YOLO face detector and the region-based watershed algorithm enables accurate localization of faces and extraction of skin regions.

A. Face Detection using YOLO

Faces were identified using a YOLO face detector from the 1382×1061 resolution raw images of the Sejong dataset, which were then scaled to 224×224 to extract the features and carry out recognition. Additionally, image augmentation is used to extend the size of a training dataset by creating new, slightly changed versions of the original data to minimize overfitting and improve the model's accuracy and generalization capacity. Image augmentation includes rotation range, wide shift range, height shift range, zoom shift range, and horizontal flip.

B. Marker-controlled watershed segmentation

We apply a marker-controlled watershed segmentation algorithm [17] to extract the skin region from disguised faces in an image of irregular forms. The algorithm consists of gradient transformation, which converts the image into a topographical representation in which the intensity values of the pixels are used to define the heights of the terrain, marker placement to identify regions that correspond to objects of interest, flooding by filling the catchment basins until the basins merge or reach an image edge, and finally segmentation. We employ the HSV and YCbCr color spaces to extract the skin region of the faces in this case. The images are pre-processed using morphological operations such as erosion and dilation, and then the region-based marker-controlled watershed technique is used to extract the skin region of the identified faces. The complete flow of skin-segmentation process is shown in Figure 3.

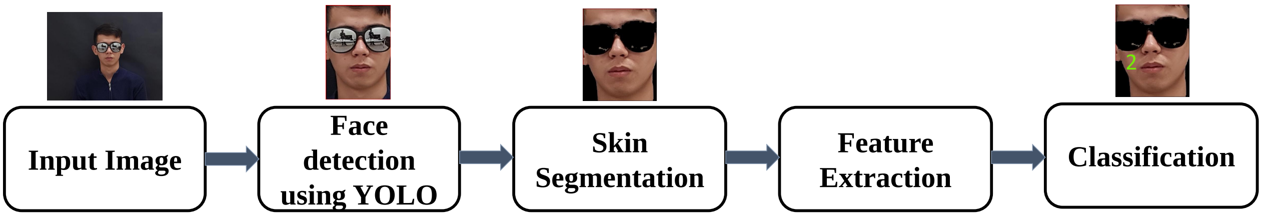


Fig. 2: General workflow of the disguised face recognition process

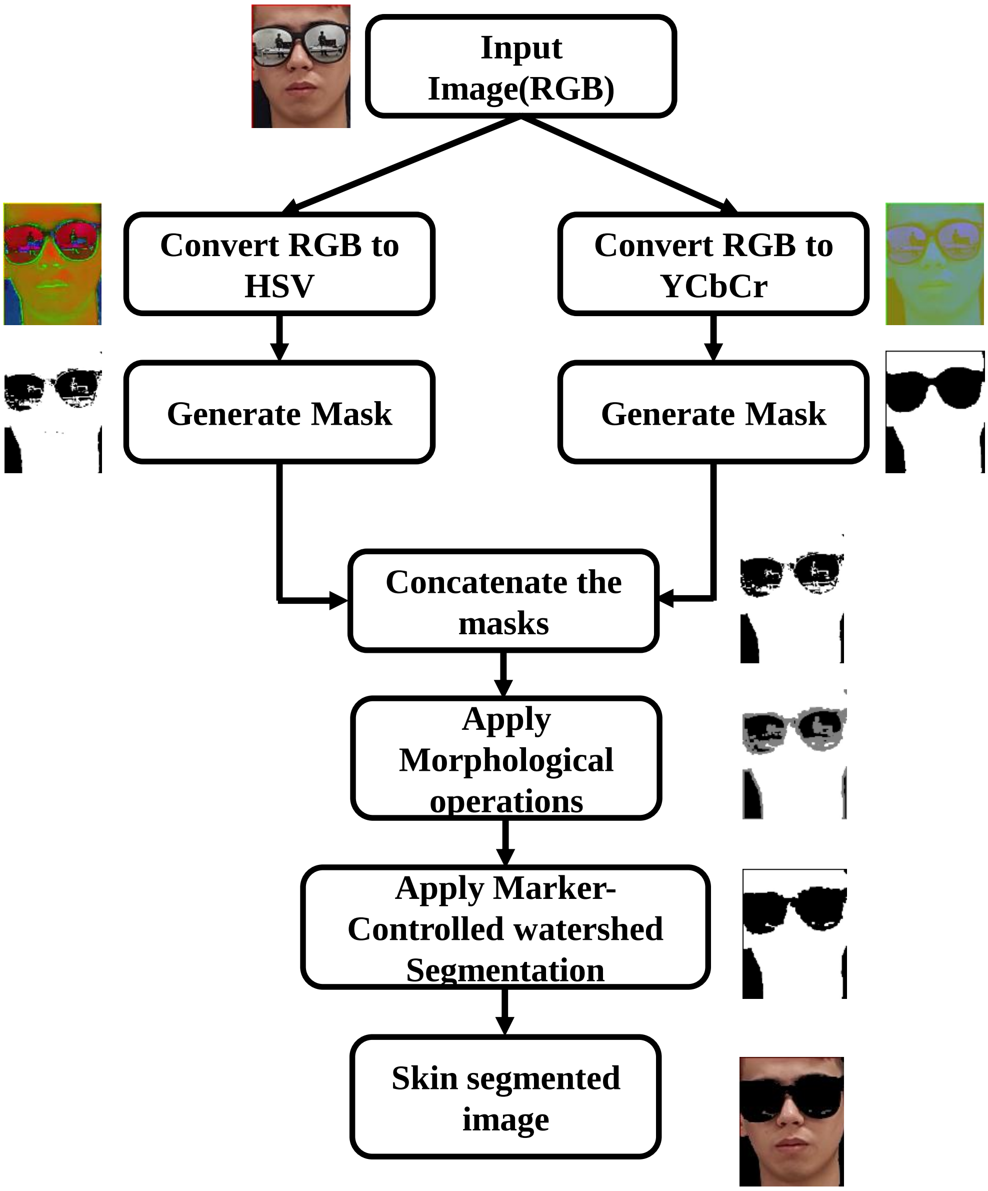


Fig. 3: Skin-segmentation process

C. Feature Extraction and classification

The entire feature extraction and classification process is visually illustrated in Figure 4. The model presented for disguised face recognition utilizes an ensemble of two powerful, pretrained convolutional neural network (CNN) architectures—ResNet101 [8] and DenseNet121—both fine-tuned to adapt to the specific task of recognizing faces with varying levels of disguise. The model has three primary components: input processing, feature extraction via transfer learning, and classification.

The model takes in skin-segmented facial images with disguises applied to them, such as glasses, masks, or other occlusions, as shown on the left side of the diagram. These images serve as input to the feature extraction pipeline. The objective here is to handle facial variations caused by different disguises while maintaining the identity information required for recognition.

Feature extraction is a critical step in the model where an ensemble approach is employed using ResNet101 [8] and DenseNet121. These models are both pretrained on large-scale datasets (e.g., ImageNet) but are now fine-tuned specifically for the task of disguised face recognition.

ResNet101 [8] is a deep CNN architecture consisting of 101 layers with residual connections. These skip connections allow the network to learn identity mappings, making it easier to train deep networks without facing the vanishing gradient problem. After processing the image through ResNet101 [8], a 2048-dimensional feature vector is obtained. To reduce dimensionality and extract more meaningful representations, this feature vector is passed through two fully connected layers: one with 4096 units and another with 512 units. In the fine-tuning process, the later layers of ResNet101 [8] are unfrozen and retrained on the disguised face dataset. This allows the model to adapt to the unique features of disguised faces while leveraging the powerful feature extraction capabilities of ResNet101 [8].

DenseNet121 features dense connectivity between layers, ensuring maximum feature reuse and efficient flow of gradients across the network. Each layer receives input from all preceding layers, resulting in more discriminative feature extraction. Similar to ResNet101 [8], DenseNet121 extracts a feature vector, but in this case, the size of the feature vector is 1024. This vector is passed through a single fully connected layer with 512 units. The fine-tuning strategy for DenseNet121 follows a similar process to that of ResNet101 [8], where the final layers of the model are retrained on the disguised face dataset to learn features unique to facial disguises.

After feature extraction, the outputs from both models (the 512-dimensional feature vectors from ResNet101 [8] and DenseNet121) are concatenated into a unified feature vector. This combined representation captures complementary information from both models. ResNet101 [8] focuses on extracting robust hierarchical features, while DenseNet121 emphasizes detailed fine-grained features through its dense connections. By combining the strengths of these two architectures, the ensemble becomes more capable of handling various levels of disguise, whether subtle (like makeup) or extreme (like masks).

