Classifiers Combination for Efficient Masked Face Recognition

Kebir Marwa, Ouni Kais
Research Laboratory Smart Electricity & ICT, SE&ICT Lab., LR18ES44
National School of Engineers of Carthage
University of Carthage, Tunisia

***Abstract*—This study was developed following the upheaval caused by the spread of the Coronavirus around the world. This global crisis greatly affects security systems based on facial recognition given the obligation to wear a mask. This latter, camouflages the entire lower part of the face, which is therefore a great source of information for the recognition operation. In this article, we have implemented three different pre-trained feature extractor models. These models have been improved by implementing the well-known Support Vector Machines (SVM) to reinforce the classification task. Among the investigated architectures, the FaceNet feature extraction model shows remarkable results on both databases with a recognition rate equal to 90% on RMFD and a little lower on SMFD with 88.57%. Following these simulations, we have proposed a combination of classifiers (SVM-KNN) that would prove a remarkable improvement and a significant increase in the accuracy rate of the selected model with almost 4%.**

Keywords*—*Masked faces; deep learning; AlexNet; ResNet50; FaceNet; classifiers combination

I. Introduction

According to the World Health Organization (WHO) [1], there have been 517.648.631 confirmed cases of COVID-19, including 6.261.708 deaths, until May 13th, 2022. That is considered an astonishing number after three years of the virus's appearance and despite all precautions taken. Therefore, the Centers for Disease Control and Prevention (CDC) [2] emphasize social distancing and obligation to wear masks in order to minimize contamination and reduce the hazardousness of this virus. Except that, wearing a mask causes the performance degradation of security systems based on the identification of people by their faces, since the mask hides a large part of the face, hence the loss of a large amount of information. Thus, existent techniques for faces recognition implemented before this crisis, must have an improvement and some adjustment to live up to the expectations of users of face-based security systems, as stated by the National Institute Of Standards and Technology (NIST) [3]. The performance of facial recognition algorithms submitted before March 2020, when the World Health Organization declared a global pandemic, was examined in a previous NIST report published in July. The error rate of these pre-pandemic algorithms was found to be between 5% and 50% in this first investigation, which confirms that these systems have become ineffective.

The degradation of identification systems, causes several problems. So, frauds and wanted persons take advantage of the mask, for illegal immigration and crimes commitment without being recognized. Community access control and face identification have become almost an impossible mission when a grand portion of the face is covered up by a mask. Due to these issues, face masks have essentially challenged existing face recognition strategies.

The epidemic situation, ensures the emergence of two new axes of research: [4]

- Face mask detection: consists of checking whether the individual is wearing a mask or not, and it is an interesting task in public squares and areas with fulls, where wearing a mask is mandatory.
- Masked face recognition (MFR): is used to identify people wearing masks on the basis of the remaining part of the face (the eyes and the forehead parts)

Our interest in this paper is the second axis. we have implemented a recent technique to identify masked faces using a deep learning-based method for extracting features. To train our model, two databases are used: Simulated Masked Faces Recognition database(SMFRD) and Real-World Masked Faces Database(RWMFD) presented in [5], specially designed to evaluate the performance of masked faces recognition methods.

An interesting preliminary phase in the recognition operation is the pre-processing of the images. However, the quality of the detection influences the accuracy of the identification, so we chose the MTCNN algorithm to have an exact and correct detection. Regarding the feature extraction task, we have opted for three pretrained models that are very recognized in the field of facial recognition, which are AlexNet, ResNet50 and FaceNet. For the classification process we have chosen a classifier which has proven a huge success in image classification, it is the Support Vector Machine(SVM). Finally, we have proposed, a classifiers combination (SVM-KNN) to enhance the performance of the classification process.

This study is organized as follows: The Section 2 provides the related works about masked faces recognition. While Section 3 highlights the motivation and the contribution of the paper. Section 4 discusses the state-of-the-art methods. Section 5 presents the used method. The remainder of the paper stated the experiments and the concluding statements based on experimental results.

II. Related Works

Obviously, the issue of partially hidden faces has existed for a long time, since there are several factors other than wearing a mask such as a beard or mustache, sunglasses or even makeup that changes the original features for masquerade reasons. But recently, the obligation to wear a mask makes the problem even worse and the dilemma of recognizing occluded faces has become at the head of research in computer vision. Consequently, significant increase in MFR research effort, extending existing MFR methods and yielding promising accuracy results.Then, search across major digital libraries to track growing research interest in the mission of occluded face recognition (OFR). A series of search strings are formulated to find leading repositories the article covers the use of deep learning techniques only in the context of face-based recognition. MFR article search results retrieved from Web of Science, Scopus, IEEE Xplore, Wiley, Ei Compendex, ACM Digital Libraries and EBSCOhost. These warehouses include articles on recent popular seminars, journals and conferences articles from five years.

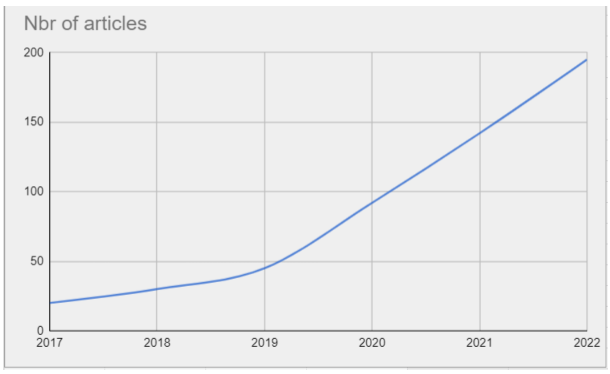

Fig. 1. Diagram of Research Efforts on MFR from 2017 to 2022.

As shown by Fig. 1, the importance of research in the field of Occluded and Masked Faces recognition (OFR and MFR) witnesses a noticeable increase in parallel to exhausting researches [6] [7] [8].

Except that, this diversity does not mean the effectiveness of the applied techniques, because until now there is no masked face recognition method whose performance exceeds or equals that of unmasked face identification techniques. As a result, the achievement of the methods used post-pandemic remain unsatisfactory for real-time systems and high-traffic sites with high security requirements. As of late, researches have proposed many useful methods. They basically consist of three categories:[9]

- Generate a typical model of the occlusion issue(restoration model),
- Occlusion removal approach,
- Deep learning-based approaches.

,

A. Restoration Model

This approach consists of generating the hidden part of the face considering that the nose, mouth and chin carry a large amount of information. For the improvement of the recognition efficiency reasons and to generate the lost features of face, there are two restoration models: robust structured error coding and robust subspace regression [9].

