IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 August 2024, accepted 16 September 2024, date of publication 24 September 2024, date of current version 7 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3467151

RESEARCH ARTICLE

A Novel Approach to Enhancing Multi-Modal Facial Recognition: Integrating Convolutional Neural Networks, Principal Component Analysis, and Sequential Neural Networks

MOHAMED ABDUL-AL[1], GEORGE KUMI KYEREMEH[1], RAMI QAHWAJI[1], (Senior Member, IEEE), NAZAR T. ALI[2], (Senior Member, IEEE), AND RAED A. ABD-ALHAMEED[1,3], (Senior Member, IEEE)

[1]Faculty of Engineering and Digital Technologies, University of Bradford, BD7 1DP Bradford, U.K.
[2]Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
[3]Al-Farqadein University College, Basrah 61004, Iraq

Corresponding authors: Mohamed Abdul-Al (m.abdul-al@bradford.ac.uk) and Raed A. Abd-Alhameed (r.a.a.abd@bradford.ac.uk)

This work was supported by European Union's Horizon-Marie Skłodowska-Curie Actions (MSCA)-RISE-2019-2024, Marie Skłodowska-Curie, Research and Innovation Staff Exchange (RISE), titled: Secure and Wireless Multimodal Biometric Scanning Device for Passenger Verification Targeting Land and Sea Border Control.

ABSTRACT Facial recognition technology is crucial for precise and rapid identity verification and security. This research delves into advancements in facial recognition technology for verification purposes, employing a combination of convolutional neural networks (CNNs), principal component analysis (PCA), and sequential neural networks. Unlike identification, our focus is on verifying an individual's identity, that is a critical distinction in the context of security applications. Our goal is to enhance the efficacy and precision of face verification using several imaging modalities, including thermal, infrared, visible light, and a combination of visible and infrared. We use the pre-trained VGG16 model on the ImageNet dataset to extract features. Feature extraction is performed using the pre-trained VGG16 model on the ImageNet dataset, complemented by PCA for dimensionality reduction. We introduce a novel method, termed VGG16-PCA-NN, aimed at improving the precision of facial authentication. This method is validated using the Sejong Face Database, with a 70% training, 15% testing, and 15% validation split. While demonstrating a remarkable approaching 100% accuracy rate across visual and thermal modalities and a combined visible-infrared modality, it is crucial to note that these results are specific to our dataset and methodology. A comparison with existing approaches highlights the innovative aspect of our work, though variations in datasets and evaluation metrics necessitate cautious interpretation of comparative performance. Our study makes significant contributions to the biometrics and security fields by developing a robust and efficient facial authentication method. This method is designed to overcome challenges posed by environmental variations and physical obstructions, thereby enhancing reliability and performance in diverse conditions. The realised accuracy rates that the approach achieves across a variety of modalities demonstrate its promise for applications that use multi-modal data. This opens the door for the creation of biometric identification

The associate editor coordinating the review of this manuscript and approving it for publication was M. Sabarimalai Manikandan.

systems that are more trustworthy and secure. It is intended that the technology will be used in real-time settings for which the new modalities can be integrated in different situations.

INDEX TERMS Facial recognition (FR), convolutional neural networks (CNNs), principal component analysis (PCA), sequential neural networks (NNs), VGG16, visible (VIS), thermal (Th), infrared (IR), visible and infrared (VIS-IR), Sejong face database, receiver operating characteristic (ROC), accuracy, recall, precision, F1-score and rank-level fusion.

I. INTRODUCTION

The field of digital identity and security is undergoing tremendous changes, with facial recognition technology at the forefront. The increasing importance of strong, safe, and efficient identification techniques in the modern world positions this technology as a potential key solution. However, the effectiveness of facial recognition technology faces numerous challenges, including changes in illumination, variations in facial expressions, and physical obstacles.

This study explores various types of face recognition technologies, including visible, thermal, infrared, and combinations of visible and infrared images. Adopting a comprehensive approach aims to address traditional methods' limitations in challenging conditions such as poor lighting or atypical facial features. The focus is specifically on face verification, which involves confirming or denying an individual's claimed identity, as opposed to identification, which involves determining the identity of an unknown individual from a set of known identities.

Furthermore, the research contributes to the critical discussion on privacy and ethical concerns associated with facial recognition technology. It emphasizes the development of security technologies that do not infringe upon individual rights, advocating for an ethical and efficient use of facial recognition across different contexts. The multi-modal approach proposed by this study offers a framework to navigate these challenges with enhanced reliability and accuracy.

This study's novelty lies in its integrated approach that combines the strengths of Convolutional Neural Networks (CNNs), Principal Component Analysis (PCA), and Sequential Neural Networks (NNs). While CNNs are proficient in feature extraction, PCA aids in dimensionality reduction, enhancing computational efficiency, and Sequential NNs excel in classification tasks. The combination of these techniques creates a robust and efficient facial recognition system that addresses the limitations of traditional methods. This approach leverages transfer learning from pre-trained CNNs to extract hierarchical features, employs PCA for efficient feature representation, and utilises Sequential NNs for accurate classification. This integration has not been previously applied to the Sejong Face Database, marking a significant advancement in multi-modal facial recognition technology.

The CNNs are now often used to address classification issues. In these networks, features are learned automatically as the network progresses from lower to higher levels through consecutive layers. The term ''transfer learning'' refers to the re-cycling of a pre-trained machine learning model to tackle a new related problem. This technique helps improve generalisation and is especially helpful when insufficient training data exists.

The paper is structured as follows: The literature review is presented in Section II, the suggested approach is explained in detail in Section III, the experiments, results, and comments are detailed in Section IV, and the conclusion is presented in Section V.

II. LITERATURE SURVEY

The purpose of this section is to showcase studies that pertain to visible and cross-spectral face identification between visible, thermal, infrared, and combinations of visible and infrared facial pictures, with a particular emphasis on subject identification verification.

Because of the ubiquitous sensors found in everyday electronic devices like smartphones and laptops, face recognition has exploded in popularity, especially in the visible spectrum. In the visible light spectrum, current face recognition technologies are more effective than humans [1]. These systems are considered the best tools now available.

Taigman et al. [2] introduced DeepFace to generalise face representation across datasets. They presented a novel design and learning approach for deep neural networks, and their evaluation of the Faces in the Wild (LFW) dataset showed an accuracy of 97.35 percent.

Wen et al. [3] suggested centre loss as a new loss function to improve the neural network's deep learning features' discriminative strength. The deep features' intraclass distances can be minimised using this method. When tested on an LFW dataset, they reached a precision of 99.28%.

Several convolutional neural networks (CNNs) architectures were examined by Parkhi et al. [4], who also presented the massive dataset VGG Face. Virtual neural networks (VGG) had a major influence. It was demonstrated that the face recognition method's most important components are the training procedure and a dataset. Additionally, they demonstrated that, with the right training procedure, 98.95% accuracy is achievable.

To help CNNs learn angular-margin discriminative facial features, Liu et al. [5] suggested an angular Softmax loss. A technique for deep hypersphere embedding is presented in their work. Using their suggested A-Softmax loss, they learned a face representation with a 99.42% success rate.

In contrast, approaches based on the visible domain have received more attention and research, while thermal-visible

face recognition has lagged in terms of both popularity and performance. Among the first studies to tackle this problem, thermal-visible face recognition was the primary emphasis [6]. Histogram of Oriented Gradients and Partial Least Squares are the foundations of the approach put out by Hu et al. Four steps make up the suggested pre-processing stage: geometric normalisation, median filtering of dead pixels, difference-of-Gaussian (DOG) filtering, and contrast enhancement. These steps are applied in the specified order. In order to test the suggested approach, we use three datasets: one from the Wright State Research Institute (WSRI), one from the University of Notre Dame (UND) Collection X1, and one from the United States Army Night Vision and Electronic Sensors Directorate (NVESD). They evaluated the effects of the workouts on the outcomes of thermal-to-visible identification and conducted trials at distances ranging from 1 to 4 m. After exercising, their Rank-1 Identification measure improved to 0.64 (equal to 64%) from 0.7 (equivalent to 70%) at a distance of 1 m.

Chen and Ross [7] showed facial identification using visual and infrared pictures using a cascaded subspace learning approach. Components of this plan include common discriminant analysis, whitening transformation, and component analysis. The identification factor was extracted as a topic across distinct spectra using a factor analysis methodology. They experimented with filtering algorithms like Self Quotient Image (SQI) and Center-Surround Divisive Normalisation (CSDN) to diminish some of the variations in appearance across different spectrums. After picture filters are implemented, features from visible and thermal facial pictures are extracted using the Pyramid Scale Invariant Feature Transform (PSIFT) and Histograms of Principal Oriented Gradients (HPOG) descriptors. In order to meet the requirements of Hidden Factor Analysis, which states that sample distributions must be isotropically Gaussian, whitening treatment is employed. Canonical Correlation Analysis (CCA) and Partial Least Squares (PLS) provide the basis of the decision function. To train their model, they combed through the PCSO dataset, and then, using the CARL dataset, they matched visible and thermal facial photos. The end findings showed that when the two feature extraction approaches were fused, the verification strategy achieved a Rank-1 Identification rate of 75.61%, a 0.1% FAR score of 27.71%, and a 1% FAR score of 51.24%.

In an effort to establish a connection between the two approaches, Sarfraz and Stiefelhagen [8] attempted to explicitly simulate the very non-linear mapping. The authors maintained the identification data while building a model that mapped the perceptual differences between the two modalities using a feedforward deep neural network. The dataset used for the study was the UND-X1 dataset from the University of Notre Dame. Rank-1 identification accuracy was 83.73% when all viewable photographs in the gallery were utilised, and 55.36% when only one visible face image per person was used.

The use of a deep autoencoder model for mapping between the visual and thermal domains was demonstrated by Kantarcıand Ekenel in [9]. They tested the suggested system using the UND-X1, CARL, and EUROCOM cross-spectral datasets. According to their study, Deep convolutional autoencoders could learn non-linear mapping between visual and thermal images for the cross-domain face recognition problem. They employed two distinct up-sampling techniques as decoders. One common method of interpolation is bilinear up sampling. Although this method shortens training time on a GPU dice and simplifies the number of trainable parameters, performance suffers due to the loss of information for the decoding stage. The second approach is a convolution that follows the U-Net's recommended 2 _ 2 filter size. This network uses mean square loss as its loss function to ensure that its output is as close as feasible to the ground truth thermal pictures. Their top Rank-1 Identification rates for the CARL dataset were 48% with a single viewable image per participant and 85% with all photos available. Rerunning the analysis using the UND-X1 dataset yielded an accuracy of 87.2% across the board and a Rank-1 recognition rate of 58.75% for the one viewable picture in the gallery. As for the EUROCOM dataset, these are the outcomes: The percentage for all photographs per subject in the collection is 88.33%, while the percentage for one image per subject is 57.91%.

The domain adaptation framework put forth by Fondje et al. [10] includes the following steps: feature extraction for both visible and thermal face images using a truncated deep neural network; the Residual Spectral Transform (RST) for features that are both visible and thermal; loss for cross-domain identification; and loss for domain invariance. The VGG16 and ResNet-50 architectures are utilised for feature extraction. By converting characteristics between the thermal and visual domains, the RST residual block preserves as much discriminability from the shortened networks as feasible. The Difference of the Gaussians filter was applied to visual and thermal facial pictures before the suggested framework's phases were executed. The CCDC Army Research Laboratory provided three distinct datasets and methodologies that were utilised for testing. Rank-1 Identification rates of 96% for ResNet-50 and 84% for VGG16 were attained for frontal face pictures, correspondingly.

Much research in this area has used generative adversarial networks (GAN) to convert images from one domain to another. Additionally, GAN networks have been used to tackle thermal-visible face identification. Building visible-like pictures from thermal recordings and matching them against a library of visible faces is what Mallat et al. called ''image synthesis'' in their proposal for cross-spectrum face identification [11]. It is possible to generate high-quality coloured visual pictures from thermal data using cascaded refinement networks in conjunction with contextual loss. To put the suggested strategy to the test, they utilised their very own EUROCOM dataset. Using OpenFace, we were able to attain a 20% accuracy rate for neutral expressions in

face photos. They were able to get 82% accuracy with the LightCNN technique on neutral face pictures. The main issue with GAN-based approaches is their lengthy processing time, which makes them unsuitable for recognition while on the go.

Combining a detector network with a generative network that is based on the CycleGAN, Wang et al. [12] created a model. With the use of unpaired training photos, the GAN network learns to translate thermal and visual images in both directions without any supervision. In order to optimise the generative network, the detection network constructs the form loss function by extracting 68-landmarks from visible faces. They conducted the study using their own dataset, which included 792 thermal and visible picture pairings of 33 participants that were aligned. The images were captured with a FLIR AX5 camera. The Facenet toolkit was used to extract features from the generated probe and gallery photos. The next step was to calculate the ''Euclidean distance'' between the probe's characteristics and the galleries. Predicting matching and checking its accuracy were both accomplished by taking the shortest distance. With their custom-generated thermal and visual photos, they were able to attain a Rank-1 rate of 91.6% utilising the Facenet approach.

Kezebou et al. [13] introduced TR-GAN (thermal to RGB Generative Adversarial Network), a system that can automatically create visible facial pictures from thermal domain photos. The TR-GAN generator uses cascade residual blocks based on the U-Net architecture. The generator accomplishes image synthesis using consistent global and local structural information. Following the conversion of thermal images to visible ones, they compared faces using ResNet-50 and a pre-trained VGG-Face recognition model. Using a TUFTS dataset, the investigation was carried out. They were able to attain an identification accuracy of 80.7% with the ResNet-50 model and an accuracy of 88.65% using VGG16.

Immidisetti et al. [14] proposed an Axial GAN architecture combining low-resolution thermal pictures with high-resolution visual images. An axial-attention layer is a defining feature of their framework. A high-efficiency method for capturing long-range relationships is an axial layer. An ARL-VTF dataset was used to conduct the investigation. They were able to attain an AUC of 91.23% by using cosine similarity between features collected from a VGG-Face model.