The unified feature vector, which now contains the information extracted by both ResNet101 [8] and DenseNet121, is passed to the classification module. This module consists of three fully connected layers. The concatenated feature vector is first passed through a layer with 1024 units, which helps in refining the combined representation from the feature extraction stage. The next layer has 400 units, where more abstraction is applied to the features, focusing on distinguishing between subtle variations in the disguised faces. The final layer with 64 units is used to further distill the features before passing them to the output layer.

The final output layer contains 64 units, which represent the number of identities in the dataset. The model assigns probabilities to each identity, predicting the correct individual even in the presence of disguises. The final classification is done using a softmax function, ensuring that the model outputs a probability distribution over the identities.

IV. Experiments and Results

A. Datasets

Subset-A and Subset-B are the two subsets of the Sejong database [13]. Subset-A comprises 30 individuals' facial images, 16 males and 14 females, recorded with one neutral and one add-on image in each modality. Frontal faces were used in all of the images. Subset-B comprises 70 individuals' facial images, 44 males and 26 females, with 15 neutral face images and 5 add-on images acquired in each modality for the remaining add-ons. In addition, 5 photos with actual beards for men and cosmetics for women were recorded. Subset-A contains $1,500$ images with 30 subjects, 4 modalities, $12-13$ addons, and 1 pose, whereas Subset-B contains $23,100$ images with 70 subjects, 4 modalities, $12-13$ addons, and $5-15$ poses. The summary of various disguised images available in the Sejong dataset is provided in Table I.

B. Experimental Setup

A series of tests are carried out to illustrate the efficiency of the suggested model. These experimental results are thoroughly covered in this section. An HP Elite Desk 800 G4 Workstation with a 48GB NVIDIA GeForce RTX A6000 GPU, 128GB RAM, and an i9 processor running at 3.7 GHz are used for all of the research. The Keras(2.3.1) Framework and Python(3.9) are used to implement the algorithm. All the experiments were carried out by employing categorical cross-entropy as the loss function and Adam as the optimizer. We utilize a batch size of 4 with an initial learning rate of 0.0001 and train our network for a maximum of 20 epochs.

C. Results analysis of the proposed model

The proposed model, which is an ensemble of DenseNet121 and ResNet101 [8], demonstrates remarkable performance across various metrics: Precision, Recall, F1-Score, Accuracy, and AUC. A precision of 99.32% implies that the model excels in accurately identifying disguised faces, particularly when there are multiple possible identities. In a security or authentication scenario, this high precision would mean that the model rarely falsely identifies an individual as someone they are not, which is crucial when face recognition systems are used in high-stakes environments like airports or banks.

The recall score of 99.23% means the model can retrieve almost all the instances where a person is disguised. This is especially useful when the system needs to ensure that no real identity is overlooked, even when disguises obscure significant parts of the face. In practical terms, this shows the system's robustness to various types of facial occlusions, such as sunglasses, masks, or other obstructions.

With an F1-score of 99.27%, the model demonstrates a fine balance between avoiding false positives and capturing all true positives. The near-perfect F1-score signifies that the model's performance is consistent across all test cases and is not overly biased toward either precision or recall. This makes the model highly reliable in scenarios where both accuracy and completeness of identification are essential, such as law enforcement or forensic investigations.

An accuracy of 99.24% indicates that the model correctly identifies the faces (disguised or undisguised) in over 99%

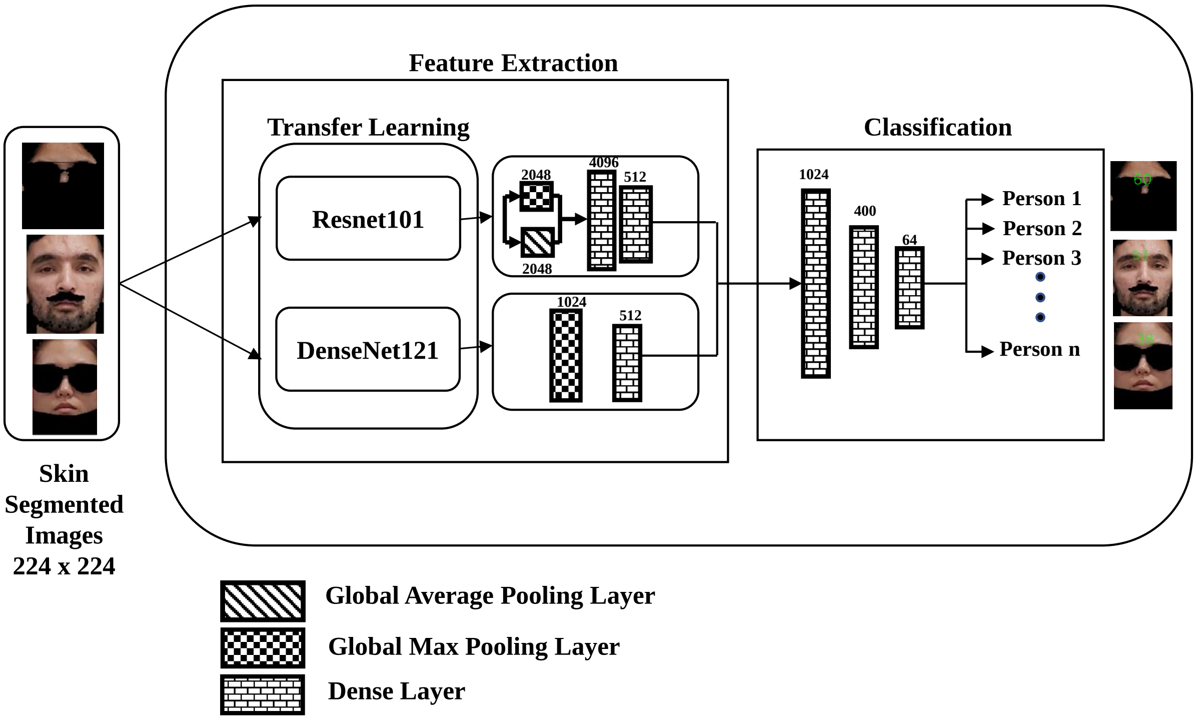


Fig. 4: Overall architecture of disguised face recognition

TABLE I: Summary of images with various disguises available in Sejong Face Dataset

	Accessories	# of Images	Gender	
			Male	Female
No Add-on	Natural Face	15	Yes	Yes
	Real Beard	10	Yes	No
Accessory Add-on	Cap	5	Yes	Yes
	Scarf	5	Yes	Yes
	Glasses	5	Yes	Yes
	Mask	5	Yes	Yes
	Makeup	5	No	Yes
Fake Add-on	Wig	10	Yes	Yes
	Fake Beard	5	Yes	No
	Fake Mustache	5	Yes	No
Combination Add-on	Wig - Glasses	5	No	Yes
	Cap - Scarf	5	No	Yes
	Glasses - Scarf	5	Yes	Yes
	Glasses - Mask	5	Yes	Yes
	Fake Beard - Cap	5	Yes	No
	Fake Beard - Glasses	5	Yes	No

of the test cases. This is a robust performance metric for real-world use, where high accuracy is often synonymous with reliability. In surveillance systems or other large-scale applications, this level of accuracy would ensure smooth operation with minimal errors, reducing the need for manual intervention or correction.

The AUC score of 99.61% signifies that the model has near-perfect separation between different individuals. This means that even subtle differences in facial features (despite disguises) are captured effectively by the model. High AUC is crucial for handling situations with high variability, where faces may appear very similar or where individuals use complex disguises. In essence, the model can confidently distinguish between a variety of individuals, even when the disguise complicates the identification process.

The loss graph in Figure 5a shows the decrease in the error rate of the model during the training and validation phases. As the number of epochs increases, the loss value decreases, indicating that the model is learning and minimizing its error over time. Typically, after a certain number of epochs, the loss stabilizes, which reflects that the model has reached an optimal state. The convergence of the loss graph demonstrates that the proposed model combination has successfully reduced the error to a minimal value, allowing for high performance.

The accuracy graph in Figure 5b represents how well the model classifies data correctly over time. The accuracy increases steadily as the model is trained, eventually plateauing when the model has fully learned from the data. A consistently high training and validation accuracy, such as the 99.24% accuracy seen in the evaluation, confirms the robustness of the proposed combination.

The AUC (Area Under the ROC Curve) is a crucial metric for evaluating the performance of the model across different classes is shown in Figure 5c. For each class, the AUC measures the model's ability to distinguish between positive and negative instances. The AUC value is remarkably high, indicating that the model has an exceptional ability to classify instances correctly across all classes. By plotting the ROC curve for each class, the graph highlights the trade-off between the true positive rate (TPR) and false positive rate (FPR) across varying threshold values. The near-perfect AUC score shows that the model achieves an optimal balance for all the classes.