- Robust structured error: The occlusion produced by the mask presents a spatial continuity. By this, an error caused has a specific spatial structure. This makes the reconstruction of the low-rank structure of the face image from the data damaged by occlusion necessary, to have a correct identification and minimize the rate of false positives and true negatives. For instances, authors in the literature [10] presented an improved robust principal component analysis (RPCA) method. At the beginning, the method consists in decomposing the learning matrix M by a lower rank matrix, which ensures obtaining a lower rank content matrix L and a sparse content matrix S. In this way, the recovery of the lower subspaces of the training sample is achieved. This method considers how to restore low-ranked structures from training samples that are error-prone but sparsely structured. This effectively suppresses the effects of sparse noise and provides strong efficiency. In order to increase interclass information between low-ranked matrices of different face categories, reference [9] has expressed all training patterns as observation matrix D. After decomposing the matrix D, we get a low-ranked matrix A with no occlusion and a sparse error matrix E. The RPCA is applied to the submatrix A and the resulting subspace is used as an occlusion dictionary for facial images. The image reconstruction was then identified and the error size was classified according to the sparse representation classification and occlusion dictionary.[9]

- Robust subspace regression: This model is generated by projecting high-dimensional feature data from different categories of facial images onto a low-dimensional subspace. Next, an independent subspace is set in the occlusion part, and the occlusion of the face image is expressed using the existing dictionary atom to realize a powerful recognition effect of the occlusion face. Currently, robust subspace solutions for occluded face detection primarily include sparse representation, collaborative representation, and obstruction dictionary learning.

B. Occlusion Removal Approach

This model aims to estimate the position of occlusion through two error indices. The first is in the form of local similarity error between the original image and the partially occluded image, while the second aims to the spatial local error caused by the occlusion. Otherwise, the method consists in locating the hidden areas of the face and eliminating them completely from the feature extraction and classification process. In this context, one of the most famous approaches is that based on segmentation. According to the literature [11], authors segmented the face in small local zones. These latter, contain the occluded part to be eliminated and which will be detected by the support vector machine (SVM). Then the last phase is to use a mean-based weight matrix for face identification.

Several techniques have been proposed to remove the concealed part of the face, including the exemplar based Image in-painting technique proposed in [12]. As well as, the structural similarity index measure and principal component analysis technique [13].

C. Deep Learning-Based Approaches

Currently, Deep learning has proven huge success in several areas, especially in FR. This comes down to the efficiency of deep features compared to others that are over shallow. The most recognized research works in occlusion face are based on this type of features, taking the example in [14], an efficient partial face recognition approach was proposed, this is Dynamic Feature Matching (DFM) approach, based on the combination of the Fully Convolutional Network (FCN) with Sparse Representation Classification (SRC). The method is intended to recognize partial faces of arbitrary size. Another model proposed to identify hidden faces called BoostGAN model [15]. The main idea of this model is to use the occluded face to elaborate the non-occluded face and this latter will then be used to recognize the person. Except that, in the case of large occluding surfaces such as face masks, GAN-based methods are hard to regenerate the details of the key points on the visage.

Table I presents the performance of some MFR methods.

TABLE I. SUMMARY OF SOME MFR METHODS PERFORMANCE

References	Model	Dataset	Accuracy
GuiLing, W.[9]	Mask separation	RMFRD	95.22
Priya, G.[11]	MBWM-SVM	GTAV	94.75
He,L.[14]	Dynamic Feature Matching	CASIA-NIR-Distance	94.96

III. MOTIVATION AND CONTRIBUTION OF THE PAPER

The failure of current methods to correctly identify masked faces in the same way as non-masked face recognition techniques motivates us to explore a new solution to overcome this shortcoming. Drawing on the significant performance and strong light resistance of CNN-based methods, facial expression variations and face occlusion. In this study, we have proposed an occlusion removal approach with transfer learning model to solve the masked face problem noticing during the COVID-19 pandemic.

To the best of our knowledge, the restoration model approach isn't truly a great selection. This procedure endures from a few issues, particularly the difficulty of implementing it and culminating the comes about. As well as, it could be a prepare that devours an expansive execution time. With that, several researchers use this method as is the case in the literature [16], contrary of our choice. The disadvantages of this approach, inspired us to base our considerations on the occlusion removal approach since it is a compelling and simple strategy. As well as, pre-trained model for feature extraction.

Other side, face detection has a huge influence on the quality of recognition. As of late, a few face tracking strategies have shown up and have demonstrated tall execution, such as the viola and Jones detector, which has gotten to be a reference in object detection and particularly for face. A bit more recently, it appears the famous MTCNN detector [17] which has proven great efficiency and speed of execution. But that the major progression of machine learning recently, gives us with unused choices, like the modern finder which made a boom in computer vision is the mediapipe algorithm [18]. MediaPipe face detector is an ultra-fast solution with six landmarks and multi-face support. It is an excellent detector for streaming videos, but it is not yet used for static images. What motives us to use the MTCNN detector.

So the contribution of this paper lies in three points:

The first consists in adapting models usually used for the recognition of unmasked faces for masked faces. This adaptation was made by training the three models AlexNet, ResNet50 and FaceNet by databases specially designed for masked faces and adjusting the parameters of each model in order to improve their performances.

The second contribution aims to improve classification task with the inserting of SVM classifier for each model.

While the third contribution consists in a classifiers combination (SVM-KNN) that considerably improves the obtained results.

IV. STATE OF-THE-ART METHODS

The ordinary and common structure for all face recognition methods is composed of a face detection, a feature extraction and a classifier. Through this paper, we have presented the races that we followed, starting with face cropping.

A. Face Detector: MTCNN

Currently, MTCNN or Multi-task convolutional neural networks, is the most popular and rigorous face detection solution. It is composed of three cascaded neural networks known by P-Net, R-Net and O-Net [17].

- **P-Net** stands for **Proposed network**: It searches for faces in frames of size 12 × 12. The task of this network is to achieve rapid results.
- **R-Net** presents **Refined network**: Its structure is deeper than Pnet. Despite all candidates originating from the previous network will be fed to R-Net, a large number of candidates are already eliminated by the first network P-Net.
- **O-Net** comes from **Output network**: briefly returns the bounding box (face area) and face landmark locations.

These three cascaded neural networks are presented in Fig. 2.

