STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–19.
Published online in International Academic Press (www.IAPress.org)

A Novel Graph-Based Approach for Masked Face Recognition Using Local Cycles and Spanning Trees

Siham Ahmam [1,*], Yazid Safiny [1], Nidal Lamghari [1], Abdelghani Ghazdali [2]

[1] *IDIA Laboratory, National School of Applied Sciences Khouribga, Sultan Moulay Slimane University, Morocco*
[2] *LaMDa Laboratory, National School of Applied Sciences Khouribga, Sultan Moulay Slimane University, Morocco*

Abstract Face recognition systems have encountered significant challenges due to the widespread adoption of face masks, particularly since most existing systems were trained primarily on unmasked facial images. Under these conditions, a substantial portion of facial information is occluded, which compromises the reliability of conventional recognition methods. To address this challenge, our work treats masked face recognition as a distinct problem that requires both appropriate data and adapted techniques. We first introduce a masked face dataset collected from students at our institute, which we refer to as ENSA-MFRD. We then evaluate several classical recognition techniques, including combinations of PCA, LDA, SVM, and LBPH, as well as a number of pretrained deep learning models, under masked conditions to better characterize their limitations. Building on these findings, we propose a graph-based descriptor, the Local Cycle Graph Structure (LCGS), that focuses on information available in visible facial regions and exploits local structural relationships that remain stable despite occlusion. A Maximum Spanning Tree is further applied to retain only the most discriminative connections within the graph. Experimental results on ENSA-MFRD show that the proposed LCGS 5×5 with Maximum Spanning Tree achieves an accuracy of 94.79%, outperforming all evaluated classical baselines and remaining competitive with several pretrained deep learning architectures, while offering a considerably lower computational cost and a more interpretable representation.

Keywords Face recognition, Masked face recognition, Graph theory, Local Cycle Graph Structure, Maximum spanning tree.

DOI: 10.19139/soic-2310-5070-4155

1. Introduction

In the past twenty years, telecommunications and information systems have grown rapidly, making digital security more important than ever. At the same time, biometric security has become common in computer access, e-commerce, workplace authentication, and traveler identification at airports and borders. As we rely more on these systems, researchers have focused on creating recognition methods that are both reliable and efficient.

Two strategies have traditionally governed person identification. The first, knowledge-based identification, depends on information the user must retain, such as a password or PIN. The second, token-based identification requires possession of a physical item, such as a badge, an identity card, or an access token, to gain access. Both approaches, however, share a fundamental weakness: the credential itself can be forgotten, lost, shared, or stolen, undermining the link it is intended to guarantee between the credential and the individual.

Documented weaknesses affect both approaches. PINs are vulnerable to observation, guessing, or extraction, while tokens can just as easily be lost, duplicated, or given to someone else. These limitations largely motivated the shift toward biometric traits, given that attributes such as fingerprints and facial geometry are universal, unique to the individual, measurable, and stable over time. As a result, biometric systems are now considered among the

*Correspondence to: Siham Ahmam (Email: siham.ahmam@usms.ac.ma). IDIA Laboratory, National School of Applied Sciences Khouribga, Sultan Moulay Slimane University, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

most accurate and most widely adopted identification technologies, though no single modality achieves perfect recognition in every situation.

Out of the numerous biometric modalities available today, facial recognition has achieved widespread adoption across a variety of environments [1]. Its popularity is due to the non-invasive and contactless nature of facial recognition technology, which can capture facial data from photographs or live video feeds without the user having to comply. Such flexibility is attributed to the unique structure of the human face, as well as the spatial configuration of facial features such as the eyes, nose, mouth, and ears, which make facial recognition an especially potent tool for identity verification.

Despite the human brain's capability of recognizing familiar faces with ease, intelligent systems are highly sensitive to the conditions in which the data is acquired [2]. A variety of factors, such as lighting, distance from the camera, camera lens quality, background noise, and facial pose can all negatively impact the accuracy of these systems [3]. This is one of the major challenges with such systems. For these reasons, preprocessing steps that help to prepare and size the facial images prior to performing analysis are an essential component of these systems.

The problem of "occlusion" adds to these environmental challenges. This was particularly true in the recent health crises when face masks were used by almost all individuals. Face masks completely block the lower half of an individual's face, making biometric recognition systems unable to identify individuals based on the features of their noses and mouths.

There are two primary reasons for focusing on the facial recognition performance of individuals who wear face masks. First, existing facial recognition databases do not properly replicate how individuals wear masks in their everyday lives, diminishing the performance of facial recognition algorithms in real-world scenarios. Second, there is a lack of research into other facial recognition strategies that avoid the need for data from the mouth and nose of individuals wearing masks by using the distinctive characteristics of the eyes and eyebrows.

The main contributions of this work can be summarized as follows:

- We introduce the **ENSA Masked Face Recognition Dataset (ENSA-MFRD)**, a real-world masked face dataset collected from students at the National School of Applied Sciences (ENSA) of our university. We acquired the facial images in an amphitheater while participants wore real face masks, providing masked face samples for studying and evaluating face recognition under mask-induced facial occlusion.
- We propose a **graph-based masked face recognition approach** that combines an extended 5×5 **Local Cycle Graph Structure (LCGS)** with **Maximum Spanning Tree (MST)** optimization. The 5×5 neighborhood captures richer local structural relationships among facial pixels, while the MST preserves the most relevant connections in the resulting graph representation.

The remainder of this paper is organized as follows. First, we review the most relevant methods reported in the literature. Next, we present the proposed approach and detail the procedure for constructing the new masked face dataset. Finally, we present an experimental analysis of the proposed method on our dataset and on public benchmarks, with particular attention to how dataset properties affect recognition performance, before concluding the paper.

2. Facial recognition techniques survey

2.1. Global methods

Global methods rely on statistical techniques to extract discriminative information from the entire face image. Rather than focusing on particular local features, such as the corners of the mouth, the centers of the eyes, or the nostrils, these methods process the facial image holistically. To retain the characteristics most useful for distinguishing between individuals, the image is typically vectorized and then projected into a lower-dimensional subspace. Simplicity and relatively low computational cost are among the primary advantages of this family of approaches. Under controlled acquisition conditions, global methods can achieve strong recognition performance, largely because they capture the joint relationships among pixels rather than treating them independently. Their

main limitation, however, lies in their sensitivity to variations in lighting, facial expression, and other appearance-related factors.

The underlying assumption of global approaches is that facial images, although originally represented in a high-dimensional image space, effectively reside in a much lower-dimensional subspace of that domain. This follows from the fact that only a small subset of all possible pixel configurations corresponds to plausible human faces. For instance, an 80×80 grayscale image contains 6,400 pixels and therefore admits 256^{6400} possible gray-level configurations, yet only a vanishingly small fraction of these configurations would ever be recognized as a face. Dimensionality reduction methods directly leverage this fact by projecting the data onto a more concise, meaningful subspace while preserving the discriminative information most relevant to identity.

Global methods for face recognition are generally divided into two main categories: linear and nonlinear ones. Linear algorithms assume that facial image data can be projected into a lower-dimensional linear subspace without losing essential identity information. In this category, the most widely used methods include Principal Component Analysis (PCA) [4], Linear Discriminant Analysis (LDA) [5] and Independent Component Analysis (ICA) [6].

When linearity assumptions fail to encode facial structure, nonlinear techniques are required and avoid working in the original image space, they look for a space that better captures the nonlinear relationships among facial samples. Some of the popular methods are kernel PCA, Local Tangent Space Alignment (LTSA) [7] and Multi-Dimensional Scaling (MDS) [8].

2.2. Local methods

There is a fundamental difference between local and global methods for facial analysis: the local methods split up the face into distinct regions of individual analysis, and the global methods analyze the face as a whole. This segmentation allows the analysis to focus on the areas that are important to classify individuals; thus, these algorithms are called feature-based or block-matching algorithms [9, 10].

Localized methods have a number of advantages, including low computational cost. These algorithms are generally fast, as they are processed independently for each facial area; they are therefore well suited for use in real-time scenarios. Moreover, it increases the robustness of the system against facial expression and head pose variations. For instance, if a person smiles or bends their head, only certain parts of the face shift, but other parts will still give consistent information for analysis.

On the other hand, there are certain difficulties in the design and implementation of the local methods, especially in locating the facial features with accuracy before the analysis. Some well-known methods are the Hidden Markov Model (HMM) [11] and the Dynamic Link Architecture and the feature-matching methods.

2.3. Hybrid methods

Hybrid methods are, as their name suggests, a fusion of global with local approaches, using full face images with local features. The primary goal is to maintain the discriminative capability of the global descriptors while taking advantage of the ability of the local features to deal with changes in facial appearance.

The complementary information sources usually combine into better overall recognition performance when a change in facial expression and/or pose or partial appearance occurs. But this stability is rarely achieved for free: hybrid methods tend to be more complex, and thus more challenging to apply. Examples of this type of approach include Local Feature Analysis (LFA) [12], Elastic Bunch Graph Matching (EBGM) [13], and Log-Gabor PCA (LG-PCA) [14].

3. Related work

Face recognition has been the subject of research for many years [15], and previous studies have proposed a wide variety of techniques, some of which have shown excellent performance under various conditions [16]. More recently, though, the widespread use of face masks has introduced an additional difficulty to the problem, since they hide the areas of the face that recognition methods have usually depended on. Several studies have proposed

specialized responses for recognizing faces when they are masked [17], aiming to maintain accuracy despite the occlusion caused by the mask.

