Neurocomputing 613 (2025) 128708

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

RI-L_1Approx: A novel Resnet-Inception-based Fast L_1-approximation method for face recognition

Supriya Bajpai [a], Gargi Mishra [b,1], Rachna Jain [c], Deepak Kumar Jain [d,e], Dharmender Saini [b], Amir Hussain [f,*]

[a] IITB-Monash Research Academy, Indian Institute of Technology Bombay, Mumbai, 400076, Maharashtra, India
[b] Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, Paschim vihar, New Delhi, 110063, India
[c] Department of Information Technology, Bhagwan Parshuram Institute of Technology, Rohini, New Delhi, 110089, India
[d] Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian, 116024, China
[e] Symbiosis Institute of Technology, Symbiosis International University, Pune, India
[f] Centre of AI and Robotics, Edinburgh Napier University, Scotland, UK

ARTICLE INFO

Communicated by N. Zeng

Keywords:
Face recognition
Linear sparse approximation
Deep learning
Transfer learning
Few shot learning

ABSTRACT

Performance of deep learning methods for face recognition often relies on abundant data, posing challenges in surveillance and security where data availability is limited and environments are unconstrained. To address this challenge, we propose a novel few shot learning approach termed Resnet Inception-based Fast L_1 approximation (RI-L_1Approx) for face recognition with limited number of image samples. The method relies on L_1 norm approximation of test sample with known class samples. Initially, facial features are extracted by leveraging ResNet-Inception hybrid network's abilities to learn rich hierarchical representations from facial images. The extracted features are subsequently employed for L_1 norm approximation of known features, which are referred to as approximation samples. The L_1 norm approximation promotes sparsity by encouraging a subset of approximation samples to possess zero coefficients. This process helps in selecting the most discriminative and informative approximation samples, leading to improved classification capabilities. The proposed method is evaluated on benchmark facial recognition datasets, demonstrating its effectiveness. Comparative experiments with state-of-the-art techniques highlight its superior recognition accuracy. Remarkably, the RI-L_1Approx model achieved high accuracy rates of 84.86% with just one sample per class and 96.144% with thirteen samples per class during experimentation. This is significantly better than existing deep learning approaches, which require a large amount of data to train the model for similar performance.

1. Introduction

Face recognition (FR) is a extensively studied task within the domain of image processing [1]. Various methods have been developed for FR tasks using images from a stored database. However, FR in various applications such as surveillance, record-keeping, identity verification, attendance tracking, marketing, and several other applications encounters several typical challenges:

1. Data Scarcity: Often, FR datasets contain only a limited number of samples for each person or class. This scarcity of data can hinder the performance of FR algorithms [2].
2. Variability in Image Conditions: While datasets typically consist of images captured in controlled environments, real-world situations introduce various challenges, including issues such as low illumination, variations in head pose, and partial occlusion of the face. These variations can make it difficult for FR systems to accurately identify individuals [3].

In the literature, various approaches have been proposed for FR application, using techniques ranging from manual feature extraction to automated deep learning methods [1]. Despite efforts to mitigate the impact of image variations, these approaches often require large databases to achieve high accuracy. In recent decades, there has been a significant rise in machine-based FR methods [1]. There has been

* Corresponding author.
E-mail address: A.Hussain@napier.ac.uk (A. Hussain).
[1] Supriya Bajpai and Gargi Mishra are first authors with equal contribution.

https://doi.org/10.1016/j.neucom.2024.128708
Received 11 June 2024; Received in revised form 12 September 2024; Accepted 1 October 2024
Available online 3 October 2024

a growing emphasis on addressing specific challenges such as face localization and feature extraction [4]. Image feature extraction involves reducing an image to a compact representation that captures important information [5]. Traditional machine learning techniques such as PCA [6], SVM [7], and ICA [8] have also been explored for enhancing FR systems, which are based on statistical analysis and follow certain assumptions about the distribution of the data. These techniques aim to produce a low-dimensional representation of the image that can be used for various tasks such as classification, recognition, and clustering. The image feature extraction approach has been found to have a significant impact on the accuracy of recognition systems [9]. Additionally, hand-engineered features like Gabor and LBP [9,10], along with their extensions, have been utilized. Despite these efforts, these representations still struggle to handle complex, non-linear variations in facial images, and real-world challenges such as lighting variations, pose changes, expressions, and occlusions.

The emergence of Deep Learning, particularly since 2012, has revolutionized FR by enabling automated feature extraction through deep neural networks [5]. The deep learning techniques can learn the compact, rich and complex features from a very large set of data that makes it worthwhile for applications involving FR [11,12]. It takes a lot of time and data to build a deep neural networks from scratch. For this reason, researchers are increasingly turning to transfer learning to extract features by using pre-trained networks [13], thus saving a considerable amount of time and resources. This approach uses the weights learned from the various layers of the pre-trained network for feature extraction [13,14]. Consequently, researchers explore diverse transfer learning approaches, leveraging pre-trained deep learning models to address Face Recognition (FR) challenges [15]. Deep learning techniques, notably convolutional neural networks (CNNs), have shown impressive capabilities in managing variations in face pose, lighting, and expressions [8].

As a result, they have become the predominant approach for FR tasks. Deep learning methods are known for their high performance, but they have a strong reliance on abundant data. They often struggle to generalize effectively when trained on limited data samples, which is a common challenge in deep learning. Moreover, in scenarios when there are very few images available for each class, FR becomes particularly difficult.

In the past few years, researchers have increasingly used linear approximation techniques for FR due to their simplicity of implementation and computational efficiency [16,17]. Several approaches based on linear approximation have emerged recently [18,19]. Nonetheless, these methods have struggled to attain high performance in FR tasks.

It is observed that humans have a remarkable ability to learn new tasks even with minimal or no prior data (known as Few-Shot Learning) [9]. It offers a solution to the common problem of limited data availability in many applications by leveraging prior knowledge and supervised learning techniques [9]. By incorporating prior knowledge into the learning process, Few-Shot Learning methods can effectively generalize for new tasks and samples, even with limited data. Few-Shot Learning has been applied across various domains and tasks, including robotics, drug discovery, item recommendation, translation, and image processing [20,21]. Extending Few-Shot Learning to scenarios with extremely scarce data, where only one or no image per class is available in the training dataset, introduces the One-Shot or Zero-Shot Learning approach [22,23]. One-Shot Learning refers to situations where there is only a single labeled example per class in the training data, while Zero-Shot Learning occurs when there are no examples available during training. In Few-Shot Learning for image recognition task, only a small number of image samples are used for training [24]. This technique mimics human visual intelligence, allowing for efficient image recognition even after minimal exposure to an image. Typically, Few-Shot image recognition systems involve feature extraction from a given image followed by identity estimation based on the extracted features [25].

Although there is not much research available on applying few-shot and one-shot learning to face recognition, some methods do utilize few-shot learning techniques [26,27]. These methods require data augmentation of single training image to convert the single image per class to multiple images per class and train the model using these augmented samples which increases the computational cost of the method. Training models for one-shot and few-shot learning in face recognition is challenging due to the complexity of facial data and the limited number of examples available. High intra-class variability, where a single person's face can look very different under various conditions (like lighting, pose, and expression), makes it hard for models to generalize from just a few samples. At the same time, low inter-class variability, where different individuals' faces may look quite similar, complicates distinguishing between people. This often leads to misclassification and overfitting, where the model memorizes specific details from the few examples it has seen, resulting in poor performance on new images. Techniques like data augmentation and synthesis attempt to address these issues by artificially increasing the diversity of training data, but they often fall short of capturing the full complexity of real-world facial variations. One of the primary limitations of deep learning models are highly data-intensive. Deep learning approaches do not typically generalize well for limited datasets with fewer samples. Although these techniques try to minimize the effect of image discrepancies, they still need huge database for accurate face recognition. Additionally, when face images are acquired in an uncontrolled, uncooperative environment or when very limited images are available for each class, face recognition task becomes extremely challenging.

To address the challenges discussed above we propose a novel methodology 'RI-L_1Approx' that employs ResNet and Inception based pre-trained network to extract meaningful high-level features. Subsequently, these extracted features are used to classify an unknown test sample's class by approximating its features using the L_1 norm with known features referred to as approximation samples.

The enforcement of the sparsity constraint is achieved by ensuring that the linear combination used in our method involves only the nearest approximation samples, with all other coefficients set to zero. This means that while our method allows for a sparse representation, it specifically identifies which coefficients are zero and how many are non-zero. The sparsity constraint in our method significantly enhances computational efficiency, particularly in large-scale applications. By leveraging only a small subset of approximation samples for both expression and classification tasks, our approach simplifies the computational process. This efficient computational requirement is advantageous for handling large-scale problems effectively.

The L_1 norm used in our method increases the accuracy of face recognition task, especially when working with limited image samples, as it effectively addresses high intra-class variability and low inter-class variability. By representing a test image as a sparse linear combination of K-nearest approximation samples, which are selected from all the available samples, the method focuses on the most relevant images, which helps in distinguishing between similar-looking classes without overfitting. This sparse representation allows the model to generalize well by selecting only a few relevant approximation samples, rather than relying on memorization or data augmentation. Most methods rely on training with large data, and in cases of small training data, overfitting occurs. However, in the proposed method, only a small subset of approximation samples is used for approximation and classification tasks, making the size of the dataset less important.

The following are the objectives of the paper:

1. To conduct the background study and enlighten literature review with regard to the limitations of current FR methods with limited data captured in uncontrolled environment.

2. To propose novel methodology 'RI-L_1Approx' for FR that specifically addresses scenarios involving limited image samples captured in uncontrolled environments. RI-L_1Approx employs ResNet and Inception based network to extract meaningful high-level features. Secondly, these extracted features are used to classify an unknown test sample's class by approximating its features using the L_1 norm with known features referred to as approximation samples.
3. To conduct detailed and extensive experiments on six open-source (publicly available) standard datasets and to test and validate the proposed technique on percentage classification accuracy metric.
4. And, to compare the proposed technique with other existing state-of-the-arts techniques.

Organization of paper: Section 2 outlines the related work for FR with limited data. Section 3 consists of material and methods with dataset details. Section 4 highlights experimentation, results and analysis. And, finally Section 5 concludes the paper with future scope.

2. Related work

In this segment, we provide an overview of recent and pertinent research in the field of FR. To address constraints stemming from a lack of training data, strategies have been devised to introduce additional variations into the training set. Proposed methods include virtual sample techniques and generic learning approaches, aiming to enhance the training set by incorporating increased variations.