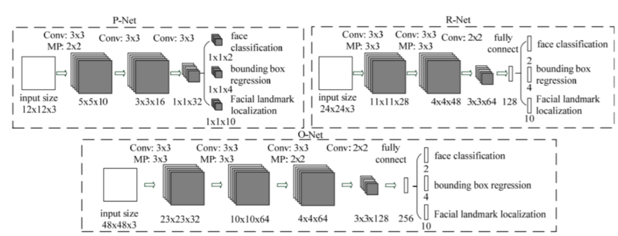

Fig. 2. Architecture of the Three Cascaded Neural Networks of MTCNN [17]

The MTCNN model detects five landmarks on the face, which are the left eye, right eye, nose, and two corners of the mouth. This model proves high face detection accuracy.

B. Masked Faces Databases

In this section, we have presented several benchmark datasets used in the literature to evaluate MFR techniques [19].

Starting with the most popular and used databases in this field which are the RWMFRD (Real-word masked faces recognition dataset), the Masked Face Detection Dataset (MFDD) and Simulated Masked Face Recognition Dataset (SMFRD) introduced in the same article [5].

As regards **MFDD**, it contains 24,771 masked face images, which allows the implemented model to accurately detect faces hidden by masks.

As for **RMFRD**, the largest existent database for MFR, since it contains 5,000 images of 525 people wearing masks, and 90,000 images of the same people without masks. This database was used in [20], in the context of face images that were not acceptable due to an incorrect match were manually removed. In addition, the right face areas have been cropped using semi-automatic annotation techniques, such as LabelImg and LabelMe, Fig.3 displays sample images from RMFRD.

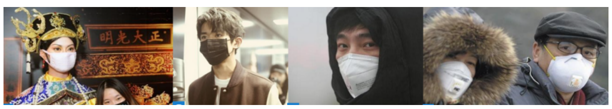

Fig. 3. Sample Images from RMFRD.

For diversification reasons, **SMFRD** has been developed, it contains 500,000 images of synthetically masked faces of 10,000 people collected from the Internet and Fig. 4 illustrate some images from this dataset.

Fig. 4. Sample Images from SMFRD.

The Synthetic face-occluded dataset **(SFOD)** [21] was elaborated using published data records from CelebA and CelebA-HQ [22]. CelebA-HQ is a large-scale facial attribute dataset containing over 30,000 celebrities. Each face image is cropped and roughly aligned based on the eye position. The occlusions were aggregated by five common non-facial objects: hands, masks, sunglasses, glasses, and microphones. Over 40 different types of objects were used in different sizes, shapes, colors and textures. In addition, non-face objects were randomly placed on the face.

The Masked Face Segmentation and Recognition dataset **(MFSRD)** [23] is composed of two parts. The first part consists of 9742 images of masked faces that have been collected from Internet with hand-labeled masked segmentation annotation. The second part contains 11,615 images of 1004 identities, of which 704 are collected from the real world and the rest of the images are collected from the Internet, where each identity has at least images with and without masks. Celebrities in Frontal Profile in the Wild **(CFP)** [24] includes the faces of 500 celebrities in face and profile view. Two verification protocols with 7000 comparisons, each presented: one compares only frontal faces (FF) and the other compares (FF) and profile faces(FP).

Both, Masked Face Verification **(MFV)** and Masked Face identification **(MFI)** are presented in [25], the first one contains 400 pairs for 200 identities while the second includes 4916 images of 669 identities.

The **LFW-SM** [26] variant database contains a simulated mask that extends the LFW dataset and contains 13,233 images from 5749 individuals. Through Fig. 5, we have presented sample images from LFW-SM.

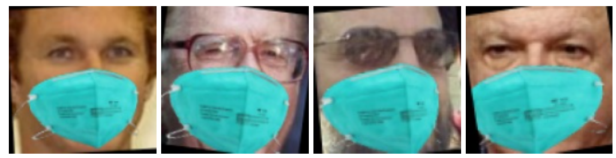

Fig. 5. Sample Images from LFW-SM.

Several MFR techniques used the VGGFace2 [27] dataset for training, which consists of 3 million images of 9131 people with over 362 images per person. From this database derives the Masked faces dataset **VGGFace2-m** [28], it contains over 3.3 million images of 9131 identities. Table II shows the main characteristics of the dataset used in the masked face recognition task.

TABLE II. SUMMARY OF THE MFR BENCHMARKING DATASETS

Database	Size	Identities	Type of masks
RMFRD	95,000	525	Real-world
SMFRD	500,000	10,000	Synthetic
MFSRD	11,615	1004	Real-world/synthetic
MFV	400	200	Synthetic
MFI	4916	669	Synthetic
LFW-SM	13,233	5749	Synthetic
VGG-Face2-m	3.3M	9131	Synthetic

C. Feature Extraction

One of the most crucial steps in the facial recognition process is feature extraction. It consists of extricating a set of features that are discriminative enough to represent and learn key facial attributes such as eyes, mouth, nose, and texture. In the presence of partial occlusion, especially that produced by the face mask, this process becomes more complex and current facial recognition systems need to be adjusted to extract representative and robust facial features. There are two categories of feature extraction, the first is a **shallow feature extraction** which is a classical technique explicitly forming a set of features fabricated with low optimization or learning mechanisms. The most popular methods of this category are Histogram of Oriented gradient (HOG), LBPs and codebooks [29]. In recognition tasks for unmasking face, these algorithms achieve considerable accuracy and remarkable robustness against a variety of facial changes such as lighting, rotation, scaling and translation. But this is not the case for masked faces, great degradation was observed.

The second category is **the deep feature extraction**. One of the most efficient neural networks in the field of face recognition is the Convolutions Neural Network (CNN), it has shown preponderance in a wide range of applications, such image classification, retrieval and detection objects. CNN-based models have been widely deployed and trained on many large-scale face datasets [30] [31] [32].

Several pre-trained architectures are recognized in the field of FR and have proven remarkable success, especially for feature extraction. To choose the most suitable extractor for our application, we have made an overview of the most cited models in the literature. AlexNet [33]is a famous model that ensures to reduce the training time and minimize the errors even on large datasets [34]. Fig. 6 shows the architecture of AlexNet.

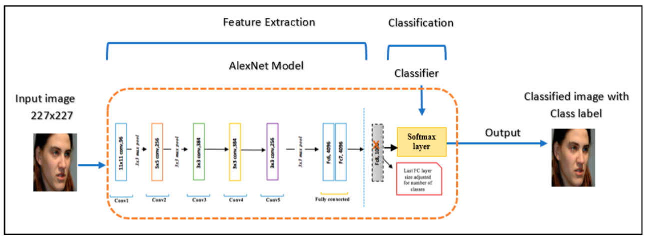

Fig. 6. Architecture of AlexNet [35].