We also present qualitative insights obtained from analyzing the outcomes produced by our disguised face recognition model on the test dataset. Through qualitative analysis, our goal was to delve into the characteristics of the model's predictions, uncover patterns or trends in its performance, and gain insights into the challenges encountered when identifying disguised faces. The test dataset encompassed a varied collection of images featuring individuals sporting a range of disguises, such as alterations in facial hair, makeup, accessories, and occlusions. Each image was meticulously selected to mirror authentic scenarios and the complexities faced in disguised face recognition endeavors.

A key observation drawn from the qualitative analysis was the model's proficiency in accurately detecting disguised faces under certain circumstances. When disguises were subtle or well-defined, the model displayed high accuracy and confidence in its predictions. However, despite its overall effectiveness, the model encountered difficulties in recognizing faces with more intricate disguises or alterations. Instances where facial features were heavily obscured or manipulated presented challenges for the model, resulting in lower accuracy rates. These observations highlight the imperative for ongoing research and development aimed at enhancing the resilience and dependability of disguised face recognition systems. Figures 6 and 7 illustrates both the qualitative results obtained from the proposed model and the misclassified predictions respectively.

V. Ablation study

A. Batch Size Impact in the Proposed Model

Batch size affects both the convergence rate of the model and its generalization performance. In this study, we tested batch sizes of 4, 8, 16, 32, and 64, providing insights into how batch size influences model performance. Figure 8 shows the performance of the proposed model with varying batch sizes.

Batch size 4 provides excellent performance across all metrics. The high precision (97.01%) and recall (96.64%) suggest that the model balances false positives and false negatives very well. The accuracy of 96.42% is one of the best results, indicating that the model can generalize well even with a smaller batch size. Small batch sizes like 4 allow for more frequent updates to the model, which can lead to better gradient estimation.

The model's performance drops significantly when the batch size is increased to 8. Precision and recall both decrease, leading to a relatively low F1-score (82.37%) and a large gap in accuracy (77.63%) compared to smaller batch sizes. The increase in batch size may be introducing noise or instability during the optimization process, resulting in degraded performance.

With batch size 16, the model recovers from the performance drop seen with batch size 8. Precision, recall, and F1-score all improve, and the accuracy rises to 91.54%.The model appears to perform well with a moderate batch size like 16, balancing computation time with good performance. This suggests that a batch size in this range allows the model to update weights more stably.

Batch size 32 yields performance metrics close to those of batch size 4, with precision, recall, and F1-score all above 96%. The accuracy of 96.36% shows that the model is capable of generalizing well even with a larger batch size.

With batch size 64, performance slightly declines again. While precision and recall remain reasonably high, the F1-score (92.04%) and accuracy (91.16%) are lower than with smaller batch sizes. Larger batch sizes like 64 may lead to less frequent updates and less variability in the learning process, potentially causing the model to converge more slowly.

The study shows that the proposed model performs best with smaller to moderate batch sizes (4 and 32), where both performance and generalization are maximized. Increasing the batch size too much (e.g., 64) results in a slight performance decline and batch size 8 demonstrates significantly poorer performance. Therefore, smaller batch sizes provide better control over model updates and help achieve superior classification results for this specific task.

B. Performance Comparison and Ensemble Selection of Pretrained Models for Disguised Face Recognition

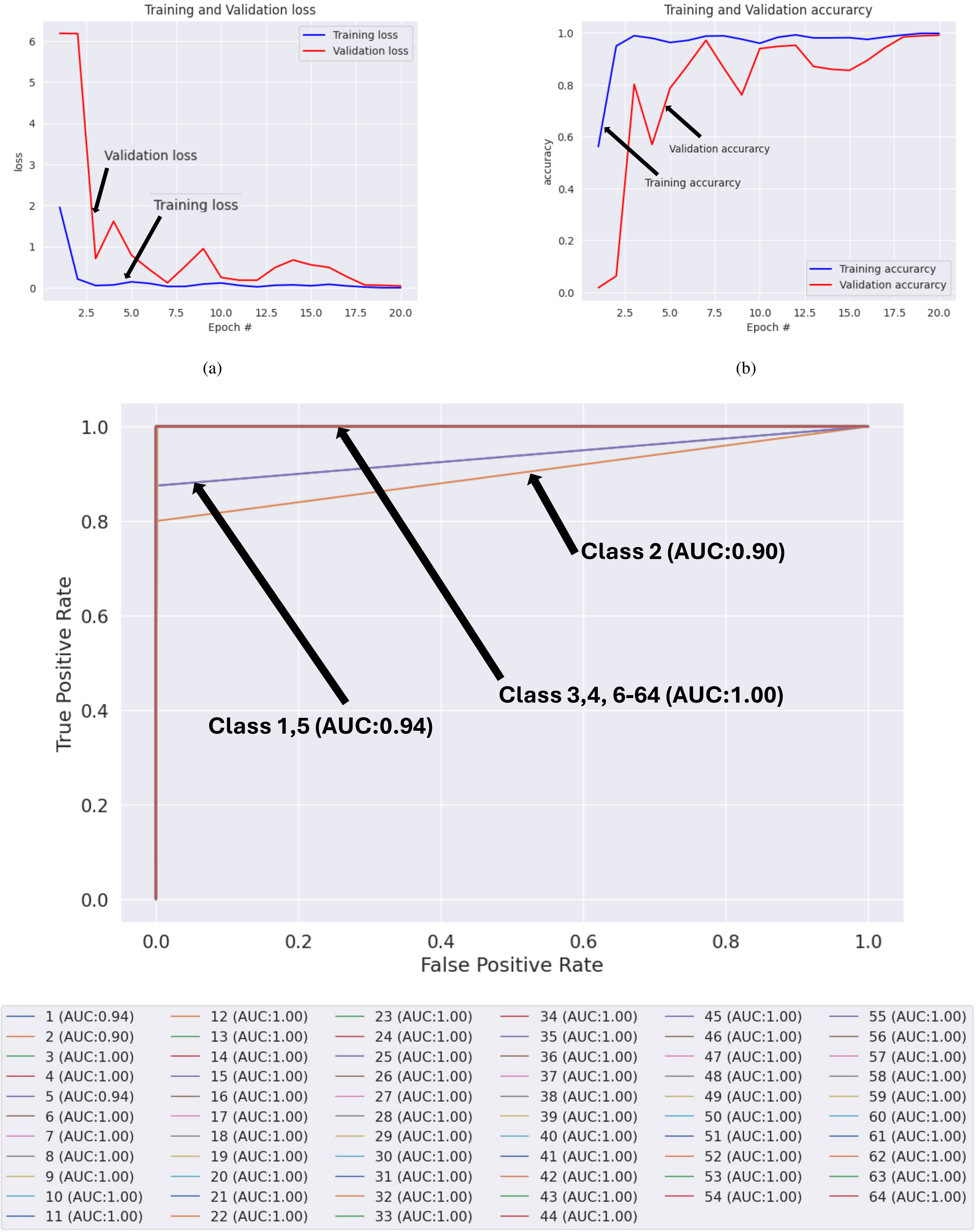
In this study, we compare several deep learning architectures—Densenet121, Inception, ResNet50 [8], Xception, ResNet152 [8], ResNet101 [8], SE-ResNeXt50 [18], [19], and VGG16 —on key performance metrics such as precision, recall, F1-score, and accuracy as depicted in Figure 9. The study highlights the contribution of each model architecture in the context of the classification task.

Densenet121 [20] achieves solid performance across all metrics, showing balanced precision and recall, which results in a high F1-score. It demonstrates strong feature extraction capability but is slightly outperformed by Inception and SE-ResNeXt50 [18], [19] in terms of accuracy.

Inception excels in this task, leading in accuracy (99.06%), which indicates exceptional generalization ability. The high precision and recall values show its effectiveness in minimizing both false positives and false negatives, resulting in the best overall performance among the models.

ResNet50 [8] shows comparatively lower performance, especially in recall, which leads to a lower F1-score. This suggests that ResNet50 [8] may have struggled with the complexity of the classification task and missed some correct predictions. The accuracy of 86.46% reflects its suboptimal performance for this task.

Xception performs well, though not at the level of Inception or Densenet121. It strikes a good balance between precision and recall, leading to a relatively high F1-score. The architecture's performance suggests strong learning capabilities but slightly lower generalization power than Inception.