Virtual sample methods generate new face samples from the available training set, contributing to the progress of multi-sample FR. These techniques produce new samples using simple image operations like rotation or multi-scale image division, leading to the creation of target-specific face images with diverse angles and scales. Zhang et al. [28] propose a novel algorithm utilizing a general deep autoencoder trained on the entire gallery to generalize intra-class variations of multi-sample subjects to single-sample subjects. Subsequently, class-specific deep autoencoders are fine-tuned for each single-sample subject. By inputting samples of the most similar multi-sample subject into the corresponding autoencoder, the algorithm reconstructs new samples. Classification involves employing minimum L_2 distance, PCA, sparse representation-based classifier, and softmax regression.

Tang et al. [29] proposed a model to address issues related to illumination variations. They proposed a multilayer generative model incorporating latent variables like albedo, surface normals, and the light source. This model, combining Deep Belief Nets (DBN) with the Lambertian reflectance assumption, enables the learning of robust priors over albedo from 2D images. This approach allows for the explanation of illumination variations by manipulating only the lighting latent variable. Numerous generative adversarial networks (GANs) have been employed in virtual sample generation to enhance the simulation of real face sample distributions. Pang et al. [30] introduced VD-GAN, a unified framework for FR, combining prototype and representation learning. VD-GAN utilizes an encoder–decoder structural generator and a multi-task discriminator to handle various variations, including single, multiple, and mixed variations. The limitation of virtual sample methods lies in their inadequate handling of significant variations in the testing set. As a result, researchers have investigated generic learning-based methodologies, utilizing an auxiliary generic set to comprehend variance information. The majority of generic learning methods are based on the $P + V$ model.

Wang et al. [31] introduced a generic learning framework for enhanced recognition with limited training samples, learning intrinsic subject properties from a diverse training database. Su et al. [23] presented an Adaptive Generic Learning (AGL) method, introducing the Coupled Linear Representation (CLR) algorithm. Deng et al. [32] introduced a parameter-free FR algorithm, Linear Regression Analysis (LRA), robust to lighting, expression, occlusion, and age variations using a single gallery sample. LRA optimally embeds equidistant prototypes, ensuring maximum one-against-the-rest margin between classes, without preserving the global or local structure of training data. The novelty includes a generic learning method enhancing LRA's generalization capability, achieving improved FR even with a few generic classes from a different database. Pang et al. introduced the Synergistic Generic Learning (SGL) method [33], enhancing prototypes for contaminated samples and refining the variation dictionary by extracting intra-subject variants from an auxiliary generic set. Additionally, they presented the Iterative Dynamic Generic Learning (IDGL) method [34], employing a dynamic label feedback network for iterative prototype recovery.

Although the $P + V$ model successfully incorporates variation information into the gallery set, it concurrently introduces disruptive facial contour details. These contour features contain identifiable information from the training dataset, hindering the precise acquisition of prototypes in testing images and leading to misclassifications.

Convolutional Neural Networks (CNNs)-based techniques [35,36] have made substantial advancements in the field of FR in addition to the approaches mentioned above, increasing accuracy and performance. For face recognition and authentication applications, CNN-based techniques have gained widespread acceptance. Recent breakthroughs in face identification and verification include DeepFace [37], FaceNet [38], ArcFace [39], and MagFace [40] which have demonstrated remarkable performance. While large-scale training data and deep network architectures have been crucial, much of the progress is attributed to the evolution of training losses for CNNs. Many approaches, including contrastive loss [41] and triplet loss [38], were initially based on metric-learning-based losses; however, these approaches encountered inefficiencies while training on big datasets. As a result, studies have turned to classification-based loss functions that are more effective, including NormFace [42] and L2-softmax [43], which boost intra-class compactness and boost overall performance. New developments include adaptive margin techniques shown in AdaptiveFace [44] and AdaCos [45], which further improved performance by dynamically adjusting hyperparameters during training, and angular margin-based losses such as CosFace [46] and ArcFace [39]. Researchers have explored alternative approaches, such as FewFaceNet [27], a lightweight, few-shot learning-based face authentication method designed for edge devices. FewFaceNet utilizes a shallow backbone model and incremental learning to operate effectively with minimal data and in low-light environments, addressing the limitations of traditional approaches that require extensive resources and well-lit conditions. However, these methods often require large datasets and substantial computational resources, posing challenges for implementation on edge devices.

In recent studies, sparse representation has exhibited efficacy in the context of limited training sample FR. The algorithms for kernel-based sparse representation [47,48] work by transforming the input data into a higher-dimensional feature space. Lu et al. [15] introduced a new approach called weighted sparse representation, which merges local information into sparse approximation. Xin et al. [49] combined local information with group sparse approximation to produce a method called weighted group sparse representation for improved label integration. In recent years, linear approximations have been used for FR which provides the simplicity of implementation and computational efficiency [16,17]. A variety of methods based on linear approximation have evolved [18,19]. These methods involve finding the closest approximation of a test image by combining a small set of approximation samples in a linear way.

The ESRC method, as documented in Ref. [50], incorporates generic learning into sparse representation, creating an intra-class variance dictionary using variational faces from the generic set. Zhu et al. [51] devise a patch variation dictionary, Pang et al. [52] create a variation dictionary from low-discriminative intra-class variants, and Gao et al. [53] suggest eliminating testing set variations with stacked supervised autoencoders (SSAE) for more discriminative features. While partially addressing intra-class variance, these methods introduce identity

Table 1
Dataset description

Specifications	Image count for each class	Classes/Categories	Image size	Total instances
ORL	10	40	92×112	400
YALE	11	15	220×175	165
AR	26	65	$165 \times 120 \times 3$	1690
Georgia Tech	15	50	$131 \times 176 \times 3$	750
FEI	14	50	$640 \times 480 \times 3$	700
LFW	14	50	$250 \times 250 \times 3$	700
CFP-FP	14	50	$250 \times 250 \times 3$	700

information from the generic set, influencing testing set classification. These methods focus on generating or eliminating facial variations.

To overcome limitations arising from insufficient training data, various strategies have been devised to introduce additional variations into the training set. Proposed methods include the utilization of virtual sample techniques and generic learning approaches, with the goal of augmenting the training set by incorporating a broader range of variations. However, virtual sample methods face a drawback in their inability to effectively handle significant variations present in the testing set. To address this issue, researchers have explored generic learning-based methodologies that leverage an auxiliary generic set to comprehend variance information.

The prevalent generic learning methods often rely on the $P + V$ model. While the $P+V$ model successfully integrates variation information into the training dataset, it simultaneously introduces disruptive facial contour details. These contour features contain identifiable information from the generic set, impeding the accurate extraction of prototypes in testing images and resulting in misclassifications. Consequently, there is a need for a methodology that balances computational efficiency and high performance, particularly when dealing with a limited dataset in unconstrained environments.

3. Materials and methods

3.1. Dataset

In this section, a comprehensive description of all the datasets utilized in the present study is provided. A summarized overview of the datasets is presented in Table 1.

- ORL dataset: The ORL dataset [54] consists of 400 facial images. It includes 40 different classes, with 10 images for each class. The resolution of each image is 92×112 pixels. The collection includes images in grayscale with dark backgrounds, frontal poses that are upright, and modest variations in lighting, stance, and facial expressions.
- YALE dataset: The YALE dataset [55] has 165 images in total, with 11 images of each of the 15 subjects. The grayscale images have subjects depicting variable expressions while being with or without glasses. Furthermore, the images have a resolution of 220×175.
- GT dataset: The Georgia Tech dataset [56] is a collection of 750 images of 50 people, with each person having 15 color images. The resolution of images in this dataset is $640 \times 480 \times 3$ pixels, but for the purpose of this study, only the cropped version with a resolution of $131 \times 176 \times 3$ pixels is used. These pictures were captured with a background that contains visual distractions, and they feature a mix of different poses (frontal and tilted), lighting conditions, facial expressions, and scales.
- AR dataset: AR dataset [57] includes 1690 colored images, with 26 images for each of the 65 subjects under consideration. The images differ in terms of gender of the subjects, their expressions, lighting, and occlusion (black glasses and face scarf). Each image has a $165 \times 120 \times 3$ pixel resolution.
- FEI dataset: The FEI dataset, [58], employed in the current study contains 700 Brazillian faces from 50 individuals. Each person has 14 images taken. The images have a $640 \times 480 \times 3$ pixel resolution and a uniform white background, showing different facial expressions and poses.
- LFW dataset: The huge image dataset called as the Labeled Faces in the Wild (LFW) [59] is made up of 13,000 total images with more than 5700 subjects. There are variable number of images for each subject. The poses, expressions, ages, places of origin, backgrounds, and resolution of the photographs in the dataset vary over a wide range. From the original dataset, 700 images were randomly chosen from 50 subjects, averaging 14 images for each subject. Additionally, every image was downsized to a resolution of $250 \times 250 \times 3$ pixels.
- CFP-FP dataset: The CFP-FP (Celebrities in Frontal-Profile) [60] dataset is a collection of facial images designed for studying face recognition across different poses. It contains images of 500 distinct celebrities, each with 10 frontal and 4 profile images, resulting in a total of 7000 images. This challenging dataset helps researchers to evaluate facial recognition systems' ability to handle pose variations. From the original dataset, 700 images were randomly chosen from 50 subjects, with 14 images for each subject.

3.2. Methods

Deep Convolutional Neural Networks (CNNs) are capable of automatically identifying complex features from images for various tasks. They are composed of several components, including multiple convolutional layers, activation layers, max-pooling layers, one or more fully connected layers, and an output layer. Convolutional layers, consisting of convolution, activation (ReLU), and max pooling, are responsible for extracting features from the image. Convolution involves applying filters of various sizes to the image to identify features as

$$\vec{Y}^m = f^m(\vec{W}^m \cdot \vec{X}^{m-1} + \vec{\beta}^m), \tag{1}$$

where, m denotes the layer under consideration from M number of layers, $\vec{Y}^m$ denotes the output feature map of the mth layer, $\vec{W}$ denotes the kernel weights of the mth layer, $\vec{\beta}$ denotes the bias of the mth layer, and $\vec{X}$ denotes the input feature map of $(m-1)^{th}$ layer.

When a pooling layer is applied, the size of the output feature maps created by each layer is reduced, but the quantity of feature maps remains unchanged. During each iteration of training a network using the backpropagation method, the weights $\vec{W}$ of the kernels and the bias $\vec{\beta}$ are adjusted.

