Taher KHADHRAOUI, PhD
khadhra.th@gmail.com
University of Tunis El Manar, Tunis, Tunisia

Fayçal HAMDAOUI, PhD (corresponding author)
faycel_hamdaoui@yahoo.fr
University of Monastir, Monastir, Tunisia

Advanced Unified Deep Learning Framework for Multi-Task Facial Recognition and Analysis

Abstract. *Face recognition is one of the most well-known topics in biometrics research. In this paper, a novel architecture based on a Single Deep Neural Network (SDNN) is presented, which can classify the variation in any given query face. Thus, this Single Deep Neural Network framework performs face recognition with real disguise, facial expression recognition, and gender classification from a given image. Our approach is to learn common features for these different tasks and also to exploit the synergy among them. The experimental results of our new approach applied on different benchmark databases are illustrated, including the AR (Augmented Reality) face database and the Labelled Faces in the Wild (LFW) database. Noteworthy is its accomplishment of an accuracy rate of 93 % and 97 % in the worst-case scenarios for the LFW and AR face databases, respectively. In the best-case scenario, it achieves a perfect 100 % accuracy for the Karolinska Directed Emotional Faces (KDEF) database. Regarding the evaluation running time, the execution is in real time, taking approximately 180 ms.*

Keywords: *biometric, face recognition, multi-task, Single Deep Neural Network, Graphic Processor Unit.*

JEL Classification: C38, C45, C88.

Received: 19 August 2025	**Revised: 4 December 2025**	**Accepted: 8 December 2025**

1. Introduction

A lot of studies have highlighted how to ameliorate the pattern recognition performance. The face is a very human expressive and communicative part. Indeed, it is currently focused in research as it enables the interface human machine to establish a dialogue between two entities.

Face Recognition has been considered as a challenging task because of the different types of large face variations, including poses, expressions, illumination, and disguise. Significantly, facial feature recognition has turned into a research hotspot in various industries (Li et al., 2022). To enhance the facial feature recognition accuracy, a facial feature recognition method based on local generic representation was suggested (Khadhraoui et al., 2018).

DOI: 10.24818/18423264/59.4.25.04

Practically, face detection and facial feature recognition operate unitedly. At first, Face detection marks faces. Then facial feature recognition finds out the identity information linked to those faces. However, the recognition effect of face feature recognition in the interference environment has not attained the expected results yet. In fact, its rate is hampered by environmental interferences linked to occlusions, side faces, and head tilts. Therefore, solutions related to environmental interference should also be proposed, beyond setting or choosing facial feature recognition models.

In this paper, we present a novel and efficient approach called the Single Deep Neural Network, which simultaneously solves the three sub-tasks: face recognition with real disguise, facial expression recognition, and gender classification. These three sub-tasks are integrated into a unified architecture which is trained jointly for exploring the interdependence between them, so that the common parameters from lower layers of the SDNN can be shared across tasks. On the other hand, the upper layers are more accurate to learn robust features, which reduces over-fitting in the shared layers. Several novelties are introduced to make the following aspects.

Firstly, one new general network optimisation method is suggested by Gradient Centralisation (GC). The latter helps quicken the SDNN training process and enhance the model generalisation performance. Then, GC is shown to reduce the loss function via a novel limitation on the weight vector. After that, this regularises the weight as well as the output spaces so as to boost the model generalisation performance.

Finally, we demonstrate that our SDNN framework achieves convincing performance in facial expression recognition, face recognition with real disguise, and gender classification. This paper is organised as follows: In section 2, we describe the regularisation approaches for Deep Learning (DL). An overview of the proposed approach is provided in Section 3. The experimental results are illustrated in Section 4. Section 5 presents the conclusion and some future work.

2. Literature review and regularisation approaches for DL

A substantial body of research has addressed various aspects of facial feature recognition. Notably, Gutta et al. in (Gutta et al., 2000) examined the problem of automatic categorisation of human faces based on gender and pose classification and they addressed discrimination using a method called ”mixture of experts”. In (Khadhraoui et al., 2016), Khadhraoui et al. put forward an algorithm based on particle swarm optimisation to extract the essential geometric content of facial features.

Ranjan et al. suggested two algorithms, one called an all-in-one CNN for face analysis (Ranjan et al., 2017), which would regularise the shared CNN parameters. The same author proposed, later in (Ranjan et al., 2017) a new approach called HyperFace, which would simultaneously perform multi-tasks, such as face detection, landmark localisation, pose estimation, and gender recognition. In another context related to the problem of human action recognition, Bayoudh et al. in (Bayoudh et al., 2022) developed an end-to-end framework based on inspiration from human visual attention mechanisms to semantically focus on relevant salient features in visual representations. Indeed, a 2D/3D hybridisation of a CNN was used for features extraction from video sequences, which were then introduced into a Long Short-Term Memory (LSTM) network to capture short- and long-term dependencies in the data structure. In terms of evaluation, the proposed model achieved an average classification accuracy of 96.8% on the public KTH database, identifying five actions: "hand waving," "jogging," "running," and "walking."