The section reviews major datasets developed for both masked and unmasked face recognition, along with the main recognition methods reported in the literature in these settings.

3.1. Masked face datasets

3.1.1. COMASK20 dataset The COMASK20 dataset [18] was compiled from data collected from 300 people, each of whom was asked to record a short video lasting between 5 and 10 seconds in a variety of indoor settings. While some of the participants filmed at home, others did so in classrooms, and still others in cafés, resulting in considerable variation in the environments. The videos were captured every 0.5 s, and the resulting frames were placed into separate folders for each individual. A manual cleaning process was then carried out, during which low-quality or non-representative samples were removed, for example, those with additional occlusions, such as hair covering part of the face, or those affected by poor lighting conditions, such as overexposure. As a result of this filtering, the number of images kept per person differs from one subject to another in the dataset. After selection and preprocessing were completed, the dataset included 2754 labeled face images representing 300 distinct identities. A summary is provided in Figure 1(a).

3.1.2. Simulated Masked Face Recognition Dataset (SMFRD) The SMFRD [19] was built by applying synthetic masks to already unmasked facial images, using the Dlib library [20] to automatically generate the masks through an auto-masking process. This procedure was applied to images drawn from well-known datasets such as LFW [21] and WebFace [22], yielding a final collection of approximately 500,000 masked face images across 10,000 distinct identities. A sample of images from this dataset is shown in Figure 1(b).

3.1.3. Real-World Masked Face Recognition Dataset (RMFRD) The RMFRD [19] contains facial images of 525 individuals, captured both with and without masks, comprising approximately 5,000 masked and 90,000 unmasked images. To build the dataset, a dedicated data-mining tool was developed to collect unmasked images of public figures, along with their corresponding masked counterparts, from large-scale internet sources. This was followed by a manual cleaning process to discard low-quality or non-representative samples, while semi-automated annotation tools, including LabelImg and LabelMe [24], were used to crop and segment the facial regions.

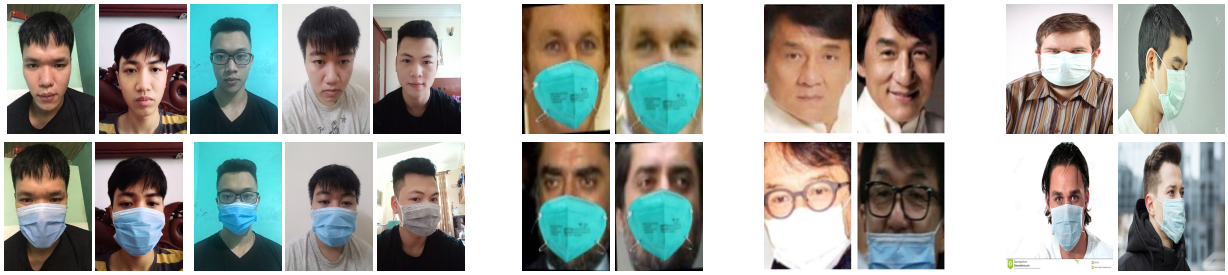
At the time of its publication, RMFRD stood out as one of the most extensive publicly available datasets for masked face recognition. An illustration of the dataset is provided in Figure 1(c).

3.1.4. Masked Face Detection Dataset (MFDD) The MFDD [19] consists of 24,771 images gathered from online sources and prior studies, annotated to distinguish between masked and unmasked faces and with bounding box coordinates for face localization. It was designed primarily for masked face detection, enabling a system to determine whether a subject is wearing a mask before recognition is even attempted. A sample of images from this dataset is shown in Figure 1(d).

3.2. Masked face recognition approaches

A number of approaches have addressed the particular difficulties posed by masked face recognition; the following paragraphs review several representative methods documented in the literature.

Kamencay et al. [25] based their architecture on a CNN, which included convolutional, max-pooling, and fully connected layers, and used inputs of size 32×32. The hierarchical features are obtained through successive applications of convolution and activation operations, and gradient descent is used to optimize the model parameters by minimizing the difference between the predicted class probabilities and the true labels. The architecture achieved an accuracy of about 98% on the ORL dataset. Chelali et al. [26] suggested an approach based on LDA: the two-dimensional face image is transformed into a vector, which results in high-dimensional data, and then LDA projects this data onto a lower-dimensional subspace that is designed to maximize class separability while at the same time retaining the discriminant information needed for the classification. The samples used for training and testing are



(a) COMASK20 (b) SMFRD (c) RMFRD (d) MFDD

Figure 1. Examples from masked face datasets: (a) original and masked faces from the COMASK20 dataset [18]; (b) simulated masked faces from the SMFRD dataset [19]; (c) masked and unmasked faces of a person from the RMFRD dataset [19]; and (d) masked faces from the MFDD dataset [19].

both projected into the same subspace, achieving 90% accuracy with a KNN classifier using Euclidean distance on a database of 7,900 face images from 395 subjects.

More recently, Zhu et al. [28] introduced FaceT, a framework built on plain Vision Transformers (ViTs) capable of handling both holistic and masked face recognition within a single model. Model parameters are initialized through a patch-reconstruction proxy task to stabilize training, while prompt-based strategies allow the model to switch between holistic and masked recognition modes. This work illustrates the broader shift toward transformer-based architectures currently underway in the field. Lata et al. [29] proposed another approach on the basis of the Eigenface technique—the use of eigenvectors of the covariance matrix of the facial images for the compact representation of faces. Their approach is twofold: first, they have to find the most salient eigenfaces in the training set, and, second, they have to project the image to be compared onto their eigenspace. Their algorithm achieved 85% accuracy on their own database and 100% on the Yale database.

Similarly, in [30] Abdeldayem et al. used a pre-trained DenseNet121 network, but this was enhanced by a few CBAM modules to better represent the features of images in challenging scenarios such as occlusion, pose, and lighting. The main objective of the algorithm was to recognize facial expressions, but the results showed that the attention modules were effective in managing occlusion, which is one of the main problems in face recognition. In the case of occlusion, Abusham et al. [31] proposed a Local Graph Structure (LGS) algorithm, which models the relationship between the center pixel and the surrounding pixels. The intensity comparisons thus obtained are then converted into a binary pattern that represents the local structure of the pixels, which results in promising performance on the ORL database. Abdullah et al. [32] extended LGS by using the Symmetric Local Graph Structure (SLGS) which expanded the size of the local graph and used symmetric comparisons of pixels.

The generative viewpoint was taken by Malakar et al. [33] who suggested an augmentation method based on CNN that reconstructs only a small portion of the face, namely the lower face region, without altering the rest and thus retaining identity-related attributes. Winarno et al. [34] compared GRU and LSTM architectures for facial expression recognition and proposed an adaptive loss selection mechanism to ensure high recognition accuracy, even in the presence of imbalanced classes, while maintaining a low computational cost. GRU was 25% more efficient than LSTM and had a similar performance at the end.

The Local Cycle Graph Structure (LCGS) was introduced in [35] which takes into consideration all the pixel relationships in a local neighborhood instead of just some. The features extracted are then reduced to lower dimensional space after PCA to be classified by an Extreme Learning Machine (ELM). LCGS was tested on both the AR and ORL databases, and it has been shown to outperform the previous graph-based methods, most notably in the case of occlusion. Vu et al. [18] introduced a three-stage system to detect a face, extract features, and recognize a face based on LBP and CNN. RetinaFace is used for detection, a CNN is used to produce high-dimensional (about 512) feature vectors, and a KNN classifier with cosine similarity is used for recognition. This system achieved F1-scores of 87% on COMASK20 and 98% on the Essex dataset. Mahmoud et al. [36] conducted a comprehensive survey of masked face recognition, detection, and unmasking techniques, together with the available datasets and



(a) (b) (c) (d)

Figure 2. Face detection with (a) Dlib, (b) Haar Cascade Classifier, (c) MTCNN, and (d) RetinaFace.

evaluation protocols, and identified open challenges such as occlusion variability and dataset realism, which are directly relevant to the present study.

Hariri [37] limited their approach to unmasked facial areas instead of the entire face. Images are first aligned using Dlib and then normalized and split into fixed-size blocks, retaining only blocks corresponding to the visible areas. A VGG-16, AlexNet, and ResNet-50 model are used to extract the features, and then the features are quantized before classification by an MLP. In a related work of attention-based facial analysis, Jaya Akeh and Putra Kusuma [38] also used the Transformer-based model, such as Vision Transformer and Swin Transformer v2, for the task of mixed facial emotion recognition, which reflects the rising importance of the attention mechanism for fine-grained facial analysis.

Finally, Septi et al. [39] suggested a system using Fisherfaces and SVM. The dataset contains original images and masked images artificially created using a GAN. Preprocessed data is then processed to extract features by PCA and LDA. SVM kernel functions like linear, polynomial, RBF, or sigmoid are used to classify. The proposed method gave the best results with high accuracy (99.79%) and high recall (99.72%) and high precision (99.72%).

4. Proposed method

4.1. Our Dataset

A sufficient amount of data is crucial for the development of a successful face recognition model. To meet this requirement, we created our own dataset with the images of the students of ENSA Khouribga. The videos were recorded in an amphitheater and featured multiple individuals in each shot. The videos feature several variations, such as individuals wearing face masks, glasses, or scarves and displaying different facial expressions. Additionally, the videos were captured from various angles to ensure a diversity of facial perspectives.