Two other popular CNN-based models were presented in [36], VGG16 and VGG19 have been used in various computer vision applications, especially facial recognition. Despite achieving considerable accuracy, they endure from training time and complexity [37].

For the most complex identification missions, as in the case of masked faces, it is preferable to be processed by deeper neural networks like residual network (ResNet) [38]. This model achieves outstanding performance and accuracy, due to the stack of additional layers. These extra layers must be determined empirically to control for any deterioration in the performance of the model. The architecture of ResNet50 is presented in Fig. 7.

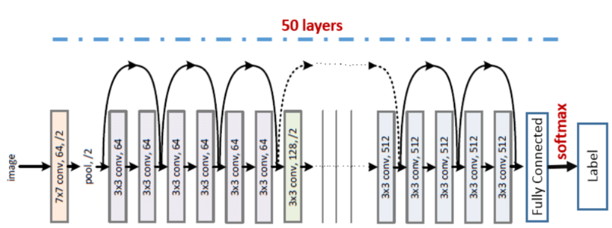

Fig. 7. Architecture of ResNet50 [39].

MobileNet [40] also is considerate as one of earliest deep neural networks, which mainly depends on a simple architecture. Its architecture exhibits high performance with hyperparameters and fast model calculations [41].

We cannot pass without mentioning Inception [42] and its variations [43], their innovation is that they use modules or blocks to build networks that contain folding layers instead of stacking them. Xception [44] is an extreme Inception version that replaces Inception modules with deeply separable convolutions.

FaceNet [45], presented by Google researches, it is a famous pre-trained model that has proven very remarkable results. Fig. 8 shows the architecture of FaceNet.

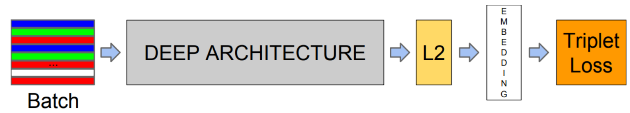

Fig. 8. Architecture of FaceNet [45].

Through Table III we have presented a summary of the pre-trained CNN-based model.

TABLE III. SUMMARY OF THE PRE-TRAINED CNN BASED MODEL

Model	Variants	Trainable param	Conv layers	Total layers
AlexNet	-	62 M	5	8
VGG	VGG16	138 M	13	16
	VGG19	143 M	16	19
ResNet	ResNet50	25 M	48	50
	ResNet101	44 M	99	101
MobileNet	MobileNet	13 M	28	30
	MobileNet-v2	3.5 M	-	53
Inception	GoogleNet	7 M	22	27
	IncepV2	56 M	22	48
	IncepV3	24 M	22	48
	IncepV4	43 M	-	164
	Incep-ResNet-V2	56 M	-	164
Xception	-	23 M	36	71
FaceNet	-	140M	22	27

D. *Classification*

Many classifiers have been mentioned through literature, given the importance of classification in improving the performance of facial recognition systems.

As far as we are aware, the two most popular classifiers in facial recognition are Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN) [46]

1) SVM: Support vector machine (SVM) is a supervised machine learning that can be used for classifications or regression problems. For the issues of multiclass classification it exists two distinguished approaches, the first is one-against-one and the second is one-against-all approach. Kernels functions are used for separation between classes for higher dimensional feature spaces. These Kernel functions are able to transform a non-linear distinguishable problem into a linear distinguishable one and projecting data into the feature space which ensure to find the optimal separating hyper plane [51].

2) KNN: K-Nearest Neighbor, a popular technique for classifying objects based on nearest training samples in feature space. This training samples are vectors with a class label for each one. The principle of the technique aims to compute the distance of the test sample to every training sample and keeping the k closest training samples (Where k designate positive integer). Then several distance functions used in the KNN algorithm, but the best methods are Euclidean distance [52].

3) SVM-KNN: Overall, the combination of classifiers is a relatively new technique. It can be considered as an optimization problem for minimizing classification errors and takes as input the outputs of M classifiers and generates the final N classes [53] Classifier combination is more efficient, especially when the classifiers are different. There are two types of combination:

Features association using similar classifiers and decision association resulted from dissimilar classifiers [54]. The second type of combination is our choice since SVM and k-NN are two dissimilar classifiers.

On the other hand, classifiers can provide three types of outputs: The measure, class and rang types [55]. Depending on these types of output the combination may be:

- In the abstract stage built on voting methods.
- In the rank stage when the outputs are labels classified by a reducing weight.
- In the measurement level if the outputs are labels combined with confidence values.

For our work, we have used the majority vote in combination.

E. Evaluation Metrics

In order to assess the cogency of implemented models for masked face recognition, we have opted for some evaluation parameters, such precision, recall, F1-score and accuracy metrics. These evaluation metrics were calculated as follow [49]:

$$Accuracy = \frac{T_N + T_P}{T_S} \tag{1}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{2}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} + score \tag{4}$$

Where T_P, T_S, F_P, T_N and F_N designate respectively True positive, total samples, false positive, true negative, and false negative. For our work we have used Accuracy, precision and recall rate.

V. USED METHOD

Our method consists of three primary scenarios (presented in Fig. 9), where we used three different pre-trained CNN-based models like feature extractors and the SVM is cascaded for classification. The model that has proven the best results will be improved by the combination of a second classifier which will be the K-NN.

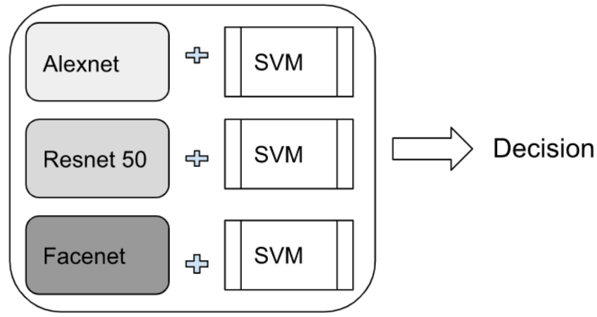

Fig. 9. Architecture of our Primary Scenarios.

A. AlexNet and SVM

We have used the well-known SMFD and RMFD databases specially designed to evaluate masked facial recognition systems. Data from each base were divided into 80% for training and 20% for testing and random distribution to avert biased results. Then, the images used are a mix of masked and unmasked faces which ensures good feedback.