(a)

(b)

(c)

Fig. 5: Training and validation performance metrics for the proposed model (a) Loss graph (b) Accuracy graph (c) AUC graph for each class demonstrates near-perfect discrimination

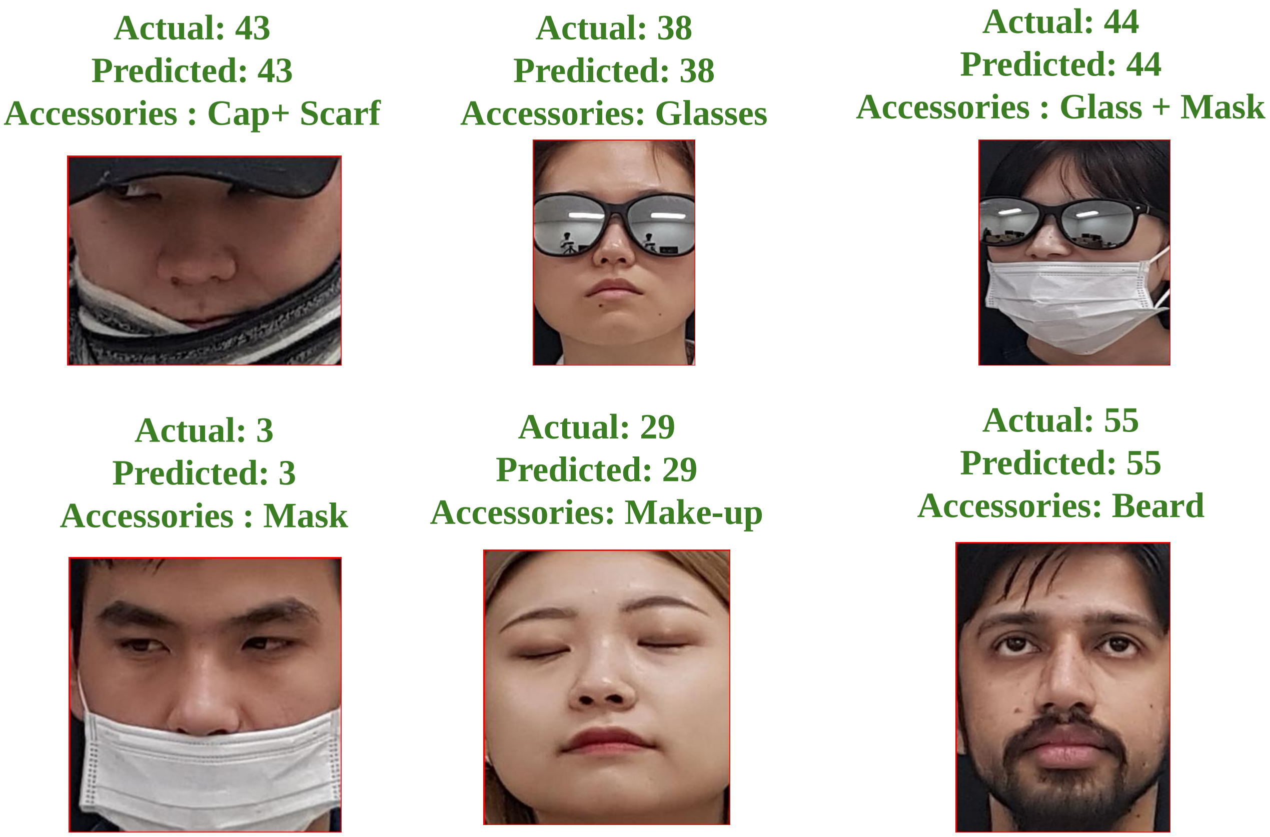


Fig. 6: Visualizations of predictions of disguised faces using the proposed model

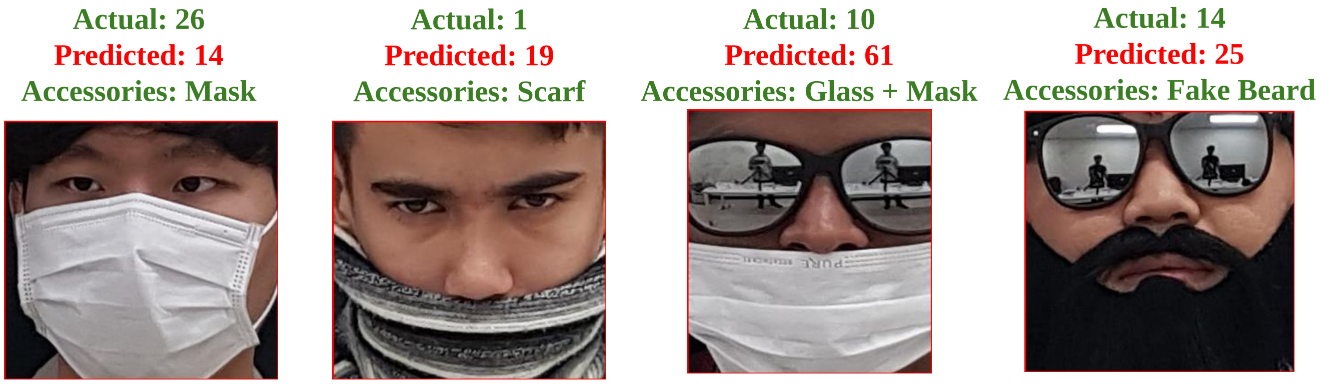


Fig. 7: Visualizations of misclassifications of disguised faces using the proposed model

ResNet152 [8] performs below average compared to the other architectures. Both its precision and recall are lower, indicating that the deeper ResNet architecture may overfit the data or have difficulty converging efficiently. Its performance suggests that simply increasing model depth does not necessarily lead to better results.

ResNet101 [8] provides a slight improvement over ResNet50 [8], with better recall and F1-score. However, its performance is still behind Densenet121, Inception, and SE-ResNeXt50 [18], [19]. This suggests that although ResNet101 [8] provides better feature extraction than ResNet50 [8], it is not the optimal architecture for this task.

SE-ResNeXt50 demonstrates very strong performance, especially in terms of precision and recall, which are both close to 98%. The F1-score and accuracy (97.74%) indicate that this model is highly effective, second only to Inception. Its performance shows the benefit of combining ResNeXt's architecture with Squeeze-and-Excitation (SE) blocks for better feature recalibration.

VGG16 demonstrates adequate but not outstanding performance across these metrics, with strengths in recall and overall accuracy, but slight weaknesses in precision. VGG16 achieves a precision of 75.38%, meaning about 75% of its positive predictions are accurate. This shows that VGG16 struggles a bit with false positives in this task. With a recall of 79.07%, it is reasonably effective at identifying positive samples, but it still misses some positive instances (false negatives).

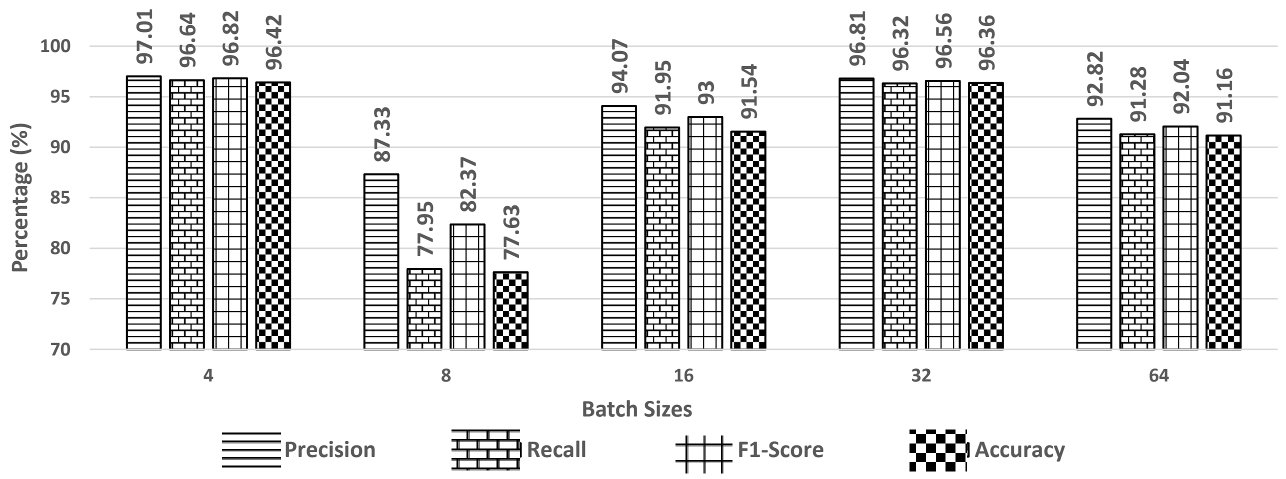


Fig. 8: Impact of Varying Batch Sizes on Model Performance Across Evaluation Metrics: Precision, recall, F1-score, and accuracy are plotted for batch sizes of 4, 8, 16, 32, and 64.