The videos collected were processed to extract individual frames from each video, with the number of frames varying according to the duration of the videos. A folder was created for each video to store the extracted frames of that particular video. The faces in each extracted frame were then detected and cropped. Four face detection techniques were used to detect the faces in the extracted frames: Dlib, Haar Cascade Classifier, MTCNN, and RetinaFace. Examples of the results obtained from these four methods can be seen in Figure 2.

A comparative evaluation revealed that Dlib, Haar Cascade Classifier, and MTCNN performed unsatisfactorily in detecting masked faces. RetinaFace, on the other hand, demonstrated greater reliability and accuracy on our images, leading to its selection for the face extraction process. For every folder of extracted frames, a corresponding folder was established to store the detected facial crops. These crops were then manually sorted into folders such that each folder represented facial images of only one individual. Extremely low-quality images were filtered out during this stage to enhance the dataset's overall quality. The entire pipeline for data collection and preprocessing is depicted

in Figure 3. Following this process, the final ENSA Masked Face Recognition Dataset (ENSA-MFRD) contained 32,186 facial images corresponding to 126 students, with the number of images per subject ranging from 27 to 528.

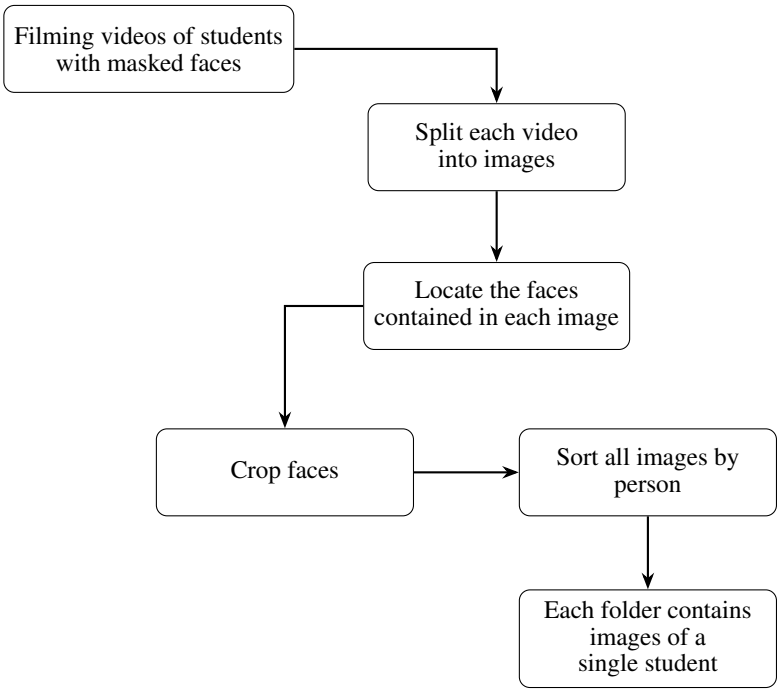


Figure 3. Masked face collection process

4.2. Masked Face Recognition Methods

4.2.1. PCA and Similarity-Based Classification As a first baseline, all masked face images from the RMFRD dataset were resized to 128×128 pixels, converted to grayscale, and then transformed into vector representations, forming a high-dimensional image space. Given the substantial redundancy in such representations, Principal Component Analysis (PCA) was used to reduce dimensionality and project the data onto a lower-dimensional eigenspace, retaining 95% of the variance. Each query image was subjected to the same preprocessing pipeline as the gallery images, that is, it underwent grayscale conversion, resizing, vectorization, and projection into the same PCA subspace before recognition; the similarity between the projected query image and all of the projected gallery images was then determined by means of two distance metrics, Euclidean and cosine distance, and the identity that corresponded to the smallest distance was taken as the predicted class for each query. Performance was based on five criteria: accuracy, precision, recall, F1-score, and execution time. Regarding Euclidean distance, the method obtained an accuracy of 57.16%, a precision of 60.11%, a recall of 44.68%, and an F1-score of 50.77%, along with an execution time of 4.302 ms; in contrast, cosine distance yielded higher values in every case, specifically an accuracy of 59.32%, a precision of 61.63%, a recall of 50.26%, and an F1-score of 55.22%, the execution time in this case being 4.290 ms. This small advantage of cosine distance over Euclidean distance is evident in all metrics, and the persistent disparity between precision and recall indicates the difficulty of consistently retrieving occluded identities.

4.2.2. Support Vector Machine (SVM) Another baseline involved training a Support Vector Machine (SVM) classifier directly on the grayscale, vectorized masked face images, skipping dimensionality reduction and instead using the original pixel representation to distinguish identities. The accuracy achieved was 57.09%, the precision 57.44%, the recall 55.13%, and the F1-score 56.22%, with an execution time of 76.076 ms. In this case, precision

and recall are well balanced, although this balance comes at the cost of a considerably longer execution time compared to PCA-based methods.

4.2.3. Linear Discriminant Analysis (LDA) A baseline approach applied Linear Discriminant Analysis (LDA) directly to the masked face images, without applying PCA for dimensionality reduction. We converted the images to grayscale and represented them as vectors, using the file names as class labels. We computed LDA-based discriminative projections from the training data and then classified the resulting representations. The accuracy was 51.87%, precision 55.54%, recall 19.99%, and F1-score 28.49%, with an execution time of 4.004 ms. The low recall relative to the precision shows that LDA struggles to generalize across identities when applied directly to high-dimensional masked face data.

4.2.4. Local Binary Pattern Histogram (LBPH) In face recognition, a texture-based method known as LBPH captures local spatial structure by comparing each pixel with its neighborhood, producing texture descriptors that remain robust to some changes. All images were converted to grayscale, and LBPH features were then extracted for classification. This approach obtained an accuracy of 55.99%, a precision of 56.65%, a recall of 51.20%, and an F1-score of 53.72%, with an execution time of 106.708 milliseconds. Although precision and recall are fairly close, LBPH's time is by far the highest among the baseline methods because it requires pixel-by-pixel computation.

4.2.5. PCA with LDA In the experiment, the images were converted to grayscale, resized to 128 by 128 pixels, and converted into a vector form, with each row of the resulting matrix representing one facial image. PCA was first used to obtain feature vectors that captured the directions of greatest variance in the data; the resulting projections were then used by LDA for classification. The query images were also subjected to the same preprocessing and projection steps before classification. The method reached an accuracy of 52.16 %, a precision of 54.88 %, a recall of 30.81 %, and an F1-score of 37.65 %, with an execution time of 3.911 ms. Although combining PCA with LDA increases recall compared to LDA alone, recall remains much lower than precision.

4.2.6. PCA with SVM In this method, masked face images were converted to grayscale and vectorized, then subjected to PCA-based dimensionality reduction with the retained variance ratio set to 0.95. The resulting low-dimensional features were then used as input to an SVM classifier. This method achieved an accuracy of 56.02%, a precision of 56.37%, a recall of 53.67%, and an F1-score of 54.95%, with an execution time of 0.768 ms. Accuracy and F1-score stay close to those of the raw-pixel SVM, but the execution time drops sharply once PCA is used for dimensionality reduction.

4.2.7. PCA with LDA and SVM In this method, masked face images were converted to grayscale and vectorized, then subjected to PCA-based dimensionality reduction, with the retained variance ratio set to 0.95. The resulting features were subsequently processed using LDA, which maximizes between-class separation while minimizing within-class dispersion, and the resulting discriminative representations were used as input to an SVM classifier. In summary, feature extraction was performed using PCA and LDA, while classification was performed using SVM. This method achieved an accuracy of 55.40%, a precision of 54.91%, a recall of 59.99%, and an F1-score of 57.29%, with a computational time of 0.765 ms. This is the only configuration where recall exceeds precision, and it also yields the highest F1-score among all baselines.

All the benchmark techniques described above were first evaluated on the masked face portion of the RMFRD dataset under a 5-fold identity-disjoint protocol and subsequently on the newly created ENSA-MFRD dataset under the same protocol. The corresponding results are reported in Table 1.

Since both datasets are evaluated under the same identity-disjoint 5-fold protocol, the substantial performance gap between them cannot be attributed to differences in the evaluation methodology. It must instead originate from the intrinsic characteristics of the datasets themselves. ENSA-MFRD was acquired under comparatively controlled conditions, with stable illumination, consistent background, and uniform image quality, which reduces the intra-class variability that classical, hand-crafted representations such as PCA, LDA, and LBPH are known

Table 1. Results of different masked face recognition methods on the ENSA-MFRD (5-fold identity-disjoint protocol).

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time (ms)
PCA + Euclidean	83.05	86.20	79.06	82.20	2.551
PCA + Cosine	85.93	86.21	85.67	85.84	2.446
LDA	72.24	84.46	55.36	66.41	2.628
SVM	84.61	88.50	79.88	83.80	6.828
LBPH	84.64	89.31	79.59	83.71	14.054
PCA + LDA	77.88	81.82	73.57	76.67	2.339
PCA + SVM	85.82	86.27	85.15	85.66	0.089
PCA + LDA + SVM	84.34	86.41	81.68	83.80	0.068

to handle poorly. This relative homogeneity likely makes identities easier to separate even under a strict identity-disjoint split. In addition, the number of images per identity in ENSA-MFRD is highly uneven, ranging from 27 to 528, which may favor recognition performance for well-represented identities and contribute to the overall higher scores observed, up to 85.93% accuracy for PCA with cosine distance. These results should therefore be read as reflecting the comparatively favorable acquisition conditions of ENSA-MFRD, rather than a difference in the intrinsic difficulty of subject-disjoint recognition across the two datasets.