Before we start, we need to preprocess the input images to fit the AlexNet model. Input image should be with a dimension equal to 227 * 277 * 3 pixels. For SMFD database, the image contains only the parts of the face we are interested in, so we don't need face detector, we just need to convert the size of the images from 128 * 128 * 3 to 227 * 227 * 3. Unlike the RMFD base, we use MTCNN to frame the faces. It is important to note that the model is made up of many layers, but not all of these layers are essential for feature extraction. For instance, the first layer deals with the extraction of features such as edges and points. As it is presented in Fig. 10.

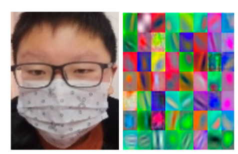

Fig. 10. Extracted Features by the First Layer of AlexNet from Original Image on RWMFD.

The table in Fig. 11 shows the different layers of the ALexnet model.

Layer	Number of Kernels	Kernel Size	Stride	Padding	Output Size
Input					[227 × 227 × 3]
Conv1	96	11 × 11 × 3	4	-	[55 × 55 × 96]
Max pool1		3 × 3	2	-	[27 × 27 × 96]
Norm1					[27 × 27 × 96]
Conv2	256	5 × 5 × 48	1	2	[27 × 27 × 256]
Maxpool2		3 × 3	2	-	[13 × 13 × 256]
Norm 2					[13 × 13 × 256]
Conv3	384	3 × 3 × 256	1	1	[13 × 13 × 384]
Conv4	384	3 × 3 × 192	1	1	[13 × 13 × 384]
Conv5	256	3 × 3 × 192	1	1	[13 × 13 × 256]
Max pool3		3 × 3	2	-	[6 × 6 × 256]
fc6 ReLU Dropout(0.5)	1				4096
fc 7 ReLU Dropout(0.5)	1				4096
fc8 softmax	1				1000

Fig. 11. Details of AlexNet Layers [47].

After features extraction, we have equipped the SVM to perform the classification task. The output from 'fc8' layer was a 4096-dimensional feature vector. For the SVM kernel function, we used a linear kernel function without optimization. Kernel functions are used to take vector data as input and convert it into the optimal format. We were inspired by this network implemented mainly for unmasked faces as displayed by Fig. 12.

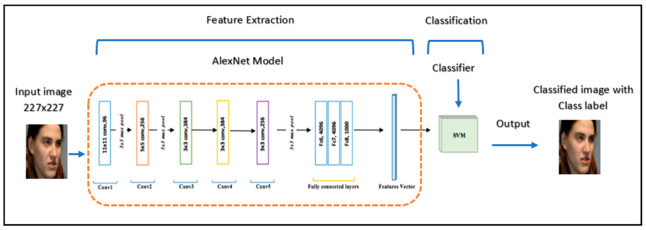

Fig. 12. AlexNet Convolutional Neural Networks with SVM.[35]

B. ResNet50 and SVM

Firsthand, Through the implementation operation, we have created an image data store that is useful for data management. Seeing that, the image was read and then loaded into the storage system. Then, the data was split into 80% training and 20% validation using random sampling for averting bias the outcomes. On the other hand, the ResNet50 network can only process images with 224 × 224 × 3, that's why we resized images to be used in this size. Regarding the next phase, which is feature extraction, there is a final layer named fc1000 found just before the classification layer was used to extract features using the activation method. The activation outputs were aligned in columns to speed up afterwards the SVM training , and the optimizer Adam was used for the training instead of Stochastic Gradient Descent. Fig. 13 presents the ResNet50 model implemented with SVM.

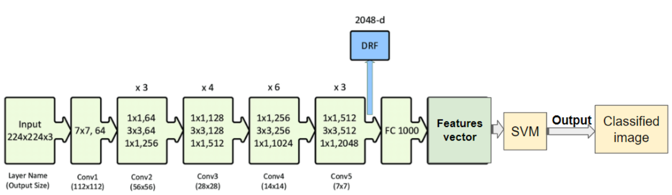

Fig. 13. ResNet50 Convolutional Neural Networks with SVM.

We have adjusted manually different hyper parameter to obtain better outcomes. Starting with the batch size, it is tuned to 32 and we run our algorithm with cross-entropy for 50 epochs.

C. FaceNet and SVM

FaceNet model accomplished state-of-the-art results in several benchmark face recognition datasets, specially Labeled Faces in the Wild (LFW) and YouTube Face Database. The model requires as input images sizes 160×160 and it contains 22 deep layers and 5 pooling layers, and a global average pooling is used at the end of the last inception module. The Fully connected layer will be used for face description. Elaborated descriptors become an embedding module for correspondence descriptors. The max operator has been applied to features to develop a one feature vector from a template. The network must be properly tuned to expect a significant boost for the particular task of face recognition and verification. To retrain the FaceNet model, we have to bring a set of masked and unmasked faces[42].

The module includes four branches, the first contains a series of 1x1 local features from the input for learning. While The second branch implements 1 x 1 convolution in order to reduce the input dimensions until 1 x 1 convolution is achieved. This greatly minimises the quantity calculation the network desires. Third branch is coherent with the second branch with 5x5 learning filters. The last branch accomplishes 3x3 max pooling with a Stride of 1x1. Finally, all branches of the Inception module converge and Channel dimensions are linked to each other before being added to the next network, as displayed by Fig. 14 [48].

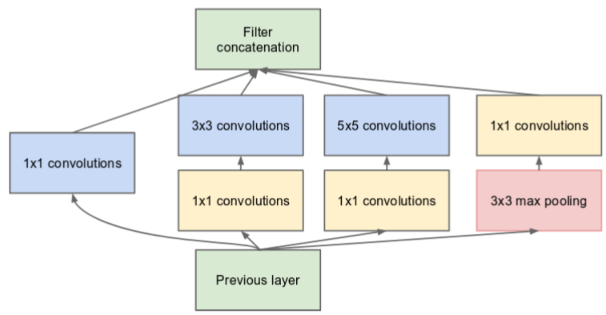

Fig. 14. Illustration of the Inception Module used for FaceNet Pre-Trained Model.

A crucial and next task in the FaceNet model is Face embedding, it consists of the representation of facial feature in the form of a vector. This latter, will be useful for comparison with the other generated vectors for identification of people. Embedding vector will be stored in order to be utilized as an input for the classifier. For this reason, we have to itemize each face in both training and testing database to have the classifier perform embedding and name prediction. It's necessary that the pixels of the image are normalized to perform the prediction operation. FaceNet architecture contains a batch layer and a deep CNN network. This latter, was supported by the normalization L2 which results face embedding. The face embedding is performed Triplet loss during the training. The triple loss has a minimum distance between an anchor and positive when the identities are the same [49]. Fig. 15 displays the triplet loss training.