DL algorithms linked to individual tasks like face detection, land- mark localisation and pose estimation, were suggested in (Scherer et al., 2010) and had very good results. Beyond a shadow of doubt, GC was embedded, in optimisation approaches which were based on gradients, like Adam and SGDM, through the use of a single line of codes. Indeed, GC would boost training processes and improve as well the generalisation performance and compatibility with the aim of refining any pre-trained model.

Zhang et al. (Zhang et al., 2016) made a notable contribution by introducing a deep cascaded multitask framework that leverages the inherent correlation between face detection and alignment to enhance their performance. This framework employs three stages of meticulously designed deep convolutional networks, which predict face and landmark locations in a coarse-to-fine manner. Additionally, they proposed a novel online hard sample mining strategy that significantly improves practical performance. This approach demonstrated superior accuracy compared to state-of-the-art techniques in challenging datasets and benchmarks, including the wider face benchmark for face detection and the annotated facial landmarks in the wild benchmark for face alignment, while maintaining real-time performance.

Kusal et al. (Kusal et al., 2024) utilise a hybrid deep learning approach with pre-trained embeddings for emotion detection in conversational texts. This method demonstrates the potential of deep learning to improve emotion recognition accuracy, which is essential for applications.

In the context of machine learning, including Deep Learning, regularisation is utilised to develop a more generalised and robust model. This heightened generalisation allows the proposed model to excel even with new and unseen data. The purpose of regularisation is to prevent overfitting by introducing constraints on the model during the training phase. These constraints serve to restrict its complexity and adaptability, preventing the memorisation of specific details from the training data.

Many regularisation approaches related to the deep CNN have been recently suggested. In 2012, Hinton (Hinton et al., 2012) put forward one clear and effective approach, called dropout, where a random subset of activations in one layer was set to zero. Figure 1 presents this dropout concept.

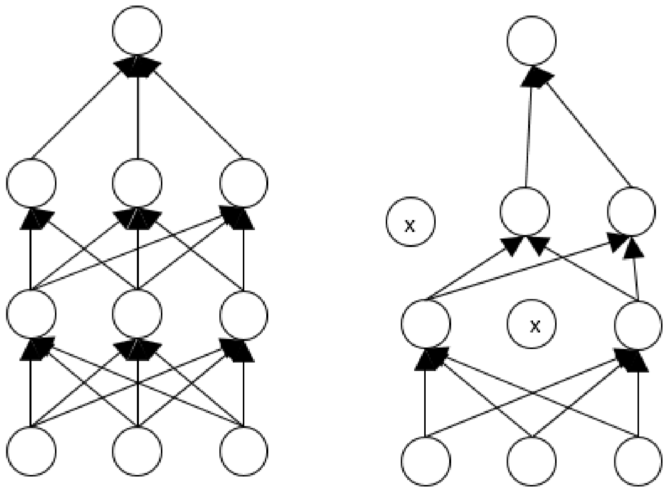

Figure 1. Representation of concept Dropout
Source: Authors' own creation.

The weight subset was set to zero in the network layers, instead of activation dropping, for dropout connections. These latter represent a different regularisation approach. For this purpose, each layer would receive a randomly selected subset of units from an immediate previous layer.

3. Motivation and methodology of the proposed approach

This section outlines the proposed SDNN framework designed for face recognition with real disguises, facial expression recognition, and gender classification from a provided image.

3.1 Motivation of the proposed approach

This proposed SDNN framework can learn common features from facial expression recognition and gender classification tasks, capitalising on the synergy between them. The key innovations addressed in this paper are summarised as follows:

1) Integrated Approach: The SDNN combines the tasks of face recognition with disguise, facial expression recognition, and gender classification within a single architecture. This promotes an effective and coordinated management of these complex tasks.

2) Task Synergy: The SDNN framework establishes synergy among various tasks related to facial analysis, exploiting potential relationships between different facets. This approach enhances the overall system performance.

3) Shared Common Features: The SDNN leverages similarities among diverse aspects of facial analysis by exploiting common features. This optimises resource utilisation and promotes better model generalisation across various scenarios.

4) Standardised Image Processing: The SDNN simplifies data preprocessing by adopting a standardised input image resolution, facilitating comparison and interoperability with other systems.

5) Refinement Capability: The SDNN utilises its ability to perform advanced refinement, suggesting a potential improvement in precision and quality across different facial analysis tasks simultaneously.

3.2 Methodology and Algorithmically Overview of the Proposed Approach

To represent the different facial analysis tasks in a successful way, the SDNN-based classification scheme is proposed. This suggested architecture is mainly trained in an SDNN framework, which makes a synergy among different tasks related to facial analysis. These tasks are divided into three groups: face recognition with real disguise, facial expression recognition and gender classification.