In order to separate pictures into their style latent code and identity latent code, Anghelone et al. [15] suggested a Latent-Guided Generative Adversarial Network (LG-GAN). It enables the acquisition of attributes that are spectral-dependent as well as spectral-invariant. LG-GAN is able to maintain the identity while undergoing spectral transformation and accomplish face recognition with an AUC rate of 96.96% in terms of visual quality. They tested using the cosine distance between ResNet-50 features.

A method for transforming images of faces from visual to thermal (V2T) and back again (T2T) with varying temperatures is described in the paper by Cao et al. [16]. A six-layer PatchGAN discriminator and a U-Net generator formed the basis of their system. Both the V2T and T2T tasks made use of databases, the Speaking Face Database for the former and the Carl database for the latter. Temperature, perceptual loss, and convolutional neural network loss are all used to train the model. This study integrates two subfields of cross-spectral recognition for synthetic and real-world pictures: thermal-to-visible and thermal-to-thermal. The three pre-trained models used for the facial recognition job are MobileNet, InceptionV3, and Xception. All the neural networks were trained with weights optimised for the ImageNet database. They supplemented each model with a classification layer, two fully connected layers with 512 units, an average pooling layer, and the removal of the final fully-connected and classification layers. Finally, when evaluated on the Speaking Face database, the approach produced a Rank-1 rate of around 78%, while on the Carl database, it reached $<96\%$.

Poster et al. [17] created a database of their own that converts thermal photos of faces to visible ones. There are 395 topics in this database. The sum of all the pictures is 549,712. There is a 2.1 m gap between the person and the camera. An RGB Basler Scout CCD camera was used to capture visible face pictures, while a FLIR Grasshooper3 CMOS camera was used for thermal face images. Five distinct approaches were employed for the aim of facial recognition. The GAN framework is the basis for four methods: Pix2Pix [18], GANVFS [19], SAGAN [20], and a naïve baseline approach called ''Raw''. The VGG-Face model was fed thermal pictures (probes) and visual images (gallery) directly, with a cosine similarity threshold applied. These are the findings for RAW, Pix2Pix, GANVFS, SAGAN, and Fondje's method: 2.77%, 6.95%, 6.69%, 84.88%, 91.55%, and 96%, respectively.

First, Regarding High-Frequency Trading (HFR), Cheema et al. [21] stand out due to their innovative method. Methods like common subspace projection, feature extraction, and image preparation have traditionally been foundational to HFR. However, these approaches have been prone to performance error accumulation and optimisation issues. When dealing with problems caused by large discrepancies between modalities, such as in visible-to-thermal identification, the Cross-Modality Discriminator Network (CMDN) has come a long way, largely thanks to its relational learning capabilities and Unit-Class Loss. This paper presents an improved and efficient framework for face recognition across many imaging modalities, which is a significant addition to the field of biometrics and security. This framework is crucial if we want to employ facial recognition for things like access control, monitoring, and other security-related tasks. Researchers and developers working on cross-modality face recognition systems may be influenced by the study's conclusions and methods.

Second, according to Alkadi et al. [22] introduce a novel machine learning-powered biometric identification system

TABLE 1. Summary of the literature survey.

Convolutional Neural Networks (CNNs) Model	Dataset	Accuracy (%)	References
DeepFace	Face in the Wild (LFW)	97.35	[2]
Center loss as a new loss function	LFW	99.28	[3]
CNN architectures	VGG Face	98.95	[4]
Angular Softmax loss for CNNs	LFW, Youtube Faces(YTF) and Mega Face Challenge1	99.42	[5]
Histogram of Oriented Gradients and Partial Least Squares based model	University of Notre Dame (UND) Collection X1, a dataset collected by the Wright State Research Institute (WSRI), and Night Vision and Electronic Sensors Directorate (NVESD)	70 for the distance 1metre 64 for the distance 4metres	[6]
Center-Surround Divisive Normalisation (CSDN) and Self Quotient Image (SQI), the Pyramid Scale Invariant Feature Transform (PSIFT) and Histograms of Principal Oriented Gradients (HPOG), Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)	PCSO CARL	75.61 51.24	[7]
Feedforward deep neural network in order to map the perceptual differences between the two modalities while preserving the identity information	University of Notre Dame UND-X1	83.73	[8]
Deep convolutional autoencoders	UND-X1, CARL, and EUROCOM cross-spectral	87.2, 85 and 88.33 respectively	[9]
ResNet-50 and VGG16	protocols compiled by the CCDC Army Research Laboratory	96 and 84 respectively	[10]
Generative adversarial networks (GAN)	EUROCOM	82	[11]
CycleGAN and detector network	-	91.6	[12]
TR-GAN (thermal to RGB Generative Adversarial Network)	TUFTS	Resnet-50 → 80.7 VGG16 → 88.65	[13]
Axial GAN framework	ARL-VTF	91.23	[14]
Latent-Guided Generative Adversarial Network (LG-GAN)	ARL-VTF	96.96	[15]
U-Net generator and a six-layer PatchGAN discriminator	Speaking Face	96	[16]
GAN framework including Pix2Pix, GANVFS, SAGAN, and "Raw"	ARL-VTF, Cityscapes, Polarimetric-Visible, and Polarimetric-Visible	96, 91.55, 84.88, 6.69, 6.95, and 2.77 respectively	[17] [18] [19] and [20]
Regarding High-Frequency Trading (HFR)	The TUFTS Face database, Collection X1 from the University of Notre Dame (UND-X1) database,	TUFTS – 98.5 (Score Fusion) UND-X1- 95.21 (Score Fusion) USTC-NVIE – 99.7 (Score Fusion)	[21]
	University of Science and Technology China, Natural Visible and Infrared facial Expression (USTC-NVIE)	Sejong Face – 92.4 (Score Fusion)	

TABLE 1. *(Continued.)* **Summary of the literature survey.**

	database, CASIA NIR-VIS, and Sejong Face Database.		
ResNet-50 Inception V3 VGG-16	I^2BVSD, BRSU Spoof, Spectral Disguise, and Sejong Face dataset	ResNet-50: Vis-99, Th-99, IR-97, Vis+IR-99%. Inception V3: Vis-84, Th-91, IR-67, Vis+IR-79%. VGG-16: Vis-99, Th-99, IR-98, Vis+IR-99%.	[22]
The triple triplet method	D4FLY Thermal and 2D Face, IOE_WAT Dataset, Speaking Faces, Sejong Face, and FaceScrub	90.61	[23]

for ATMs. The system's objective is to precisely recognise people, even in cases when they are wearing equipment that modifies their facial characteristics, such as glasses or masks. This study emphasises the need to maintain a harmonious equilibrium between ease and security in consumer bank transactions. Conventional ways of authenticating ATMs, such as PINs, provide security vulnerabilities, whereas biometric systems provide a more secure option by utilising distinctive bodily attributes, such as facial features, for authentication. The suggested method employs multi-modal imaging, including visible light, thermal, and infrared imaging, to enhance the accuracy of identification and mitigate typical obstacles such as presentation assaults and disguises. The system underwent training and testing using many datasets, with particular emphasis on the Sejong multi-modal disguised face dataset. This dataset was chosen for its wide range of photographs showcasing different facial add-ons, ensuring a varied and balanced collection. The study revealed that some machine learning models, specifically ResNet-50, accurately distinguished users with and without add-ons. The study also examined the performance of the authentication system when faced with various cultural and gender-specific disguises, such as hijabs and ghutras. This demonstrated the system's adaptability and dependability in a wide range of situations. This research makes a substantial contribution to the area by showing that it is possible to create a more secure and user-friendly ATM authentication system. This has the potential to completely change the way banking transactions are protected.

When dealing with challenging low-light or on-the-go scenarios, Kowalski et al. [23] suggested a novel method for thermal-visible face verification. While previous work on face recognition using visible spectrum data has achieved impressive accuracy, it has struggled to work in low-light conditions. Although earlier proposed solutions have shown only moderate success, the thermal-visible approach fills this need. The triple triplet approach is an innovative method for handling the different optical and thermal spectra, which uses multiple convolutional neural networks (CNNs). The outcome is a considerable improvement in the recognition accuracy (90.61%). This approach is particularly innovative since it increases the identification process's durability and reliability by integrating data from the visible and thermal spectra. This means that future studies in this area will have to measure up to this mark. The summary of the literature survey is demonstrated in Table 1.

A. KNOWLEDGE GAP

Taigman et al. [2] introduce notable progress in face verification technology by utilising 3D modelling and deep learning. Nevertheless, it presents some areas of incomplete understanding that require additional research. Importantly, it fails to investigate the model's resilience to adversarial attacks, a critical aspect for guaranteeing the security and dependability of face verification systems in hostile contexts. Moreover, the research does not include an in-depth analysis of the system's ability to apply its knowledge to different and more difficult datasets, except from LFW and YouTube Faces. This raises concerns about its effectiveness in more diversified situations. The computational efficacy and scalability of implementing such a model in practical applications, where processing resources and latency are crucial considerations, are acknowledged but not extensively examined. Moreover, the ethical and privacy concerns associated with implementing sophisticated facial recognition technology, such as permission, data protection, and potential misuse, have not been acknowledged or resolved. Furthermore, the possibility of incorporating temporal dynamics from video sequences to improve verification accuracy and resilience is disregarded.

In their study, Wen et al. [3] introduce a new approach to improve the ability of deep learning features in face recognition by incorporating a centre loss function. Although the research has made important contributions, several areas still require more exploration. Significantly, the proposed method does not include the ability to handle massive and diverse datasets in a scalable and computationally efficient manner. This is particularly important for real-time applications and managing the continuously growing number of digital photographs. Furthermore, the method's ability to withstand adversarial attacks, a significant concern for applications requiring high security, has not been investigated. Further

work is needed to explore the possible use of the centre loss function in various domains beyond face recognition, such as other computer vision tasks or even non-visual domains. In addition, the research does not thoroughly examine how data variety and bias affect the model's performance and fairness. It also fails to compare the suggested method with new and developing deep learning methodologies and architectures.

Parkhi et al. [4] present significant progress in facial recognition using deep learning and a comprehensive compilation of datasets. Although it has made significant contributions, there are still some areas where information is lacking. The examination of the model's generalisability across different and challenging real-world settings is currently restricted. The text did not discuss the optimisation and adaptability of implementing these models in real-time applications, namely those that handle considerably larger datasets or continuous data streams. Furthermore, the research fails to examine the resilience of the proposed system against adversarial assaults, which is a crucial factor for applications that prioritise security. The potential for their methods to be applied to other domains within and beyond computer vision has not been explored. Moreover, there is a dearth of thorough analysis regarding the potential influence of dataset variety and inherent biases on model fairness and performance. Furthermore, there is a lack of comparisons with the most recent advancements in deep learning technologies that could potentially improve the performance or efficiency of the model.

Liu et al. [5] propose a novel A-Softmax loss function to improve face recognition using CNNs by training the network to learn features that are discriminative in terms of angles. Although it demonstrates notable progress in facial recognition, there is still a lack of understanding regarding its wider range of uses. The research does not explore the possibility of applying the A-Softmax loss to other domains or data types that may have similar manifold structures. Additionally, it does not discuss the method's resilience against adversarial attacks or its performance when dealing with noisy data. Furthermore, the full examination of the computational efficiency and scalability of the A-Softmax loss in larger datasets or real-time applications is lacking. Additional empirical investigation on the selection and influence of the hyperparameter $\boldsymbol{m}$ in various scenarios could offer more profound insights for practical implementation. Furthermore, the incorporation of A-Softmax loss into various neural network designs or its amalgamation with other losses and regularisation procedures has the potential to create new opportunities for investigation, amplifying the adaptability and efficacy of the method in a broader spectrum of deep learning applications.

Hu et al. [6] conducted a comprehensive study on cross-modal face recognition, although various aspects still need to be explored in the future. The extent to which the proposed strategy may be applied to real-world situations with unforeseen factors, such as harsh environmental conditions or occlusions, has not been thoroughly investigated. Furthermore, the research fails to explore the capabilities of the approach to handle large-scale or real-time applications, which is essential for its practical implementation. The model's performance and fairness are not adequately addressed in relation to the impact of dataset variety and inherent biases, which raises concerns about the system's inclusion across various populations. In addition, the method's ability to withstand sophisticated adversarial attacks, which is becoming increasingly crucial in the implementation of secure biometric systems, is not considered. Furthermore, the analysis does not explore the possibility of combining this method with other biometric techniques or utilising emerging deep learning technology to improve the accuracy and resilience of recognition.

Chen and Ross [7] propose a novel method for recognising faces from several sources, known as heterogeneous face recognition (HFR), which leads to significant enhancements in performance. Nevertheless, it provides numerous opportunities for more investigation. The method's ability to withstand environmental fluctuations, such as occlusions, facial expressions, and head postures, has not been properly assessed, which may limit its suitability for real-world applications. The study's emphasis on thermal and visual spectra for face recognition might be broadened to incorporate integration with additional biometric modalities, thereby boosting security and practicality. Furthermore, it is necessary to conduct further evaluation of the suggested method's computing efficiency and scalability, which are essential for real-time and large-scale applications. Furthermore, although the research performs cross-database tests to assess the model's potential to generalise, a more extensive analysis of its performance on a wider variety of datasets could offer more thorough insights into its adaptability.

Sarfraz and Stiefelhagen [8] propose a novel method that uses deep neural networks to minimise the difference in characteristics between thermal and visual images to improve face identification. Although the research has made significant contributions, some areas still require more exploration. The text does not thoroughly examine the model's capacity to generalise across various ambient circumstances, camera technology, and subject groups, which is essential for practical use. The model's capacity to handle other cross-domain issues, such as NIR-to-visible mapping, has not been addressed. In addition, the research does not thoroughly analyse the scalability and computing efficiency of the system, which are crucial for large-scale or real-time surveillance applications. Examining the most recent advancements in deep learning architectures could offer valuable insights regarding possible enhancements. Moreover, it is crucial to pay attention to ethical and privacy concerns, particularly in situations involving monitoring, to ensure responsible utilisation.