Each input image in deep CNN is represented as a tensor $\vec{X}$ of dimension $\left[H\ W\ C\right]$, where the image's height is denoted by H, width by W, and the number of color channels by C. An L-layer pre-trained CNN technique is represented as M functions in the sequence $f^1, f^2, \ldots, f^M$. Eq. (1) denotes the output $\vec{Y}^m$ of the layer m. During network training, the CNN learns the weights of the layers and subsequently the features, and each layer learns a separate set of features. Complex traits can be learned through the deeper layers. The images are then classified using the newly-learned attributes. Any layer of the pre-trained CNN can compute image features by supplying the function f^m, learnt weights $\vec{W}^m$, and input image data $\vec{X}$ such that

$$\vec{Y}^m = f^m(\vec{X}^{m-1} : \vec{W}^m). \tag{2}$$

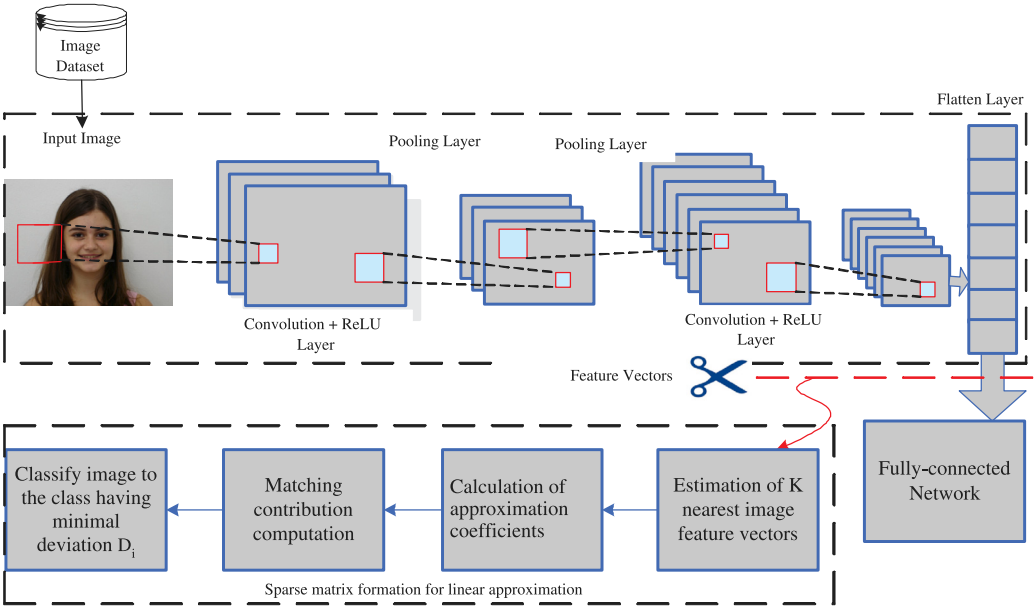

Fig. 1. Basic Framework of $RI\text{-}L_1$Approx model.

4. Proposed methodology

The proposed method is based on the L_1 norm approximation of a test sample using known class samples. Initially, facial features are extracted by utilizing the capabilities of the ResNet-Inception hybrid network to learn intricate hierarchical representations from facial images. These extracted features are then used for L_1 norm approximation with known features referred to as approximation samples.

4.1. System model

The $RI\text{-}L_1$Approx model follows a structure as depicted in Fig. 1. It involves FR through a two-step approach. Initially, it employs pre-trained layers from a deep Convolutional Neural Network (CNN) to extract features. Subsequently, it conducts L_1-norm approximation between approximation samples and test sample for calculation of matching contribution. The subsequent sections will delve into a more comprehensive explanation.

4.2. Feature extraction

Image feature extraction is a process of taking an image as input and using a deep-learning neural network such as CNNs to extract important features from it. These features are then used as a compact representation of the image, capturing the important information and details about the image.

We use a deep convolutional neural network that combines the concepts of both the Inception network and the ResNet (Residual Network) architecture for facial feature extraction. The detailed model architecture is provided in Fig. 2. The Inception-ResNet-v1 architecture ingeniously combines elements from both the Inception and ResNet models to create a powerful and efficient network for image recognition tasks. At its core, the network uses modified Inception modules, which are augmented with residual connections inspired by ResNet. These hybrid modules, named Inception-ResNet blocks, appear in three variants: A, B, and C. Each block processes input through multiple parallel paths with different convolutional filter sizes, similar to traditional Inception modules. However, the key addition is a shortcut connection that bypasses these parallel paths. The outputs from the parallel convolutions are concatenated, passed through a 1×1 convolution to adjust the channel dimensions, and then added to the input via the shortcut connection. This residual approach helps in maintaining good gradient flow through the deep network. The full architecture alternates these Inception-ResNet blocks with traditional convolutional layers and pooling operations, gradually reducing spatial dimensions while increasing feature depth.

The facial characteristics are extracted from the version of Inception-ResNet architecture [61] described above (Fig. 2), pre-trained with FaceNet model [38], on VGGFace2 [62], and the Casia-Webfaces [63] databases. Each face image is translated into a Euclidean space, with the distances in that space corresponding to the similarity of the faces. The network is trained so that face similarity corresponds to the squared L2 distance between the embeddings. The idea is to make sure that an image x_i^a (anchor) of person under consideration is closer to all the other images x_i^p (positive) of the same person in comparison to an image x_i^n (negative) of another person. Hence, we require that

$$\begin{aligned} &\left\| f\left(x_i^a\right) - f\left(x_i^n\right) \right\|_2^2 \; > \; \left\| f\left(x_i^a\right) - f\left(x_i^p\right) \right\|_2^2 + \lambda \;, \\ &\forall\, f\left(x_i^a\right),\; f\left(x_i^n\right),\; f\left(x_i^p\right) \;\in\; \omega, \end{aligned} \tag{3}$$

where λ denotes the margin which is enforced between the positive and the negative pairs, ω (having cardinality M) is the set of all of the possible triplets within the training set. The loss Δ to be minimized can be expressed as

$$\Delta = \sum_i^M \left[\left\| f\left(x_i^a\right) - f\left(x_i^p\right) \right\|_2^2 \; - \; \left\| f\left(x_i^a\right) - f\left(x_i^n\right) \right\|_2^2 \; + \; \lambda \right]_+ . \tag{4}$$

It is essential to choose triplets that defy the triplet limitation in Eq. (3) to achieve rapid convergence. This essentially implies that, given x_i^a, it is desirable to choose an x_i^p (hard positive) so that $\arg\max_{x_i^p} \left\| f\left(x_i^a\right) - f\left(x_i^p\right) \right\|_2^2$. Likewise, it is desirable to select an x_i^n (hard negative) so that $\arg\min_{x_i^n} \left\| f\left(x_i^a\right) - f\left(x_i^n\right) \right\|_2^2$. In practise, choosing the hardest negatives can result in poor local minima early in the training, more particularly, it can cause a collapsed model, i.e. $f(x) = 0$. Selecting x_i^n such that

$$\left\| f\left(x_i^a\right) - f\left(x_i^n\right) \right\|_2^2 \; > \; \left\| f\left(x_i^a\right) - f\left(x_i^p\right) \right\|_2^2 \;, \tag{5}$$

helps to mitigate the aforementioned issue. The Stochastic Gradient Descent (SGD) with conventional backprop and AdaGrad was used to train CNNs. ReLU was selected as the activation function, λ set to 0.2,

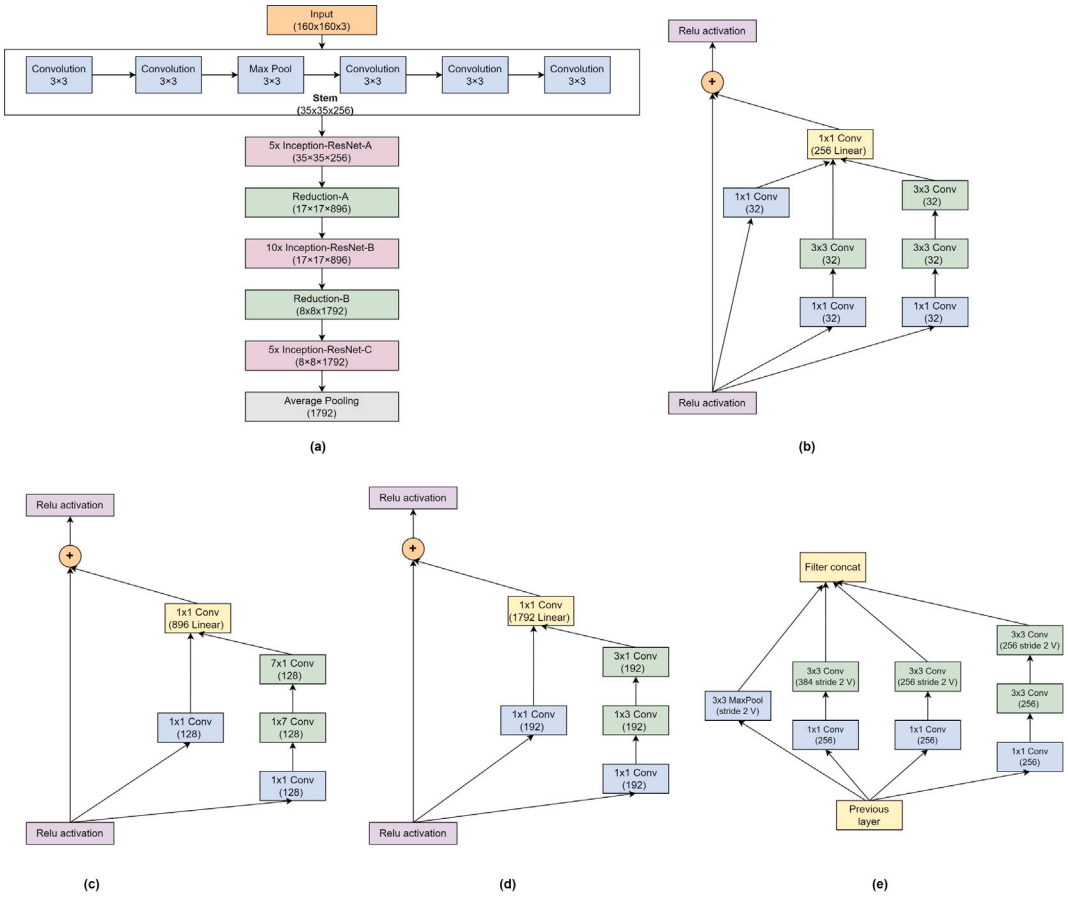

Fig. 2. The figure shows (a) Basic Framework of Inception-ResNet-v1 network (b) Inception-ResNet-A module (c) Inception-ResNet-B module (d) Inception-ResNet-C module (e) Reduction-B module of Inception-ResNet-v1 network.

and the initial learning rate set to 0.05 [38]. The feature vectors for the images are provided by the network's final layer (average pooling) (Fig. 1).