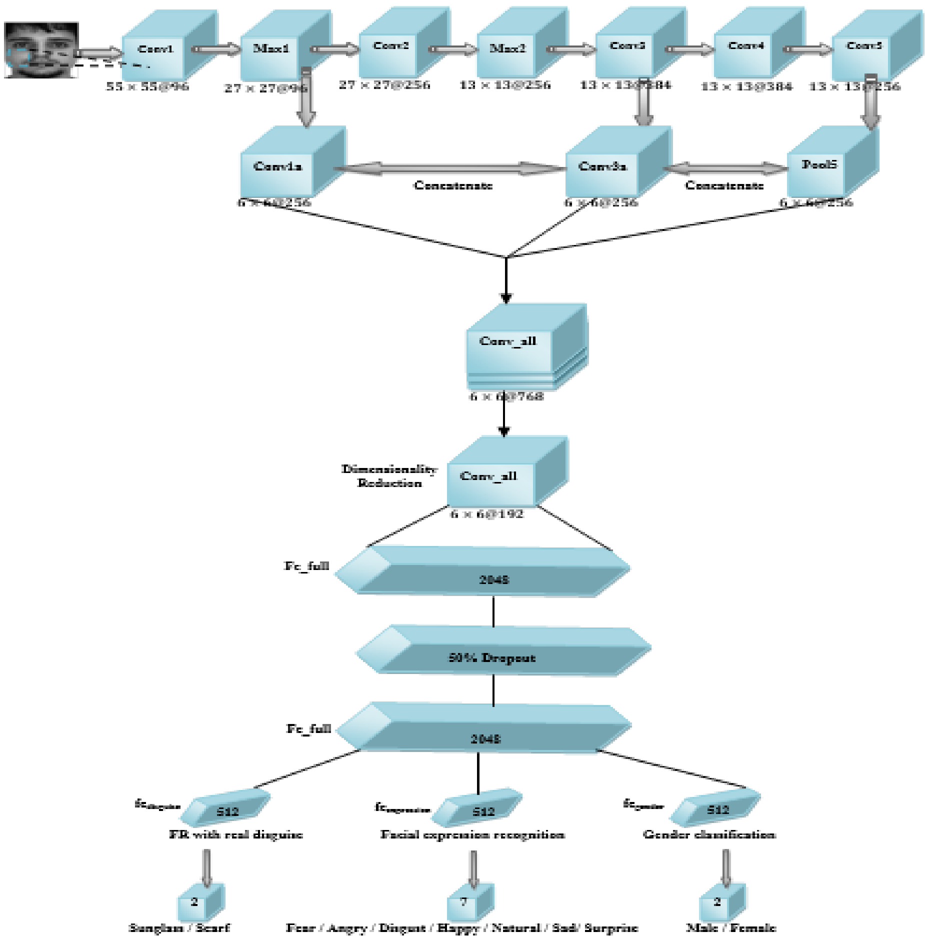

Figure 2. Overview of SDNN architecture

Source: Authors' own creation.

In summary, this can be described as a cropped face region, and pixels are scaled to 227x227 pixels as an input image, followed by a powerful refinement of jointly multiple tasks, and assuming that all tasks share some common features.

The network is made up of five convolutional layers with three fully-connected layers. The $Max1$, $Conv3$ and $Pool5$ layers of Alexnet are fused, through the use of a separate network. Naive fusion directly concatenates features. Since the feature maps of these layers have multiple dimensions (27x27x96, 13x13x384 and 6x6x256), they cannot be concatenated. Thus, the $Conv1a$ and $Conv3a$ convolutional layers are added to the $Max1$ and $Conv3$ layers to obtain at the output consistent 6x6x256 feature maps. Then, the output of such layers is concatenated along with $Pool5$ in order to form the 6x6x768 feature maps. This dimension is quite high for training a multi- task framework. Accordingly, we add one 1x1 kernel convolutional layer $Conv\_all$ to reduce dimensions (Szegedy et al., 2015) into 6x6x192. After that, we add a fully-connected layer Fc_full to $Conv\_all$. The latter will output a vector with a 2,048-dimensional feature. It is followed by a 50% dropout layer to reduce over-fitting. This is followed by a fully-connected layer of dimension 2,048. Consequently, we split the network into three separate branches related to different tasks. The fully-connected layers $fc_{disguise}$, $fc_{expression}$ and fc_{gender} are added to f call, each of dimension 512.

Finally, a fully-connected layer is added to every branch with the goal of predicting individual task labels. After each convolution or fully-connected layer, we deploy the ReLU non-linearity. Then we use task-specific loss functions to learn the network weights.

Our algorithm uses a multi-task feature learning method to share features and make the classification problem convex. In practice, rather than looking for an exact solution, we look for an approximate solution, which is obtained by restricting the solution space to a well-behaved class of signals. In the prediction phase of gender recognition, we apply the same feature extraction process to the new images. For a candidate region with an overlap of 0 : 5 with the ground truth, we have the reproducing formula in Eq. (1) of the softmax loss as follows:

$$loss_G = -(1-g) * log(1-p_0) - g * log(p_1)$$

where (p_0, p_1) is a two-dimensional probability vector computed from the network, and $g = 0$ if the face is classified as a male, or $g = 1$ if the face is classified as a female.

In the prediction phase of facial expression recognition, we use the Euclidean loss to train the facial expression estimates of fear (p_1), angry (p_2), disgust (p_3), happy (p_4), natural (p_5), sad (p_6) and surprise (p_7). We compute the loss for a candidate region having an overlap more than 0.5 with the ground truth, in Eq. (2) below:

$$loss_{exp} = \frac{(\hat{p}_1-p_1)^2+(\hat{p}_2-p_2)^2+(\hat{p}_3-p_3)^2+(\hat{p}_4-p_4)^2}{7} + \frac{(\hat{p}_5-p_5)^2+(\hat{p}_6-p_6)^2+(\hat{p}_7-p_7)^2}{7}$$

where $(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \hat{p}_6, \hat{p}_7)$ are the estimated facial expression labels.