Kantarcıand Ekenel [9] propose a highly effective method for thermal-to-visible face identification by utilising deep autoencoders to overcome the differences between thermal and visible modalities. Nevertheless, the model's

performance in terms of facial variations such as expressions, occlusions, and head positions, which are crucial for real-world applications, remains unexamined. The study excludes an examination of the proposed system's real-time processing capabilities and scalability, which are crucial for implementing it in large-scale or dynamic contexts. In addition, the article does not investigate the model's adaptation to different imaging modalities, which could increase its usefulness in surveillance and biometrics. Furthermore, the research does not include a comparative examination with alternative deep learning architectures, which could potentially uncover additional enhancements. Furthermore, the ethical and privacy concerns related to the implementation of thermal imaging for facial recognition in surveillance are not discussed, which is becoming increasingly significant as these technologies become more widespread.

The study conducted by Fondje et al. [10] presents a notable breakthrough in addressing the disparity between thermal and visual domains in face recognition. Nevertheless, it presents numerous areas that can be further investigated in the future. The assessment of the system in real-world scenarios, particularly in dynamic surroundings with fluctuating illumination and weather conditions, is still restricted. The framework's capacity to handle other cross-domain issues, such as NIR-to-visible or SWIR-to-visible recognition, is not discussed. This limitation hinders its potential use in surveillance and security fields. In addition, the paper does not investigate the incorporation or comparison with state-of-the-art deep learning frameworks, such as GANs or transformers, which could potentially enhance the accuracy of recognition. The proposed method's scalability and computing efficiency, which are essential for implementation in large-scale or real-time applications, have not been properly explored. Finally, the ethical and privacy concerns associated with implementing sophisticated facial recognition technologies, especially in surveillance, require additional dialogue and investigation to guarantee responsible usage and mitigate any prejudices.

Mallat et al. [11] present a novel method for creating visible-like images from thermal data to improve cross-spectrum face identification. Although the research has made valuable contributions, some areas still require more exploration. More precisely, the process of synthesising images sometimes fails to depict certain characteristics, including gender and ethnicity effectively. This highlights the necessity for enhancing the accuracy of certain features in created images. The study's ability to be applied to different datasets has not been evaluated, indicating the possibility of wider applicability and the need for evaluation of its robustness. Furthermore, integrating the synthesis model directly with deep learning-based facial recognition technologies has the potential to improve and simplify the system's overall performance. The ethical and privacy aspects of implementing this technology, particularly in surveillance situations, require thorough examination and the development of ways to resolve any concerns. Finally, the research does not explore the suggested method's computational efficiency or real-time processing capabilities, which are essential for practical implementation.

Wang et al. [12] propose a remarkable approach for converting thermal face photos into visible ones by utilising an adversarial network with a generative algorithm environment. Although the research has made progress, it fails to adequately assess the model's capacity to apply its knowledge to multiple thermal imaging circumstances, differences in subjects, and environments. This raises doubts about its usefulness in diverse real-world situations. Furthermore, the computing requirements and effectiveness for processing in real-time, which is essential for actual implementations, have not been investigated. The study does not include a comparison with the most recent advancements in deep learning, which have the potential to enhance both efficacy and performance. Moreover, the ethical and privacy concerns associated with employing thermal imaging for facial recognition in surveillance or other sensitive contexts are not being acknowledged, a matter of growing significance in guaranteeing the appropriate utilisation of technology. Furthermore, the possibility of applying this technology to other types of images or wider applications has not been explored, which has the potential to improve its usefulness beyond just facial identification greatly.

Immidisetti et al. [14] present a novel method to improve thermal-to-visible face verification, specifically for low-resolution thermal pictures. Although the research has made valuable contributions, there are still unknown areas that could enhance and improve the technology even further. The efficacy of the suggested approach in numerous real-world circumstances remains uncertain due to the lack of comprehensive assessment regarding its resilience under different climatic conditions and thermal imaging technology. Furthermore, the discussion of Axial-GAN's scalability and real-time processing capabilities, which are essential for its implementation in surveillance systems or mobile devices, is lacking. The possibility of this strategy to apply to additional cross-domain difficulties, such as converting NIR to visible, has not been explored. Moreover, the exploration of integrating synthesised photos with sophisticated facial recognition systems to enhance end-to-end verification performance is lacking. Furthermore, the research fails to address the ethical and privacy concerns arising from thermal-to-visible face verification methods, especially in sensitive monitoring situations.

Anghelone et al. [15] introduce a novel method for thermal-to-visible face recognition. Although it has made significant progress, some domains still necessitate additional investigation. The comprehensive assessment of the suggested method's applicability in varied climatic circumstances and with multiple thermal imaging technologies has not been adequately examined, raising concerns about its resilience in diverse scenarios. The LG-GAN's scalability and real-time processing capabilities, which are crucial for its application in surveillance systems or on mobile devices, are not addressed. Moreover, the potential

of this framework to address other cross-spectral difficulties, which may include NIR-to-visible or SWIR-to-visible facial recognition, has not been investigated. The possibility of integrating advanced facial recognition systems directly to improve overall performance has not been explored. Furthermore, the ethical and privacy considerations associated with utilising thermal imaging for facial recognition, particularly in surveillance contexts, have not been acknowledged, underscoring the importance of employing technology responsibly.

Cao et al. [16] introduce a new method that utilises conditional Generative Adversarial Networks (cGAN) to translate thermal images into visible facial images. Although the work shows promising outcomes in converting visible images into thermal images that are visually and identifiably consistent, it fails to investigate numerous crucial aspects. More precisely, the model does not explore how well it can adapt to diverse climatic circumstances or thermal imaging methods, which is crucial for its practical use in real-world situations. The text does not address computing efficiency and appropriateness for real-time processing, which are crucial factors for surveillance and security applications. Furthermore, the report fails to discuss the possible ethical and privacy issues that may arise from implementing this technology, especially in the context of monitoring. Furthermore, the potential for expanding this paradigm to encompass other spectral bands or imaging modalities has not been explored. Furthermore, doing a comprehensive examination of how the conversion process impacts different facial traits or attributes could yield significant observations regarding the model's abilities and constraints.

Poster et al. [17] present the ARL-VTF dataset, which makes a substantial contribution to the field of thermal and visual face recognition research. Although the study has made significant progress, numerous crucial areas still have not been investigated. A comprehensive evaluation of the resilience of established models under diverse environmental circumstances and various thermal imaging technologies is necessary to guarantee their extensive applicability. Furthermore, the lack of discussion regarding the computing efficiency and scalability of real-time applications, such as surveillance, raises concerns about the actual practicality of their deployment. The dataset's capacity to be applied to different spectral bands or imaging modalities, as well as the ethical and privacy concerns associated with its utilisation in surveillance, have not been addressed. Furthermore, doing a thorough examination of the impact of eyewear and other accessories on algorithm performance could yield a more profound understanding of the obstacles and remedies in thermal-to-visible face identification.

Zhang et al. [19] present an innovative approach for generating visible face images from thermal images using GANs. The method achieves remarkable results in terms of producing realistic photos and preserving the identity of the individuals. Nevertheless, the research does not investigate the model's adaptability to different climatic conditions or other thermal imaging technologies, which is essential for broader applications. Furthermore, the study does not address the computational efficiency and the practicality of real-time processing, which are crucial for implementing the technology in surveillance and security applications. Neglecting ethical and privacy concerns, especially with thermal imaging's surveillance capabilities, raises doubts about responsible use. The potential to apply this technique to additional imaging modalities or spectral bands has not been investigated. Furthermore, conducting a comprehensive examination that examines the model's capacity to maintain identification despite differences such as facial expressions or obstructions will enhance comprehension.

Di et al. [20] provide a novel method for thermal-to-visible face identification, which makes notable progress in image synthesis and verification. Nevertheless, there are other areas that necessitate additional investigation. The method's application in multiple real-world scenarios has not been evaluated for the generalisation of the model across a broader range of environmental conditions, thermal imaging technologies, and diverse subject demographics. This testing is vital for judging the method's effectiveness. The text does not address the important aspects of computational efficiency and the potential for real-time processing, which are crucial for practical surveillance and security applications. The ethical and privacy concerns regarding the creation of visible faces from thermal images, especially in surveillance settings, have not been discussed, raising concerns about responsible usage and potential privacy violations. Furthermore, the method's potential to be applied to other imaging modalities apart from polarimetric thermal pictures has not been investigated, which may restrict its usefulness in many applications. Furthermore, the strategy does not consider the longitudinal robustness, encompassing elements like ageing, that could affect its effectiveness as time progresses.

Kowalski et al. [23] present a new method for addressing the disparity between thermal and visual modalities. Although the study has made valuable contributions, some areas have not been thoroughly investigated. The method's robustness across varying climatic conditions, subject demographics, or alternative thermal imaging technologies is not thoroughly assessed, which is essential for its implementation in various real-world circumstances. The computing efficiency and the feasibility of real-time implementation are uncertain, raising doubts about its applicability to identity verification systems used while in motion. In addition, the research does not explore the possibility of applying the strategy to additional cross-domain difficulties outside of thermal-to-visible recognition. The ethical and privacy concerns associated with implementing biometric technologies in surveillance settings are not acknowledged, disregarding the importance of responsible technology usage and privacy protections. Furthermore, the discussion does not address the method's long-term dependability considering potential alterations in people's looks over time.

III. METHODOLOGY

The article aims to improve the accuracy and precision of facial verification by utilising multiple imaging modalities such as visible, thermal, infrared, and a combination of visible and infrared. The study presents a method named VGG16-PCA-NN to enhance classification accuracy using the Sejong Face Database for both training and testing as illustrated in Figure 1. This technology excels in several areas, enhancing biometrics and security industries with a strong and effective facial verification system that overcomes hurdles from environmental changes and physical obstacles.

Identifying faces in photos has three key steps: pre-processing photos, extraction of features, and classification. Normalising lighting and removing noise and background noise are both done during pre-processing. Using techniques like CNNs and PCAs, features are collected from pre-processed facial photos. Classifiers, like NNs, are used to carry out the classification process. To train input face photos, we employ the structure of VGG16 in the proposed research. This design has already been pre-trained on a huge ImageNet database with over 1 million images from 1000 distinct categories. A fully connected layer with Softmax activation is used for classification. In order to increase photo accuracy in classification compared to utilising VGG16 features alone or PCA alone, this paper contributes by merging two models with NNs using the Sejong Face database, which is our innovation in using this data since nobody has used our method before.

The suggested method uses a pre-trained VGG16 CNN for feature extraction, PCA for dimensionality reduction, and a sequential neural network for classification. The ImageNet-trained VGG16 model extracts hierarchical features from input photos. The extracted characteristics are reduced in dimensionality via PCA, improving computing efficiency. To learn sophisticated patterns in the restricted feature space, a sequential neural network with dense layers is built. A rank-level fusion technique combines neural network and cosine similarity index predictions to classify and recognise images accurately once the model is trained on the dataset. Accuracy, precision, recall, and ROC curves evaluate the model's performance. The suggested method uses transfer learning, dimensionality reduction, and neural networks for robust and efficient picture recognition.

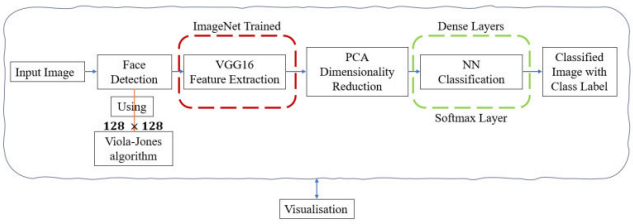


FIGURE 1. A flowchart for our proposed approach.

A. FACE DETECTION

The approach used to find the face in the image largely determines the efficacy of biometric systems based on face authentication. To that end, we employ the Viola-Jones algorithm, which effectively detects a wide range of facial features, including those of the mouth, eyes, nose, eyebrows, lips, ears, and so on [24]. The vision Cascade Object Detector function in MATLAB has been used to create this approach. When it comes to face feature recognition, Viola-Jones employs three key methods:

i One way to obtain an Integral picture is to employ rectangular Haar-like features for feature extraction [25].
ii The Ada Boost algorithm uses machine learning and artificial intelligence to recognise faces. A notion that unites several algorithms that depend on sets of binary classifiers is defined by the word "boosted" [26].
iii Third, a Cascade classifier, which can effectively merge many characteristics, is the last phase. A classifier's "cascade" feature specifies the many filters used for the final product.

After the pre-trained Viola-Jones face, we manually selected the participants who had faulty face identification findings in order to guarantee that every participant would have the opportunity to provide feedback on the subsequent evaluation of the methodology that was recommended. The pictures of the face were scaled to (128×128) pixels to reduce the noises of the images like background, hair, and so on. In Figure 2, we can see a Viola-Jones approach in action.

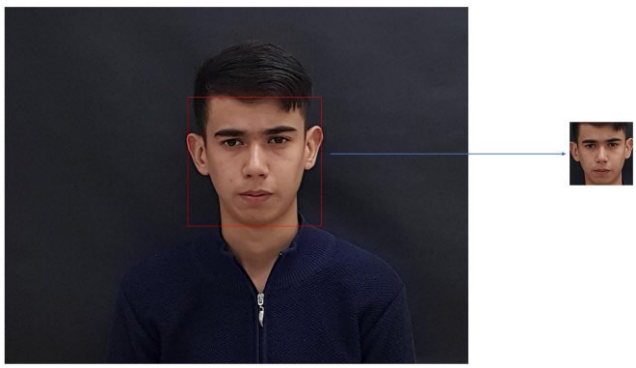


FIGURE 2. Face and facial components detection using the Viola-Jones technique.