4.3. Sparse matrix formation for linear approximation

The features that have been extracted are employed to categorize the class of an unfamiliar test sample. This classification is achieved by approximating the test Sample features with the known sample features termed as approximation samples using the L_1-norm. The incorporation of the L_1-norm approximation encourages sparsity by pushing certain approximation samples to possess coefficients of zero. This process aids in identifying the most relevant samples that share similar attributes with the test sample.

In this model, the classification is conducted through a sequence of three steps:

- One nearest approximation sample that is closest to the test sample is selected from each class. As a result, the number of nearest approximation samples selected equals κ number of classes.
- The second stage involves representing the test sample as a linear sparse combination of the κ nearest approximation samples and calculating the coefficient values needed for L_1-norm approximation.
- The third step involves calculating the matching contribution for each nearest approximation samples.
- Finally, the test sample is assigned to the class of the nearest approximation sample, which contributes the highest level of similarity.

Following is a discussion of the steps employed in sparse matrix formation for linear approximation:

Step 1 *Determination of the nearest approximation samples:* By employing the squared Euclidean distance, the nearest approximation samples are identified for each class. The ith approximation sample $\vec{t}_i^{\,j}$ pertaining to the jth class, such that $j = 1, 2, \ldots, \kappa$ and $i = 1, 2, \ldots, \tau$, are represented together as the elements of approximation sample set $\vec{T}$, such that $\vec{t}_i^{\,j} \in \mathbb{R}^{1792\times 1}$ and $\vec{T} \in \mathbb{R}^{\kappa\times\tau}$ as

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^1 & \mathbf{t}_2^1 & \ldots & \mathbf{t}_\tau^1 \\ \mathbf{t}_1^2 & \mathbf{t}_1^2 & \ldots & \mathbf{t}_\tau^2 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{t}_1^\kappa & \mathbf{t}_1^\kappa & \ldots & \mathbf{t}_\tau^\kappa \end{bmatrix}. \tag{6}$$

Additionally, $\vec{q}_r$, for $r = 1, 2, \ldots, (N-\tau)\times\kappa$ such that $\vec{q}_r \in \mathbb{R}^{1792\times 1}$, is the test sample, N is the total number of image samples per class, τ is the number of approximation samples per class and $(N-\tau)$ is the number of test samples per class. Using Eq. (7), the distance between $\vec{q}_r$ and each approximation sample of every class can be computed for every value of r as

$$d_i^{\,j} = \|\vec{t}_i^{\,j} - \vec{q}_r\|^2, \tag{7}$$

and represented as

$$\mathbf{D}_r = \begin{bmatrix} d_1^1 & d_2^1 & \ldots & d_\tau^1 \\ d_1^2 & d_2^2 & \ldots & d_\tau^2 \\ \vdots & \vdots & \vdots & \vdots \\ d_1^\kappa & d_2^\kappa & \ldots & d_\tau^\kappa \end{bmatrix} \tag{8}$$

Table 2
Important parameters used in the Proposed algorithm, and their significance.

Parameters	Significance
L	Number of layers
N	Total number of images samples per class
κ	Total number of classes
τ	Number of approximation image samples per class
$\mathbf{t}_i^j$	Each ith approximation image sample from jth class expressed as a feature vector
$\mathbf{T}$	Approximation sample set containing feature vectors from each class
$(N-\tau)$	Number of test image samples per class
$(N-\tau)\times\kappa$	Total number of test image samples
$\mathbf{q}_r$	Each r^{th} test image sample expressed as a feature vector
$\mathbf{D}_r$	Matrix of distances between $\vec{q}_r$ and each $\mathbf{t}_i^j$ computed for every value of r
$\mathbf{d}$	Vector of minimum distances for each of the κ classes
$\boldsymbol{\Phi}$	Matrix of nearest approximation samples for each of the κ classes
$\hat{\mathbf{t}}_j$	nearest approximation sample corresponding to the $min\{d_j^i\}$ for each class
$\boldsymbol{\alpha}$	Coefficient vector for L_1-norm approximation
$\mathbf{c}_j$	Matching contribution of each class for every nearest approximation sample
$\hat{d}_j$	Deviation between matching contribution of jth class and test sample
PCA	percentage classification accuracy

such that every $\mathbf{D}_r \in \mathbb{R}^{\kappa\times\tau}$. From Eq. (8), the minimum distance $min\{d_i^j\}$ is selected for every jth class such that $\vec{d} = \left[d_1, d_2, \ldots, d_\kappa\right]^T$, where $d_j = min\{d_j^i\}$ for $j = 1, 2, \ldots, \kappa$ and $\mathbf{d} \in \mathbb{R}^{\kappa\times 1}$. Each of these distances d_j are used to identify the nearest approximation sample corresponding to the $min\{d_j^i\}$ for each class.

As a result, a collection of κ nearest approximation samples is created by choosing one nearest approximation sample from every class. Finally, a matrix $\boldsymbol{\Phi} = \left[\vec{\hat{t}}_1, \vec{\hat{t}}_2, \ldots, \vec{\hat{t}}_\kappa\right]$ such that $\boldsymbol{\Phi} \in \mathbb{R}^{1792\times\kappa}$ is used to represent all κ nearest approximation samples $\vec{\hat{t}}_1, \vec{\hat{t}}_2, \ldots, \vec{\hat{t}}_\kappa$.

Step 2 *Computation of coefficient vector for L_1-approximation:* The test sample $\vec{q}_r$ is approximated as a linear sparse combination of the selected κ nearest approximation samples as

$$\vec{q}_r = \alpha_1 \vec{\hat{t}}_1 + \alpha_2 \vec{\hat{t}}_2 + \cdots + \alpha_\kappa \vec{\hat{t}}_\kappa. \tag{9}$$

In Eq. (9), the κ nearest approximation samples are denoted as $\vec{t}_j$ and the corresponding coefficients needed for sparse L_1 approximation as α_j where $j = 1, 2, \ldots, \kappa$. Thus, we can write,

$$\vec{q}_r = \Sigma_{j=1}^{\kappa} \alpha_j \, \boldsymbol{\Phi}(:, j). \tag{10}$$

We can rewrite Eq. (9) in matrix form as

$$\vec{q}_r = \boldsymbol{\Phi}\, \boldsymbol{\alpha}, \tag{11}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_\kappa]^T$ such that $\boldsymbol{\alpha} \in \mathbb{R}^{\kappa\times 1}$. Additionally, a singularity test is run on $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$. If it is non-singular, the formula in Eq. (11) is used to solve the Eq. (10). Alternatively,

$$\boldsymbol{\alpha} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\, \vec{q}_r\ . \tag{12}$$

In case the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is proved to be singular, the coefficient vector $\boldsymbol{\alpha}$ can be solved for using Eq. (12) as

$$\boldsymbol{\alpha} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mu\vec{I})^{-1}\, \boldsymbol{\Phi}^T\vec{q}_r\ , \tag{13}$$

where μ is a small-valued real number such that $\mu > 0$ and $\vec{I} \in \mathbb{R}^{\kappa\times\kappa}$ denotes the identity matrix. The value of μ is fixed to 0.01 in this work, in accordance with the existing applications of the sparse approximation in FR.

Step 3 *Final classification by computation of matching contribution:* Each nearest approximation sample $\vec{\hat{t}}_j$ included in $\boldsymbol{\Phi}$ is clearly drawn from a different class, as was previously discussed. Each class's matching contribution $\vec{c}_j$ for every $\vec{\hat{t}}_j$ in $\boldsymbol{\Phi}$ is computed for the final classification as

$$\boldsymbol{c}_j = \alpha_j\, \vec{\hat{t}}_j, \tag{14}$$

such that $\vec{c}_j \in \mathbb{R}^{1792\times 1}$ and $j = 1, 2, \ldots, \kappa$. The norm-2 distance, as expressed in Eq. (15), is used to determine the deviation between the matching contribution $\vec{c}_j$ of every $\vec{\hat{t}}_j$ and the test sample $\vec{q}_r$ as

$$\hat{d}_j = \|\vec{q}_r - \vec{c}_j\|^2, \tag{15}$$

where $\vec{\hat{d}} = \left[\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_\kappa\right]^T$ such that $\vec{\hat{d}} \in \mathbb{R}^{\kappa\times 1}$. In this case, it is obvious that a smaller value of $\hat{d}_j$ indicates a better match between the test sample $\vec{q}_r$ and the jth nearest approximation sample.

Table 2 describes the important parameters of the proposed algorithm. Algorithms 1, 2 and flowchart in Fig. 3 illustrate the entire work-flow of the proposed RI-L_1Approx model, replete with necessary programming specifics.[2]

Algorithm 1 Feature extraction

Input: Set of images $\{I_i\}$ where I_i is either a grayscale or colored image
Output: Set of feature vectors $\{F_i\}$ where $F_i \in \mathbb{R}^{1792}$
Initialize an empty set to store feature vectors: $\{F_i\} = \emptyset$
for each image $I_i \in \{I_i\}$ **do**
  if I_i is a grayscale image **then**
    Perform image scaling $I_i \leftarrow \text{Resize}(I_i, (160, 160, 1))$
  else if I_i is a colored image **then**
    Perform image scaling $I_i \leftarrow \text{Resize}(I_i, (160, 160, 3))$
  end if
  Let M be the pre-trained Inception-ResNet-v1 model
  Let L be the average pooling layer of M
  Forward propagate I_i through M up to layer L:
  $F_i = L(M(I_i))$ where $F_i \in \mathbb{R}^{1792}$
  Add F_i to the set of feature vectors $\{F_i\}$
end for
return $\{F_i\}$

5. Experimentation, results and analysis

In this study, we perform two separate experiments to evaluate the efficiency of the RI-L_1Approx model in an unrestricted setting with a limited number of training images.

5.1. Experimental setup

All experimental procedures and model training were carried out on a computational system featuring the following specifications: 11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHZ, 2496 Mhz, 8 Core(s) 32 GB

[2] From the different image databases used to test the proposed algorithm, some have gray-scale images while others have colored images. Resizing all images to a small size would have resulted in the loss of information from colored images. On the other hand, resizing images to a large size value would have resulted in higher computational complexity of the system. Due to these factors, the size 160 × 160 × 3 pixels was selected.