Similar to gender recognition prediction, we compute the softmax loss, for the task of face recognition with real disguise, given in the formula below below:

$$loss_D = -(1-d) * log(1-p_0) - d * log(p_1)$$

where $d = 0$ if the face is classified as a face with sunglasses, or $d = 1$ if the face is classified as a face with a scarf.

The full loss is calculated as the weighted sum of the three individual losses as presented in the formula below:

$$loss_{full} = \sum_{t=1}^{t=3} \lambda_t loss_t$$

Where λ_t is an appropriately chosen weight parameter, and $loss_t$ is the individual loss corresponding to the t^{th} element from the set of tasks $T = \{G, exp, D\}$. In all the experiments, after several attempts, we fix $\lambda_G = 2$, $\lambda_{exp} = 5$ and $\lambda_D = 1$.

The full algorithm of SDNN-Training can be given as shown in Table 1.

Table 1. Training a multi-task SDNN

Initialize Θ

for epoch in 1 ... N **do**
  for iteration in 1 ... L **do**
    1. Pick a task t randomly
    2. Pick sample(s) from task t
    3. Compute loss:
      3.1. $loss_G$ = Eq.(1) for gender prediction
      3.2. $loss_{exp}$ = Eq.(2) for facial expression recognition
      3.3. $loss_D$ = Eq.(3) for face recognition with real disguise
    4. Calculate gradient $\nabla(\Theta)$
    5. Update Θ by $\Theta - \varepsilon * \nabla(\Theta)$
  end for
end for

Source: Proposed by authors.

The details of the SDNN architecture, used in our experiments, are shown in Figure 3.

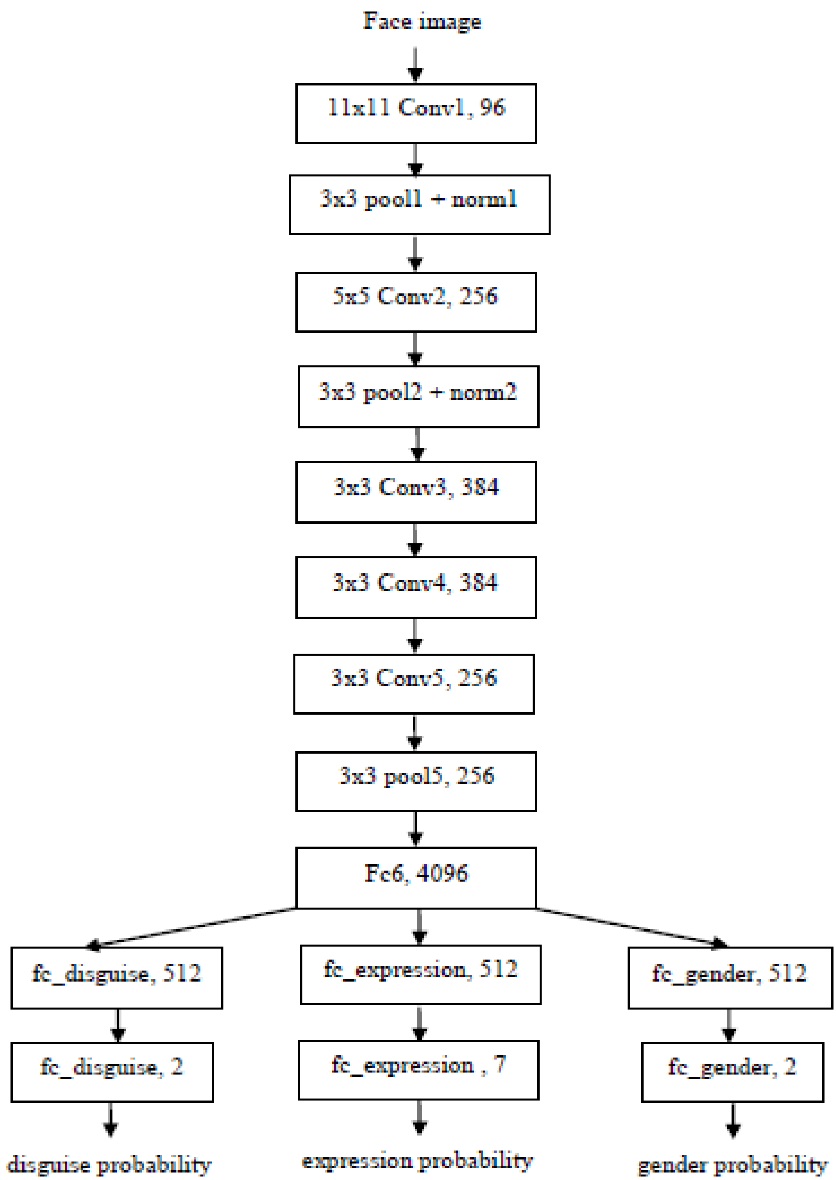

Figure 3. SDNN architecture of multi-task face analysis. In each box, the left numbers with the size of n×n denote the kernel size, and the right numbers denote the cardinality of feature maps for a given layer
Source: Authors' own creation.