B. THE USE OF TRANSFER LEARNING AND CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Among the many subfields of machine learning, "deep learning" is known for its emphasis on building progressively meaningful representations from the ground up through several layers of training data. One measure of a model's depth is the total number of layers that make up its data. Each neural network has an input layer, a number of hidden layers, and an output layer. "Depth of the model" refers to the total number of data-contributing layers. Contrast convolutional neural networks (CNNs) with traditional neural networks that use matrix multiplication in their convolutional layers. For further information about transfer learning and CNN, the reader is referred to [25], [26], [27], [28], and [29].

In our research, we adopted the VGG-16 model, a complex deep convolutional neural network that has been trained on the ImageNet dataset. We employed this model to extract characteristics from various forms of imaging, such as visible, thermal, infrared, and a combination of visible and infrared. The VGG-16 model was selected due to its exceptional ability to capture intricate details that are crucial for facial recognition tasks. By utilising the pre-existing weights, we were able to extract superior features from a small amount of training data. This is particularly advantageous for datasets with specific limitations, such as the Sejong Face Database.

The proposed method uses a pre-trained VGG16 CNN for feature extraction, PCA for dimensionality reduction, and a Sequential NN for classification. This combination is novel because it integrates deep learning with classical machine learning techniques in a way that enhances the overall performance and efficiency of FR systems. By combining these methods, we are able to address the challenges posed by different imaging modalities and environmental variations. This integration is validated using the Sejong Face Database, demonstrating its effectiveness in improving FR accuracy across multiple modalities, including VIS, Th, IR, and a combination of VIS-IR spectra.

Furthermore, we improved our method by combining VGG-16 with PCA to reduce the number of dimensions and then using a sequential neural network for classification. This approach utilised the advantages of deep learning for extracting features and classical machine learning for effectively managing data and accurately classifying it. This blend exhibited remarkable versatility and effectiveness in handling a wide range of data presentations from different imaging technologies, establishing a new standard in multi-modal facial recognition, and making a substantial contribution to the progress of biometrics and security domains.

The innovative VGG16-PCA-NN framework is a major step forward in our study. It enhances the accuracy of classifying by prioritising important characteristics and reducing unnecessary repetition of data, effectively tackling the difficulties posed by variations in the environment and physical aspects of facial recognition tasks. This comprehensive approach not only enhances the precision of facial recognition in many settings but also advances the fields of biometrics and security by proposing a method that skillfully overcomes the main challenges faced in these areas.

In this work, we employ the VGG-16 architecture as our deep learning model. VGG-16 comprises a total of 19 layers, consisting of 13 convolutional layers followed by three fully connected layers, and culminating in a Softmax output layer. All hidden layers utilise the Rectified Linear Unit (ReLU) activation function, providing non-linearity to the model [27]. The convolutional layers employ 3×3 convolution kernels, expanding the number of channels to capture complex and expressive features. Zero-padding is applied to maintain the size of the output data. Between the layers, Maxpooling with a 2×2 window size and a stride of 2 is utilised to extract detailed information. The convolutional layer depths progressively increase as follows: 64 -> 128 -> 256 -> 512 -> 512 as represented in Figure 3.

Figure 4 shows that the revised network, VGG16-PCA - NN (Neural Network), uses an input image size of $128 \times 128 \times 3$ and produces feature dimensions of $4 \times 4 \times 512$. In order to extract detailed characteristics from pictures, we use VGG-16 and the VGG16-PCA-NN architecture. Also, dimensionality is reduced using Principal Component Analysis (PCA), and then a neural network is trained to learn and identify the features that have been reduced. In comparison to using VGG-16 features alone or PCA alone, the performance of picture classification is better when these two methods are combined. The VGG-16 model has a total of 14,714,688 parameters, and all parameters in the convolutional layers are non-trainable since they are initialised with pre-trained weights from ImageNet. The model size is approximately 56.13 MB. In addition, by using NN then the VGG-16 model has a total of 225,339 parameters and all parameters in the convolutional layers are trainable after we modified the VGG-16.

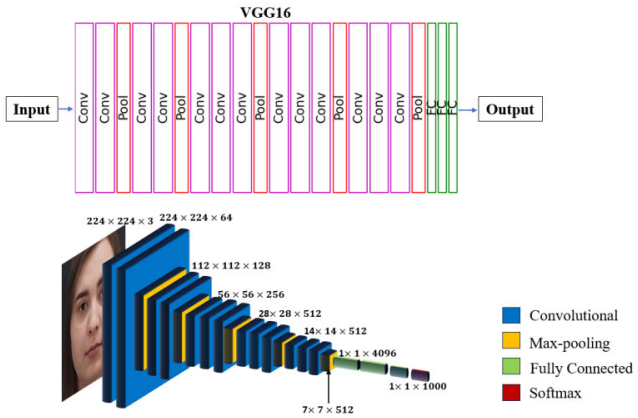


FIGURE 3. An overview of the VGG16 convolutional neural network design.

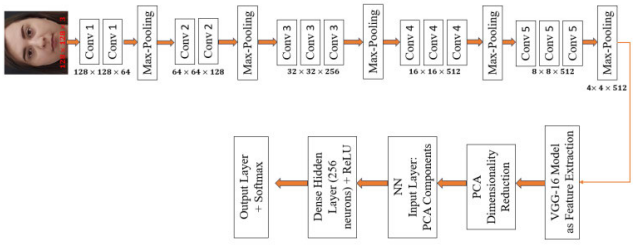


FIGURE 4. Enhanced recurrent neural network.

In the improved architecture that we have developed, which is called VGG16-PCA-NN, we have used ReLU as the activation function in the network, and dropout is strategically applied in order to remedy the problem of overfitting. We use the Softmax classifier [28], [29] as the final output layer for the classification job. This allows us to estimate the likelihood

of each class label in X different classes. The improvements that we have made are in line with our overarching objective, which is to not only solve the problems associated with overfitting but also to improve the classification accuracy of the network.

C. PRINCIPAL COMPONENT ANALYSIS (PCA)

Applying Principal Component Analysis (PCA) to the face dataset is crucial for improving the accuracy and efficiency of facial recognition technologies in various imaging modalities. By utilising a pre-trained VGG16 Convolutional Neural Network (CNN) to extract initial features from facial photographs captured using different imaging techniques. This approach leverages the strong capabilities of the VGG16 model to detect complex features accurately. PCA is critical for lowering the dimensionality of facial features, eliminating redundancies, and improving computing efficiency without compromising the fundamental properties for accurate facial recognition. The reduced characteristics are subsequently employed to train a specialised sequential neural network, which is specifically designed to identify and categorise the subtle patterns in the data. The integration method, known as the VGG16-PCA-NN method, is highly advanced and complex. PCA is crucial in this context as it not only simplifies the feature space but also preserves the integrity of the data, which is essential for achieving high accuracy rates. Consequently, it greatly enhances the performance of facial recognition in biometric and security technologies.

PCA is applied to the VGG16-extracted features to reduce their dimensionality while retaining essential information. For example, the modified VGG16 model, optimised by integrating PCA for dimensionality reduction, showed a considerable decrease in the number of parameters from 14,714,688 to 225,339. This reduction not only decreases the computational load but also addresses the overfitting issue by reducing the complexity of the model. The model size was approximately reduced to 56.13 MB, highlighting the impact of PCA in making the model more efficient without compromising its facial recognition accuracy.

Moreover, the application of PCA contributed to an enhancement in the cosine similarity indices across different modalities, indicating a closer match between the test and training data features. This improvement suggests that PCA effectively retains the most significant features necessary for accurate facial recognition, thereby reducing the computational complexity and noise in the data.

D. NEURAL NETWORKS

Utilising a pre-trained VGG16 model's features and principal component analysis (PCA) to reduce dimensionality, our study implements a neural network model for image categorisation.

In the NN, there is an input layer whose form is dictated by the 'n_PCA_components', a hidden dense layer that uses ReLU activation and has 256 neurons. The output layer, which uses softmax activation, completes the network. With the characteristics acquired by PCA fed into the input layer, the output layer generates class-specific probabilities.

Consider the following: $\boldsymbol{I}$ is the input with shape (n_PCA_components,), $\boldsymbol{A_1}$ and $\boldsymbol{B_1}$ are the hidden layer's weights and biases, and $\boldsymbol{A_2}$ and $\boldsymbol{B_2}$ are the output layer's weights and biases. Here are the formulas (1) and (2) for the hidden layer's output $(\boldsymbol{H})$ and the final output $(\boldsymbol{O})$, respectively:

$$\boldsymbol{H} = \boldsymbol{ReLU}\,(\boldsymbol{A_1} \times \boldsymbol{I}) + \boldsymbol{B_1} \tag{1}$$

$$\boldsymbol{O} = \boldsymbol{Softmax}\,(\boldsymbol{A_2} \times \boldsymbol{H}) + \boldsymbol{B_2} \tag{2}$$

In order to classify images, the NN makes use of features retrieved from the VGG16 model and dimensionality reduction via PCA. The NN's role as a classifier is critical for the end classification since it maps the retrieved characteristics to class probabilities. The goal of this study is to categorise photos using characteristics extracted from a pre-trained VGG16 model; this neural network offers a flexible and efficient method for solving image identification problems.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Sejong Face datasets were utilised for training, assessment, and testing in this work. All databases must contain pictures of faces captured in the visible plus infrared (VIS-IR), visible (VIS), thermal (Th), and infrared (IR) domains, which are utilised separately in our approach, VGG16-PCA-NN. This part provides a concise overview of all the datasets used in the study.

A. DATABASE

All of the face data was sourced from the Sejong Face Database (SFD) [30]. Both Group-A (subset-A) and Group-B (subset-B) are classified parts of the database. A year before the collection of Group B, Group A was compiled. Among the 30 people whose faces makeup Group A, 16 are male and 14 are female. We used frontal faces for all photographs and caught each person's face in both an add-on and neutral condition. In contrast, there are 70 people represented by the faces in Group B; of these, 44 are men and 26 are women. Each modality in Group B received 15 neutral face shots, whereas the other add-ons each received 5 images. Plus, five pictures of males with real beards and five women sporting makeup were shot [30].

Two cycles, separated by 14 days, were used to gather individual pictures [30]. The initial batch of photos included men with a variety of accessories, including clean-shaven faces and neutral expressions. The second batch of photographs was taken of the identical subjects 14 days later, but this time they had facial hair that had developed over that time. For the female contestants, the first round was all about taking images without makeup or with accessories, while the second round was all about getting photos with cosmetics. Thirteen distinct disguise accessories were available for males and twelve for females [30].

Figure 5 shows the results of our investigation using the SFD, which consists of 60 persons without any additional

features and each of them has pictures in the VIS, IR, Th, and combined VIS-IR spectrums [30]. Every face photo is captured using a combination of four different spectra. The VIS, VIS-IR, IR, and Th spectra are all included in this set, with dimensions of 4032 × 3024, 1680 × 1050, and 768 × 756 respectively.

FIGURE 5. The database we present contains four types of images that can be classified into several categories. The first type is infrared (IR) images, denoted by category A, followed by visible images with IR, denoted by category B. The third type is thermal images, represented by category C. Categories D-H include makeup images, while categories I-L are similar to categories A-D. Finally, categories M-P comprise images that do not have any makeup.

B. EXPERIMENTAL SETTING AND MODEL DESIGN

This article's testing setup included a desktop and a laptop computer with 11th Gen Intel(R) Iris(R) Xe Graphics, Core (TM) i7-1165G7 @ 2.80 GHz and 16.0 GB of RAM. The TensorFlow framework, built on top of the Python 3.11 language, is utilised for the development of our network architecture and code. Each network's optimizer was set to rmsprop during training; Keras uses the optimizer's default learning rate. The model.fit() and epochs parameters were each set to 200, with a maximum limit of 1000 for dataset combinations such as VIS, Th, IR, or VIS-IR. The proposed code makes use of a hybrid neural network, consisting of both a pre-trained VGG16 model and an individualised dense neural network. Before PCA uses the VGG16 model to extract features, it loads it with weights that have already been trained on the ImageNet dataset. The VGG16 model is enhanced for further processing by adding the custom dense neural network on top of it.

The VGG16 model was selected due to its well-established performance in image classification tasks. It comprises 19 layers, including 13 convolutional layers and 3 fully connected layers, culminating in a Softmax output layer. This architecture allows for the extraction of complex features from the input images, making it suitable for facial recognition tasks. The Rectified Linear Unit (ReLU) activation function was used in all hidden layers to introduce non-linearity and improve the model's ability to learn complex patterns. The final layer uses the Softmax activation function to output class probabilities. The model was initialised with pre-trained weights from the ImageNet dataset, leveraging the knowledge gained from a large dataset to improve feature extraction capabilities even with limited training data.

To further enhance the model, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the feature space while retaining essential information. This step helps in reducing computational complexity and addressing the overfitting issue by limiting the number of parameters. The number of PCA components was chosen based on the cumulative explained variance, with a total of 820 components selected to balance the trade-off between retaining significant variance and reducing dimensionality.

The neural network consists of an input layer matching the number of PCA components, a hidden dense layer with 256 neurons using ReLU activation, and an output layer with 59 neurons (corresponding to the number of classes) using Softmax activation. Dropout regularisation was strategically applied to mitigate overfitting, ensuring the model generalises well to unseen data.

The rmsprop optimizer was chosen for its effectiveness in training deep neural networks by adjusting the learning rate dynamically, achieving faster convergence and better performance. A batch size of 32 was used, balancing the trade-off between computational efficiency and the stability of gradient updates. The model was trained for 500 epochs, a number selected based on preliminary experiments to ensure sufficient training without overfitting. Early stopping or validation-based adjustments were employed to monitor the training process and prevent overfitting.

The categorical cross-entropy loss function was used for multi-class classification, effectively measuring the performance of the model in predicting the correct class probabilities. The primary metric used for evaluation was categorical accuracy, which measures the proportion of correct predictions out of all predictions made. Additionally, precision, recall, and F1-score were used to provide a comprehensive evaluation of the model's performance across different classes.