The ablation study shows that model architecture significantly impacts performance. The study emphasizes the importance of considering not only depth but also architectural innovations like multi-scale feature extraction and channel recalibration when selecting models for specific tasks.

C. Ensemble Models for Disguised Face Recognition

To further enhance the performance of our proposed model for disguised face recognition, we conducted an ablation study by systematically experimenting with various combinations of pre-trained models. The objective was to identify the best ensemble of two models to deliver superior accuracy while capturing the intricate features of disguised faces.

The ablation study focused on three high-performing pre-trained models: DenseNet121, Inception, and Xception. These models were selected due to their proven effectiveness in complex visual recognition tasks. We evaluated the performance of these models in pairs, combining their strengths to determine which ensemble model yields the best results across multiple performance metrics, including precision, recall, F1-score, accuracy, and AUC (Area Under the Curve) and are depicted in the Figure 10.

DenseNet121 and ResNet50 [8] achieved a solid performance, with an accuracy of 96.42%, precision of 97.01%, recall of 96.64%, and F1-score of 96.82%. The AUC was 98.29%, indicating high discrimination ability. However, while the performance was notable, it was surpassed by other ensembles in the study.

DenseNet121 and Inception resulted in slightly lower performance, with an accuracy of 95.86%, precision of 97.3%, recall of 95.74%, and an F1-score of 96.51%. The AUC of 97.84% showed reasonable capability, though it fell short compared to other pairs.

DenseNet121 and Xception pairing demonstrated a significant improvement, with an accuracy of 97.74%, precision of 98.34%, recall of 97.83%, and an F1-score of 98.08%. The AUC value of 98.9% made this one of the top-performing ensembles, showcasing its ability to effectively identify disguised faces.

DenseNet121 and ResNet152 [8] ensemble showed a performance dip, with accuracy dropping to 93.42%, precision to 94.67%, recall to 93.6%, and an F1-score of 94.13%. The AUC of 96.74% confirmed that this combination was less effective in this task.

DenseNet121 and ResNet101 [8] combination emerged as the best-performing model across all metrics. It achieved an impressive accuracy of 99.24%, precision of 99.32%, recall of 99.23%, and an F1-score of 99.27%. Additionally, the AUC of 99.61% set this combination apart as the most robust model for disguised face recognition.

Inception and ResNet152 [8] combination yielded decent results, with an accuracy of 93.79%, precision of 94.64%, recall of 93.89%, and an F1-score of 94.26%. The AUC was 96.89%, placing this ensemble in the lower-performing category compared to others.

Inception and ResNet101 [8] ensemble showed a substantial drop in performance, with accuracy at 87.4%, precision of 90.11%, recall of 86.49%, and an F1-score of 88.26%. The AUC was 93.14%, indicating weaker performance.

Inception and Xception combination displayed competitive performance, achieving an accuracy of 96.42%, precision of 97.17%, recall of 96.49%, and an F1-score of 96.83%. The AUC of 98.21% ranked it as one of the better-performing ensembles.

With accuracy of 87.96%, precision of 91.7%, recall of 87.73%, and an F1-score of 89.67%, Inception and ResNet50 [8] pairing did not perform as well. The AUC of 93.77% confirmed its relatively weaker capability.

Xception and ResNet152 [8] combination performed decently, with an accuracy of 95.3%, precision of 95.59%, recall of 95.08%, and an F1-score of 95.33%. The AUC of 97.5% suggested strong but not top-tier performance.

Xception and ResNet101 [8] combination showed strong results, with an accuracy of 95.3%, precision of 96.29%, recall of 95.39%, and an F1-score of 95.84%. The AUC of 97.65% positioned it among the higher-performing ensembles.

The results for Xception and ResNet50 [8] pair were similar to other top-performing combinations, with an accu-

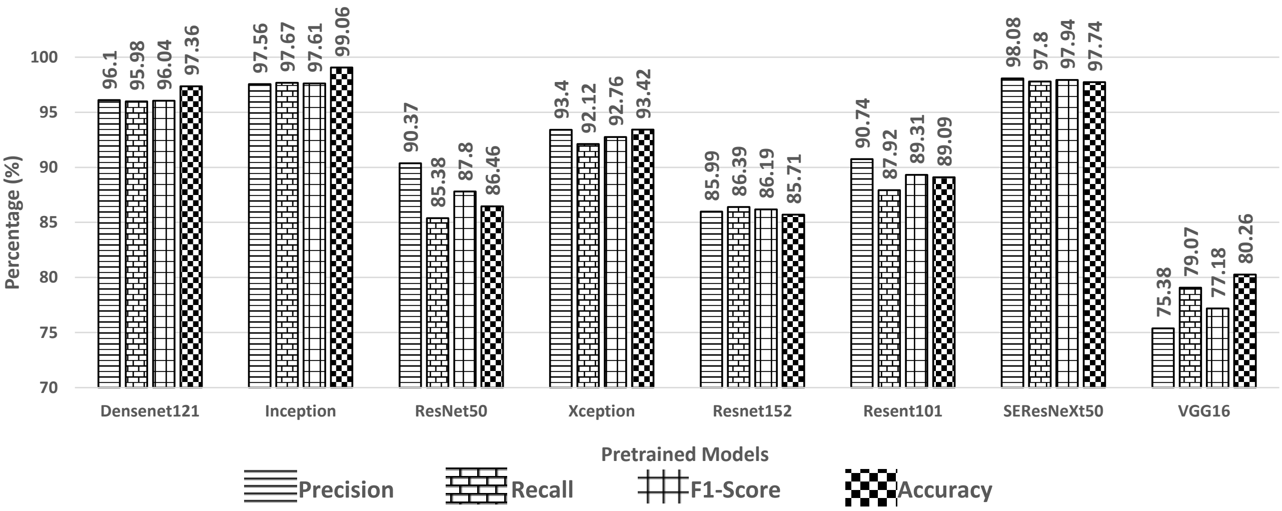


Fig. 9: Comparative Evaluation of Disguised Face Recognition Models with the Pre-trained Models

racy of 95.11%, precision of 96.02%, recall of 95.08%, and an F1-score of 95.55%. The AUC of 97.5% showed solid performance but fell short of the top results.

The ablation study clearly demonstrated that the DenseNet121 and ResNet101 [8] combination outperformed all other ensembles, with superior results across all performance metrics. The combination of DenseNet121 and Xception also showed exceptional performance and emerged as a close second. These findings underscore the importance of carefully selecting and combining pre-trained models to leverage their unique strengths for complex feature extraction tasks, such as disguised face recognition.

This ablation study offers valuable insights into how the ensemble of different pre-trained models can substantially improve model accuracy and overall performance. The models that perform best, identified here, DenseNet121 and ResNet101 [8], and DenseNet121 and Xception, are the best choices for the proposed task.

D. Comparative Evaluation of Loss Functions in StealthFace for Disguised Face Recognition

In this section, we compare the performance of the proposed ensemble model with three state-of-the-art face recognition models: CosFace [21], ArcFace [22], and SphereFace [23]. CosFace, ArcFace, and SphereFace each introduce unique modifications to the loss function to enhance the discriminative power and separability of the feature space. These modifications, such as angular margins and non-linear transformations, are designed to improve the performance of the models in recognizing disguised faces. These models have shown promising results in enhancing face recognition accuracy by incorporating angular margin-based loss functions. The proposed model achieved an accuracy of 99.28% and a validation loss of 0.0473 as shown in Figure 11. The exceptional accuracy and significantly low validation loss demonstrate the superiority of the proposed model. It surpasses all other models in terms of accuracy, indicating its ability to correctly identify all disguised faces in the dataset. The extremely low validation loss suggests that the model's predictions have minimal error or uncertainty, making it highly reliable for disguised face recognition.