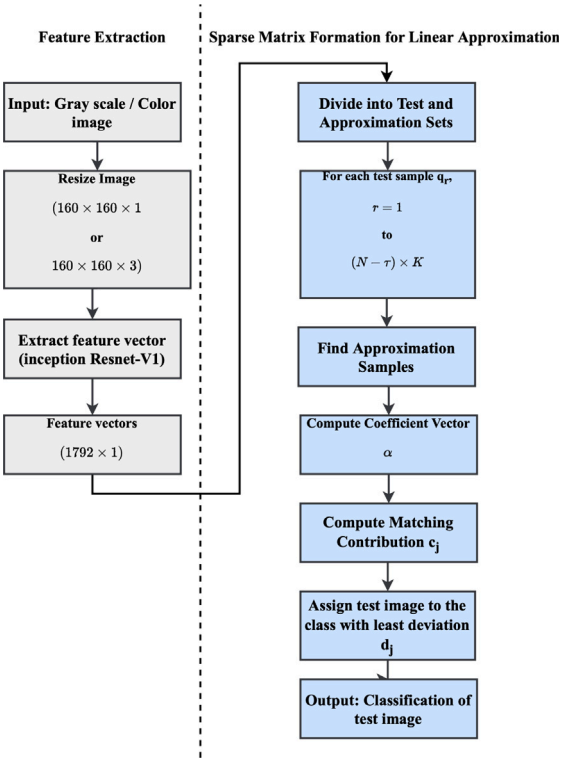

Fig. 3. Simplified flowchart of $RI\text{-}L_1$Approx.

RAM, Nvidia 3060 RTX GPU, running Python 3.10, and operating on Windows 10.

5.2. Performance metrices

The experiments focus on exploring the FR performance of the $RI\text{-}L_1$Approx model on multiple datasets. This model uses a pre-trained convolutional neural network architecture (Inception-ResNet-v1) to extract features from faces. The VGGFace2 and CASIA-Webface datasets were used to pre-train the network. Two experiments were conducted to evaluate the $RI\text{-}L_1$Approx model's performance on different datasets. The first experiment used feature vectors extracted from the network pre-trained using the VGGFace2 database, while the second experiment used network pre-trained using CASIA-Webface dataset.

The feature vectors that have been obtained are divided into two separate sets as test image samples and approximation image samples. These sets are mutually exclusive i.e. do not overlap with each other. This division is done to carry out experiments 1 and 2. Each image is represented by a single feature vector. For each dataset, a certain number, τ, of these image samples are selected from the N total number of image samples for each class to be used as the approximation samples. The remaining image samples for each class, amounting to $(N-\tau)$, are utilized as the test image samples. This leads to a total of $\frac{N}{\tau!(N-\tau)!}$ feasible combinations of the approximation sets. Table 3 shows the details on the number of tests carried out on each dataset. It is important to note that the testing and approximation samples used in Experiments 1 and 2 are distinct from each other to avoid overfitting. The results for different values of approximation samples per class (τ) are presented as the mean, minimum, and maximum percentage classification accuracy (PCA). The PCA may be calculated as

$$\text{PCA} = \frac{1}{(N-\tau)\times\kappa}\sum_{p=1}^{(N-\tau)\times\kappa} w_p \times 100, \tag{18}$$

Algorithm 2 Sparse Matrix Formation for Linear Approximation

Input: Image feature collection $\mathcal{F}$, Number of classes κ, Number of approximation samples per class τ
Output: Class assignments for test samples
Step 1: Determination of Nearest Approximation Samples
for each test sample $\vec{q}_r$, $r = 1, 2, \ldots, (N-\tau)\times\kappa$ **do**
  for each class $j = 1, 2, \ldots, \kappa$ **do**
    for each approximation sample $\vec{t}_i^{\,j}$, $i = 1, 2, \ldots, \tau$ **do**
      Compute distance $d_i^j = \|\vec{t}_i^{\,j} - \vec{q}_r\|^2$ (Eq. (7))
    end for
    Determine minimum distance $d_j = \min\{d_i^j\}_{i=1}^{\tau}$
    Select nearest approximation sample $\hat{\vec{t}}_j = \vec{t}_i^{\,j}$ where $d_i^j = d_j$
  end for
  Form matrix $\boldsymbol{\Phi} = [\hat{\vec{t}}_1, \hat{\vec{t}}_2, \ldots, \hat{\vec{t}}_\kappa] \in \mathbb{R}^{1792\times\kappa}$
  Step 2: Computation of Coefficient Vector for L_1-Approximation
  Represent $\vec{q}_r$ as a sparse linear combination of κ nearest approximation samples:

$$\vec{q}_r = \sum_{j=1}^{\kappa}\alpha_j\hat{\vec{t}}_j \implies \vec{q}_r = \boldsymbol{\Phi}\boldsymbol{\alpha} \tag{16}$$

  Compute $\boldsymbol{\alpha}$:
  if $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is non-singular **then**
    $\boldsymbol{\alpha} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\vec{q}_r$ (Eq. (12))
  else
    $\boldsymbol{\alpha} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mu\mathbf{I})^{-1}\boldsymbol{\Phi}^T\vec{q}_r$ (Eq. (13))
  end if
  Step 3: Final Classification by Computation of Matching Contribution
  for each class $j = 1, 2, \ldots, \kappa$ **do**
    Compute matching contribution $\vec{c}_j = \alpha_j\hat{\vec{t}}_j$ (Eq. (14))
    Compute deviation distance $\hat{d}_j = \|\vec{q}_r - \vec{c}_j\|^2$ (Eq. (15))
  end for
  Assign class of $\vec{q}_r$ to class with minimum deviation $\hat{d}_j$:

$$\hat{d}_j = \min\left\{\|\vec{q}_r - \alpha_j\hat{\vec{t}}_j\|^2\right\}_{j=1}^{\kappa} \tag{17}$$

end for
Return: Class assignments for test samples

such that

$$w_p = \begin{cases} 1 & \text{if } \psi(q_p) = Z_p \\ 0 & \text{else} \end{cases} \tag{19}$$

The set of class labels is represented by $Z = 1, 2, \ldots, \kappa$, $\vec{q}_1, \vec{q}_2, \ldots \vec{q}_p$ belongs to the set of test sample, and Z_i represents true class label of q_i. Moreover, Z_i is a number from set $\vec{Z}$ such that for the classifier ψ and the elements q_i of the test sample, $\psi(q_i) = Z_i$ is the label prediction.

5.3. Results

5.3.1. Experiment 1: L1 approximation of feature vectors extracted using pre-trained VGGFace2

Table 3 and Fig. 4 facilitate the comparison of classification accuracy of the proposed $RI\text{-}L_1$Approx model pre-trained using VGGFace2 in terms of mean PCA, for all datasets under consideration. The performance of the proposed $RI\text{-}L_1$Approx model, when using $\tau = 1$ or $\tau = 2$, is significantly better compared to other techniques based on sparse approximation or CNN. For all the considered datasets, it is evident that mean PCA rises as τ increases. The ORL (for $\tau = 9$), YALE (for $\tau = 6$), and GT (for $\tau = 9$) datasets achieve the highest value of mean PCA of 100 percent. This exceptional accuracy can be attributed to the limited face photos and uniform background. The greatest mean PCA value of 99.88 percent is obtained for the AR dataset at $\tau = 25$ and 95.14 percent for the FEI dataset at $\tau = 13$. The results show that when τ is set to a low value, the mean PCA score for the FEI dataset is better than that for the AR dataset. But when τ is set to a high value, the AR dataset performs

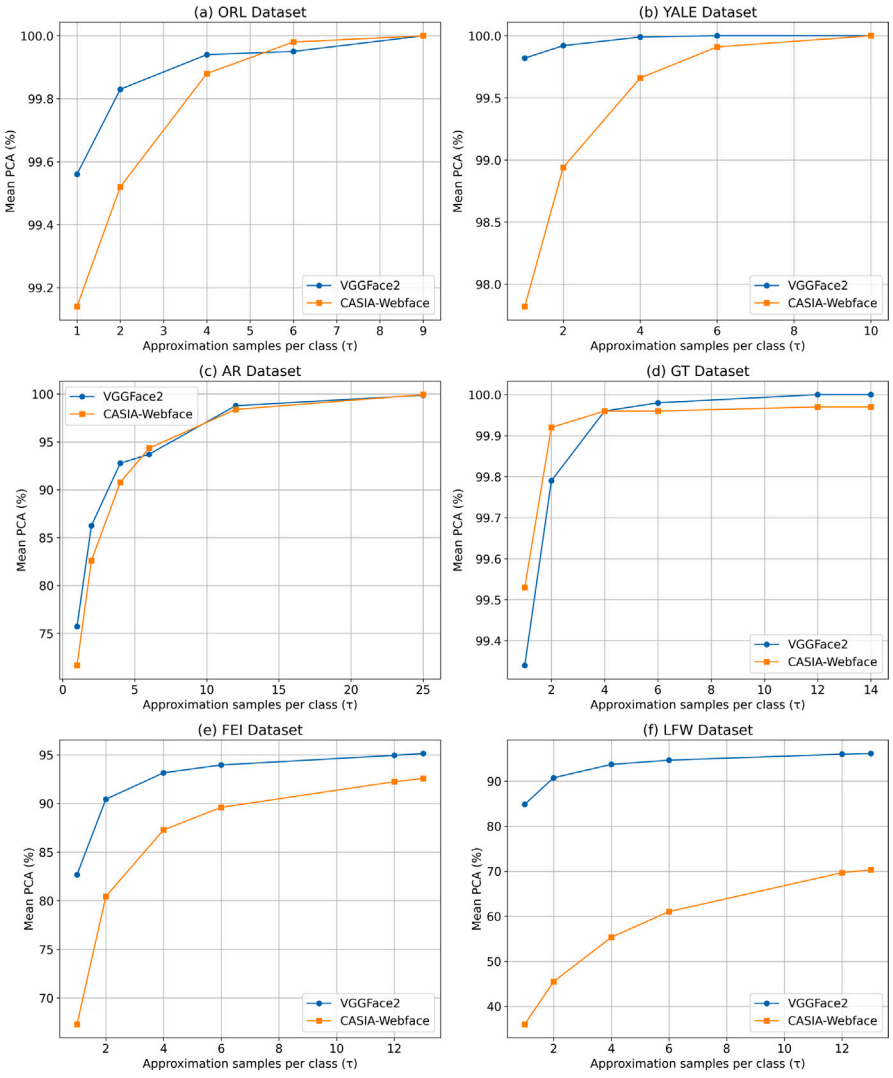

Fig. 4. Mean PCA vs τ for different datasets under consideration.