This asymmetry motivates a further experiment. Pooling RMFRD and ENSA-MFRD into a single evaluation set tests whether the two datasets are simply of unequal intrinsic difficulty or whether their combination changes the classification problem itself by exposing the methods to a broader and more heterogeneous identity pool. The results of this experiment using the combined dataset are given in Table 2, with the experiment carried out under the same 5-fold identity-disjoint protocol.

Table 2. Results of different masked face recognition methods on the Combined dataset (RMFRD + ENSA-MFRD), 5-fold identity-disjoint protocol.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time (ms)
PCA + Euclidean	61.20	66.40	46.02	54.10	4.550
PCA + Cosine	63.09	67.68	51.07	57.83	4.609
LDA	50.69	52.18	24.85	32.12	4.740
SVM	64.95	66.49	60.27	63.21	61.520
LBPH	67.21	68.52	63.62	65.96	158.228
PCA + LDA	55.85	71.96	19.68	30.73	4.629
PCA + SVM	64.07	67.65	53.94	59.99	0.897
PCA + LDA + SVM	64.55	62.80	71.49	66.85	0.955

A consistent, direction-dependent effect appears when the two datasets are merged. Compared with ENSA-MFRD alone, all methods show decreased accuracy and F1-score when RMFRD's more heterogeneous samples are added to the identity pool. This confirms that the higher scores on ENSA-MFRD alone stem from its more favorable acquisition conditions, a benefit that diminishes as the identity set becomes more diverse. Conversely, for RMFRD alone, seven of the eight methods show increased accuracy, with gains ranging from about 4 points for the PCA-based similarity methods and for PCA with LDA to over 9 points for LBPH and for PCA with LDA and SVM. These findings indicate that including additional identities and images, while reducing ENSA-MFRD's individual scores, expands the discriminative feature space available to each method beyond what RMFRD alone offers. The exceptions concern the two LDA-based configurations without SVM. Standalone LDA is the only method whose accuracy does not improve, 51.87% on RMFRD versus 50.69% on the combined set, although its F1-score rises slightly from 28.49% to 32.12%. PCA with LDA shows the opposite pattern: its accuracy increases from 52.16% to 55.85%, but its F1-score drops from 37.65% to 30.73%, as its recall falls to 19.68%, the lowest value in Table 2. These two configurations yield the lowest recall in both settings, in line with the known difficulty of LDA with high-dimensional data as the identity pool expands.

To better understand this behavior, a complementary closed-set identification experiment was carried out on the combined dataset using a single train/test split, in which each identity appears in both partitions. Per-class results

for a subset of identities are reported in Table 3. P27, P28, and P29 specifically correspond to identities drawn

Table 3. Per-class results of a complementary closed-set identification experiment on the combined RMFRD + ENSA-MFRD dataset (single train/test split).

Identities	Precision	Recall	F1-score
P27	1.00	1.00	1.00
P28	1.00	1.00	1.00
P29	1.00	1.00	1.00
anhu	0.00	0.00	0.00
aoziyi	0.00	0.00	0.00
baike	0.00	0.00	0.00
caihancen	0.00	0.00	0.00

from ENSA-MFRD, whereas anhu, aoziyi, baike, and caihancen originate from RMFRD. In this experiment, all metrics for the RMFRD identities are equal to zero, whereas the ENSA-MFRD identities are perfectly recognized. Although this experiment follows a different protocol from Table 2, it is consistent with the hypothesis that the quality and characteristics of the RMFRD dataset adversely affect the performance of the evaluated recognition methods.

4.3. Our method

4.3.1. Maximum spanning tree Within graph theory, a *maximum spanning tree* (MST) of a connected, undirected, weighted graph is defined as a spanning tree, that is, a subgraph connecting all vertices without cycles, whose total edge weight is maximal among all spanning trees of the graph. This method offers the best way to retain the strongest links between vertices while ensuring the graph remains fully connected, and in different cases, it can be seen as representing the most informative or significant tree structure obtainable from the original weighted graph. In practice, if a graph represents relationships, similarities, or interactions among a set of elements, the MST keeps the most relevant connections while reducing structural redundancy.

Figure 4 illustrates an example of a weighted graph. This graph admits several possible spanning trees, each of which connects all vertices of the graph G using the minimum number of edges required to maintain connectivity. Among these possible spanning trees, the *maximum spanning tree* is the one whose total edge weight is the highest. Figure 5 illustrates the maximum spanning tree corresponding to graph G.

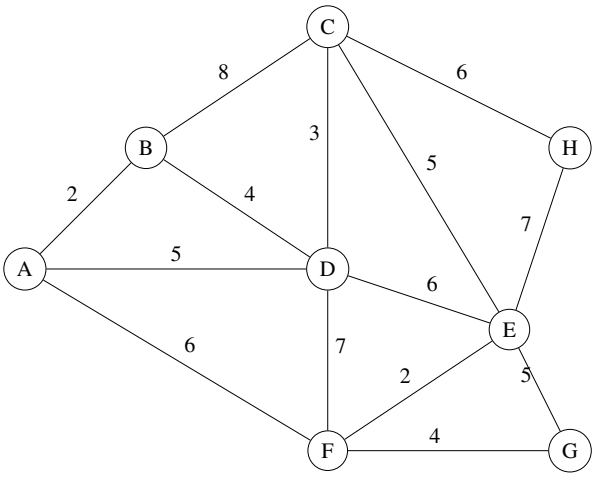


Figure 4. A weighted graph G

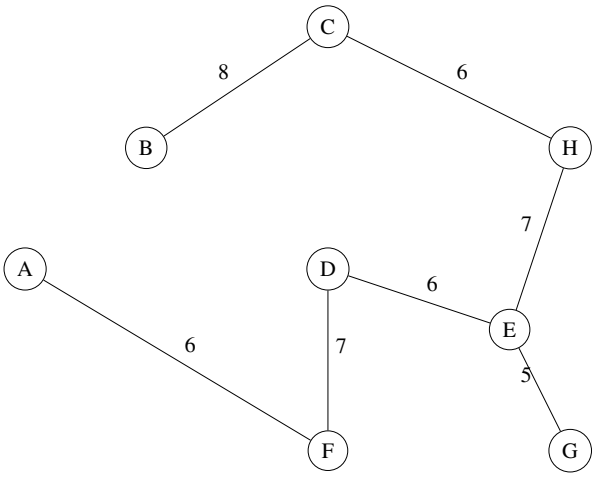


Figure 5. The maximum spanning tree of G

4.3.2. Proposed method In this section, we present the proposed method for masked face recognition using the newly constructed ENSA-MFRD dataset, which contains only masked facial images. As described previously, following the data collection stage, all images underwent preprocessing to prepare them for subsequent feature extraction and recognition.

Preprocessing. In our case, since the facial images were manually selected and verified during dataset construction, no additional conventional preprocessing steps, such as denoising or face alignment, were required. Nevertheless, a specific preprocessing stage was applied to adapt the images to the proposed feature extraction method. Because the dataset contains only masked faces, the lower facial region is largely occluded, providing limited discriminative information. For this reason, this region was removed, resulting in cropped facial images of size 64×128 pixels. In addition, since the proposed descriptor relies on local neighborhood relationships and edge information, a 2-pixel padding was added to all four sides of each image to preserve neighborhood consistency near the image boundaries.

Feature extraction. A crucial stage of the recognition process, feature extraction has been informed here by the analysis of previous studies, which points to graph-based descriptors as particularly effective for handling occluded facial images. Among these, the Local Cycle Graph Structure (LCGS) stands out for its promising performance. However, the original LCGS method relies on a limited 3×3 neighborhood. To capture richer local structural information, we propose extending this representation to a larger 5×5 neighborhood, as illustrated in Figure 6.

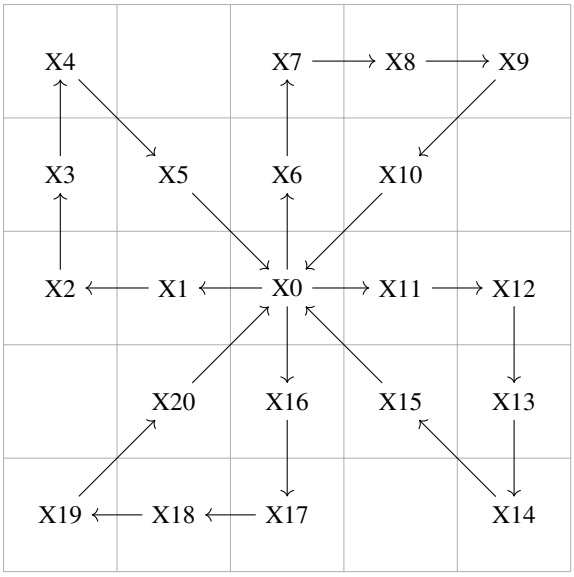


Figure 6. Neighbors considered for feature extraction.