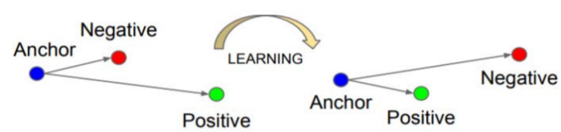

Fig. 15. The Triplet Loss Training[45]

Then, the validation process is integrated to recognize a candidate's face by performing a classification task within an integrated support vector machine (SVM). Since its inception, the SVM algorithm has been effectively applied to various classification-related problems . The SVM finds an Hyperplane that performs the classification task of the optimization issues. This maximizes the boundaries between the two classes of a particular input and target pair. The Classification is the result of certain robustness against over-fitting and margins represent class separation efficiency.

VI. Experiments and Results

A. Implementation

Through this section, we have presented the experimental results acquired by face recognition using the three deep convolutional neural networks previously mentioned, after they are chained by the famous SVM classifiers, these implementations are based on both MFRD and SMFD databases. Three major experiments in our study were performed to compare performance differences between pre-trained CNN architectures. First, we evaluated the performance when extracting the learned image features from a pre-trained CNN AlexNet, followed by SVM as a classifier. Second, we have realised the same experience with ResNet50. Third, we have evaluated the performance of FaceNet model with SVM. The analysis and evaluation were carried out on the basis of the performance recognition accuracy, precision and recall rate.

Before beginning the training process for the convolutional neural network architectures, a previous pre-processing is required. For all datasets, a rescale is applied to resize the images to a 227×227 as input for AlexNet, 224×224 as input for ResNet50 and 160×160 for FaceNet model.

All experiments were conducted using the platform of Windows with the configuration of AMD Ryzen5-GPU with 16 GB of NVIDIA GEFORCE RTX 3050 TI. Python tool was used to evaluate the method and perform the feature selection and classification task.

Table IV and Table V display the results obtained from the pre-trained models using two different databases.

TABLE IV. Evaluation of Implemented Models using RMFD

Model	Accuracy%	Precision%	Recall%
AlexNet	88.89	90.00	90.00
ResNet50	84.20	83.650	85.310
FaceNet	90	90	91.5

TABLE V. Evaluation of Implemented Models using SMFD

Model	Accuracy%	Precision%	Recall %
AlexNet	85.210	88.12	88.12
ResNet50	83.304	81.870	84.10
FaceNet	88.57	88.5	88.6

Since FaceNet has demonstrated considerable robustness with masked faces, we have opted this model to improve this performance by combining the used SVM classifier with the well-known K-NN classifier. The results obtained are presented in Table VI.

TABLE VI. FaceNet Model with Classifiers Combination

Model	RMFRD	SMFRD
FaceNet with SVM	90%	88.57%
FaceNet with SVM-KNN	94.46%	91.87%

B. Discussion

As a first observation, the two databases are not simulated in the same way, RMFD provides more considerable results than the SMFD database for AlexNet, ResNet50 and FaceNet. The difference of results between the two databases comes down to the fact that in the SMFD database, mostly, masks are not really well placed on the face, since the masks are synthetic and are not real-world as in the case of the database RMFD database.

Regarding the models performance, FaceNet performed better on both databases. Although, AlexNet and ResNet50 models show significant results in unmasked face recognition, they present a remarkable degradation with masked faces. Even with the adjustment of some parameters for each architecture, as well as with the training of the models by a large number of masked faces.

According the authors in [35], AlexNet model reached higher accuracy of 100% on YTF datasets, 99.55% and 99.17% for the GTAV face and ORL datasets respectively. While ResNet50 has achieved high accuracy of 100% on GTAV face and YTF datasets. Similarly, FaceNet model presents a degradation compared to the results provided with the unmasked faces, let us quote the example of literature [50] where authors mentionned that FaceNet is highly efficient in non-masked faces recognition that it can reach 100% accuracy on YALE, JAFFE, AT&T datasets. Although the masked faces influenced the performance of the FaceNet model, it remains robust relatively to other models.

We have improved the classification process by a combination of voting-based classifiers which greatly improves the performance of the model, where we have reached an accuracy rate equals to 94.46% with RMFRD. This combination has been used in several works for different objects recognition other than masked faces recognition. Let's cite the example of the article [56], where the authors obtained an accuracy rate equals to 97.11% for Arabic-word handwriting recognition. In literature [57], classifiers combination for recognition of Arabic literal provides an accuracy rate equals to 98%.

It is obvious that these values exceed ours, this excess in values comes down to the fact that the models are less complex, the stains are easier and even the databases are much smaller.

To further appraise the proposed method, we have compared our model with several techniques destined specially for recognizing masked faces, such as the model implemented in [58] that just combined FaceNet with SVM and they got an accuracy rate equals to 91.304%. In the study presented in [59], authors used VGG16, the Multilayer Perceptron Classifier (MLP) and Bag-of-Features (BoF) paradigm, the accuracy rate obtained is equals to 91.3% . An efficient face recognition method presented in [35] using Transfer learning (ResNet50 and AlexNet) to fine-tune pre-trained models to the masked face detection dilemma using an SVM classifier, authors have obtained an accuracy equal to 87%. Table VII summarizes this comparative study.

TABLE VII. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS

Model	Dataset	Accuracy
FaceNet+Svm	RWMFD	91.304%
VGG16+BOF	RWMFD	91.3%
CNN+Svm	RWMFD	87%
Our model	RWMFD	94.46%

These methods achieve lower values than obtained by the proposed technique, which confirms the effectiveness of the classifiers combination. Then, the complementarity between the two classifiers increases the recognition accuracy and robustness of the model, this comes down to the fact that in image classification, the concept of ambiguity is particularly related to the presence of noisy pixels, mixed pixels and pixels from regions that have undergone changes. If some pixels are between two classes (such as those located on the boundaries of homogeneous regions), those pixels should be classified into a union of two classes rather than a single class. This case can be important when the spatial resolution of the sensor is high. Mixed pixels intervene in the image modeling of a single source whenever that source cannot distinguish between the two classes. In this case, only class related information is available. Pixels in areas with little change are difficult to distinguish from pixels in stable areas and must therefore using two classifiers to get more accurate results.