Table 3
Details of tests conducted on Proposed $RI\text{-}L_1$Approx model and the mean PCA achieved.

Dataset	Approximation samples per class (τ)	Number of test sets	Number of tests/ set	Total number of tests	Mean PCA of the Proposed $RI\text{-}L_1$Approx Model Pre-trained with VGGFace2 dataset	Pre-trained with CASIA-Webface dataset
ORL	1	10	360	3,600	99.56	99.14
	2	45	320	14,400	99.83	99.52
	4	210	240	50,400	99.94	99.88
	6	210	160	33,600	99.95	99.98
	9	10	40	400	**100**	**100**
YALE	1	11	15	1,650	99.82	97.82
	2	55	135	7,425	99.92	98.94
	4	330	105	34,650	99.99	99.66
	6	462	75	34,650	**100**	99.91
	10	11	15	165	**100**	**100**
AR	1	26	1,625	42,250	75.71	71.65
	2	325	1,560	5,07,000	86.25	82.6
	4	14,950	1,430	21,378,500	92.77	90.77
	6	2,30,230	1,300	29,92,99,000	93.71	94.35
	12	96,57,700	910	8,78,85,07,000	98.78	98.39
	25	26	65	1,690	**99.88**	**99.94**

(*continued on next page*)

Table 3 (*continued*).

Dataset	Approximation samples per class (τ)	Number of test sets	Number of tests/ set	Total number of tests	Mean PCA of the Proposed RI-L_1Approx Model Pre-trained with VGGFace2 dataset	Pre-trained with CASIA-Webface dataset
GT	1	15	700	10,500	99.34	99.53
	2	105	650	68,250	99.79	99.92
	4	1,365	550	7,50,750	99.96	99.96
	6	5,005	450	22,52,250	99.98	99.96
	12	455	150	68,250	**100**	99.97
	14	15	50	750	**100**	**99.97**
FEI	1	14	650	9,100	82.67	67.29
	2	91	600	54,600	90.44	80.44
	4	1,001	500	5,00,500	93.15	87.28
	6	3,003	400	12,01,200	93.96	89.60
	12	91	100	9,100	94.95	92.24
	13	14	50	700	**95.14**	**92.57**
LFW	1	14	650	9,100	84.86	36.01
	2	91	600	54,600	90.75	45.47
	4	1,001	500	5,00,500	93.71	55.32
	6	3,003	400	12,01,200	94.66	61.04
	12	91	100	9,100	95.99	69.67
	13	14	50	700	**96.14**	**70.29**

Table 4
Comparison of the RI-L_1Approx model with the state-of-the-art FR models.[a]

(a) Comparison of models on ORL, YALE, YaleB-Extended, Georgia Tech, AR and FEI datasets.

Datasets	References	Models	τ	mean PCA
ORL	**Proposed model**	RI-$L1$Approx	**6**	**99.95%**
	S. Y. Wang et al. (2020) [14]	SAMPSR	5	94.14%
	S. Almabdy et al. (2019) [64]	ResNet-50 + SVM	8	100%
	S. Guo et al. (2016) [14]	CNN + SVM	7	97.5%
YALE	**Proposed model**	RI-$L1$Approx	**6**	**100%**
YaleB-Extended	S. Zhu et al. (2017) [65]	Histogram-basedfeature representation	NA	99.37%
Georgia Tech	**Proposed model**	RI-$L1$Approx	**12**	**100%**
AR	**Proposed model**	RI-$L1$Approx	**6**	**94.35%**
	S. Y. Wang et al. (2020) [14]	SAMPSR	8	72.49%
	Proposed model	RI-$L1$Approx	**4**	92.77%
	N. Zhu et al. (2014) [47]	KSR method	4	91.61%
FEI	**Proposed model**	RI-$L1$Approx	**13**	95.14%
	Proposed model	RI-$L1$Approx	**1**	82.67%
	J. Cai et al. (2015) [66]	Sparse representation	1	61.31%
	S. Almabdy et al. (2019) [64]	Transfer learning(AlexNet)	11	98.7%

(b) Comparison of models on LFW dataset.

References	Models	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 6$	$\tau = 12$	$\tau = 13$
Proposed model	RI-L_1Approx	84.86%	90.75%	93.71%	94.66%	95.99%	96.14%
Qiang Meng et al. (2021)[40]	MagFace	79.33%	86.15%	90.90%	92.88%	94.75%	94.98%
F. Schroff et al. (2015)[38]	FaceNet	77.46%	84.50%	89.12%	91.33%	93.66%	93.92%
Abu Sufian et al. (2023)[27]	FewFaceNet	71.20%	74.18%	76.56%	80.22%	83.10%	83.77%

(c) Comparison of models on CFP-FP dataset.

References	Models	$\tau = 1$	$\tau = 2$	$\tau = 4$	$\tau = 6$	$\tau = 12$	$\tau = 13$
Proposed model	RI-L1Approx	83.76%	88.63%	92.51%	95.28%	97.89%	98.42%
Qiang Meng et al. (2021) [40]	MagFace	79.21%	87.18%	90.43%	94.56%	96.97%	98.50%
F. Schroff et al. (2015) [38]	FaceNet	77.02%	85.33%	88.98%	92.31%	94.23%	95.63%
Abu Sufian et al. (2023) [27]	FewFaceNet	68.27%	77.54%	81.61%	84.91%	85.81%	85.81%

[a] Table 4a, and 4b contains only the selected relevant results from Table 3 for comparison with the existing state-of-the-art.

better than the FEI dataset. This is because the large occluded areas in the images in the AR dataset result in a reduced performance with small number of approximation image samples. With $\tau = 13$, the maximum value of mean PCA for the LFW dataset is greater than that for the FEI dataset, at 96.14 percent. The Proposed RI-L_1Approx model performed better on the LFW dataset compared to the CNN architectures that are commonly used. [64]

The PCA of the tests is analyzed in Figs. 5 and 6 by showing the minimum PCA, maximum PCA, and standard deviation. The results of the analysis indicate that the lowest PCA for VGGFace2 on the ORL dataset is 100% at $\tau = 9$. For the YALE dataset, the lowest PCA is 100% when $\tau \geq 5$, and for the GT dataset, it is 100% when $\tau \geq 9$. For the AR, FEI, and LFW datasets, the minimum PCA value never reaches 100% regardless of the value of τ. The maximum PCA for the YALE and GT datasets is 100% for all possible values of τ. For the ORL dataset, the maximum PCA is 100% for $\tau \geq 2$. The maximum PCA for the AR dataset is 100% for $\tau \geq 12$, and for the FEI and LFW datasets it is 100% for $\tau \geq 13$.

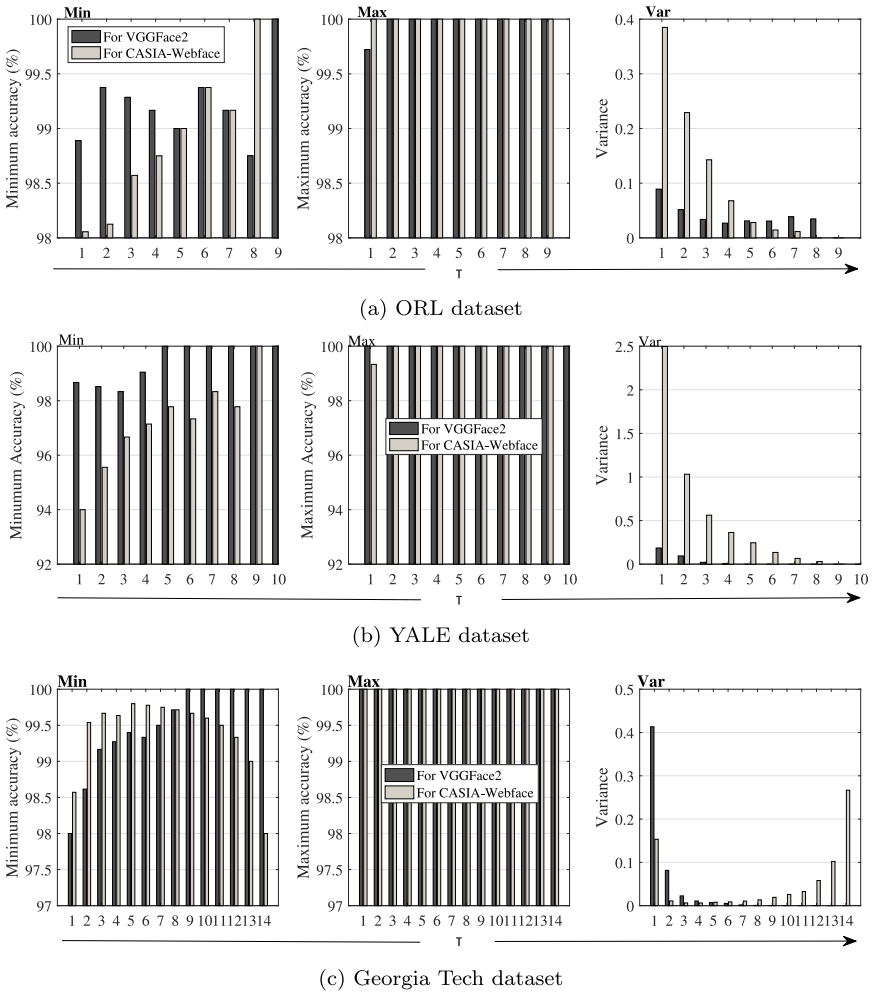

Fig. 5. Comparing the minimum, maximum, and variance of classification accuracy for each dataset.

5.3.2. Experiment 2: $L1$ approximation of feature vectors extracted using pre-trained CASIA Webface

Table 3 and Fig. 4 compare the classification accuracy using mean PCA, for the proposed model pre-trained on CASIA-Webface, across all datasets considered in this study. The ORL $\tau = 9$ and the YALE $\tau = 10$ datasets yield the maximum value of mean PCA of 100 percent. The highest value of mean PCA for AR, GT, FEI, and LFW datasets is found to be 99.94% when using a value of $\tau = 25$, 99.97% when using $\tau = 14$, 92.57% when using $\tau = 13$, and 79.29% when using $\tau = 13$, respectively. It is discovered that the accuracy of the LFW dataset is significantly lower compared to the PCA from Experiment 1, regardless of the different values of τ considered.