By carefully selecting these parameters and hyperparameters, the model aims to achieve high accuracy and robustness in multi-modal facial recognition tasks.

C. RESULTS

The datasets specified in Section IV-A were divided into three parts: 70% for training, 15% for validation (development set), and 15% for testing. The training photos, stored in subdirectories representing various classes, are imported and categorised. The validation and testing photos are processed with matching labels. All datasets are normalised and a VGG16 model with pre-trained weights is used for feature extraction, followed by Principal Component Analysis (PCA)

for reducing dimensionality. The neural network model is built and trained using dense layers on the reduced PCA features from the training set. The validation set is utilised to tune the model's hyperparameters and prevent overfitting. The algorithm assesses the model's performance by utilising criteria including accuracy, precision, recall, F1-score, confusion matrix, and ROC curves. Cosine similarity is calculated between the characteristics of the test and training sets. The code ends with a rank-level fusion method that merges predictions from the neural network with cosine similarity for a comprehensive assessment. The outcomes are displayed via graphs, providing insights into the model's accuracy over epochs, confusion matrix, ROC curves, and rank-level fusion curves. The algorithm offers a thorough strategy for picture categorisation by combining neural network and similarity-based techniques to enhance accuracy.

1) ACCURACY

Having reliable assessment indicators that represent our model's problem-solving capability is crucial for assessing the model's quality. When dealing with face authentication, we often classify the instance as either beneficial or detrimental because it is a binary problem. True positives (TP), false negatives (FN), true negatives (TN), and false positives (FP) are the four frequent instances represented by the confusion matrix in real classification. TP is one of them; it is the model's accurate prediction of a positive category sample. The model accurately predicts that samples will be negative categories when TN is present. FP occurs when the model makes the mistake of classifying a sample from a negative category as positive. FN occurs when a sample that should be a positive category is predicted to be a negative category by the model.

Standard statistical test assessment metrics, such as accuracy, precision, recall, and F1_score, were used to analyse the performance of all experiments. Accuracy was determined by the proportion of correct predicted labels in Eq. (3):

$$\textit{Accuracy}\ (\boldsymbol{A}) = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FN} + \boldsymbol{FP}} \tag{3}$$

The performance of a system's algorithm with different modalities was evaluated using accuracy. Table 2 provides a summary of the accuracy results obtained from various algorithms or models across different modalities, including VIS, Th, IR, and a combination of VIS-IR. The algorithms demonstrated highly impressive accuracy levels, achieving 100% accuracy on both VIS and Th modalities, which shows their ability to verify samples within these categories correctly. The IR modality had a slightly lower accuracy of 95%, indicating that the algorithm could still classify the majority of samples accurately. Furthermore, the algorithm maintained perfect accuracy when dealing with a combined dataset of VIS-IR information. These results highlight the algorithm's effectiveness and versatility across diverse modalities, making it a compelling choice for applications involving multi-modal data. It is essential to consider various evaluation metrics when assessing a dataset's accuracy. While accuracy is a crucial metric, other factors such as potential challenges and variations in the dataset should be taken into account. For instance, lighting can have an impact on accuracy, particularly when dealing with infrared images that feature different lighting conditions, as illustrated in Figure 6.

TABLE 2. The performance of the VGG16-PCA-NN algorithm in different modalities.

Algorithm	Visible (VIS)	Thermal (Th)	Infrared (IR)	Visible and IR (VIS-IR)
Accuracy	100	100	95	100

FIGURE 6. An example of different lighting conditions.

2) RECALL

Focuses on the ability of the model to identify all relevant instances correctly. Recall is particularly important in scenarios where missing a positive detection has serious implications. The model's recall performance across different image types is outlined, showing perfection in VIS, Th, and VIS-IR images as illustrated in Figure 7.

The recall is the ratio of TP cases to the sum of TP and FN as represented in Eq. (4):

$$\textit{Recall}\ (\boldsymbol{R}) = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{TN}} \tag{4}$$

The data presented in Figure 7 shows the recall percentages for an algorithm tested on different image types, including VIS, Th, IR, and a combined set of VIS-IR images. The algorithm performs exceptionally well on VIS and Th images, achieving a high recall score, approaching 100%, in both cases, which indicates its ability to capture all relevant instances of the target class effectively. Although the recall percentage on IR images is slightly lower at 95%, it still reflects strong performance in identifying relevant instances within this specific dataset. Notably, the algorithm maintains its high performance on the joint dataset of VIS-IR images, achieving a near-perfect recall score of 100%. Overall, the algorithm shows robust capabilities across diverse image types, particularly excelling in the VIS and Th domains.

3) PRECISION

Examines the proportion of true positive predictions in the total positive predictions made. Precision is crucial when

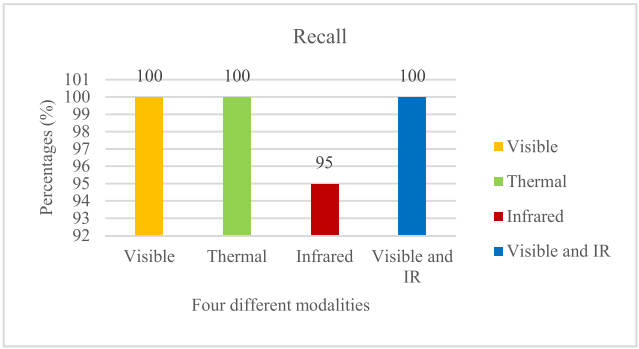


FIGURE 7. Algorithm recall performance across different image modalities.

false positives are costly. Figure 8 details precision scores for various datasets, indicating the model's precision efficiency.

The proportion of genuine positives relative to the total number of positives was called precision. Precision was characterised in Eq. (5) as:

$$\textbf{\textit{Precision}}\ (\boldsymbol{P}) = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \tag{5}$$

In Figure 8, we can see the precision analysis of different algorithms applied to various datasets. The results show that the algorithms performed exceptionally well. They achieved a perfect precision score of nearly 100% for both the VIS and Th datasets, which indicates that all positive predictions for these datasets were correct. For the IR dataset, the precision score was still high at 97.5%, suggesting that there were mostly accurate positive predictions, but with a small percentage of misclassification (2.5%). It is worth mentioning that when the algorithms were tested on a combination of VIS-IR data, they maintained a perfect precision score, which implies that the positive predictions were accurate. Overall, the algorithms demonstrated robust performance across diverse datasets with minor variations in precision, which proves their effectiveness in handling various types of data.

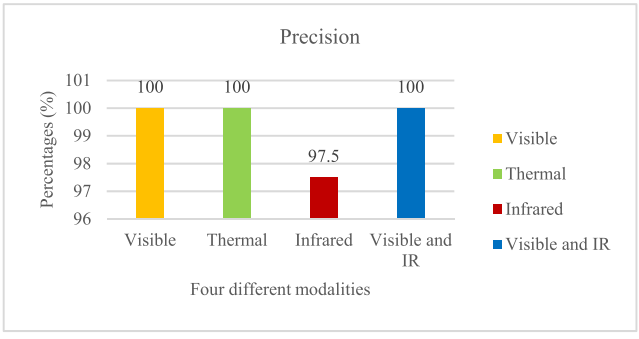


FIGURE 8. Precision analysis across multiple modalities for algorithm performance.

4) F1-SCORE

Combines precision and recall into a single metric that provides a balanced view of the model's performance. The F1-Score is particularly useful when dealing with imbalanced datasets. The model's F1-score across different spectral datasets demonstrates its balanced capability in classification.

The F1 Score is a metric that combines accuracy and recall. The F1 Score was characterised in Eq. (6) as:

$$\boldsymbol{F1\_Score} = \frac{\boldsymbol{P} \times \boldsymbol{R}}{\boldsymbol{P} + \boldsymbol{R}} \tag{6}$$

After analysing the F1-Score percentages shown in Figure 9, it can be concluded that the algorithm performs well on different spectral datasets. The algorithm obtained a near-perfect F1-Score of 100% on both the VIS and Th datasets, indicating an accurate and reliable classification in those spectral domains. Additionally, the algorithm demonstrated strong performance on the IR dataset, achieving a high F1-Score of 95.57%. This suggests that the model is effective in classifying images from the IR spectrum. Moreover, when dealing with a combined dataset of VIS-IR images, the algorithm maintained a near-perfect F1-Score of nearly 100%, showcasing its capability to distinguish and classify accurately across diverse image types. In conclusion, the algorithm is versatile and effective in handling images from different spectral domains, making it suitable for multi-spectral image classification tasks.

After evaluating the performance of the algorithms across different spectral datasets, it was found that they are highly effective in multi-spectral image verification. The models were able to achieve perfect accuracy, recall, precision, and an F1-Score of nearly 100% for the VIS and Th datasets, which highlights their robustness in accurately classifying images within these spectral domains. Similarly, the combined VIS-IR dataset also showed flawless scores, indicating the algorithms' adaptability to diverse spectral information. Although the IR dataset had a slightly lower accuracy of 95% and F1-Score of 95.57%, it still maintained high precision and recall, which signifies a balanced trade-off between false positives and false negatives. Overall, the consistently high scores in accuracy, recall, precision, and F1-Score emphasize the algorithms' proficiency in multi-spectral image classification, showcasing their suitability for a broad range of applications across various spectral domains.

5) ERROR RATE

Measures the rate at which the model makes incorrect predictions. A lower error rate signifies a more accurate model. Figure 10 represents the model's low error rates across different imaging modalities, highlighting its robustness.

Error rate measures inaccuracies or losses in a system. Reducing errors simplifies the system. Eq. (7) quantifies the error rate.

$$\textbf{\textit{ErrorRate}}\ (\boldsymbol{ER}) = \frac{\boldsymbol{AE} - \boldsymbol{EE}}{\boldsymbol{EE}} \times 100 \tag{7}$$

The error rate of the system is indicated by ER, the actual error rate of the system is represented by AE, and the estimated error value is designated by EE.

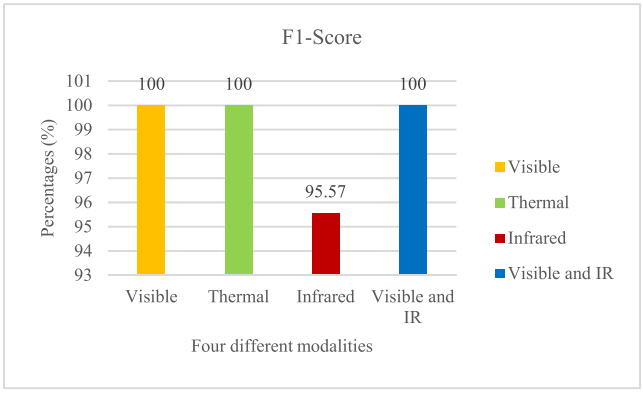


FIGURE 9. Analysis of F1-Score performance for multi-spectral image classification across four different modalities.

In Figure 10, the performance of algorithms is demonstrated across a variety of imaging modalities, such as VIS, IR, and Th imaging, as well as a combination of VIS-IR. The algorithms exhibit accuracy, with error rates of 0% in both the VIS, Th and a combination of VIS-IR modalities, showing that they work flawlessly in these domains concerning performance. It has been noted that the IR modality has a low error rate of 5%, which indicates that it has a high level of accuracy but only a tiny percentage of misclassifications.

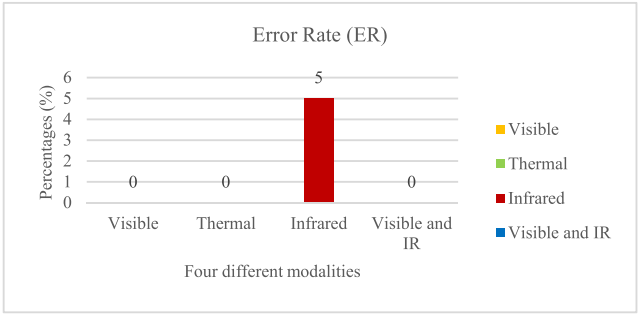


FIGURE 10. Analysis of error rates for algorithm performance across different imaging modalities.

6) COSINE SIMILARITY INDEX

PCA transforms the original feature space into a new space defined by the principal components (PCs). Each PC is a linear combination of the original features, with coefficients (or loadings) that indicate the weight of each original feature in that component. The inner product between the PCs and the original features can be understood through these coefficients. For a given PC, its inner product with the original feature set essentially reproduces the PC itself, because the PC is defined as a linear combination (inner product) of the features with its coefficients.

CS (Cosine Similarity) is a measure of similarity between two vectors that calculates the cosine of the angle between them. It is widely used in high-dimensional space to compare the orientation (but not the magnitude) of vectors. After applying PCA, the features of both the test and training sets are transformed into a reduced-dimensional space where each dimension corresponds to a principal component. This transformation retains the most significant variance directions of the data. Given 'test_PCA' and 'train_PCA', which are the PCA-reduced representations of your test and training data respectively:

i Standardise Data: Ensure both test and training sets are standardised (mean = 0, variance = 1) before PCA application, as PCA is sensitive to the scale of the data.
ii Apply PCA: Apply PCA to both the training and test datasets to reduce dimensionality. This process involves calculating the eigenvectors (PCs) and eigenvalues from the covariance matrix of the data.
iii Calculate Cosine Similarity:
a. Use the 'cosine_similarity' function from libraries such as 'sklearn.metrics.pairwise' in Python.
b. The input to this function will be 'test_PCA' and 'train_PCA'.
c. The output is a matrix where the element at (i, j) represents the cosine similarity between the i^{th} test PCA feature vector and the j^{th} training PCA feature vector.

Once the matrix of CSI has been generated, it is utilised for a variety of reasons, including the determination of the indices that are most similar, the computation of the mean CSI, and the utilisation of the matrix for additional study or evaluation. According to our methodology, the angle between feature vectors is more important than their size when it comes to determining the degree of similarity between them as represented in Figure 11. The CSI is particularly valuable in this regard. The results of the CSI tests performed on the algorithms in each of the four modalities are presented in Table 3.