E. Impact of Data Augmentation on Disguised Face Recognition

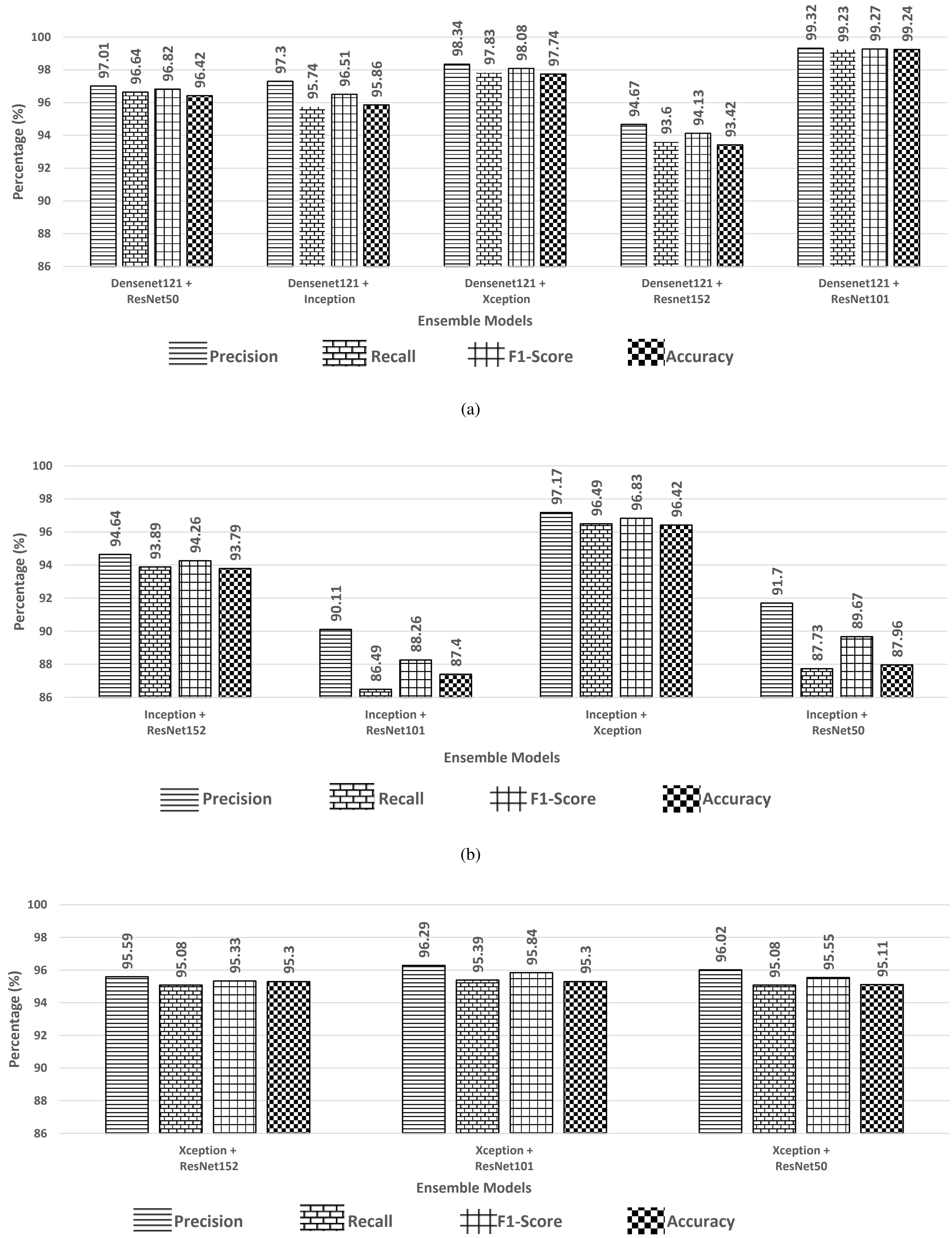
To thoroughly analyze the effect of data augmentation techniques on the classification performance of disguised faces, we conducted an ablation study comparing the model's results when trained with and without augmentation. The augmentation techniques applied in the experimental setup include horizontal flip, rotation, brightness adjustment, and zooming. These augmentations were designed to introduce diversity into the training dataset, enabling the model to generalize across unseen data variations.

The comparison between the results with and without augmentation is visually depicted in Figure 12. It can be observed that, while the augmentation techniques slightly improve the robustness of the model, the proposed method without augmentation yields superior results across all evaluation metrics.

With augmentation, the model achieved a precision of 91.98%, recall of 90.65%, and F1-Score of 90%, with an overall accuracy of 90.6% and an AUC of 95.25%. This suggests that augmentations introduced variability to the data, allowing the model to learn more generalized patterns, albeit at a slightly lower performance.

When trained on the original data set without augmentation, the proposed method achieved significantly better results. The precision increased to 97.01%, recall to 96.64%, and F1-Score to 96.82%, with an accuracy of 96.42% and an AUC of 98.29%. These improvements can be attributed to the robustness of the proposed model architecture and its ability to effectively capture discriminative features of disguised faces.

The ablation study demonstrates that while augmentations improve generalization in some scenarios, the proposed architecture is capable of achieving exceptional performance without reliance on synthetic data augmentation techniques. This highlights the importance of model design and feature

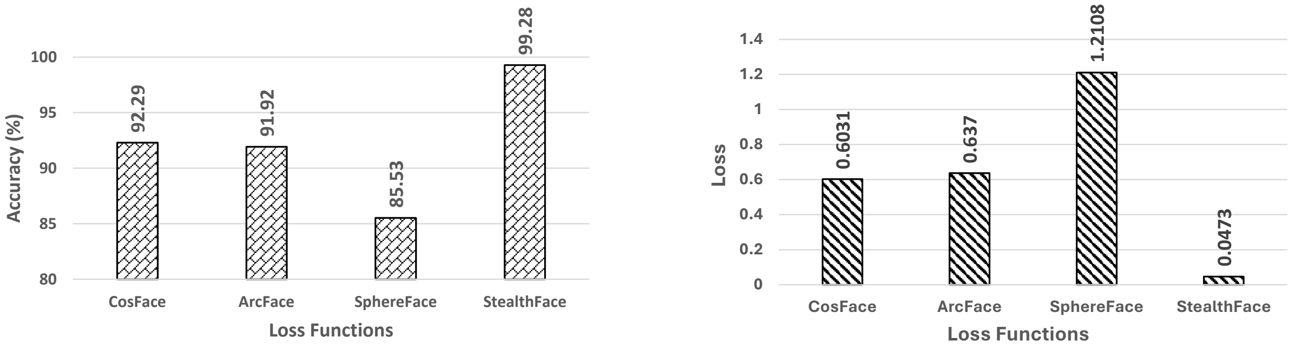


(a)

(b)

(c)

Fig. 10: Performance Comparison of Various Ensemble Models for Disguised Face Recognition. Figures (a), (b), and (c) present precision, recall, F1-score, and accuracy for different ensemble combinations of pre-trained models. These results demonstrate the impact of selecting appropriate model pairs to enhance performance in recognizing disguised faces.



(a)

(b)

Fig. 11: Visualization of comparative Evaluation of Loss Functions in StealthFace for Disguised Face Recognition (a) Accuracy Values (b) Loss Values

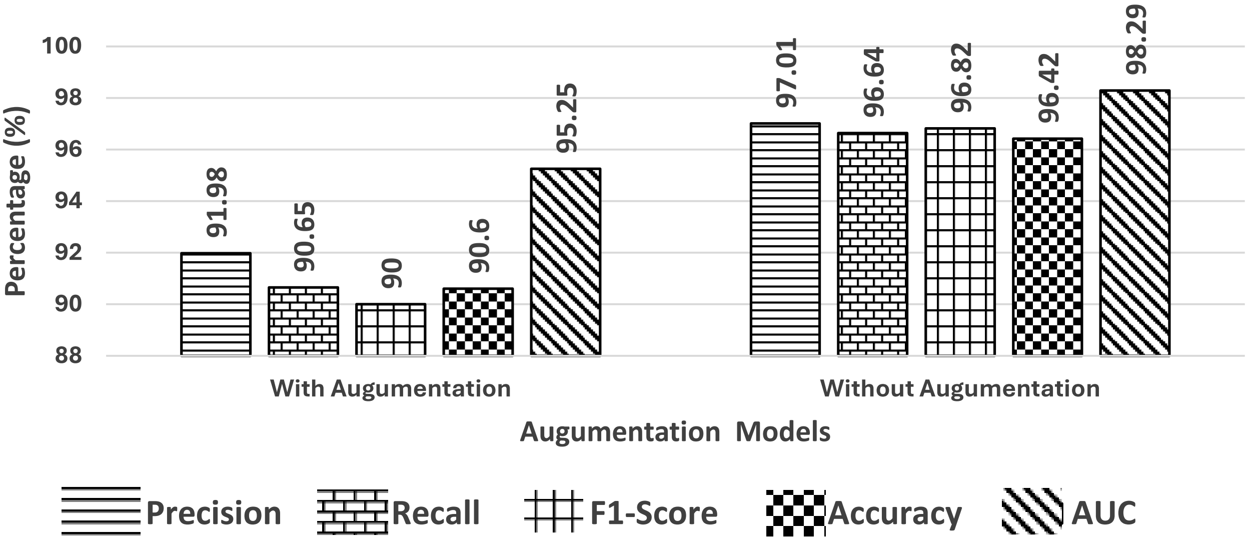


Fig. 12: Comparative performance analysis of the model with augmentation and the proposed method (without augmentation).

learning strategies in addressing the challenges of disguised face classification.