4. Results and discussion

For the validation of our proposed approach and assessment of its performance, we conduct extensive numerical experiments on benchmark face databases for face recognition problems, with real face disguise and gender classification. Our algorithm is implemented through the use of the Caffe deep learning library (Jia et al., 2014). The suggested method is compared to other state-of-the-art methods including Collaborative Representation Classification with Regularised Least Squares (CRC-RLS) (Zhang et al., 2011), Sparse Representation Classification (SRC) (Wright et al., 2008), SVM (Cortes et al., 1995), FaceTracer (Kumar et al., 2008), PANDA-w (Zhang et al., 2014), PANDA-1 (Zhang et al., 2014), ANet (Li et al., 2013), LNets+Anet (Liu et al., 2015), RCNN Gender (Kumar et al., 2008) and Multitask Face and HyperFace (Ranjan et al., 2017). Two face databases, including the AR (Martinez et al., 1998) and LFW databases (Huang et al., 2008), are used for testing the SDNN performance and compare it to competing methods.

4.1 Experiment 1: FR with real face disguise: AR database

For this experiment, a subset of the AR database (Martinez et al., 1998) is selected, with a real face disguise (scarf and sunglasses). The database consists of over 3000 colour images displaying frontal views of faces. These images represent the faces of 116 individuals (63 men and 53 women) captured under various conditions, including different facial expressions, different illumination conditions, and with different characteristic changes (Figure 4).

Figure 4. Some representative subjects (3 images with scarves, 3 images with sunglasses) from AR database in our experiment
Source: AR database.

This database was divided into two distinct parts, 80% for training and validation, and 20% for testing. To assess the classification quality of the proposed framework, the public AR database with templates of various resolutions and colors was utilised. In this work, the AR database images were resized to 227 x 227 pixels before being introduced into the SDNN model. Similarly, regularisation techniques were adopted to enhance the representation capability of the proposed model and address the imbalance in the distribution of different data types in the database, using a coefficient of 0.5.

For the first experimental case, we start by evaluating the performance of the SDNN framework on the AR database. The accuracy results are consolidated in Table 2. This metric is computed based on the following expression:

$$Accuracy = \frac{VP + VN}{VP + FP + VN + FN}$$

with:

- TP (True Positives): Represents the number of faces correctly classified.
- TN (True Negatives): Represents the number of non-faces correctly classified.
- FP (False Positives): Represents the number of faces incorrectly classified.
- FN (False Negatives): Represents the number of non-faces incorrectly classified.

The results reported in Table 2 demonstrate that our SDNN method provides the best performance with real face disguise (scarf and sunglasses) compared to those given by other techniques. Although the results vary depending on Adornment, our SDNN method consistently achieves the highest accuracy for both types of images: with sunglasses (98.16%) and with a scarf (97.00%). Indeed, our method exhibits a slight advantage over the SRC (partitioned) (Wright et al., 2008) method by 0.66% and 3.5% for sunglasses and scarves, respectively. In comparison with the CRC-RLS (Zhang et al., 2011) method, the differences are more significant, standing at 6.66% and 2%, respectively, for sunglasses and scarves.

Table 2. Face recognition performance with real disguise using the AR database

Methods	**Sunglass (%)**	**Scarf (%)**
SVM (Cortes et al., 1995)	66.50	16.50
SRC (Wright et al., 2008)	87.00	59.50
SRC (partitioned) (Wright et al., 2008)	97.50	93.50
CRC-RLS (Zhang et al., 2011)	68.50	90.50
Improved CRC-RLS (Zhang et al., 2011)	91.50	95.00
Our (SDNN)	98.16	97.00

Source: Authors' processing.

As depicted in Table 2, and as previously demonstrated, the SDNN leads to a substantial improvement in the FR rate compared to other methods for sunglasses and scarves. Moreover, the proposed model based on the SDNN framework has led to higher accuracy for images with sunglasses compared to those with scarves. Indeed, this is because when sunglasses are worn, they typically cover a smaller portion of the face, leading to reduced occlusion and preserving more facial features essential for recognition. Conversely, However, scarves can introduce more variations, such as different draping styles, leading to increased complexity for the SDNN. Additionally, the training data may have included a more diverse representation of facial expressions and features in the case of sunglasses, contributing to enhanced accuracy in those scenarios.

4.2 Experiment 2: Facial expression recognition

For this second experimental case, the utilised database is the renowned KDEF database (Lundqvist et al., 1998). This database comprises a set of 4900 RGB color images (32 bits: 16.7 million colours) with size of 562 * 762 pixels. This collection is sourced from seven different emotional expressions (neutral, happy, angry, sad, disgusted, surprised, and fear) exhibited by 70 individuals (35 females and 35 males).

Each expression is captured from five distinct angles (full left profile, half left profile, frontal, half right profile, and full right profile). Based on this database, we will begin by evaluating the model's performance in recognising facial expressions (the seven expressions) using the following metrics: training and test accuracy, training and test loss, and the confusion matrix. The results of these metrics are consolidated into three figures, namely Fig. 5, Fig. 6, and Fig. 7.