TABLE 3. Face and facial components detection using the Viola-Jones technique.

Algorithm	Visible (VIS)	Thermal (Th)	Infrared (IR)	Visible and IR (VIS-IR)
Cosine Similarity Index (%)	89.4	87.03	82.9	89.39

Algorithms used to VIS, Th, IR, and VIS-IR are presented in Table 3 with their respective cosine similarity indices (%). The method demonstrated outstanding performance in the VIS spectrum, with a maximum CS of 89.4%, indicating a high degree of similarity between the training and test data. At 87.03%, the Th modality was somewhat more comparable to the IR modality, which was 82.9% weaker. The VIS-IR modalities showed a strong cosine similarity of 89.39%, indicating that the two spectra may work together and complement each other. These results suggest that visual information spectrum pattern recognition is successful and that combining visual and infrared information is

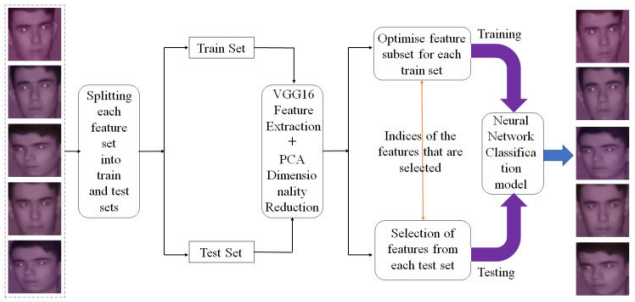


FIGURE 11. Flowchart of cosine similarity indices.

useful for obtaining strong similarity within the framework of the provided techniques and modalities. Understanding domain-specific needs and interpreting similarity measures according to data properties and application goals are of utmost importance.

7) CONFUSION MATRIX

When evaluating the effectiveness of a classification system, the confusion matrix (CM) is an essential assessment tool that should not be overlooked. It offers a tabular representation of the predictions made by the model, including particular information on the quantity of TP, TN, FP, and FN. Within the parameters of our discussion, each row of the matrix refers to examples that belong to a real class, but each column is associated with instances that belong to a class that is anticipated. These elements can be expressed mathematically in the Eq. (8):

$$\mathbf{CM} = \begin{bmatrix} \mathbf{TN} & \mathbf{FP} \\ \mathbf{FN} & \mathbf{TP} \end{bmatrix} \tag{8}$$

where:

TN is the total number of negative predictions that were accurate.

FP stands for the number of positive predictions that were erroneous.

FN represents the total number of negative predictions that were made with an error.

TP stands for the number of positive cases that were accurately anticipated.

To produce the CM by utilising the code that has been supplied, it is necessary to utilise the CM function that is available in the scikit-learn package. After using the model to make predictions about the labels on the test dataset, the function constructs the matrix. The matrix may then be visualised through the use of a heatmap by utilising the Seaborn library.

The components of the CM that are diagonal indicate that the predictions are accurate, whereas the off-diagonal elements suggest that the forecasts are inaccurate. The intensity of the colour or the values of the cells in the heatmap transmits the number of forecasts or the percentage of predictions in each category. Figure 12 represents the CM for four different modalities, such as VIS, Th, IR, and VIS-IR.

In this confusion matrix, we can see the outcomes of four modalities' classifications over several classes. The representation of Figures 12a, b, and d is a 20 × 20 matrix. If the values on the diagonal are non-zero, then the matrix is diagonal-dominant. All three examples in each class were correctly predicted by the model, as shown by the diagonal. The symmetrical nature of the matrix indicates that class performance is fairly even. This evaluation's visual picture classifications are perfectly accurate, as seen by this symmetric pattern, which means the model is doing well. Every class has been accurately predicted by the model, demonstrating its outstanding performance. If the model's predictions are accurate and dependable across all classes, it indicates that it has learned and applied patterns from the VIS, Th, and VIS-IR pictures successfully.

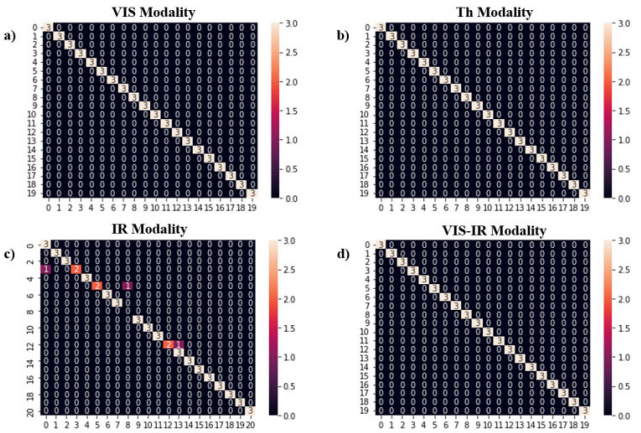


FIGURE 12. Confusion matrix for four different modalities a) VIS, b) Th, c) IR, and d) VIS-IR.

Figure 12c presents a matrix with dimensions of 21 × 21. With three TPs in the first class and zeros for every other element in the row, there were no misclassifications for this class, as seen in the top-left quadrant of the matrix. A high degree of precision is demonstrated by the pattern's repetition across the diagonal. A small number of off-diagonal items show signs of misclassification. Take the fourth row and first column as an example. There is one case of misclassification, which indicates a FN for the fourth row (Class 4). This suggests that the real person is 4, but the algorithm is forecasting them as 1. The ninth column and sixth row both show similar instances. In one case, the algorithm incorrectly predicted that a person in the sixth row (Class 6) would be a 9, when in fact, it was a 6. This pattern repeats itself in the thirteenth column and thirteenth row as well. In one case, the algorithm incorrectly predicted a person's age as 14, whereas the real identified person is 13. This occurred in the thirteenth row (Class 13). Beyond these three incorrect classifications in row 9 and column 9, the algorithm was unable to determine the individual's identity. This might be due to factors such as the lighting conditions, background noise, hair, or the various positions of the head at different angles.

Based on the results, it can be concluded that all four confusion matrices work admirably. Predictions for the majority of classes are accurate, as indicated by the strong diagonal lines in each matrix. It is notable that the models linked to these matrices can accurately categorise examples within each class. Even though all the matrices are very accurate, the last one has a little different structure because of mistakes in the fourth, sixth, and thirteenth classes. Another thing: the system has absolutely no idea who was in [9, 9]. The total performance, nevertheless, is unaffected by these few misclassifications.

8) RECEIVER OPERATING CHARACTERISTIC CURVE AND AREA UNDER CURVE

The model's performance across multiple classes is evaluated by plotting the Receiver Operating Characteristic (ROC) curve. One popular way to see the trade-off between sensitivity and specificity at different categorisation thresholds is via a ROC curve. Here, we determine the Area Under the Curve (AUC) for every one of the 59 classes by computing ROC curves. By selecting the classes with the highest AUC, we can see how well the model discriminates against those classes by plotting their ROC curves. The definitions of FPR (False Positive Rate) and TPR (True Positive Rate) are shown in Eqs. (9) and (10) [31].

$$\textit{False Positive Rate } (FPR) = \frac{FP}{FP + TN} \tag{9}$$

$$7\textit{True Positive Rate } (TPR) = \frac{TP}{TP + FN} \tag{10}$$

A predetermined probability threshold can determine the label for the final prediction, and the probability is the total output value for a classification model. When you change the threshold value, the TPR and FPR will change as well. If the ROC curve is as close to the top left corner of the graph as possible a sign of a successful model and TPR = 1 and FPR = 0 then the classification scenario is excellent [31].

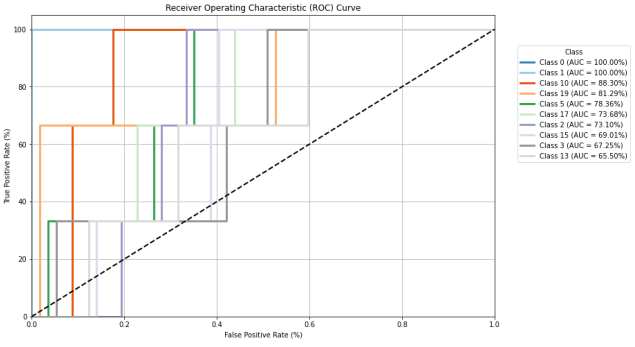


FIGURE 13. Evaluating the efficacy of visible facial recognition algorithms through ROC analysis.

The ROC curve, shown in Figure 13, is a useful tool for comparing a classifier's performance across various users or scenarios. A perfect score of 1.0 for a particular class indicates flawless facial recognition, and it is used as a measure of the classifier's accuracy. For example, a perfect score of 100% indicates that all faces were accurately matched to their respective identities, with no cases of misidentification. The results of a model that uses many classes for categorisation are shown by this ROC curve. Some classes are predicted with 100% accuracy, while others are predicted with less precision, demonstrating varying levels of discriminating ability across different classes. Class 0 and Class 1 are the best-case scenarios when the face recognition algorithm accurately detects these specific face classes, with AUCs of 100%.

The system's accuracy varies among classes, as seen by the range of AUC values in Figure 13. Factors like as lighting, facial expressions, backdrops, occlusions, or underlying similarities between faces may contribute to the system's less consistent face recognition for classes like 15, 3, 13, and others, as indicated by lower AUC values for these classes.

The classifier's performance in differentiating a specific face from the entire dataset may be more easily evaluated with the help of the color-coded curves for each class. These curves appear to be measuring the system's performance at distinct threshold levels, given their shape with steps and edges. Maybe this level of detail is a reflection of how sensitive the face recognition system is to changes in the probability criteria that determine if a face is a match or not.

Reliability and security in multi-class face recognition systems depend on the correct recognition of each class. Differences in AUC and ROC curves reveal which face recognition classes the system excels at and fails miserably at detecting; this information might inform future training and development efforts.

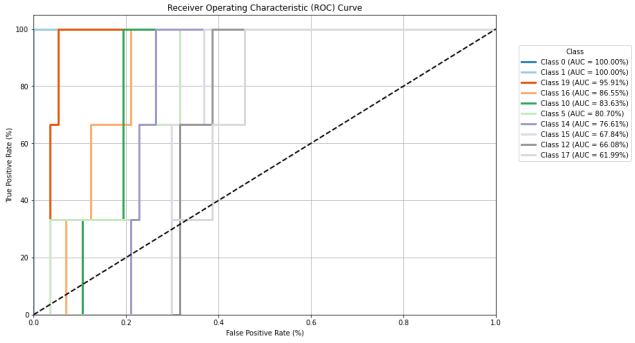


FIGURE 14. Evaluation of thermal facial recognition: ROC curves and class performance.

Figure 14 demonstrates a variety of curves that are associated with different people or facial expressions as recorded by thermal imaging. The x-axis displays the FPR, which is the proportion of faces that do not match that were mistakenly marked as matches; the curves, on the other hand, indicate the various classes. The TPR, or the proportion of properly detected real-matching thermal pictures, is shown on the y-axis. A better fit between the curve and the ROC space's left and top borders indicates a more accurate categorisation.

The dotted diagonal line on the graph depicts an AUC of 0.5 for a random estimate. A point in the top left corner

would represent an ideal classifier with an area under the curve (AUC) of 1. Figure 14 shows that the thermal facial recognition system successfully distinguished between Class 0 and Class 1 with an AUC score of 100%.

On the other hand, class recognition becomes more challenging as we go to further classes and the AUC values drop. While Class 16 and Class 10 both have AUCs above 80%, suggesting great accuracy, Class 19 has an AUC of 95.91%, which is still exceptional. Class 15, Class 12, and Class 17 all have AUC values that indicate average performance. The system's facial recognition capabilities may be compromised in certain classes. This might be due to the fact that thermal imaging has some limitations, such as a reduced ability to show facial features compared to visible light imaging clearly.

In conclusion, our findings indicate that the thermal facial recognition system excels in some classes but might benefit from further training data or tuning to achieve better accuracy in all classes. Different facial thermal signatures, changes in heat patterns caused by expressions or the environment, or sensor resolution issues could all explain the different performances seen between classes.

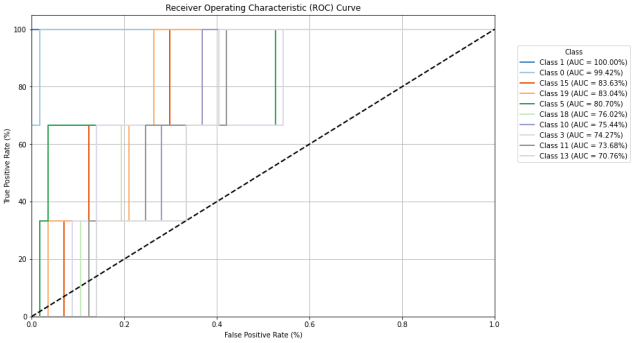


FIGURE 15. Evaluation of Infrared face recognition system performance using ROC curve analysis.

Figure 15 represents a ROC curve likely utilised for assessing an infrared facial recognition system. Each curve represents distinct classes, which might correlate to different persons or circumstances observed in infrared imaging. These classes are crucial for detecting characteristics not apparent in the regular spectrum.

Each line on the graph reflects a distinct class, presumably indicating a single individual or a distinctive facial expression captured under infrared imaging settings. The AUC is a single numerical measure that ranks classifiers based on their ability to differentiate between positive and negative classes. An AUC of 1.0 indicates flawless identification, whereas an AUC of 0.5 implies a lack of discriminatory ability.

Class 1 shows an optimal situation with an AUC of 100%, indicating that the infrared facial recognition system has accurately categorised these faces. Class 0 demonstrates exceptional performance with an AUC of 99.42%, while there is a little margin for error.