The results of all the tests performed on the datasets in this study can be seen in 5 and 6, which provide a statistical analysis of the results. The plot shows that the minimum PCA for the CASIA Webface dataset is 100 % when τ is greater than or equal to 8 for the ORL dataset and greater than or equal to 9 for the YALE dataset. Additionally, the value of minimum PCA for the LFW, AR, FEI, and GT datasets is never 100 percent for any value of τ. On the other hand, the highest PCA that can be achieved on the ORL and GT datasets is 100 percent for all possible values of τ. Moreover, it is also 100 percent for $\tau \geq 2$ for the YALE dataset, $\tau \geq 12$ for the AR dataset, and $\tau \geq 13$ for the FEI dataset. Additionally, it is not possible to attain a maximum PCA of 100 percent on the LFW dataset.

5.4. Comparative analysis

From the aforementioned trials in Experiments 1 and 2, it can be seen that models that have been pre-trained using the VGGFace2 database perform better overall in terms of mean PCA for all the datasets considered in this work. Additionally, for the GT dataset, the model trained using CASIA-Webface performs better at a lower value of τ, whereas the model trained with VGGFace2 performs better at a higher value of τ. Additionally, the performance of the proposed RI-L_1Approx model surpasses the existing FR application approaches (Table 4a). The diversity in training and testing datasets presents a significant challenge in comparing our results with prior work in the area of FR. Compared to the approximation data used in our experiments, the training datasets utilized in prior works utilizing deep learning CNN models are much larger.

There is limited research on the use of few-shot and one-shot learning techniques in face recognition. For comparison with FaceNet, MagFace, and FewFaceNet models we tested the models with the same samples, same number of samples and used the same setting for estimating the mean percentage classification accuracy (PCA) as the proposed method. These comparisons demonstrate the superior performance of our method in terms of mean PCA. This comprehensive comparison further validates the effectiveness and robustness of our approach.

We have performed paired t-tests to validate the performance improvements of our proposed model over baseline approaches. Specifically, paired t-tests are conducted to compare the proposed model with other state-of-the-art models, such as MagFace, FaceNet, and

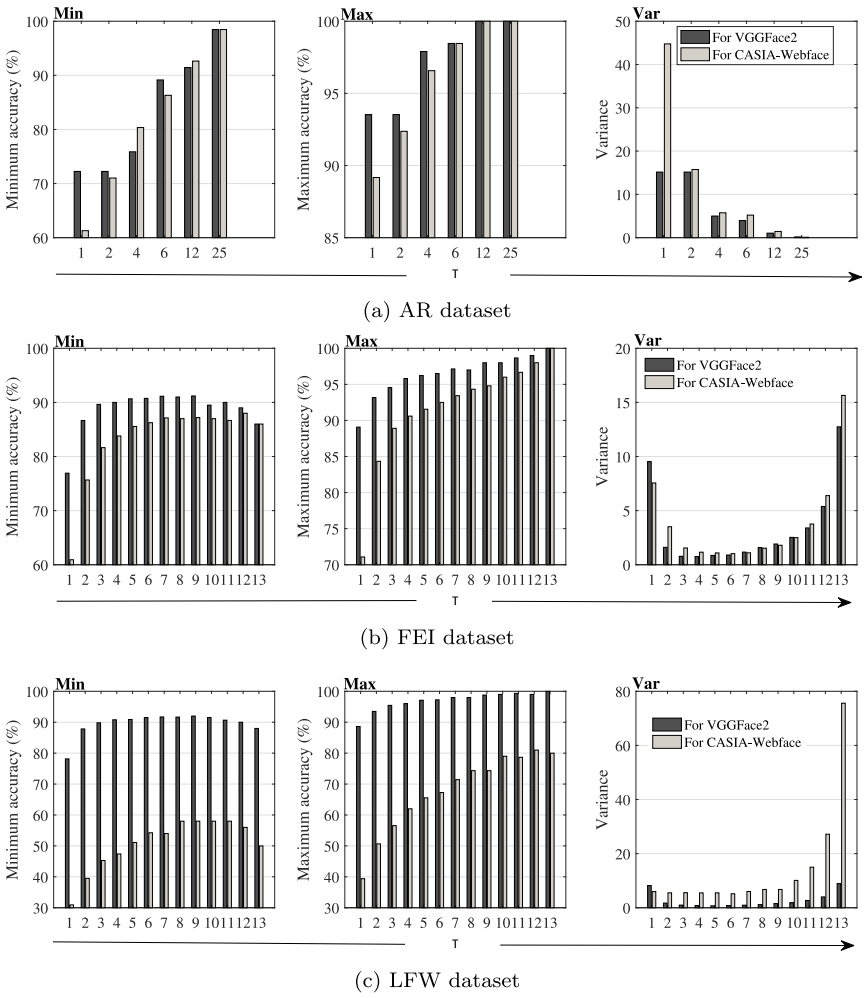

Fig. 6. Comparing the minimum, maximum, and variance of classification accuracy for each dataset.

FewFaceNet, across various approximation sample sizes using the LFW dataset. The p-values obtained from these tests are as follows: 0.01252 for MagFace, 0.004061 for FaceNet, and for FewFaceNet 0.00000928. These results demonstrate that the improvements achieved by our proposed model are statistically significant, providing evidence of its superiority over the baseline models.

5.5. Computational complexity

The model's high computational efficiency is one of its main benefits. The total amount of time needed by the deep CNN layers and the L_1-norm approximation layer can be used to determine the time complexity required by this model. Due to the pre-training of the deep CNN layers used in this model, the time needed for feature extraction is minimal. Hence, the time complexity of the suggested model is primarily due to the time needed by the L_1-norm approximation layer, which is $O(K^2M + K^3 + KM)$, where M denotes the feature vector dimension.

6. Conclusion and future scope

In conclusion, we have introduced a novel approach, Resnet-Inception-based Fast L_1-approximation (RI-L_1Approx), for enhancing face recognition with limited images. Our method leverages a hybrid ResNet-Inception network to extract facial features. We perform L_1 approximation to promote sparsity in the extracted features and predict the class of unfamiliar test samples by approximating the L_1 norms of approximation samples. This process encourages a subset of approximation samples to have zero coefficients, selecting the most discriminative ones and enhancing classification capabilities. Our approach has been rigorously evaluated on benchmark facial recognition datasets, demonstrating its effectiveness. Comparisons against state-of-the-art techniques have highlighted the advantages of our approach. Notably, our RI-L_1Approx model has achieved high accuracy rates of 84.86% and 96.144% with just one sample per class and thirteen samples per class respectively during experimentation. This performance significantly surpasses existing deep learning approaches, which typically demand larger datasets for comparable results. The future scope of the work includes refining the feature representation to capture more discriminative and robust facial features. Additionally, incorporating attention mechanisms or meta-learning strategies could further enhance the model's ability to adapt and generalize effectively with limited labeled data, potentially leading to even higher accuracy and efficiency in face recognition tasks.

Ethics approval

No Human subject or animals are involved in the research.

Funding statement

The authors received no specific funding for this study.

CRediT authorship contribution statement

Supriya Bajpai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Gargi Mishra:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Data curation, Conceptualization. **Rachna Jain:** Writing – review & editing. **Deepak Kumar Jain:** Writing – review & editing. **Dharmender Saini:** Writing – review & editing. **Amir Hussain:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used in this research is available publicly.

References

[1] M. Taskiran, N. Kahraman, C.E. Erdem, Face recognition: Past, present and future (A review), Digit. Signal Process. 106 (2020) 102809.

[2] Y. Mei, P. Guo, V.M. Patel, Escaping data scarcity for high-resolution heterogeneous face hallucination, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 18676–18686.

[3] M.K. Rusia, D.K. Singh, A comprehensive survey on techniques to handle face identity threats: challenges and opportunities, Multimedia Tools Appl. 82 (2) (2023) 1669–1748.

[4] V. Tomar, N. Kumar, A.R. Srivastava, Single sample face recognition using deep learning: A survey, Artif. Intell. Rev. 56 (Suppl 1) (2023) 1063–1111.

[5] M. Wang, W. Deng, Deep face recognition: A survey, Neurocomputing 429 (2021) 215–244.

[6] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 1991, pp. 586–587.

[7] S.A. Hannan, Pushparaj, M.W. Ashfaque, A. Lamba, A. Kumar, Analysis of detection and recognition of human face using support vector machine, in: International Conference on Artificial Intelligence of Things, Springer, 2023, pp. 86–98.

[8] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Netw. 13 (6) (2002) 1450–1464.

[9] Q. Xu, N. Zhao, A facial expression recognition algorithm based on CNN and lbp feature, in: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, ITNEC, vol. 1, IEEE, 2020, pp. 2304–2308.

[10] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, T. Tan, Gabor ordinal measures for face recognition, IEEE Trans. Inf. Forensics Secur. 9 (1) (2013) 14–26.

[11] Y.-X. Yang, C. Wen, K. Xie, F.-Q. Wen, G.-Q. Sheng, X.-G. Tang, Face recognition using the SR-CNN model, Sensors 18 (12) (2018) 4237.

[12] T. Goel, R. Murugan, Classifier for face recognition based on deep convolutional-optimized kernel extreme learning machine, Comput. Electr. Eng. 85 (2020) 106640.

[13] A. Phornchaicharoen, P. Padungweang, Face recognition using transferred deep learning for feature extraction, in: 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, ECTI DAMT-NCON, IEEE, 2019, pp. 304–309.

[14] S. Guo, S. Chen, Y. Li, Face recognition based on convolutional neural network and support vector machine, in: 2016 IEEE International Conference on Information and Automation, ICIA, IEEE, 2016, pp. 1787–1792.

[15] Z. Lu, X. Jiang, A. Kot, Deep coupled resnet for low-resolution face recognition, IEEE Signal Process. Lett. 25 (4) (2018) 526–530.

[16] X. Deng, F. Da, H. Shao, Y. Jiang, A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors, Comput. Electr. Eng. 85 (2020) 106700.

[17] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.

[18] Y. Wang, Y. Peng, S. Liu, J. Li, X. Wang, Sparsity adaptive matching pursuit for face recognition, J. Vis. Commun. Image Represent. 67 (2020) 102764.

[19] J.X. Mi, Y. Sun, J. Lu, H. Kong, Robust supervised sparse representation for face recognition, Cogn. Syst. Res. 62 (2020) 10–22.