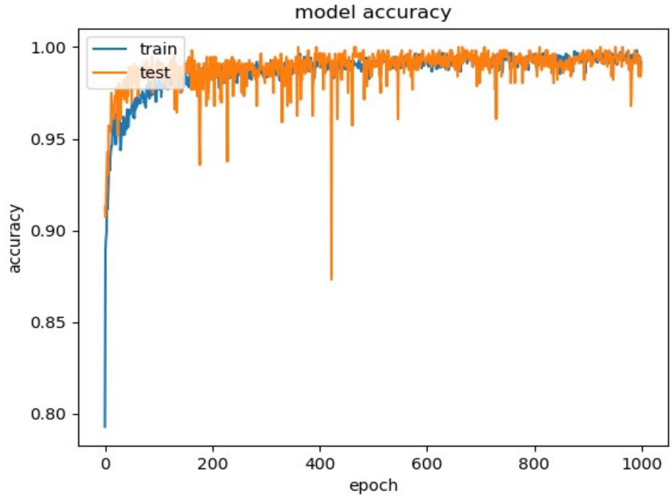

Figure 5. Training and testing accuracy curve of facial expression
Source: Authors' own creation.

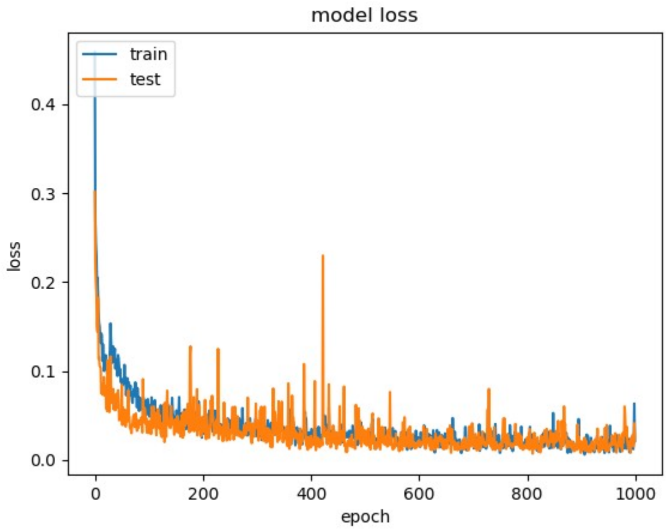

Figure 6. Training and testing loss curve of facial expression
Source: Authors' own creation.

According to Figures 5 and 6, the training and test accuracy values for our SDNN model stabilise at 99.95% and 99.13%, respectively, starting from the five hundredth epoch. Simultaneously, the training loss and test loss values decrease rapidly, reaching values of 1% for training and less than 0.5% for testing, respectively, from the six hundredth epoch onwards. This indicates that our model has effectively overcome the issue of overfitting, as the test loss values are lower than the training loss.

Similarly, according to Figure 7, illustrating the confusion matrix, our SDNN facial expression recognition model predicted all expressions of the "Disgust" and "Sad" classes as true positives. However, for the "Angry" and "Happy" classes, 99.99% were correctly predicted, with 1% false negatives misclassified as "Disgust" instead of the "Angry" class and 1% false positives categorised as "Neutral" instead of "Happy." For the "Surprise" class, the true positive rate was 99.98% of the prediction. The last two classes, "Neutral" and "Fear," have true positive rates of 99.86% and 99.82%, respectively, representing the lowest rates.

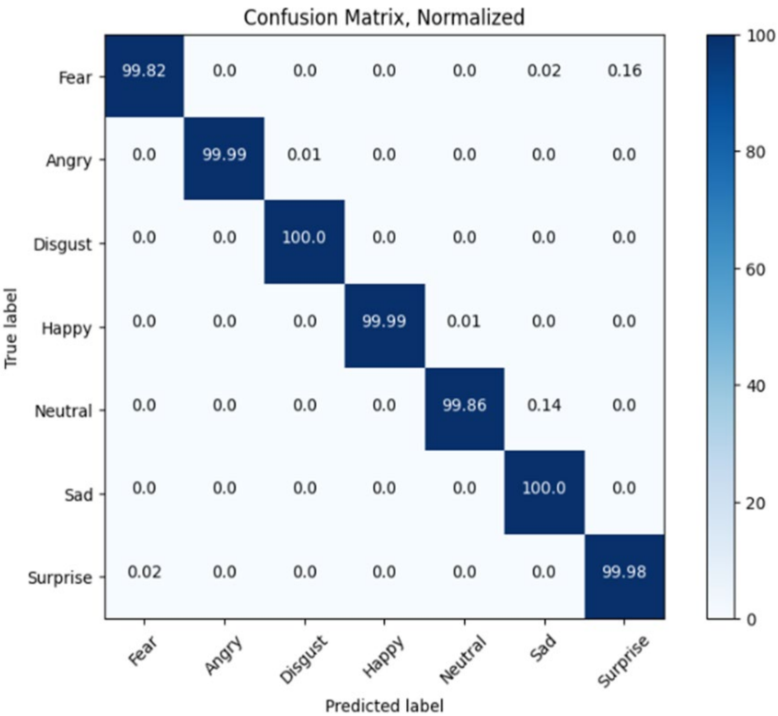

Figure 7. Confusion matrix on KDEF database
Source: Authors’ own creation.