Figure 15 indicates different degrees of accuracy for additional classes as well. Classes with AUCs of around 80%, including Class 15, 19, and 5, demonstrate strong classification performance. Classes 18, 10, and 3, with AUCs ranging from 70% to 80%, have intermediate classification ability. The differences may be caused by the distinct difficulties associated with infrared imaging, such as fluctuating heat patterns resulting from physiological or environmental factors. The color-coded curves provide a rapid visual comparison of each class's performance, emphasising the infrared system's capacity to distinguish one class from the rest efficiently. Steep shifts in the curve indicate distinct threshold values employed in classification, affecting the system's ability to distinguish between genuine and false positives.

The ROC curve study of an IR face recognition system demonstrates its superior performance in low-light or nighttime circumstances compared to typical visible-light facial recognition systems. The variation in AUC values among different classes indicates inconsistent performance of the system, possibly influenced by factors like image quality, facial heat signature distinctiveness, environmental conditions, and algorithm sensitivity to infrared image features. Overall, our findings suggest that the diversity of the approach is very beneficial for certain classes but also highlights areas that require development. Additional examination indicates possible improvements, like refining the classification methods, modifying threshold values, or augmenting the training dataset with a wider variety of infrared face photos. These actions will enhance the system's precision and durability in the infrared spectrum.

Figure 16 displays several classes, some of which have an Area Under Curve (AUC) of 100%, indicating perfect classification with no false positives or false negatives for those specific conditions or individuals. Classes 10 and 19 have AUCs close to 100%, signifying excellent classification performance.

On the other hand, classes like 7, 2, 17, 12, and 16 have lower AUCs, indicating a less accurate performance. A lower AUC reflects a higher chance of misclassification. In the context of face recognition, this could mean that the system is less able to distinguish between different individuals or the same individual under varying conditions (e.g., different lighting or angles).

A detailed analysis would consider the specifics of the face recognition system being evaluated, including the technology used (visible light, infrared), the database of faces, and the operational requirements of the system (e.g., whether it is used for verification or identification, and in what sort of environmental conditions it must operate). The high-performing classes could represent more distinct individuals or conditions that are easier for the system to identify, while the lower-performing classes may represent more challenging cases for the system.

The steps and borders of the curves give information about the categorisation threshold levels. At any given threshold, the system's capacity to distinguish between positive and negative classes is proportional to the steepness of the corresponding curve. Our results conclude that this ROC curve

adequately assesses the face recognition system's performance over a range of VIS-IR pictures for different classes. Class identification is successful for some, but not all; this might be because it is difficult to combine visible and infrared data, which could require fine-tuning the system or using sophisticated computational techniques.

Every classification algorithm in this novel fusion technique links a rank with every registered characteristic in the network, when it has a higher rank that means a good match. It combines numerous UBS matcher outcomes and calculates a new rank to assist in predicting the final choice [32], [33]. For identification rather than verification, rank-level fusion is commonly used. The working techniques, in this case, are as follows: firstly, construct a rank of IDs ordered by all modalities. Secondly, the rating for every person is provided for different fused modalities using any type of fusion. Lastly, the identification with the lowest score is recognised as the right one [34].

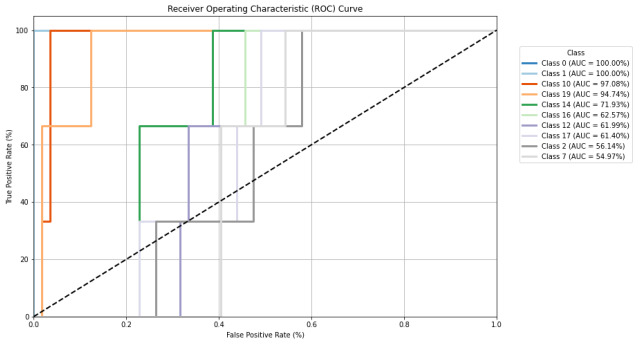


FIGURE 16. Evaluating the performance of visible-infrared face recognition systems using ROC curve measurements.

Apart from ordering the IDs depending on similarity/distance, it does not require any normalising technique [35]. This strategy is more accurate than just picking the appropriate match with one modality. In contrast to match score level fusion, it is simple to compare the rankings from multiple biometric modalities. Therefore, making a choice is quite simple [34].

Nevertheless, there is one drawback to this form of fusion. In the case of MBSs, when several identities are generated from several matching modules, some identities of just one matcher emerge, and incorrect findings pose a danger of reaching rank-level fusion [36]. Rank level fusion, in contrast to match score level fusion, delivers less data. It is superior since it assigns a rank to various matches and allows weights to be provided to specific classifiers [34]. Earlier research on rank-level fusion using fusion techniques and modalities merged is detailed in [32], [36], [37], [38], and [39] as an instance. In general, it is still vastly understudied.

Our approach also investigates Rank-level Fusion, which merges neural networks and Cosine Similarity predictions. It scores all of the predictions and uses accuracy as an evaluation metric for performance. To demonstrate how well the model identifies relevant occurrences, the Rank-level Fusion accuracy is shown against the number of top matches or ranks.

Rank-level fusion curves are used to measure the success of biometric recognition systems. The following graphs show the relationship between rank and the recognition rate, as shown in Figure 17. There are four distinct modalities for face recognition: VIS, Th, IR, and a combined VIS-IR model.

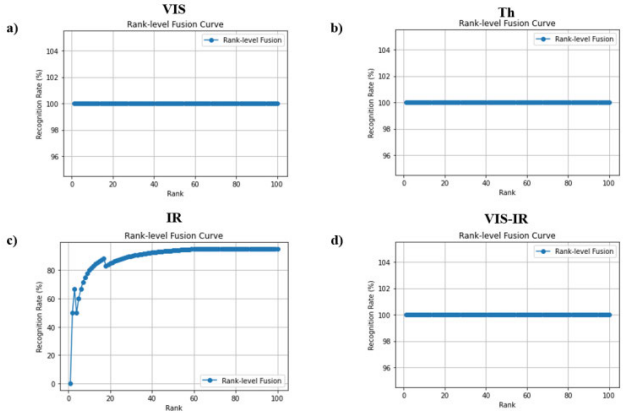


FIGURE 17. Evaluation of rank-level fusion in facial recognition using visible, thermal, infrared, and a combination of visible and infrared modalities.

Figures 17a and b illustrate that the recognition rate is at or just under 100% across all levels, maintaining a consistently high level. Regardless of rank, this shows that the recognition rate is nearly flawless, with nearly every query image accurately matching its associated identity in the database. The system's efficacy in these two modalities under the testing settings is demonstrated by the flatness of the curve, which shows that the top-ranked findings are frequently true. Figure 17c displays a different trend, but the identification rate is low at the beginning and rapidly climbs until it reaches a plateau at about the 20th rank. This indicates that the best match for an IR image might not always be at the top of the results but could be farther down the page. Possible causes for the system's less reliable performance in this modality compared to VIS and Th modalities include poor IR picture quality, unfavourable ambient circumstances, or algorithmic constraints while processing IR data.

A high, flat slope similar to the VIS and Th plots may be observed in Figure 17d as well. So, it is clear that merging visible and infrared data produces excellent identification rates at every level. This area shows promising results, which may indicate that the system's person identification accuracy is improved by using a more complete picture derived from visible and infrared spectra.

Our research shows that both the VIS and Th modalities consistently produce high-quality results with reliable identification rates. Nevertheless, there is room for development as the IR modality alone displays variability. Conversely, it appears that the VIS-IR combination modality compensates for the IR modality's limitations, leading to very accurate

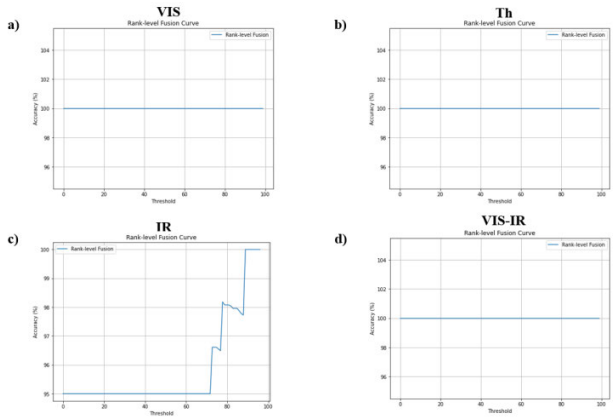


FIGURE 18. Accuracy performance of rank-level fusion in face recognition: visible, thermal, infrared, and a combination of visible and infrared modalities with threshold variation.

identification. This is why multimodal biometric systems are so beneficial.

Lastly, the threshold for combined predictions is varied to construct a Rank-level Fusion Curve. You can see the model's identification rate under various settings, as well as how the accuracy varies when the threshold is increased, on the curve. This comprehensive review sheds light on the model's efficiency and adaptability to various dataset circumstances.

Figure 18 demonstrates the accuracy of face identification for several modalities, shown versus the threshold parameter. These modalities include VIS, Th, IR, and VIS-IR.

The presence of flat, almost 100% accuracy lines in Figures 18a and b indicates that, for these modalities, nearly all genuine matches are reliably identified regardless of the threshold value. This demonstrates that the system is robust and dependable over different threshold levels, which is indicative of a high degree of accuracy in face recognition.

Figure 18c shows that when the threshold rises, the accuracy varies significantly. At certain threshold values, accuracy improves dramatically, as shown by the multiple steps in the curve. There appears to be a correlation between the threshold parameter and the system's IR performance, as seen by the progressive rise. Accuracy may be greatly improved with the correct threshold setting, however calibration may be necessary to prevent misidentification.

In addition, a flat line that is highly accurate, like the VIS and Th modalities, is shown in Figure 18d. This suggests that a robust system, less vulnerable to changes in threshold and maintaining high accuracy across varied threshold levels, is produced by merging VIS and IR data.

One possible interpretation of the threshold parameter is a cutoff value for determining the truth or falseness of a match. When it comes to face recognition systems, a lower threshold might lead to a larger false match rate (more candidates being deemed matches) and a higher false non-match rate (more actual matches being excluded) than a lower criterion. An important finding for the dataset we used, where lighting conditions and face angles might fluctuate, is that the systems are resistant to these variations, producing correct matches over a broad range of threshold settings. This is supported by the flat curves for VIS, Th, and VIS-IR, according to our study. On the other hand, the IR system's performance is more unpredictable, which suggests that finer-grained threshold adjustment is necessary for peak performance.

TABLE 4. Comparative analysis of facial recognition technologies across diverse methodologies and applications.

References	Method(s)	Dataset	Accuracy
[21]	An integrated Cross Modality Discriminator Network (CMDN) designed for High Frame Rate (HFR).	Sejong Face	92.4%
[22]	Utilises photos obtained through various media such as visible light, thermal imaging, and infrared imaging.	Sejong multi-modal disguised face dataset.	Almost 99% accuracy
[23]	Triple triplet method using multiple CNNs for Thermal-Visible face recognition.	Sejong Face	90.61%
Proposed Method	Pre-trained VGG16 CNN, PCA for dimensionality reduction, and sequential neural network for classification. (VGG16-PCA-NN)	Sejong Face Database (SFD), focusing on Visible, Thermal, Infrared, and a combination of Visible and Infrared spectrums.	Approaching 100% for VIS, Th, and VIS-IR 95% for IR

D. COMPARATIVE ANALYSIS USING THE LATEST MODELS

In this part, we offer a statistical evaluation that compares the results with those of modern modelsusing the Sejong Face Database as a dataset. Reliability served as the basis for the statistical study. The results of our model compared to the modern models, including the Sejong Face Database, are summarised in Table 4.

Our proposal presents a novel approach to image recognition using a sequential neural network for classification, a VGG16 CNN for feature extraction, and PCA for

dimensionality reduction. We highlight the importance of transfer learning, dimensionality reduction, and neural networks in achieving robust and efficient image recognition that performs well across different modalities. Our proposed approach achieves 100% accuracy in visible (VIS) and thermal (Th) modalities, as well as in a combination of visible and infrared (VIS-IR). Infrared (IR) accuracy is slightly lower.

V. CONCLUSION AND FUTURE WORK

The integration of CNNs, PCA, and sequential neural networks to improve facial verification across different imaging modalities has been presented and analysed. The imaging modalities included visible light, thermal, infrared, and a combination of visible light and infrared images from the Sejong face database. The proposed approach utilises the pre-trained VGG16 framework on the ImageNet dataset to extract features. PCA is subsequently employed to decrease the dimensionality, resulting in quicker calculations. The method is improved by using a sequential neural network model for classification, combining the benefits of transfer learning and dimensionality reduction.

Conclusion:

Our study demonstrates that the VGG16-PCA-NN method achieves high accuracy across multiple modalities, particularly excelling in combined VIS-IR scenarios. While these results are promising, further testing on diverse datasets and real-world conditions is essential to validate these findings comprehensively.

This work enhances facial recognition by introducing a robust approach that addresses variations in lighting and physical obstacles. With high accuracy rates, it shows significant potential for multi-modal applications. It will significantly contribute to the advancement of more secure and dependable biometric verification systems.

An analysis of the constraints of the suggested facial recognition system, emphasising opportunities for enhancement. One of the shortcomings of the system is its inconsistent accuracy when dealing with different classes. This inconsistency can be related to several aspects such as lighting conditions, facial emotions, face angle, backdrops, occlusions, or similarities across faces. This heterogeneity indicates the necessity for further training and development initiatives to improve performance consistently in all categories. In certain situations, the thermal facial recognition system is effective, but it has obstacles in other cases. These challenges may arise from the limits of thermal imaging in accurately capturing facial characteristics, as contrasted to visible light imaging. This suggests that accuracy might be enhanced by the utilisation of further tuning or training data. The diversity observed in the IR modality performance can be attributed to factors such as suboptimal IR picture quality, adverse environmental circumstances, or computational constraints in IR data processing. This emphasises the necessity for improvement in the IR modality to attain recognition rates that are as dependable as those of the VIS and Th modalities. In addition, our work examines the difficulties related to rank-level fusion, namely the incorporation of numerous identities produced by different matching modules. Inaccurate results in this process might potentially affect the outcomes of rank-level fusion. Although this fusion approach requires less data compared to match score level fusion, it still poses issues that need to be further examined. These limits highlight the necessity for continuous research and development to tackle these issues, enhance system performance in different modalities, and guarantee the dependability of face recognition technology.