VI. Impact of Regularization Techniques on Disguised Face Recognition

To understand the influence of different regularization techniques on the performance of the disguised face classification model, we conducted a systematic ablation study. Regularization plays a crucial role in improving generalization by mitigating overfitting and enhancing the robustness of deep learning models. The following regularization methods were evaluated:

- **L1 Regularization (LASSO)**: Adds a penalty proportional to the sum of the absolute values of weights, encouraging sparsity in the model.
- **L2 Regularization (Ridge)**: Adds a penalty proportional to the sum of the squared weights, which helps prevent large weight values and enhances stability.
- **Dropout**: Randomly deactivates a fraction of neurons during each forward pass, forcing the model to learn robust features.
- **L1 + Dropout**: Combines L1 regularization and dropout to encourage both sparsity and generalization.
- **L2 + Dropout**: Combines L2 regularization and dropout for better regularization of weights and neuron activation.
- **Proposed Method**: Our model baseline architecture without additional combined regularization strategies.

Figure 13 visually compares the performance of different regularization techniques in terms of Precision, Recall, F1-Score, Accuracy, and AUC and emphasizes the need for carefully selecting regularization techniques to balance generalization and stability in disguised face recognition tasks.

L1 regularization achieved a precision of 98.43%, recall of 98.13%, and an AUC of 99.05%, indicating good performance. The sparsity enforced by L1 helps reduce unnecessary weights, improving precision. However, the slight drop in recall suggests that the strict sparsity constraint may cause the model to ignore some relevant features, leading to missed predictions.

L2 regularization delivered the best overall performance, with an accuracy of 98.68% and an AUC of 99.32%. The model demonstrated high precision (98.83%) and recall

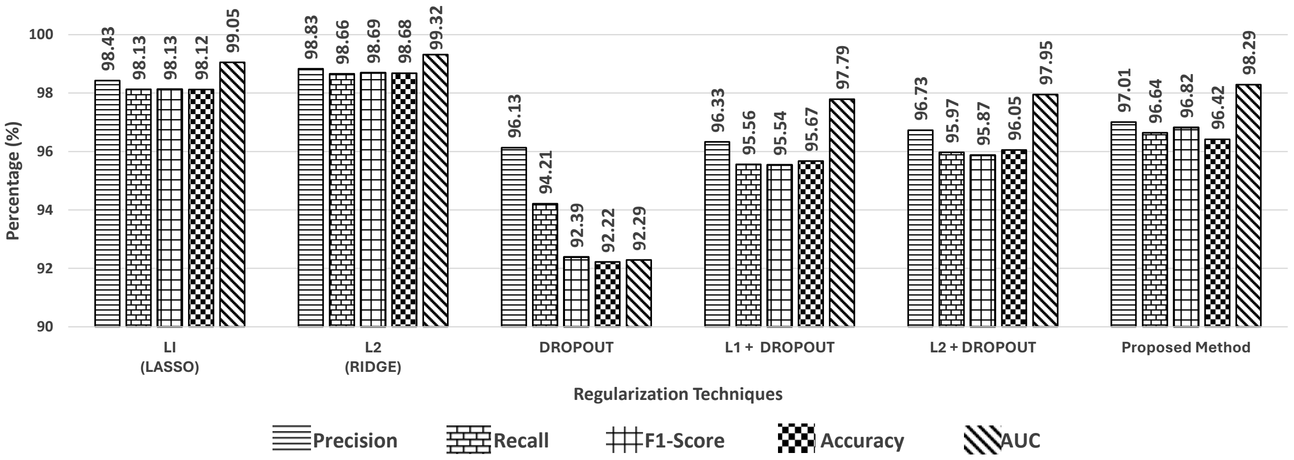


Fig. 13: Comparative Performance of Regularization Techniques.

(98.66%), resulting in a strong F1-Score. This highlights the stability of L2 regularization in preventing overfitting by penalizing large weights, leading to balanced feature learning.

Dropout alone resulted in the lowest performance across all metrics, with an accuracy of 92.22% and an AUC of 92.29%. The precision (96.13%) and recall (94.21%) were significantly reduced. This indicates that relying solely on dropout can lead to excessive neuron deactivation, which may hinder the learning of complex feature representations required for disguised face classification.

Combining L1 regularization with dropout improved performance compared to dropout alone, achieving an accuracy of 95.67% and an AUC of 97.79%. The precision and recall values (96.33% and 95.56%, respectively) highlight the benefit of introducing sparsity alongside neuron regularization. The improvement shows that this combination helps the model generalize better without overfitting.

The combination of L2 regularization and dropout further improved performance to an accuracy of 96.05% and an AUC of 97.95%. Precision (96.73%) and recall (95.97%) also increased, suggesting that the weight penalization introduced by L2 complements dropout effectively. This combination strikes a good balance between stability and generalization.

The proposed method achieved competitive performance, with an accuracy of 96.42%, AUC of 98.29%, and an F1-Score of 96.82%. Although it did not surpass L2 regularization, the results demonstrate the robustness of the model architecture and optimization strategies employed in the proposed method. The slight improvement in recall (96.64%) compared to other combined regularization techniques highlights its effectiveness in minimizing misclassifications.

Among all techniques, L2 regularization delivered the highest performance across all metrics. This emphasizes its ability to prevent overfitting by ensuring smoother and more stable weight updates. Dropout, when used independently, led to significant performance degradation, highlighting that excessive regularization can disrupt learning, particularly for complex disguised face datasets. Combining L1 or L2 with dropout resulted in substantial improvements over dropout alone, suggesting that regularizing both weights and neuron activations provides complementary benefits. The proposed method showcased a well-balanced performance, demonstrating competitive results without the need for additional combined regularization. This validates the model's inherent strength and optimization strategy.

VII. Impact of Different Optimizers on Model Performance

In addition to regularization techniques, we conducted an ablation study to investigate the impact of various optimizers on the performance of the disguised face classification model. Optimizers play a critical role in determining how the model updates weights during training, influencing convergence speed, stability, and overall performance. We evaluated six optimizers: SGD, RMSPROP, ADAMW, ADAGRAD, ADAMAX, and ADAM (Proposed Method).

Figure 14 compares the precision, recall, F1 score, accuracy and AUC values in SGD, RMSPROP, ADAMW, ADAGRAD, ADAMAX, and ADAM. ADAMAX achieves the highest performance, while ADAM (Proposed Method) remains highly competitive.

SGD achieved the lowest performance among modern optimizers, with an accuracy of 82.33% and an AUC of 90.4%. While SGD is a foundational optimizer, its slower convergence and susceptibility to local minima explain the reduced precision (83.49%) and recall (81.09%). It lacks the adaptive learning rate adjustments required for complex tasks such as disguised face recognition.

RMSPROP delivered significantly better results, with an accuracy of 96.61% and an AUC of 98.34%. Precision (96.52%) and recall (96.73%) highlight its ability to adapt the learning rate for each parameter. RMSPROP effectively balances convergence and generalization, making it suitable for this classification task.

ADAMW achieved moderate performance, with an accuracy of 89.09% and an AUC of 94.48%. Although it performed better than SGD and ADAGRAD, its recall (89. 15%) lagged behind the RMSPROP and ADAM-based optimizers. This indicates that ADAMW may not handle the complexity of the dataset as efficiently, despite its weight-decay mechanism.

ADAGRAD yielded the lowest results overall, with an accuracy of 66.54% and an AUC of 82.42%. The optimizer struggles to maintain long-term training learning rates due to its aggressive rate reduction, leading to poor generalization.