VII. CONCLUSION

Today, the task of recognizing a masked face is a challenging process which makes it a focus of interest for scientific committees, given the importance of facial recognition for the security of various organizations and applications around the world. Models that intended for the recognition of unmasked faces have become helpless and unable to provide satisfactory performance to the expectations of security systems and real-time applications. Over the last two years, several techniques have emerged for this purpose but these techniques always remain less effective than the models destined of unmasked faces. This fact comes to several factors. First, the loss of the majority of facial details. Second, all the proposed techniques are based on models previously intended for unmasked faces by adjusting a few parameters to render the model adaptable with the new task. Third factor, the databases designed for the study of masked faces recognition systems, require more improvement. In this regard, we have proposed a model based on FaceNet with combination of classifiers (SVM-KNN) in order to have satisfactory results. This combination gives better results compared to the classification by a single classifier given the complementarity between SVM and K-NN classifiers. Finally, safety is vital at all levels (social, industrial, services, etc.) and security systems must reach a certain level of robustness, that's why our future work aims to develop a new model which does not consider masked faces as a barrier to having excellent results.

REFERENCES

[1] Organization WH, et al,*"Advice on the use of masks in the context of Covid-19, interim guidance"*, Tech. rep., World Health Organization, 2022.

[2] Centers For Disease Control (CDC),*"Your health"*, 2022.

[3] Ngan, M. L. and Grother, P. J. and Hanaoka, K. K,*"Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms*, Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, USA, 2020.

[4] Hariri. W,*"Efficient masked face recognition method during the COVID-19 pandemic"*,Signal Image Video Process, Vol.16(3), pp.605-612, 2022, 10.1007/s11760-021-02050-w.

[5] Wang, Zhongyuan, Wang, Guangcheng, Huang, Baojin, Xiong, Zhangyang and Hong, Qi and Wu, Hao. et al, *"Masked Face Recognition Dataset and Application"*, arXiv, 2020.

[6] Ligang Zhang, Brijesh K. Verma, Dian Tjondronegoro and Vinod Chandran, *"Facial Expression Analysis under Partial Occlusion"*, ACM Computing Surveys (CSUR), Vol.51, pp.1-49, 2018.

[7] Lahasan. B, Lutfi, S.L. and San-Segundo, R. *"A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression"*. Artif Intell, Rev 52, pp.949–979, 2019.

[8] Zeng Dan, Veldhuis Raymond and Spreeuwers Luuk, *"A survey of face recognition techniques under occlusion"*, IET Biometrics, Vol. 10(6), pp.581-606, https://doi.org/10.1049/bme2.12029, 2021.

[9] GuiLing. W, *"Masked Face Recognition Algorithm for a Contactless Distribution Cabinet"*, Mathematical Problems in Engineering, 10.1155/2021/5591020, 2021.

[10] Wang. S., Xia. K., Wang. L., Zhang, J. and Yang. H., *"Improved RPCA method via non-convex regularisation for image denoising"*, IET Signal Process, pp. 269-277, https://doi.org/10.1049/iet-spr.2019.0365,2020.

[11] Priya. GN. and Banu. RW., *"Occlusion invariant face recognition using mean based weight matrix and support vector machine"*, Sadhana, Vol.39(2), pp.303-315, 10.1007/s12046-013-0216-3, 2014

[12] Andesh Khamele. V. and Mundada. S. G.,*"An Approach for Removal of Occlusion using Examplar based Image-Inpainting Technique: A Review"*, International Journal of Engineering research & Technology (IJERT), vol.4, 2015.

[13] Rajeswari.G. and Ithaya Rani. P., *"Face Occlusion Removal for Face Recognition Using the Related Face by Structural Similarity Index Measure and Principal Component Analysis"*, Journal of Intelligent & Fuzzy Systems, Vol.42(6), pp. 5335 – 5350, 2022.

[14] He. L., Li. H., Zhang, Q.and Sun, Z., *"Dynamic feature learning for partial face recognition"*, The IEEE conference on computer vision and pattern recognition, Vol.42, pp. 7054–7063, 2018.

[15] Duan. Q and Zhang, L, *"Look more into occlusion: Realistic face frontalization and recognition with boostgan"*, IEEE Transactions on Neural Networks, Vol.42(6), pp. 1–15, 2020.

[16] Nizam Ud Din, Kamran Javed, Seho Bae and Juneho Yi, *"A Novel GAN-Based Network for Unmasking of Masked Face"*, IEEE Access, Vol.8, pp. 44276-44287, 2020.

[17] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li and Yu Qiao, *"Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks"*, Institute of Electrical and Electronics Engineers (IEEE), IEEE Signal Processing Letters, Vol.23(10), pp. 1499–1503, 10.1109/lsp.2016.2603342, oct, 2016.

[18] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays. et al, *"MediaPipe: A Framework for Perceiving and Processing Reality"*, Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR), 2019.

[19] Alzu'bi Ahmad, Albalas Firas, AL-Hadhrami Tawfik, Younis Lojin Bani and Bashayreh Amjad, *"Masked Face Recognition Using Deep Learning: A Review"*, Electronics, Vol.10(21), 2021.

[20] Yalavarthi Bharat Chandra and G. K. Reddy, *"A Comparative Analysis Of Face Recognition Models On Masked Faces"*, International Journal of Scientific & Technology Research, Vol.9, pp.175-178, 2020.

[21] Nizam Ud Din, Kamran Javed, Seho Bae and Juneho Yi, *"Effective Removal of User-Selected Foreground Object From Facial Images*

Using a Novel GAN-Based Network", IEEE Access, Vol.8, pp. 109648-109661, 2020.

[22] Tero Karras, Timo Aila, Samuli Laine and Jaakko Lehtinen, *"Progressive Growing of GANs for Improved Quality, Stability, and Variation"*, ArXiv, 2018.

[23] Mengyue Geng, Peixi Peng, Yangru Huang and Yonghong Tian, *"Masked Face Recognition with Generative Data Augmentation and Domain Constrained Ranking"*, Proceedings of the 28th ACM International Conference on Multimedia, 2020.

[24] Sengupta Soumyadip, Chen Jun-Cheng, Castillo Carlos, Patel Vishal M, and Chellappa, Rama, Jacobs, David W., *"Frontal to profile face verification in the wild"*, IEEE Winter Conference on Applications of Computer Vision (WACV), pp.1-9, 10.1109/WACV.2016.7477558, 2016.

[25] Feifei Ding, Peixi Peng, Yangru Huang, Mengyue Geng, Yonghong Tian, *"Masked Face Recognition with Latent Part Detection"*, Proceedings of the 28th ACM International Conference on Multimedia, 2020.

[26] Aqeel Anwar and Arijit Raychowdhury, *"Masked Face Recognition for Secure Authentication"*, ArXiv, 2020.