One critical area for future research is improving the system's robustness against presentation attacks, such as spoofing. Integrating techniques like liveness detection or adversarial training into the VGG16-PCA-NN framework will strengthen security and resilience in real-world applications, making the system more robust to potential threats.

Future research in multi-modal facial recognition will focus on (i) integrating additional modalities, comparing them with individual modalities, (ii) refining algorithms for increased accuracy, (iii) applying the technology in real-time scenarios, (iv) utilising PCA for eigenvalues, and eigenvectors, as well as VGG16 for feature extraction and neural network classification, and (v) three different algorithms will be used and compared, namely VGG16-RF (Random Forest)-NN, VGG16-LR (Logistic Regression)-NN, and VGG16-SVM (Support Vector Machine)-NN.

REFERENCES

[1] P. J. Phillips, A. N. Yates, Y. Hu, C. A. Hahn, E. Noyes, K. Jackson, J. G. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J.-C. Chen, C. D. Castillo, R. Chellappa, D. White, and A. J. O'Toole, "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 24, pp. 6171–6176, Jun. 2018.

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[3] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016.

[4] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–12.

[5] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6738–6746.

[6] S. Hu, J. Choi, A. L. Chan, and W. R. Schwartz, "Thermal-to-visible face recognition using partial least squares," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 3, pp. 431–442, 2015.

[7] C. Chen and A. Ross, "Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework," *Pattern Recognit. Lett.*, vol. 72, pp. 25–32, Mar. 2016.

[8] M. Saquib Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for thermal to visible face recognition," 2015, *arXiv:1507.02879*.

[9] A. Kantarci and H. K. Ekenel, "Thermal to visible face recognition using deep autoencoders," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–5.

[10] C. N. Fondje, S. Hu, N. J. Short, and B. S. Riggan, "Cross-domain identification for thermal-to-visible face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep. 2020, pp. 1–9.

[11] K. Mallat, N. Damer, F. Boutros, A. Kuijper, and J.-L. Dugelay, "Cross-spectrum thermal to visible face recognition based on cascaded image synthesis," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

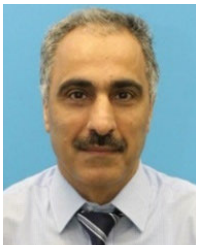
[12] Z. Wang, Z. Chen, and F. Wu, "Thermal to visible facial image translation using generative adversarial networks," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1161–1165, Aug. 2018.

[13] L. Kezebou, V. Oludare, K. Panetta, and S. Agaian, "TR-GAN: Thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition," *Proc. SPIE*, vol. 11399, pp. 158–168, Apr. 2020.
[14] R. Immidisetti, S. Hu, and V. M. Patel, "Simultaneous face hallucination and translation for thermal to visible face verification using axial-GAN," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–8.
[15] D. Anghelone, C. Chen, P. Faure, A. Ross, and A. Dantcheva, "Explainable thermal to visible face recognition using latent-guided generative adversarial network," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.
[16] X. Cao, K. Lai, G. J. Hsu, M. Smith, and S. N. Yanushkevich, "Cross-spectrum thermal face pattern generator," *IEEE Access*, vol. 10, pp. 9576–9586, 2022.
[17] D. Poster, M. Thielke, R. Nguyen, S. Rajaraman, X. Di, C. N. Fondje, V. M. Patel, N. J. Short, B. S. Riggan, N. M. Nasrabadi, and S. Hu, "A large-scale, time-synchronized visible and thermal face dataset," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1558–1567.
[18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
[19] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, "Generative adversarial network-based synthesis of visible faces from polarimetrie thermal faces," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 100–107.
[20] X. Di, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Polarimetric thermal to visible face verification via self-attention guided synthesis," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.
[21] U. Cheema, M. Ahmad, D. Han, and S. Moon, "Heterogeneous visible-thermal and visible-infrared face recognition using unit-class loss and cross-modality discriminator," 2021, *arXiv:2111.14339*.
[22] A. M. Alkadi, R. A. AlMahdawi, S. M. Anwahi, and B. Soudan, "Biometric authentication based on multi-modal facial recognition using machine learning," in *Proc. Adv. Sci. Eng. Technol. Int. Conferences (ASET)*, vol. 10, Feb. 2023, pp. 1–6.
[23] M. Kowalski, A. Grudzień, and K. Mierzejewski, "Thermal–visible face recognition based on CNN features and triple triplet configuration for on-the-move identity verification," *Sensors*, vol. 22, no. 13, p. 5012, Jul. 2022.
[24] H. Filali, J. Riffi, A. M. Mahraz, and H. Tairi, "Multiple face detection based on machine learning," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, Apr. 2018, pp. 1–8.
[25] B. R. Ilyas, B. Mohammed, M. Khaled, and A. T. Ahmed, "Wavelet-based facial recognition," in *Proc. 6th Int. Conf. Control Eng. Inf. Technol. (CEIT)*, Oct. 2018, pp. 1–6.
[26] R. Saikia and A. B. Kandali, "DWT-ELBP based model for face recognition," in *Proc. Int. Conf. Energy, Commun., Data Analytics Soft Comput. (ICECDS)*, Aug. 2017, pp. 1348–1352.
[27] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[28] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.
[29] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1335–1344.
[30] U. Cheema and S. Moon, "Sejong face database: A multi-modal disguise face database," *Comput. Vis. Image Understand.*, vols. 208–209, Jul. 2021, Art. no. 103218.
[31] X. Li and H. Niu, "Feature extraction based on deep-convolutional neural network for face recognition," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 22, p. 1, 2020.
[32] N. Radha and A. Kavitha, "Rank level fusion using fingerprint and iris biometrics," *Indian J. Comput. Sci. Eng.*, vol. 2, no. 6, pp. 917–923, 2012.
[33] G. Amirthalingam and G. Radhamani, "A multimodal approach for face and ear biometric system," *Int. J. Comput. Sci. Issues (IJCSI)*, vol. 10, no. 5, p. 234, 2013.
[34] D. T. Meva and C. K. Kumbharana, "Comparative study of different fusion techniques in multimodal biometric authentication," *Int. J. Comput. Appl.*, vol. 66, no. 19, pp. 16–19, 2013.
[35] M. L. Gavrilova and M. M. Monwar, "Current trends in multimodal biometric system-rank level fusion, in pattern recognition," in *Machine Intelligence and Biometrics*. Berlin, Germany: Springer, 2011, pp. 657–673.
[36] M. M. Monwar and M. L. Gavrilova, "Multimodal biometric system using rank-level fusion approach," *IEEE Trans. Syst., Man, Cybern., B*, vol. 39, no. 4, pp. 867–878, Aug. 2009.
[37] M. Monwar and M. Gavrilova, "Secured access control through Markov chain based rank level fusion method," in *Proc. 5th Int. Conf. Comput. Vis. Theory Applications (VISAPP)*, Angres, France, 2010, pp. 458–463.
[38] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *IEEE Trans. Syst., Man, Cybern., C*, vol. 41, no. 5, pp. 743–752, Sep. 2011.
[39] M. D. Monwar, "A multimodal biometric system based on rank level fusion," Doctoral dissertation, Dept. Elect. Comput. Eng., Univ. Calgary, Calgary, AB, Canada, 2013.

MOHAMED ABDUL-AL was born in Sidon, Lebanon. He received the B.Sc. degree in computer and communication engineering with a minor in biomedical engineering and biomedical sciences from American University of Science and Technology, Beirut, Lebanon, and the M.Sc. degree in advanced biomedical engineering and the Ph.D. degree in the development of a novel multimodal biometrics system by fusing face and infrared information from the University of Bradford, U.K., in 2020. He was a Research Fellow in the European Union's Horizon-MSCA-RISE-2019-2024, Marie Skłodowska-Curie, Research, and Innovation Staff Exchange (RISE) Program, titled: Secure and Wireless Multimodal Biometric Scanning Device for Passenger Verification Targeting Land and Sea Border Control with the University of Bradford. His research interests include face recognition using different modalities, such as visible, infrared, thermal, and visible and infrared. He also contributes to wireless communication and radio frequencies. In addition, he is involved in regenerative medicine and tissue engineering, specifically, the development of biomaterials for breast reconstruction, stem cell niche microenvironment, glioblastoma, and encapsulation techniques in the ocular and respiratory systems.

GEORGE KUMI KYEREMEH was born in Ghana. He received the B.Sc. degree in electrical engineering from the Kwame Nkrumah University of Science and Technology, Ghana, and the M.Sc. degree in advanced biomedical engineering and the Ph.D. degree in the development of a novel multimodal biometrics system by fusing fingerprint and fingers vein from the University of Bradford, U.K., in 2020. He was a Research Fellow in European Union's Horizon-MSCA-RISE-2019-2024, Marie Skłodowska-Curie, Research, and Innovation Staff Exchange (RISE) Program, titled: Secure and Wireless Multimodal Biometric Scanning Device for Passenger Verification Targeting Land and Sea Border Control with the University of Bradford. His research interests include fingerprint and finger vein recognition. He also contributes to regenerative medicine and tissue engineering, specifically, stem cell niche microenvironment.

RAMI QAHWAJI (Senior Member, IEEE) received the M.Sc. degree in control and computer engineering and the Ph.D. degree in AI and signal/image processing. He is currently a Professor of visual computing with the University of Bradford. He has been working with different industries in the fields of satellite/space imaging, communications, remote sensing, digital health and imaging, biometrics, AI, and data visualization developing intelligent systems in collaboration with NASA, ESA, NHS, and different SMEs. He attracted millions of pounds in research funding from various U.K. and European funding agencies. He has more than 140 refereed journal and conference publications and has been invited to deliver many keynote speeches at national and international conferences. He has supervised 31 completed Ph.D. projects and is an external examiner for several U.K. and international universities. He is heavily involved in the organization of international activities and public engagement events. He is a fellow of the Institution of Engineering and Technology, a Chartered Engineer, a fellow of the Higher Education Academy, and an IET Technical Assessor and sets on the IET's Healthcare Sector Executive Committee.



NAZAR T. ALI (Senior Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from the University of Bradford, U.K., in 1990. From 1990 to 2000, he held various posts with the University of Bradford as a Researcher and a Lecturer. He worked on many collaborative research projects in U.K., under the umbrella of the Centre of Research Excellence, Department of Trade and Industry (DTI), and EPSERC. This involved a consortium of a number of universities and industrial companies. He is currently an Associate Professor with Khalifa University, United Arab Emirates. His current research interests include antennas and RF circuits and systems, indoor and outdoor localization techniques, and RF measurements. He has more than 100 papers published in peer-reviewed high-quality journals and conferences.

RAED A. ABD-ALHAMEED (Senior Member, IEEE) has been a Research Visitor with Wrexham University, Wales, since 2009, covering the wireless and communications research areas, and an Adjunct Professor with the College of Electronics Engineering, Ninevah University, since 2019, and the Department of Information and Communication Engineering, College of Science and Technology, Basrah University, Basrah, Iraq. He is currently a Professor of electromagnetic and radiofrequency engineering with the University of Bradford, U.K. He is also the Leader of radiofrequency, propagation, sensor design, and signal processing with the School of Engineering and Informatics, University of Bradford, where he is also leading the Communications Research Group. He is a Chartered Engineer. He has many years of research experience in the areas of radio frequency, signal processing, propagations, antennas, and electromagnetic computational techniques. He is a Principal Investigator for several funded applications to EPSRCs, Innovate U.K., and British Council, and the Leader of several successful knowledge transfer programs, such as Arris (previously known as Pace plc), Yorkshire Water plc, Harvard Engineering plc, IETG Ltd., Seven Technologies Group, Emkay Ltd., and Two World Ltd. He has also been a co-investigator in several funded research projects, including H2020-MSCA-RISE-2024-2028, Marie Skłodowska Curie, Research and Innovation Staff Exchange (RISE), titled "6G Terahertz Communications for Future Heterogeneous Wireless Network;" HORIZON-MSCA-2021-SE-01-01, Type of Action: HORIZON-TMA-MSCA-SE 2023-2027: ROBUST: Proposal titled "Ubiquitous eHealth Solution for Fracture Orthopaedic Rehabilitation;" Horizon 2020 Research and Innovation Program; H2020 Marie Skłodowska-Curie Actions: Innovative Training Networks Secure Network Coding for Next Generation Mobile Small Cells 5G-US; European Space Agency: Satellite Network of Experts V, Work Item 2.6: Frequency Selectivity in Phase-Only Beamformed User Terminal Direct Radiating Arrays; Nonlinear and Demodulation Mechanisms in Biological Tissue (Department of Health, Mobile Telecommunications and Health Research Program); and Assessment of the Potential Direct Effects of Cellular Phones on the Nervous System (EU: collaboration with six other major research organizations across Europe). He has published more than 800 academic journals and conference papers; in addition, he has co-authored seven books and several book chapters including seven patents. His research interests include computational methods and optimizations, wireless and mobile communications, sensor design, EMC, beam steering antennas, energy-efficient PAs, and RF predistorter design applications. He is a fellow of the Institution of Engineering and Technology and the Higher Education Academy. He was a recipient of the Business Innovation Award for the successful KTP with Pace and Datong companies on the design and implementation of MIMO sensor systems and antenna array design for service localizations. He is the chair of several successful workshops on energy-efficient and reconfigurable transceivers: approach toward energy conservation and CO2 reduction that addresses the biggest challenges for future wireless systems. He has been the General Chair of the IMDC-IST International Conference, since 2020; a Co-Editor of *Electronics* (MDPI), since 2019; and a Guest Editor of *IET Science, Measurement and Technology*, since 2009.