To further validate the proposed system, we compared it with two other methods from the literature: "Mixtures of Experts" implemented with ensembles of radial basis functions (ERBF), inductive decision trees (DTs), and support vector machines (SVMs) (Gutta et al., 2000), and Feedforward ANN (Hinton et al., 2012) based on an ANN architecture incorporating dropout techniques. The accuracy results applied to the KDEF database for each method are presented in Table 3. Referring to the results, we observe that our model outperforms the others with an average accuracy of 99.95%, surpassing the Feedforward ANN (Hinton et al., 2012) method by a margin of 39.02%. Similarly, our method exhibits superior performance when compared to the "Mixtures of Experts: ERBF, DTs, and SVM" method (Gutta et al., 2000), surpassing it for all seven facial expressions.

Table 3. Performance comparison on KDEF database

	Fear	Angry	Disgust	Happy	Neutral	Sad	Surprise	Average
Mixtures of Experts: ERBF, DTs and SVM (Gutta et al., 2000)	85.33	85.19	97.74	-	98.55	79.76	98.80	90.89
Feedforward ANN (Hinton et al., 2012)	46	53	70	80.50	51.50	63	62.50	60.92
Our SDNN method	99.82	99.99	100	99.99	99.86	100	99.98	99.94

Source: Authors' processing.

4.3 Experiment 3: Gender classification

For the last experimental case, we conducted a study on gender classification based on facial images, utilising two databases. The first is derived from the AR database (Martinez et al., 1998), where we selected a non-occluded subset (14 images per subject) from the original database that initially comprised 63 male subjects and 53 female subjects. The images of the first 50 males and 42 females were used for the training phase, with the remaining images reserved for the testing phase.

Figure 8. Some representative subjects (3 males, 3 females) from AR database in our experiment

Source: AR database.

The second database employed is the public Labelled Faces in the Wild (LFW) database (Huang et al., 2008), consisting of over 13000 facial images collected from the internet. Each face in this database is labelled with the name of the person pictured, and 1680 individuals have two or more different images available for analysis.

In this experiment, we compare the proposed approach to other state-of-the-art methods such as the FaceTracer (Kumar et al., 2008), PANDA-w (Zhang et al., 2014), PANDA-1 (Zhang et al., 2014), ANet (Li et al., 2013), LNets+ANet (Liu et al., 2015), RCNN Gender (Kumar et al., 2008), Multitask Face and HyperFace (Zhang et al., 2016), and the results are shown in Table 4. The table indicates that the SDNN achieves the third best performance after HyperFace and LNets+ANet on the LFW databases.

Table 4. Gender classification performance on AR database and LFW databases for different methods

Method	LFW	AR
FaceTracer (Kumar et al., 2008)	84.00	-
PANDA-w (Zhang et al., 2014)	86.00	-
PANDA-1 (Zhang et al., 2014)	92.00	-
ANet (Li et al., 2013)	91.00	-
LNets+ANet (Liu et al., 2015)	94.00	-
RCNN Gender (Kumar et al., 2008)	91.00	-
Multitask Face (Zhang et al., 2016)	93.00	-
HyperFace (Ranjan et al., 2017)	94.00	-
Our (SDNN)	93.00	97.00

Source: Authors' processing.

Based on the results presented in Table 4 for the LFW database, it can be concluded that the proposed SDNN model exhibits a highly acceptable accuracy of 93%. Indeed, our model ranks third, trailing behind the LNets+Anet (Liu et al., 2015) and HyperFace (Ranjan et al., 2017) models, both achieving the same accuracy of 94%, surpassing our model by only 1%. Notably, our model outperforms five other models, demonstrating the robustness of our approach. For the AR database, our method achieves a prediction accuracy of 97%.

4.4 Running time

Our network is trained and evaluated on an ASUS PC Intel(R) Core(TM) i7 Processor, clock frequency @2.6 GHz and 16 Go RAM) with NVIDIA GTX 950M. Training the network takes about four hours, while the evaluation for each task takes about 180ms.

5. Conclusions

In this paper, we have presented an SDNN for multi-task of facial analysis. The experimental tests run on multiple face databases have demonstrated that the SDNN performs very favourably compared to other methods based on state-of-the-art face recognition, face recognition with real disguise, and gender classification. Few crucial experimental observations have been discussed. Firstly, all the face related tasks can take advantage of using the multitask learning framework. In fact, such a gain is mainly caused by the networks capability of learning many more discriminative characteristics, as well as post-processing methods. Secondly, the fusion of these intermediate layers will improve, in general, the performance of structure dependent tasks. We can particularly cite facial expression recognition and face recognition with a real disguise. In this case, their characteristics become geometry invariant within deeper CNN layers. In addition to that, the SDNN exploits these observations with the aim of improving the performance of the three sub-tasks. However, further enhancements can be proposed to augment the value of this framework, such as deploying it on an embedded GPU system to utilise it in detection applications that require real-time responses. Additionally, implement SDNN in edge computing and IoT environments for real-time face recognition on devices with limited resources.