[27] Cao Qiong, Shen Li, Xie Weidi, Parkhi Omkar and Zisserman Andrew, *"VGGFace2: A Dataset for Recognising Faces across Pose and Age"*, pp. 67-74, 10.1109/FG.2018.00020, 05.2018.

[28] Deng Hongxia, Feng Zijian, Qian Guanyu, Lv Xindong, Li Haifang and LiGang, *"MFCosface: A Masked Face Recognition Algorithm Based on Large Margin Cosine Loss"*, Applied Sciences, Vol.11(16), 2021.

[29] Yuan Xiaowei and Park In Kyu, *"Face De-Occlusion Using 3D Morphable Model and Generative Adversarial Network"*, IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10061-10070, 10.1109/ICCV.2019.01016, 2019.

[30] Parkhi O.M., Vedaldi, A. and Zisserman, A., *"Deep Face Recognition"*, the British Machine Vision Conference (BMVC), Swansea, pp. 41.1–41.12, 2015.

[31] Yi Dong, Lei Zhen, Liao Shengcai and Li Stan Z., *"Learning Face Representation from Scratch"*, arXiv, 2014.

[32] Aaron Nech, Ira Kemelmacher-Shlizerman, *"Level Playing Field for Million Scale Face Recognition"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3406-3415, 2017.

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, *"ImageNet classification with deep convolutional neural networks"*. Communications of the ACM, Vol.60, pp. 84–90, https://doi.org/10.1145/3065386, 2017.

[34] Lu Yang, *"Image Classification Algorithm Based on Improved AlexNet in Cloud Computing Environment"*, 2020 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI), pp. 250-253, 10.1109/IAAI51705.2020.9332891, 2020.

[35] Almabdy Soad and Elrefaei Lamiaa, *"Deep Convolutional Neural Network-Based Approaches for Face Recognition"*, Applied Sciences, Vol.9(20), 10.3390/app9204397, 2019.

[36] Simonyan Karen and Zisserman Andrew, *"Very Deep Convolutional Networks for Large-Scale Image Recognition"*, arXiv, 10.48550/ARXIV.1409.1556, 2014

[37] Tony Gwyn, Roy Kaushik and Mustafa Atay, *"Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study"*, Future Internet, Vol.13, p.164, 2021.

[38] He, K., Zhang, X., Ren, S. and Sun, J., *"Deep Residual Learning for Image Recognition"*, The IEEE Conference on Computer Vision and Pattern Recognition", pp. 770-778, Las Vegas, NV, 27-30 June 2016.

[39] Tavakolian Niloofar, Nazemi Azadeh, Azimifar Zohreh and Murray Iain, *"Face recognition under occlusion for user authentication and invigilation in remotely distributed online assessments"*, International Journal of Intelligent Defence Support Systems, Vol.5, p.277, 10.1504/IJIDSS.2018.099889, 12.2018.

[40] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang , Tobias Weyand. et al, *"MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications"*, ArXiv, Vol. abs/1704.04861, 2017.

[41] Tony Gwyn, Roy Kaushik and Mustafa Atay, *"Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study"*, Future Internet, Vol.13, p.164, 2021.

[42] Szegedy Christian, Vanhoucke Vincent, Ioffe Sergey, Shlens Jon and Wojna Zbigniew, *"Rethinking the Inception Architecture for Computer Vision"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 10.1109/CVPR.2016.308, 2016.

[43] Szegedy Christian, Ioffe Sergey, Vanhoucke Vincent and Alemi Alex, *"Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning"*, arXiv, 10.48550/ARXIV.1602.07261, 2016.

[44] Chollet François, *"Xception: Deep Learning with Depthwise Separable Convolutions"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800-1807, 10.1109/CVPR.2017.195, 2017.

[45] Florian Schroff, Dmitry Kalenichenko and James Philbin, *FaceNet: A unified embedding for face recognition and clustering*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 1-10, 10.1109/cvpr.2015.7298682, 2015.

[46] Marcelo Beckmann, Nelson F. F. Ebecken and Beatriz S. L. Pires de Lima, *"A KNN Undersampling Approach for Data Balancing"*, Journal of Intelligent Learning Systems and Applications, Vol.7, November 11, 2015.

[47] Krizhevsky, A., Sutskever, I. and Hinton, G.E., *"ImageNet Classification with Deep Convolutional Neural Networks"*, ACM, Vol.60, pp. 84-90, 2017.

[48] Le Khanh, *"A study of face embedding in face recognition"*, 2019.

[49] Gregory R. Koch, *"Siamese Neural Networks for One-Shot Image Recognition"*, 2015.

[50] William Ivan, Ignatius Moses Setiadi, De Rosal and Rachmawanto, Eko Hari, Santoso and Heru Agus. et al, *"Face Recognition using FaceNet (Survey, Performance Test, and Comparison"*,Fourth International Conference on Informatics and Computing (ICIC), pp. 1-6, 10.1109/ICIC47613.2019.8985786, 2019.

[51] Jean-Pierre PECUCHET, *"Evolutionary Optimisation of Kernel Functions for SVMS"*, 2015.

[52] Wirdiani Ayu, Hridayami Praba, Widiari Ayu, Rismawan Komang, Candradinata Putu and Jayantha, I., *"Face Identification Based on K-Nearest Neighbor"*. Scientific Journal of Informatics. Vol.6. pp.150-159. 2019.

[53] M. Mohandes, M. Deriche and S. O. Aliyu, *"Classifiers Combination Techniques: A Comprehensive Review"*, in IEEE Access, vol. 6, pp. 19626-19639, 2018.

[54] Ahmad H. A. Eid, *"Combined Classifiers for Invariant Face Recognition"*, arXiv, 2016

[55] Li Yixiao, Fang Yuan, Haipeng Peng and Yang Yixian, *"Bagged Tree Based Frame-Wise Beforehand Prediction Approach for HEVC Intra-Coding Unit Partitioning"*. Electronics. Vol.9. 1523. 10.3390/electronics9091523, 2020.

[56] Karim Alia and Alawi Mohammed, *"Combine SVM and KNN Classifiers for Handwriting Arabic Word Recognition based on Multi feature"*,2019.

[57] Zaghdoudi Rachid and Seridi Hamid.*"Combination of Multiple Classifiers for Off-Line Handwritten Arabic Word Recognition"* . International Journal of Information Technology. Vol.14. pp.713-720, 2017.

[58] Hamdi, Houssamddine and Yurtkan, Kamil, *"Masked Face Recognition Based on FaceNet Pré-Trained Model"*,2022.

[59] walid hariri, *"Efficient Masked Face Recognition Method"*, p.8, 2020.