[20] D. Vella, J.P. Ebejer, Few-shot learning for low-data drug discovery, J. Chem. Inf. Model. 63 (1) (2022) 27–42.

[21] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Comput. Surv. 53 (3) (2020) 1–34.

[22] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C.P. Lim, Q.J. Wu, A review of generalized zero-shot learning methods, IEEE Trans. Pattern Anal. Mach. Intell. (2022).

[23] Y. Su, S. Shan, X. Chen, W. Gao, Adaptive generic learning for face recognition from a single sample per person, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2699–2706.

[24] Z. Peng, Z. Li, J. Zhang, Y. Li, G.J. Qi, J. Tang, Few-shot image recognition with knowledge transfer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 441–449.

[25] Q. Zhu, Q. Mao, H. Jia, O.E.N. Noi, J. Tu, Convolutional relation network for facial expression recognition in the wild with few-shot learning, Expert Syst. Appl. 189 (2022) 116046.

[26] A. Holkar, R. Walambe, K. Kotecha, Few-shot learning for face recognition in the presence of image discrepancies for limited multi-class datasets, Image Vis. Comput. 120 (2022) 104420.

[27] A. Sufian, A. Ghosh, D. Barman, M. Leo, C. Distante, B. Li, FewFaceNet: A lightweight few-shot learning-based incremental face authentication for edge cameras, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2018–2027.

[28] Y. Zhang, H. Peng, Sample reconstruction with deep autoencoder for one sample per person face recognition, IET Comput. Vis. 11 (6) (2017) 471–478.

[29] Y. Tang, R. Salakhutdinov, G. Hinton, Deep lambertian networks, 2012, arXiv preprint arXiv:1206.6445.

[30] M. Pang, B. Wang, Y.M. Cheung, Y. Chen, B. Wen, VD-GAN: A unified framework for joint prototype and representation learning from contaminated single sample per person, IEEE Trans. Inf. Forensics Secur. 16 (2021) 2246–2259.

[31] J. Wang, K.N. Plataniotis, J. Lu, A.N. Venetsanopoulos, On solving the face recognition problem with one training sample per subject, Pattern Recognit. 39 (9) (2006) 1746–1762.

[32] W. Deng, J. Hu, X. Zhou, J. Guo, Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning, Pattern Recognit. 47 (12) (2014) 3738–3749.

[33] M. Pang, Y.M. Cheung, B. Wang, J. Lou, Synergistic generic learning for face recognition from a contaminated single sample per person, IEEE Trans. Inf. Forensics Secur. 15 (2019) 195–209.

[34] M. Pang, Y.M. Cheung, Q. Shi, M. Li, Iterative dynamic generic learning for face recognition from a contaminated single-sample per person, IEEE Trans. Neural Netw. Learn. Syst. 32 (4) (2020) 1560–1574.

[35] L. Hu, Z. Wang, H. Li, P. Wu, J. Mao, N. Zeng, ℓ-DARTS: Light-weight differentiable architecture search with robustness enhancement strategy, Knowl.-Based Syst. 288 (2024) 111466.

[36] P. Wu, H. Li, L. Hu, J. Ge, N. Zeng, A local-global attention fusion framework with tensor decomposition for medical diagnosis, IEEE/CAA J. Autom. Sin. 11 (6) (2024) 1536–1538.

[37] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.

[38] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.

[39] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4690–4699.

[40] Q. Meng, S. Zhao, Z. Huang, F. Zhou, Magface: A universal representation for face recognition and quality assessment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14225–14234.

[41] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, vol. 1, IEEE, 2005, pp. 539–546.

[42] F. Wang, X. Xiang, J. Cheng, A.L. Yuille, Normface: L2 hypersphere embedding for face verification, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 1041–1049.

[43] R. Ranjan, C.D. Castillo, R. Chellappa, L2-constrained softmax loss for discriminative face verification, 2017, arXiv preprint arXiv:1703.09507.

[44] H. Liu, X. Zhu, Z. Lei, S.Z. Li, Adaptiveface: Adaptive margin and sampling for face recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11947–11956.

[45] X. Zhang, R. Zhao, Y. Qiao, X. Wang, H. Li, Adacos: Adaptively scaling cosine logits for effectively learning deep face representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10823–10832.

[46] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: Large margin cosine loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5265–5274.

[47] N. Zhu, S. Li, A kernel-based sparse representation method for face recognition, Neural Comput. Appl. 24 (2014) 845–852.

[48] Y. Zhou, K. Liu, R.E. Carrillo, K.E. Barner, F. Kiamilev, Kernel-based sparse representation for gesture recognition, Pattern Recognit. 46 (12) (2013) 3208–3222.
[49] X. Tang, G. Feng, J. Cai, Weighted group sparse representation for undersampled face recognition, Neurocomputing 145 (2014) 402–415.
[50] W. Deng, J. Hu, J. Guo, Extended SRC: Undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
[51] P. Zhu, M. Yang, L. Zhang, I.Y. Lee, Local generic representation for face recognition with single sample per person, in: Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III, vol. 12, Springer International Publishing, 2015, pp. 34–50.
[52] M. Pang, Y.M. Cheung, B. Wang, J. Lou, Synergistic generic learning for face recognition from a contaminated single sample per person, IEEE Trans. Inf. Forensics Secur. 15 (2019) 195–209.
[53] S. Gao, Y. Zhang, K. Jia, J. Lu, Y. Zhang, Single sample face recognition via learning deep supervised autoencoders, IEEE Trans. Inf. Forensics Secur. 10 (10) (2015) 2108–2118.
[54] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, IEEE, 1994, pp. 138–142.
[55] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[56] Y. Xu, Z. Zhong, J. Yang, J. You, D. Zhang, A new discriminative sparse representation method for robust face recognition via l_2 regularization, IEEE Trans. Neural Netw. Learn. Syst. 28 (10) (2016) 2233–2242.
[57] A. Martinez, R. Benavente, The AR Face Database: CVC Technical Report, Tech. rep., CVC Technical Report, 24, 1998.
[58] C.E. Thomaz, G.A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis, Image Vis. Comput. 28 (6) (2010) 902–913.
[59] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, in: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.
[60] S. Sengupta, J.-C. Chen, C. Castillo, V.M. Patel, R. Chellappa, D.W. Jacobs, Frontal to profile face verification in the wild, in: 2016 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2016, pp. 1–9.
[61] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, (1) 2017.
[62] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[63] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning Face Representation from Scratch, Tech. rep., 2014, arXiv preprint arXiv:1411.7923.
[64] S. Almabdy, L. Elrefaei, Deep convolutional neural network-based approaches for face recognition, Appl. Sci. 9 (20) (2019) 4397.
[65] J.Y. Zhu, W.S. Zheng, F. Lu, J.H. Lai, Illumination invariant single face image recognition under heterogeneous lighting condition, Pattern Recognit. 66 (2017) 313–327.
[66] J. Cai, J. Chen, X. Liang, Single-sample face recognition based on intra-class differences in a variation model, Sensors 15 (1) (2015) 1071–1087.

Supriya Bajpai holds a joint Ph.D. degree from the Indian Institute of Technology Bombay (India) and Monash University (Melbourne, Australia), and a Master's degree from the Indian Institute of Technology Roorkee. Her research expertise lies in the areas of Artificial Intelligence, Machine Learning, Computer Vision, Natural Language Processing, Generative AI, and Healthcare AI.

Gargi Mishra received her Master's degree from the Indian Institute of Technology, Roorkee and completed her Ph.D. in Machine Learning from GGSIP University, New Delhi. She is currently an Associate Professor in the Department of Computer Science and Engineering at Bharati Vidyapeeth's College of Engineering, New Delhi. Her research interests include artificial intelligence, machine learning, and computer vision, with a particular focus on face recognition technologies.

Dr. Rachna Jain, currently working as Associate Professor in Bhagwan Parshuram Institute of Technology (GGSIPU) since Aug 2021. She has worked as Assistant Professor(Computer Science Department) in Bharati Vidyapeeth's College of Engineering (GGSIPU) from Aug 2007–Aug 2021. She did her PHD from Banasthali Vidyapith (Computer Science) in 2017. She received ME degree in year 2011 from Delhi college of engineering (Delhi University) with specialization in Computer Technology and Applications. She did her B.tech (Computer Science) in 2006 from N.C College of Engineering, Kurukshetra University. Her current research interests are Cloud Computing, Fuzzy Logic, Network and information security, Swarm Intelligence, Big Data and IoT, Deep Learning and Machine Learning.

Dr. Deepak Kumar Jain currently serves as an Associate Professor at the School of Artificial Intelligence at Dalian University of Technology in Dalian, China. Prior to this, he held the position of Assistant Professor at Chongqing University of Posts and Telecommunications in Chongqing, China. Dr. Jain's educational journey began with a Bachelor of Engineering degree, which he earned in 2010 from Rajiv Gandhi Proudyogiki Vishwavidyalaya in India. Subsequently, in 2012, he obtained a Master of Technology degree from Jaypee University of Engineering and Technology, also in India. Later, he completed his Ph.D. at the Institute of Automation, University of Chinese Academy of Sciences, Beijing, China.

Dharmender Saini graduated in Computer Science and Engineering from The Technological Institute of Textile & Sciences, Bhiwani, Haryana. He received the degree of M.Tech in Computer Science and Engineering from Department of CSE, Guru Jhambeshawar University, Hissar and received the degree of Ph.D. in Computer Science and Engineering from Jamia Millia Islamia, Delhi. He has been a Professor of Computer Science & Engineering since 2013. He has also been serving as a Principal of Bharati Vidyapeeth's College of Engineering since Jan, 2014. His research interests include Cryptography, Foundation of Computer Science, and artificial intelligence.

Amir Hussain received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1992 and 1997, respectively. He is the Founding Director of the Centre of AI and Robotics, Edinburgh Napier University, Edinburgh, U.K. He has led major national and international projects and supervised over 40 Ph.D. students. He has authored several patents and over 600 publications, including around 300 journal articles and 20 books/monographs. His research interests are cross-disciplinary and industry-led and aimed at developing responsible artificial intelligence (AI) and cognitive data science technologies to engineer the smart healthcare and industrial systems of tomorrow. Dr. Hussain is the Founding Chief Editor of Springer's Cognitive Computation journal and an invited editor for various other journals. Among other distinguished roles, he has served as the General Chair for the flagship 2020 IEEE World Congress on Computational Intelligence (WCCI) and the 2023 IEEE Smart World Congress (SWC). He is the Chair of the IEEE U.K. and Ireland Chapter of the IEEE Industry Applications Society.