For a central pixel X_0, the proposed 5×5 structure uses 20 of the 24 surrounding pixels. Each pair of successive pixels in the predefined traversal order is connected by an edge, as illustrated in Figure 7. Each edge is assigned a binary value, either 0 or 1, based on the intensity relationship between two consecutive pixels: if the current pixel's intensity is greater than that of the preceding pixel, the corresponding edge is assigned 1; otherwise, it is assigned 0. In this way, the local neighborhood around the central pixel is represented as a binary graph structure that encodes the relative intensity transitions between adjacent pixels.

Since the proposed neighborhood structure contains several cycles, a cycle-free representation is required to obtain a more compact and informative graph. In particular, each local configuration around the central pixel contains multiple cyclic connections, which introduce structural redundancy. To address this issue, a graph-theoretic optimization step is applied using the *maximum spanning tree* (MST) algorithm. The use of MST allows the removal of redundant edges that form cycles while preserving the most informative connections within the local graph structure. As a result, the local graph is reduced from 24 to 20 edges. Each of the 24 edges is assigned a

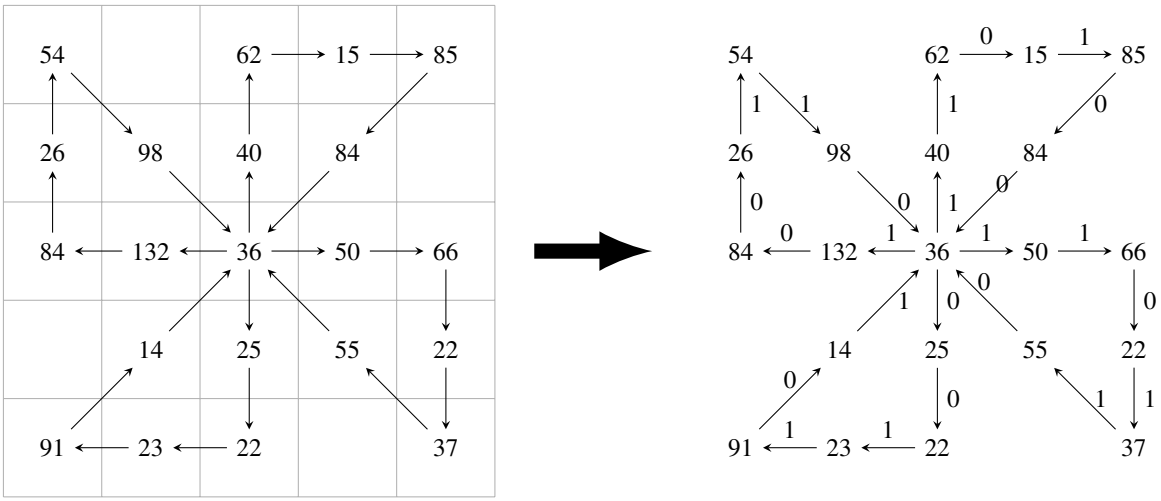


Figure 7. Association of weights to edges.

fixed positional value 2^p, where p is its index in a traversal order that always starts from the edge connecting X_0 and X_1, and the descriptor value of the central pixel is the sum of the positional values of the 20 edges retained by the MST. To ensure this process can also be applied to boundary pixels, a padding operation was introduced during preprocessing, as described previously. Algorithm 1 summarizes this procedure formally.

Algorithm 1 LCGS 5×5 + MST local descriptor for one pixel

Input: Grayscale image I padded by 2 pixels on each side; center position (i, j); vertex set $\mathcal{V} = \{X_0, X_1, \ldots, X_{20}\}$ and ordered edge list $\mathcal{E} = \{e_p = (a_p, b_p)\}_{p=0}^{23}$ defined in Figure 6, forming four 6-edge cycles $H_1, \ldots, H_4$ that share only X_0, where a_p is the current pixel and b_p its preceding pixel in the traversal order

Output: Descriptor value $D(i, j)$

Step 1: Edge weighting

for $p \leftarrow 0$ **to** 23 **do**
  $w_p \leftarrow 1$ if $I(a_p) > I(b_p)$, else $w_p \leftarrow 0$ ▷ binary intensity transition
  $v_p \leftarrow 2^p$ ▷ fixed positional value
end for

Step 2: Maximum spanning tree (Kruskal)

$\mathcal{L} \leftarrow$ edges of $\mathcal{E}$ sorted by w_p in non-increasing order, ties broken by increasing p
MAKESET(x) for each $x \in \mathcal{V}$
$S \leftarrow 0$
for all $e_p = (a_p, b_p) \in \mathcal{L}$ **do**
  if FIND$(a_p) \neq$ FIND(b_p) **then** ▷ e_p joins two components
    UNION(a_p, b_p)
    $S \leftarrow S + v_p$ ▷ e_p retained in the tree
  end if
end for

Step 3: Output

$D(i, j) \leftarrow S$ ▷ 20 of the 24 edges are retained
return $D(i, j)$

Because each of the four cycles $H_1, \ldots, H_4$ forms a closed loop of 6 edges over 6 pixels, exactly one edge per cycle is always discarded as redundant during the maximum spanning tree construction, regardless of the specific pixel intensities involved (4 discarded edges out of 24 in total, leaving 20 retained edges). However, *which* of the 6 edges within a given cycle is discarded depends on the actual weight pattern of that cycle rather than on a fixed position, since Kruskal's algorithm always processes edges of weight 1 before edges of weight 0 and, within each weight class, preserves their original index order p.

Data preparation for classification. After applying the proposed feature extraction method, a new feature matrix was obtained for each image. This matrix was subsequently converted into a one-dimensional feature vector to facilitate classification, and all resulting vectors were assembled into a global feature matrix, with each row corresponding to a single facial image. It should be noted that the dataset was not balanced with respect to the number of images per identity, as each subject was represented by between 27 and 528 images.

Two complementary evaluation protocols were used to split this global feature matrix. In the first, a closed-set 80/20 split was applied, with 80% of the images used for training and the remaining 20% for testing, with no constraints on identity overlap between the two sets. In the second, a stricter identity-disjoint protocol was adopted, using GroupKFold cross-validation (5-fold and 10-fold) to ensure that no identity present in the training set for a given fold also appeared in the corresponding test set, with disjointness verified programmatically for every fold.

Dimensionality reduction. The proposed graph-based descriptor is used to extract facial features from each image. As a result, the resulting feature matrix is generally high-dimensional, which may increase computational cost during both training and testing. To resolve this issue, a dimensionality reduction stage was incorporated into the pipeline. As a first step, PCA projects the extracted features onto a lower-dimensional subspace while retaining the most informative variance, with a retention threshold of 0.95 in our experiments. Linear Discriminant Analysis (LDA) is then applied to the resulting feature vectors to learn a discriminative representation of the data. More specifically, LDA seeks a subspace in which samples belonging to the same class are grouped as closely as possible, while samples from different classes remain well separated.

Classification phase. In the final recognition stage, an SVM with a radial basis function (RBF) kernel was trained and tested on our dataset, with the regularization parameter set to $C = 1.0$.

5. Results and Discussion

This section reports the experimental evaluation of the proposed method under four complementary protocols, presented in order of increasing difficulty: a kernel sensitivity analysis on the primary dataset, closed-set classification under a conventional random split, identity-disjoint pairwise verification using 5-fold and 10-fold cross-validation across all datasets, and cross-dataset generalization using a single model trained once and evaluated without retraining. For readability, since two ORL variants appear throughout this section, both built from the same masked ORL images, we abbreviate the unaugmented, synthetically masked variant as ORL-M and the augmented, synthetically masked variant as ORL-MA; all other dataset names are kept as is.

5.1. Kernel Selection

Before deciding on the classifier configuration for later experiments, we compared polynomial, radial basis function (RBF), sigmoid, and linear support vector machine (SVM) kernels on the main dataset using a closed-set 80/20 split, along with the PCA-to-LDA feature pipeline. To check the stability of these results, we trained and evaluated each configuration independently twice, and all repetitions yielded identical results.

Accuracy exceeds 98.5% for all four kernels, and precision, recall, and F1-score remain closely aligned with it, a pattern suggesting that the LDA-reduced feature space is highly separable, with this separability holding consistently across identities rather than resulting from a handful of dominant classes. The linear and RBF kernels edge out sigmoid and polynomial by a small margin, in line with the fact that LDA already optimizes for linear class separability, leaving little room for a non-linear kernel to add much. RBF was therefore retained as the default configuration for the remainder of this study.

Table 4. Kernel comparison for closed-set classification on the primary dataset (80/20 split)

Kernel	Accuracy	Precision	Recall	F1-score
Linear	0.9964	0.9958	0.9945	0.9951
RBF	**0.9967**	**0.9971**	**0.9959**	**0.9967**
Sigmoid	0.9902	0.9899	0.9879	0.9888
Polynomial	0.9851	0.9924	0.9810	0.9853

5.2. Closed-Set Classification

We first report closed-set, multi-class classification performance under a single random 80/20 split, where the classifier predicts the identity directly rather than performing pairwise verification. In this setting, the same identities can appear in both the training and test partitions, so results should be interpreted as an upper bound on within-distribution recognition rather than as a test of generalization. Precision, recall, and F1-score are macro-averaged, computed independently for each identity, and then averaged with equal weight so that identities with few images are not overshadowed by identities with many. This makes the three metrics meaningfully distinct from overall accuracy and exposes disparities across identities that accuracy alone would hide.