References

[1] Bayoudh, K., Hamdaoui, F., Mtibaa, A. (2022), *An attention-based hybrid 2D/3D CNN-LSTM for human action recognition. In 2022 2nd international conference on computing and information technology (ICCIT), Tabuk, Saudi Arabia, 25-27 January 2022*, IEEE, 97-103.

[2] Bulò, S.R., Porzi, L., Kontschieder, P. (2016), *Dropout distillation. In International Conference on Machine Learning, PMLR, New York, NY, USA, 20-22 June 2016*, 99-107.

[3] Cortes, C., Vapnik, V. (1995), *Support-vector networks. Machine learning*, 20(3), 273-297.

[4] Gutta, S., Huang, J.R., Jonathon, P., Wechsler, H. (2000), *Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE Transactions on neural networks*, 11(4), 948-960.

[5] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R. (2012), *Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv*:1207.0580.

[6] Huang, B.G., Ramesh, M., Berg, T., Learned-Miller, E. (2008), *Labeled faces in the wild: A database for studying face recognition. Unconstrained Environments. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct 2008, Marseille, France.*

[7] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Darrell, T. (2014), *Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, November 3-7, 2014*, 675-678.

[8] Khadhraoui, T., Borgi, M.A., Benzarti, F., Ben Amar, C., Amiri, H. (2018), *Local generic representation for patch uLBP-based face recognition with single training sample per subject. Multimedia Tools and Applications*, 77(18), 24203-24222.

[9] Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D. (2024), *Transfer learning for emotion detection in conversational text: a hybrid deep learning approach with pre-trained embeddings*. *International Journal of Information Technology*, 1-18.

[10] Khadhraoui, T., Ktata, S., Benzarti, F., Amiri, H. (2016), *Features selection based on modified PSO algorithm for 2D face recognition. In 2016 13th international conference on computer graphics, imaging and visualization (CGiV), Beni Mellal, Morocco, 29 March - 01 April 2016*, IEEE, 99-104.

[11] Kumar, N., Belhumeur, P., Nayar, S. (2008), *Facetracer: A search engine for large collections of images with faces. In European conference on computer vision* (ECCV 2008), *Marseille, France, October 12-18, 2008*, Proceedings, Part IV, 340-353.

[12] Li, M., Huang, B., Tian, G. (2022), *A comprehensive survey on 3D face recognition methods*. *Engineering Applications of Artificial Intelligence*, 110, 104669.

[13] Li, J., Zhang, Y. (2013), *Learning surf cascade for fast and accurate object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23-28 June 2013*, 3468-3475.

[14] Liu, Z., Luo, P., Wang, X., Tang, X. (2015), *Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 07-13 December 2015,* 3730-3738.

[15] Lundqvist, D., Flykt, A., Ohman, A. (1998), *Karolinska directed emotional faces [database of standardized facial images]*. *Psychology Section*, Department of Clinical Neuroscience, Karolinska Hospital, Sweden, S-171, 76.

[16] Martinez, A., Benavente, R. (1998), *The ar face database: Cvc technical report, Robot Vision Lab, Purdue University, http://www.cat.uab.cat/Public/Publications/1998/MaB1998/CVCReport24.pdf*, 24.

[17] Ranjan, R., Sankaranarayanan, S., Castillo, C.D., Chellappa, R. (2017), *An all-in-one convolutional neural network for face analysis. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017), Washington, DC, USA, 30 May - 03 June 2017, IEEE,* 17-24, https://doi.org/10.48550/arXiv.1611.00851.

[18] Ranjan, R., Patel, V.M., Chellappa, R. (2017), *Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition*. *IEEE transactions on pattern analysis and machine intelligence*, 41(1), 121-135.

[19] Scherer, D., Schulz, H., Behnke, S. (2010), *Accelerating large-scale convolutional neural networks with parallel graphics multiprocessors. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds) Artificial Neural Networks – ICANN 2010. ICANN 2010. Lecture Notes in Computer Science, vol 6354. Springer,* Berlin, Germany, https://doi.org/10.1007/978-3-642-15825-4_9.

[20] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Van-Houcke, E., Rabinovich, A. (2015), *Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition*, *Boston, MA, USA, 07-12 June 2015,* 1-9, 10.1109/CVPR.2015.7298594.

[21] Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., Fergus, R. (2013), *Regularization of neural networks using dropconnect. In International conference on machine learning,* PMLR, *New York City, NY, USA, June 19-24 2016*, 1058-1066.

[22] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y. (2008), *Robust face recognition via sparse representation*. *IEEE transactions on pattern analysis and machine intelligence*, 31(2), 210-227.

[23] Zhang, K., Zhang, Z., Li, Z., Qiao, Y. (2016), *Joint face detection and alignment using multitask cascaded convolutional networks*. *IEEE signal processing letters*, 23(10), 1499-1503.

[24] Zhang, L., Yang, M., Feng, X. (2011), *Sparse representation or collaborative representation: Which helps face recognition?*. *In 2011 IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011,* 471-478.

[25] Zhang, N., Paluri, M., Ranzato, M.A., Darrell, T., Bourdev, L. (2014), *Panda: Pose aligned networks for deep attribute modeling*. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, *Columbus, OH, USA, 23-28 June 2014*, 1637-1644.