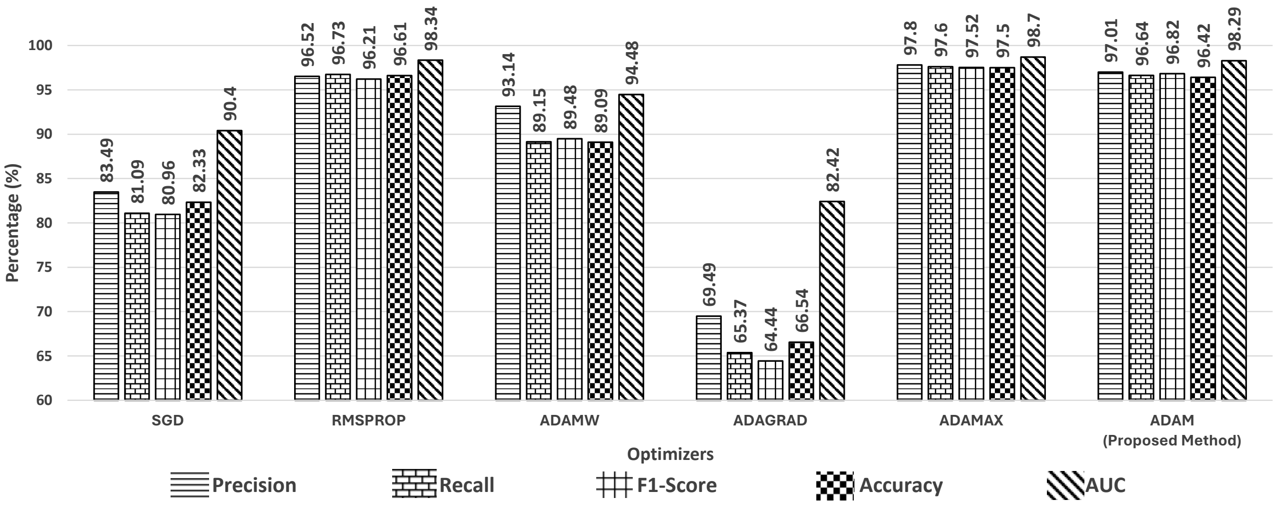


Fig. 14: Comparative Performance of Different Optimizers for Disguised Face Recognition.

The precision (69. 49%) and the recall (65. 37%) further emphasize its limited utility for this problem.

ADAMAX demonstrated the best overall performance, with an accuracy of 97.5%, and a precision of 97. 8% and an AUC of 98.7%. This optimizer, a variant of ADAM, performs particularly well when gradients are sparse or when models require higher stability during training. Its ability to adaptively adjust learning rates ensures efficient convergence and robust generalization.

The ADAM optimizer (Proposed Method) achieved strong performance, with an accuracy of 96.42% and an AUC of 98.29%. Precision (97.01%) and recall (96.64%) reflect its ability to balance adaptive learning rates and momentum, leading to stable training and robust performance. Although ADAMAX slightly outperformed ADAM, the proposed method provides near-optimal results with reliable convergence.

Both SGD and ADAGRAD performed poorly compared to modern optimizers. The slower convergence of SGD and ADAGRAD's diminishing learning rates make them unsuitable for complex datasets requiring high generalization. RMSPROP and ADAM showed significant improvements, demonstrating their ability to adapt learning rates during training. These optimizers are efficient for tasks that involve large, dynamic datasets. ADAMAX delivered the best performance, indicating its suitability for disguised face classification. Its robustness to sparse gradients and improved stability during weight updates ensure superior generalization. The ADAM optimizer, used in the proposed method, performed very close to ADAMAX, confirming its reliability and robustness for this problem. The study underscores the limitations of traditional optimizers such as SGD and ADAGRAD, while emphasizing the advantages of adaptive optimizers such as RMSPROP and ADAM-based variants.

VIII. Benchmarking the StealthFace against State-of-the-Art Techniques:A Comparative Analysis

In this section, we compare the performance of our proposed model, StealthFace with state-of-the-art techniques in the field of disguised face recognition. We evaluate the effectiveness and robustness of our model by considering accuracy as the key performance metric. Table II presents the performance of the proposed disguised face recognition model along with the baseline results of the Sejong face dataset[6]. The efficiency of the suggested system can be observed when it achieves state-of-the-art performance with a Precision, Recall, F1-score, and Accuracy of 99.28%. The comparison with state-of-the-art techniques not only showcases the advancements made in the field of disguised face recognition but also reinforces the significance and effectiveness of our proposed model. It establishes our model as a reliable and state-of-the-art solution for addressing the challenges of disguised face recognition, paving the way for enhanced security and improved authentication systems.

TABLE II: Comparison of Disguised Face Verification Accuracy: State-of-the-Art vs. StealthFace

Models	**Accuracy(%)**
[13]	92.6
StealthFace	99.28

IX. Conclusion

In this work, we presented a novel approach to the problem of face recognition under disguise, leveraging deep ensemble neural networks and transfer learning. By integrating pretrained models such as DenseNet121 and ResNet101 [8], and incorporating feature extraction techniques, we significantly improved the model's ability to detect and classify disguised faces with high precision, recall, and accuracy. Our model demonstrated robust performance across multiple evaluation metrics, achieving an impressive accuracy of 99.24

The ensemble architecture, which combined DenseNet121 with ResNet101 [8], outperformed other ensemble combinations, highlighting the importance of selecting complementary models to maximize recognition accuracy in challenging scenarios. The comparative analysis with other state-of-the-art approaches further validated the efficacy of our method in handling occluded and partially disguised faces. The ablation study provided insights into the performance contribution of different model combinations, confirming that ensemble learning significantly enhances recognition capabilities, especially in complex face recognition tasks.

Moreover, our model's adaptability makes it suitable for various real-world applications, including security, surveillance, and identity verification. Future work could explore extending the ensemble strategy to include more diverse networks and applying the proposed approach to other challenging image recognition tasks, such as age progression and expression-invariant face recognition. Additionally, further research can incorporate advanced techniques for handling variations in lighting, pose, and expression, to further improve model robustness.

”BibTeXtran.bst”)

REFERENCES

[1] Tejas Indulal Dhamecha, Richa Singh, Mayank Vatsa, and Ajay Kumar. Recognizing disguised faces: Human and machine evaluation. *PloS one*, 9(7):e99212, 2014.

[2] Maneet Singh, Mohit Chawla, Richa Singh, Mayank Vatsa, and Rama Chellappa. Disguised faces in the wild 2019. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.

[3] Maneet Singh, Richa Singh, Mayank Vatsa, Nalini K Ratha, and Rama Chellappa. Recognizing disguised faces in the wild. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(2):97–108, 2019.

[4] Gaurav Goswami, Mayank Vatsa, and Richa Singh. Face verification via learned representation on feature-rich video frames. *IEEE Transactions on Information Forensics and Security*, 12(7):1686–1698, 2017.

[5] Yueqi Duan, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Context-aware local binary feature learning for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5):1139–1153, 2017.

[6] Kaipeng Zhang, Ya-Liang Chang, and Winston Hsu. Deep disguised faces recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 32–36, 2018.

[7] Ankan Bansal, Rajeev Ranjan, Carlos D Castillo, and Rama Chellappa. Deep features for recognizing disguised faces in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 10–16, 2018.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[9] Naman Kohli, Daksha Yadav, and Afzel Noore. Face verification with disguise variations via deep disguise recognizer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 17–24, 2018.

[10] Skand Vishwanath Peri and Abhinav Dhall. Disguisenet: A contrastive approach for disguised face verification in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 25–256. IEEE, 2018.

[11] Anshuman Suri, Mayank Vatsa, and Richa Singh. A2-link: recognizing disguised faces via active learning and adversarial noise based inter-domain knowledge. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):326–336, 2020.

[12] Maneet Singh, Shruti Nagpal, Richa Singh, and Mayank Vatsa. Disguise resilient face verification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[13] Usman Cheema and Seungbin Moon. Sejong face database: A multi-modal disguise face database. *Computer Vision and Image Understanding*, 208:103218, 2021.

[14] Muhammad Junaid Khan, Muhammad Jaleed Khan, Adil Masood Siddiqui, and Khurram Khurshid. An automated and efficient convolutional architecture for disguise-invariant face recognition using noise-based data augmentation and deep transfer learning. *The Visual Computer*, 38(2):509–523, 2022.

[15] Baojin Huang, Zhongyuan Wang, Kui Jiang, Qin Zou, Xin Tian, Tao Lu, and Zhen Han. Joint segmentation and identification feature learning for occlusion face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[16] Ankit Sharma, Neeru Jindal, Abhishek Thakur, Prashant Singh Rana, Bharat Garg, and Rajesh Mehta. Multimodal biometric for person identification using deep learning approach. *Wireless Personal Communications*, pages 1–21, 2022.

[17] Frerk Saxen and Ayoub Al-Hamadi. Color-based skin segmentation: An evaluation of the state of the art. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4467–4471. IEEE, 2014.

[18] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1492–1500, 2017.

[19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[20] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[21] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[22] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[23] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recaognition*, pages 212–220, 2017.