The proposed method was first evaluated on the ENSA-MFRD dataset. To further assess its effectiveness and generalization ability, we also applied it to the ORL dataset [27]. Since ORL is relatively limited in size, a data augmentation strategy was applied to increase the number and variability of training samples: horizontal and vertical translations of 3, 6, and 9 pixels in both directions, followed by rotations of 3°, 6°, 9°, -3°, -6°, and -9° applied to the translated images. Because ORL originally contains only unmasked faces, an additional occlusion step was applied to simulate mask-like facial occlusions and make the dataset suitable for masked face recognition evaluation; this synthetic masking was applied to both ORL variants reported below, and the augmentation procedure was applied on top of it for ORL-MA only, so the two rows differ solely in whether augmentation was used, not in whether masking was applied. For COMASK20, non-masked and masked images for each identity were pooled before the random split, so both the training and the test partitions contain a mixture of masked and non-masked samples of the same individuals. LFW-masked refers to the LFW dataset [21] in which synthetic masks were added to the face images using the MaskTheFace tool [23]. The results of this closed-set test are summarized in Table 5.

Table 5. Closed-set classification accuracy under a random 80/20 split

Dataset	Accuracy	Precision	Recall	F1-score
ENSA-MFRD	0.9967	0.9971	0.9959	0.9967
COMASK20	**0.9983**	**0.9983**	**0.9963**	**0.9973**
ORL-MA	0.9868	0.9869	0.9868	0.9868
ORL-M	0.7073	0.6859	0.6996	0.6564
LFW-masked	0.5954	0.5202	0.4709	0.4943
RMFRD	0.5620	0.3920	0.3790	0.3853

For ENSA-MFRD, COMASK20, and ORL-MA, accuracy and macro F1-score are less than one percentage point apart, indicating that performance is not driven by a small number of well-represented identities, even though the number of images per identity in ENSA-MFRD is highly uneven. For ORL-M, LFW-masked, and RMFRD, the gap is much larger, with the macro F1-score about 5, 10, and 17 percentage points behind accuracy, respectively. For LFW-masked and RMFRD, this is consistent with an uneven number of images per identity, where a handful of well-represented identities inflate the accuracy figure while poorly represented ones are hard to learn under a random split and pull the macro-averaged metrics down. ORL-M, in contrast, contains exactly 10 images per identity, so its gap is more likely due to the small number of training samples per identity under the 80/20 split,

which leaves some identities poorly recognized. Augmentation appears to close much of this gap, since accuracy rises from 0.7073 for ORL-M to 0.9868 for ORL-MA, and the accuracy–F1 gap nearly disappears once translation and rotation increase both the volume and the diversity of training samples per identity.

5.3. Identity-Disjoint Pairwise Verification

The closed-set protocol above establishes an upper bound on achievable performance, but it does not test the model's ability to recognize identities absent from training, since the same individuals can appear on both sides of the split. To address this, we adopt a pairwise verification protocol: for each fold, PCA and LDA are fitted exclusively on the training identities, a binary SVM is trained to discriminate between genuine (same-identity) and impostor (different-identity) pairs, and evaluation is performed on pairs drawn only from held-out identities. For every run, the disjointness of the identities between folds was verified programmatically. Tables 6 and 7 report results under 5-fold and 10-fold cross-validation, respectively.

Table 6. 5-fold identity-disjoint pairwise verification

Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	EER
ENSA-MFRD	**0.9479±0.0403**	**0.9489±0.0364**	**0.9466±0.0563**	**0.9477±0.0418**	**0.9560±0.0289**	**0.0702±0.0300**
COMASK20	0.9118±0.0374	0.9339±0.0467	0.8810±0.0434	0.9067±0.0381	0.8736±0.0307	0.1201±0.0430
ORL-M	0.9000±0.0568	0.8516±0.0753	0.9404±0.0390	0.8938±0.0382	0.9255±0.0163	0.1907±0.0237
LFW-masked	0.5720±0.0151	0.5652±0.0157	0.6281±0.0220	0.5947±0.0122	0.5956±0.0148	0.4317±0.0207
RMFRD	0.5456±0.0234	0.5375±0.0203	0.6555±0.0542	0.5897±0.0291	0.5772±0.0300	0.4417±0.0232

Table 7. 10-fold identity-disjoint pairwise verification

Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	EER
ENSA-MFRD	**0.9231±0.0447**	0.9295±0.0465	**0.9156±0.0680**	**0.9225±0.0476**	**0.9459±0.0357**	**0.0801±0.0402**
COMASK20	0.9083±0.0287	**0.9313±0.0753**	0.8919±0.1003	0.9112±0.0375	0.9169±0.0453	0.1124±0.0426
ORL-M	0.9125±0.0918	0.8277±0.1037	0.9017±0.1379	0.8631±0.0997	0.9144±0.0700	0.1508±0.1036
LFW-masked	0.5659±0.0260	0.5588±0.0231	0.6187±0.0785	0.5851±0.0488	0.5941±0.0262	0.4259±0.0271
RMFRD	0.5342±0.0443	0.5303±0.0399	0.6143±0.0622	0.5681±0.0431	0.5396±0.0492	0.4563±0.0323

Two patterns stand out from Tables 6 and 7. First, ENSA-MFRD, COMASK20, and ORL-M consistently reach AUC-ROC scores above 0.85, while RMFRD and LFW-masked plateau around 0.54 to 0.60, barely above chance level. The gap cannot be explained by either the number of identities or dataset size, as ORL—the smallest dataset in the evaluation—still matches or outperforms COMASK20 despite containing far fewer images. A potentially more credible answer can be found in the acquisition conditions: ENSA-MFRD, COMASK20, and ORL were all collected under relatively uniform backgrounds, with minimal variation in pose or illumination. RMFRD and LFW, instead, collate images scraped from uncontrolled real-world sources. The proposed graph-based descriptor is therefore sensitive to this kind of variation, as non-uniform illumination, pose changes, and inconsistent cropping alter the ordinal intensity relationships within each local neighborhood on which it relies.

The balance between precision and recall, however, does not follow the same acquisition-condition split described above. Precision exceeds recall for ENSA-MFRD and COMASK20, whereas the reverse holds for ORL-M, LFW-masked, and RMFRD, a pattern consistent across the 5-fold and 10-fold protocols. For ORL-M, LFW-masked, and RMFRD, this elevated recall points to a bias toward predicting genuine pairs, likely because different identities sharing similar local occlusion patterns can yield small feature differences and be misclassified as genuine. The opposite pattern for ENSA-MFRD and COMASK20 suggests that, under more controlled acquisition conditions, pairwise features become sufficiently discriminative that the classifier becomes comparatively conservative, prioritizing confident genuine predictions over broader recall. Shifting from 5-fold to 10-fold cross-validation slightly reduces mean performance and increases variance across nearly all datasets, which is unsurprising, as 10-fold leaves fewer identities and genuine pairs per test fold, making per-fold estimates noisier,

although dataset ranking remains stable regardless of the protocol. This same protocol still fits PCA, LDA, and the SVM independently on each target dataset, and the next experiment removes this degree of freedom entirely.

5.4. Cross-Dataset Generalization

To test whether the discriminative space learned on one dataset transfers to another without retraining, the most demanding setting considered in this study, we trained a single model once on the full training partition of ENSA-MFRD and evaluated it, without any fine-tuning, on pairs drawn from four external datasets never seen during training. For COMASK20, genuine pairs consist of a non-masked and a masked image of the same identity, reflecting the realistic scenario of matching an unmasked reference to a masked probe. For ORL, images were fitted with a synthetic mask overlay before feature extraction. The cross-dataset results in Table 8 mirror the pattern already seen in the internal identity-disjoint experiments: a model trained only on ENSA-MFRD transfers strongly to COMASK20, with an AUC of 0.9102, transfers reasonably to ORL-M, with an AUC of 0.8601, but degrades sharply on RMFRD and LFW-masked, both below an AUC of 0.59. Seeing the same split appear under two independent protocols, once where each dataset trains its own model and once where a single ENSA-MFRD model is frozen and applied elsewhere, strengthens the case that acquisition condition, shared between ENSA-MFRD, COMASK20, and ORL, is doing most of the work here, rather than per-dataset hyperparameter fitting. Unlike the identity-disjoint experiments above, precision now exceeds recall on three of the four external datasets (ORL-M, RMFRD, and LFW-masked), COMASK20 being the exception, and this asymmetry is most pronounced for RMFRD and LFW-masked, where recall drops below 0.17 while precision remains above 0.60. A frozen model, evaluated on data it has never adapted to, appears to fall back on a conservative decision boundary that correctly rejects most impostor pairs but also misses a large share of genuine pairs on the harder, uncontrolled datasets.

Table 8. Cross-dataset generalization of a single ENSA-MFRD-trained model, evaluated without retraining

Test dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	EER
COMASK20	**0.8331**	**0.8228**	**0.8490**	**0.8357**	**0.9102**	**0.1636**
ORL-M	0.7683	0.7895	0.7317	0.7595	0.8601	0.2341
RMFRD	0.5429	0.6820	0.1609	0.2604	0.5844	0.4439
LFW-masked	0.5298	0.6066	0.1693	0.2647	0.5507	0.4624

5.5. Comparison with state of the art

The performance of various graph-based configurations based on the proposed Local Cycle Graph Structure (LCGS) descriptor is shown in Table 9 when tested on ENSA MFRD using a five-fold identity-disjoint protocol. The purpose of this comparison is to investigate the effects of graph granularity and spanning-tree-based optimization on masked face recognition performance.

Table 9. Performance comparison of different LCGS-based graph configurations on ENSA-MFRD, 5-fold identity-disjoint protocol.

Method	Accuracy	Precision	Recall	F1-score	AUC-ROC	EER
LCGS 3×3	0.8070	0.8068	0.8117	0.8079	0.9014	0.1880
LCGS 5×5 + MinST	0.8642	0.8653	0.8613	0.8632	0.9223	0.1826
LCGS 5×5 (no MST)	0.8614	0.8630	0.8550	0.8590	0.9277	0.1722
LCGS 5×5 + MST (proposed)	**0.9479**	**0.9489**	**0.9466**	**0.9477**	**0.9560**	**0.0702**

The LCGS configuration with dimensions 3×3 achieved an accuracy of 80.70% and an F1 score of 80.79%. With only nine nodes, this set captures limited local structure and thus provides insufficient spatial information for localization, leaving many subtle facial features underrepresented. When the graph resolution is increased to 5×5 without spanning tree reduction, accuracy rises to 86.14%, and the F1 score reaches 85.90%. This

improvement shows that a finer graph yields a more detailed facial representation that encodes more precise structural relationships between regions, and it confirms that spatial granularity alone enhances the descriptive power of the LCGS representation.

The impact of graph topology was further examined using methods on the 5×5 graph, namely the Minimum Spanning Tree and the Maximum Spanning Tree (MST). The MinST achieved an accuracy of 86.42% and an F1 score of 86.32%, representing only a small improvement over the simple 5×5 graph. By retaining only the lowest-cost edges necessary for connectivity, the MinST reduces redundancy and produces a more compact graph, but it also sacrifices some of the stronger connections between facial regions. Of all the graph-based configurations, the proposed LCGS 5×5 using MST performs the best, attaining an accuracy of 94.79%, a precision of 94.89%, a recall of 94.66%, an F1 score of 94.77%, an AUC-ROC of 95.60%, and the lowest equal error rate of 7.02%.

The difference between MinST and MST shows that the strongest connections between regions—not the weakest—have the greatest discriminatory value in masked face recognition. Because the mask removes much of the visible face, preserving the most informative structural links among the remaining parts is essential to achieve strong recognition performance. By preserving these strong connections, the Maximum Spanning Tree focuses the graph on the relationships most important for distinguishing identity, resulting in significant increases in both accuracy and F1 score compared with the simple 5×5 graph and its MinST version.

Overall, these results show that both graph resolution and spanning tree selection have a substantial effect on recognition performance. Increasing granularity from 3×3 to 5×5 provides a first improvement, while the choice of spanning tree strategy proves even more decisive: the Minimum Spanning Tree only slightly refines the baseline, whereas the Maximum Spanning Tree, by preserving the strongest structural links, delivers the largest performance gain among all graph-based configurations.

Table 10 reports the performance and processing time of several pretrained deep learning models evaluated on the ENSA-MFRD dataset under the same five-fold identity-disjoint protocol, alongside the proposed LCGS 5×5 + MST method. All timing measurements were carried out on the same Kaggle compute instance, with the deep learning models running on GPU and the proposed method running entirely on CPU.

Table 10. Performance and processing time comparison of pretrained deep learning models and the proposed method on the ENSA-MFRD dataset, 5-fold identity-disjoint protocol (mean $\pm$ std over folds; timings measured on the same compute instance, 32,186 images).

Model	**Acc (%)**	**Prec (%)**	**Rec (%)**	**F1 (%)**	**Time/img**	**FPS**	**HW**
ArcFace[a]	98.67 ± 0.84	99.73 ± 0.53	97.60 ± 1.77	98.64 ± 0.87	52.86 ms	18.9	GPU
FaceNet[b]	96.14 ± 2.05	97.35 ± 2.66	94.94 ± 3.32	96.08 ± 2.09	27.24 ms	36.7	GPU
ResNet50	91.78 ± 2.42	93.27 ± 2.58	90.15 ± 4.77	91.60 ± 2.58	13.13 ms	76.2	GPU
VGG19	90.19 ± 1.51	90.70 ± 1.98	89.64 ± 2.80	90.13 ± 1.55	15.35 ms	65.1	GPU
InceptionV3	87.29 ± 1.86	91.57 ± 3.29	82.46 ± 5.79	86.57 ± 2.38	17.22 ms	58.1	GPU
Proposed	$\mathbf{94.79 \pm 4.03}$	$\mathbf{94.89 \pm 3.64}$	$\mathbf{94.66 \pm 5.63}$	$\mathbf{94.77 \pm 4.18}$	**20.39 ms**	**49.0**	**CPU**

[a]IResNet100 backbone. [b]InceptionResnetV1 backbone.

ArcFace achieved the best recognition performance, with 98.67% accuracy and 98.64% F1 score, but also the highest processing cost at 52.86 ms per image (18.9 FPS), despite running on GPU. FaceNet followed a similar pattern, ranking second in accuracy (96.14%) and second in processing time (27.24 ms). Among the general-purpose convolutional networks, ResNet50 achieved the highest accuracy (91.78%) and the fastest processing time overall (76.2 FPS), while VGG19 and InceptionV3 trailed in both accuracy and speed. The proposed LCGS 5×5 + MST method reaches 94.79% accuracy, ranking below ArcFace and FaceNet but above ResNet50, VGG19, and InceptionV3, despite the latter three running on GPU. Its variability across folds is somewhat higher than that of the deep learning models, with a standard deviation of 4.03 points on accuracy compared to 0.84 to 2.42 points for the pretrained networks, which suggests that the classical pipeline is more sensitive to the specific identity split used in each fold. Regarding speed, the proposed method takes 20.39 ms (49.0 FPS) to process each image on the CPU, which is faster than both ArcFace and FaceNet and only moderately slower than the convolutional networks

that use a GPU. Since the deep learning models benefit from GPU acceleration while the proposed method runs on CPU only, this comparison is conservative with respect to the proposed method; even so, ArcFace requires nearly 2.6 times as much processing time per image.

In general, the proposed method achieves a good balance between accuracy and computational cost. Even though its performance is not equal to that of specialized deep metric learning models such as ArcFace or FaceNet, it is more accurate than conventional convolutional networks, even when it is running entirely on a CPU, and this is why it is particularly suitable for use in situations involving limited resources or edge deployment, where access to GPUs and model transparency are often practical problems.

6. Conclusion

Owing to its non-intrusive nature and its applicability in diverse real-world scenarios, face recognition has become one of the most widely accepted biometric solutions. However, the large-scale use of face masks poses inherent difficulties for traditional face recognition systems, as a large portion of the facial features is obscured.

In this study, we present an LCGS descriptor over a 5×5 grid, which we use to stabilize recognition against masked faces, together with Maximum Spanning Tree (MST), forming the LCGS 5×5 + MST framework. It models the structural connections between visible facial regions and uses the relationships among facial regions with strong inter-region connections to form discriminative features for partially occluded faces. To position the method with respect to existing approaches, we compared it with eight classical baselines (PCA with Euclidean or cosine distance, LDA, SVM, LBPH, PCA + LDA, PCA + SVM, and PCA + LDA + SVM), with five pretrained deep learning models, and with alternative graph configurations (a coarser 3×3 grid, the 5×5 graph without spanning tree reduction, and a Minimum Spanning Tree variant).

The experimental evaluation of the proposed LCGS 5×5 + MST framework on the ENSA-MFRD dataset confirms its effectiveness. This was found to outperform all tested graph configurations, achieving a mean accuracy of 94.79%, and it stresses the importance of maintaining the strongest structural connections among facial regions rather than preserving all edges or only the weakest ones. Among the pretrained deep learning models, ArcFace achieved the highest accuracy, followed by FaceNet. The LCGS 5×5 + MST method outperformed the three general-purpose convolutional networks (ResNet50, VGG19, and InceptionV3) while running entirely on CPU, and it offers a form of interpretability absent in deep neural networks.

These results highlight, once more, the importance of graph topology, resolution, and dataset quality in masked face recognition. On ENSA-MFRD, most identity discrimination can be achieved using only the strongest structural links retained by the spanning tree. However, performance drops markedly on datasets acquired under uncontrolled conditions, such as RMFRD and LFW-masked, which indicates that the descriptor remains sensitive to variations in illumination, pose, and cropping.

Future work will investigate hybrid architectures combining the interpretability of the proposed graph-based framework with the representational power of deep neural networks. To further evaluate the method's robustness, practicality, and generalizability, it will also be tested on more challenging cross-dataset scenarios and under real-world acquisition conditions.

REFERENCES

1. P. Shende, and Y. Dandawate, *Convolutional neural network-based feature extraction using multimodal for high-security application*, Evol. Intel., 2021, 14, 1023–1033, doi: 10.1007/s12065-020-00522-5.
2. T.S. Akheel, V.U. Shree, and S.A. Mastani, *Stochastic gradient descent linear collaborative discriminant regression classification based face recognition*, Evol. Intel. 15, 1729–1743 (2022), doi: 10.1007/s12065-021-00585-y.
3. W.N.I. Al-Obaydy, Z.M. Fadhil, and B.H. Ali, *Patch-based pose invariant features for single sample face recognition*, Evol. Intel., 2022, 15, 585–591, doi: 10.1007/s12065-020-00531-4.
4. M. Turk and A. Pentland, *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991, doi: 10.1162/jocn.1991.3.1.71.
5. X. Tan, L. Deng, Y. Yang, et al., *Optimized regularized linear discriminant analysis for feature extraction in face recognition*, Evol. Intel., 2019, 12, 73–82, doi: 10.1007/s12065-018-0190-0.

6. F. Anowar, S. Sadaoui, and B. Selim, *Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)*, Computer Science Review, 2021, Volume 40, 100378, doi: 10.1016/j.cosrev.2021.100378.
7. X. Zheng, Z. Ma, and L. Li, *Local tangent space alignment based on Hilbert–Schmidt independence criterion regularization*, Pattern Anal Applic 23, 2020, 855–868, doi: 10.1007/s10044-019-00810-6.
8. F. Yang, W. Yang, R. Gao, and Q. Liao, *Discriminative multidimensional scaling for low-resolution face recognition*, IEEE Signal Processing Letters, vol. 25, no. 3, pp. 388–392, 2018, doi: 10.1109/LSP.2017.2746658.
9. M. Taskiran, N. Kahraman, and C.E. Erdem, *Face recognition: Past, present, and future (a review)*, Digital Signal Processing, 2020, 106, 102809, doi: 10.1016/j.dsp.2020.102809.
10. B.T. Chinimilli, T. Anjali, A. Kotturi, V.R. Kaipu, and J.V. Mandapati, *Face Recognition-Based Attendance System Using Haar Cascade and Local Binary Pattern Histogram Algorithm*, 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), Tirunelveli, India, 2020, pp. 701-704, doi: 10.1109/ICOEI48184.2020.9143046.
11. D. Biswas, S. Jayan, S. Nadipalli, S. B, and S. R, *Attendance Tracking with Face Recognition Through Hidden Markov Models*, 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1640-1645, doi: 10.1109/ICEARS53579.2022.9751888.
12. T. Sabharwal, R. Gupta, L.H. Son, R. Kumar, and S. Jha, *Recognition of surgically altered face images: An empirical analysis on recent advances*, Artificial Intelligence Review, 2019, 52(2), 1009–1040, doi: 10.1007/s10462-018-9660-0.
13. L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Malsburg, *Face recognition by elastic bunch graph matching*, In: Sommer, G., Daniilidis, K., Pauli, J. (eds) Computer Analysis of Images and Patterns. CAIP 1997. Lecture Notes in Computer Science, 1997, vol 1296. Springer, Berlin, Heidelberg, doi: 10.1007/3-540-63460-6_150.
14. H. Yang, Y. Fu, J. Zeng, and C. Xu, *Face Recognition Based on Log-Gabor Feature and Collaborative Representation*, 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 2017, pp. 423-427, doi: 10.1109/IHMSC.2017.207.
15. N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar, and K. Sharma, *Comparative study of machine learning and deep learning algorithm for face recognition*, Jordanian Journal of Computers and Information Technology, 2021, vol. 7, no. 3, pp. 313–325, doi: 10.5455/jjcit.71-1624859356.
16. S. Ahmam, N. Lamghari, and A. Ghazdali, *Masked Face Recognition: A Comprehensive Review of Techniques, and Datasets*, In International Conference on Mathematics Data Science (pp. 203-212), Cham: Springer Nature Switzerland, (2024, October), doi:10.1007/978-3-031-94802-2_20.
17. S. Ahmam, N. Lamghari, H. Khalfi, and A. Ourdou, *Masked Facial Recognition Using Deep Metric Learning*, In: Aboutabit, N., Lazaar, M., Hafidi, I. (eds) Advances in Machine Intelligence and Computer Science Applications. ICMICSA 2022. Lecture Notes in Networks and Systems, vol 656, 2023, Springer, Cham, doi: 10.1007/978-3-031-29313-9_17.
18. H.N. Vu, M.H. Nguyen, and C. Pham, *Masked face recognition with convolutional neural networks and local binary patterns*, Appl Intell 52, 2022, 5497–5512, doi: 10.1007/s10489-021-02728-1.
19. Z. Wang, B. Huang, G. Wang, P. Yi, K. Jiang, *Masked Face Recognition Dataset and Application*, IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 5, no. 2, pp. 298-304, April 2023, doi: 10.1109/TBIOM.2023.3242085.
20. D.E. King, *Dlib-ml: A Machine Learning Toolkit*, Journal of Machine Learning Research, vol. 10, pp. 1755–1758, 2009.
21. G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, Technical Report 07-49, University of Massachusetts, Amherst, 2007.
22. D. Yi, Z. Lei, S. Liao, and S. Z. Li, *Learning Face Representation from Scratch*, arXiv preprint arXiv:1411.7923, 2014.
23. A. Anwar and A. Raychowdhury, *Masked face recognition for secure authentication*, arXiv preprint arXiv:2008.11104, 2020.
24. Hmrishav, B., Labeling with LabelMe: Step-by-step Guide [Alternatives + Datasets]. Available on: https://www.v7labs.com/blog/labelme-guide, 2022.
25. P. Kamencay, M. Benco, T. Mizdos and R. Radil, *A new method for face recognition using convolutional neural network*, Advances in Electrical and Electronic Engineering, 2017, 15(4), 663-672, doi: 10.15598/aeee.v15i4.2389.
26. F. Z. Chelali, A. Djeradi, and R. Djeradi, *Linear discriminant analysis for face recognition*, 2009 International Conference on Multimedia Computing and Systems, Ouarzazate, Morocco, 2009, pp. 1-10, doi: 10.1109/MMCS.2009.5256630.
27. F.S. Samaria and A.C. Harter, *Parameterisation of a stochastic model for human face identification*, Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 1994, pp. 138–142, doi: 10.1109/ACV.1994.341300.
28. Y. Zhu, M. Ren, H. Jing, L. Dai, Z. Sun, and P. Li, *Joint holistic and masked face recognition*, IEEE Transactions on Information Forensics and Security, vol. 18, pp. 3388–3400, 2023.
29. Y.V. Lata, C.K.B. Tungathurthi, H.R.M. Rao, A. Govardhan and L.P. Reddy, *Facial recognition using eigenfaces by PCA*, International Journal of Recent Trends in Engineering, 1(1), 587, 2009.
30. M. Abdeldayem, W. Badawy, H. F. A. Hamed, and A. M. Nagy, *Improving facial expression recognition in real-world environments*, Statistics, Optimization & Information Computing, vol. 14, no. 6, pp. 3546–3564, 2025.
31. E.E.A. Abusham, and H.K. Bashir, *Face Recognition Using Local Graph Structure (LGS)*, In: Jacko, J.A. (eds) Human-Computer Interaction. Interaction Techniques and Environments. HCI 2011. Lecture Notes in Computer Science, 2011, vol. 6762. Springer, Berlin, Heidelberg, doi: 10.1007/978-3-642-21605-3_19.
32. M.F.A. Abdullah, M.S. Sayeed, K.S. Muthu, H.K. Bashier, A. Azman, and S.Z. Ibrahim, *Face recognition with symmetric local graph structure (SLGS)*, Expert Systems with Applications, 2014, 41(14), 6131-6137, doi: 10.1016/j.eswa.2014.04.006.
33. S. Malakar, W. Chiracharit, and K. Chamnongthai, *Masked face recognition with generated occluded part using image augmentation and CNN maintaining face identity*, IEEE Access, vol. 12, pp. 126356–126375, 2024.
34. S. Winarno, F. Alzami, D. A. Santoso, M. Naufal, H. Al Azies, R. M. Brilianto, and K. A. S. Muthu, *Efficient GRU-based facial expression recognition with adaptive loss selection*, Statistics, Optimization & Information Computing, vol. 14, no. 6, pp. 3468–3499, 2025.
35. J. Yang, L. Zhang, M. Li, T. Zhao, Y. Chen, J. Liu, and N. Liu, *Face Recognition with Facial Occlusion Based on Local Cycle Graph Structure Operator*, In Machine Learning and Biometrics, IntechOpen, 2018, doi: 10.5772/intechopen.78597.

36. M. Mahmoud, M. S. Kasem, and H.-S. Kang, *A comprehensive survey of masked faces: Recognition, detection, and unmasking*, arXiv preprint arXiv:2405.05900, 2024.
37. W. Hariri, *Efficient masked face recognition method during the COVID-19 pandemic*, SIViP 16, 2022, 605–612, doi: 10.1007/s11760-021-02050-w.
38. L. Jaya Akeh and G. Putra Kusuma, *Mixed emotion recognition through facial expression using transformer-based model*, Statistics, Optimization & Information Computing, vol. 13, no. 2, pp. 531–546, 2024.
39. S.Q.N. Septi, I.N. Yulita, and H. Napitupulu, *Face Recognition Using Fisherface and Support Vector Machine Method*, 2021 International Conference on Artificial Intelligence and Big Data Analytics, Bandung, Indonesia, 2021, pp. 50-55, doi: 10.1109/ICAIBDA53487.2021.9689738.