Journal of Engineering Research

Volume 7 | Issue 2 Article 29

2023

Deeply Smile Detection Based on Discriminative Features with Modified LeNet-5 Network

Hend Maher Obaya

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

Deeply Smile Detection Based on Discriminative Features with Modified LeNet-5 Network

Hend Maher Obaya, Amany Mahmoud Sarhan, Marwa Mahmoud Badr

Dept. Computer and Control engineering, Faculty of Engineering, Tanta University, Egypt
Emails: hend.maher@f-eng.tanta.edu.eg, amany_sarhan@f-eng.tanta.edu.eg, marwa.badr@f-eng.tanta.edu.eg

***Abstract*- Facial expressions are caused by specific movements of the face muscles; they are regarded as a visible manifestation of a person's inner thought process, internal emotional states, and intentions. A smile is a facial expression that often indicates happiness, satisfaction, or agreement. Many applications use smile detection such as automatic image capture, distance learning systems, interactive systems, video conferencing, patient monitoring, and product rating. The smile detection system is divided into two stages: feature extraction and classification. As a result, the accuracy of smile detection is dependent on both phases. In recent years, numerous researchers and scholars have identified various approaches to smile detection, however, their accuracy is still under the desired level. To this end, we propose an effective Convolutional Neural Network (CNN) architecture based on modified LeNet-5 Network (MLeNet-5) for detecting smiles in images. The proposed system generates low-level face identifiers and detect smiles using a strong binary classifier. In our experiments, the proposed MLenet-5 system used the SMILEsmilesD and (GENKI-4 K) databases in which the smile detection rate of the proposed method improves the accuracy by 2% on SMILEsmilesD database and 5% on GENKI-4 K database relative to LeNet-5-based CNN network. In addition, the proposed system decreases the number of parameters compared to LeNet-5-based CNN network and most of the existing models while maintaining the robustness and effectiveness of the results.**

***Keywords*: Convolutional Neural Networks, Facial expressions, Feature extraction, Smile detection.**

I. INTRODUCTION

Our daily communications depend greatly on facial expressions. The smile is one of the most common facial expressions. A smile expresses our happiness, satisfaction, or relaxation, among other emotions. Smile recognition is beneficial in a variety of fields, including mental health monitoring, human-computer interaction, smile payment, camera shutter control, and patient monitoring. As a result, smile recognition has evolved into an active and valuable research field that has received significant attention from researchers in recent years [2].

With the emergence of virtual agents and other human-centric applications over the last decade, detection systems for nonverbal expressions, particularly smiles and laughter, have captured the attention of the research community. This is due to the importance of these expressions in human communication. When we interact with others, we are more likely to smile than when we are alone. They can direct the flow of a conversation by changing the current topic or encouraging someone to continue speaking. The feature representation step is critical for smile recognition. Traditional feature representation methods can be divided into two categories: appearance-based and shape-based approaches. Appearance-based methods, such as local binary pattern (LBP) [3], scale-invariant feature transform (SIFT) [4] and histogram of oriented gradients (HOG) [5], extract textual features from images. Shape-based methods, on the other hand, use fiducial points or facial landmarks as discriminating features, which are then fed into classifiers to perform recognition, such as Facial Action Units [23] and Temporal Phases of Facial Actions [34]. However, designing and selecting a distinguishing feature for smile recognition remains a significant challenge, Figure 1 shows basic Architecture of smile detection System.

The convolutional neural network (CNN) can be used to solve the feature representation problem [18] CNN eliminates the unnecessary features to reduce the dimensionality and keep the importance features that effect on the output, also CNN have self-adaptive by change the values of weights and bias according to cost function. Based on the objective function, this method can automatically extract the optimal feature from raw data. The excellent performance of the CNN has attracted a large number of researchers from all areas of computer vision, including vehicle detection [25], face recognition [30], and human action recognition [21]. A CNN combines the feature extraction and the classification processes, whereas the traditional classification uses features extracted by other algorithms. Because of the benefits of CNNs mentioned above, this approach typically outperforms handcrafted features.

Furthermore, CNNs have made significant advances in the domain of expression recognition. As a result, in this study, we chose CNN for smile detection.

A subfield of expression recognition is smiling recognition. Recognizing smiles in unconstrained scenarios is difficult. The variety of facial sizes, lighting conditions, head postures, occlusions, and other factors raises the difficulty level. As a result, a model that has been well trained on images obtained in a laboratory setting always performs poorly when applied to real-world images [35]. Figure 2 shows samples of the smile face images from SMILEsmilesD [43] while Figure 3 shows samples of non-smiling face images from the same database.

Most of the related work algorithms depend on a complex architecture of CNN with high depth of layers. This leads to high performance but with high dimensional features. In addition,

DOI: 10.21608/ERJENG.2023.194599.1156

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

their complex architecture increases the computational overhead in terms of the used memory and processing resources. Sometimes, it influences the efficiency of dealing with pattern recognition problems. The key aim of recent researches is to reduce the feature dimension using the best mapping function with maintaining the nature of the original data.

In this paper, we present an effective deep learning approach for smile detection. Deep learning has the advantage of not only being able to perform classification effectively, but also of learning some high-level abstract representations from raw inputs due to the hierarchical multiple layers of a deep learning model. As learned abstract representations, we can extract the activations of the hidden layers. These models can be used to train a classifier such as SoftMax. The main objectives of the proposed system are to reduce the number of parameters while maintaining the robustness and effectiveness of the results. We try to look into the following issues through our work: (1) creation of a deep convolutional neural network, Smile Detection-CNN Model, to detect smiles; (2) extraction of the activations of the last hidden layer as learned representations; (3) training of SoftMax on these learned features to assess discriminative power.

Our model will be trained and tested on two databases: SMILEsmilesD [43], which consists of positive images that represent smile face images and negative images that represent natural face images, in addition to GENKI-4K database [32], which contains an unconstrained scenario database.

The proposed work has made the following contributions:

1. The proposed model demonstrates that the simple and straightforward model structure with efficient objective function can reduce the number of parameters while maintaining the robustness and efficacy of the results.
2. An effective deep learning-based approach is built by using a proposed modified version of the well-known LeNet-5 convolutional network (MLeNet-5) to provide accurate detection and address the various challenges of smile detection.
3. Use multiple performance indicators to assess the proposed method using the two different datasets, SMILEsmilesD [43] and GENKI-4K datasets [32], which demonstrate that the results of our method outperform those of existing models.
4. When compared to LeNet-5 convolutional network, the proposed model (MLeNet-5) performs more accurately.

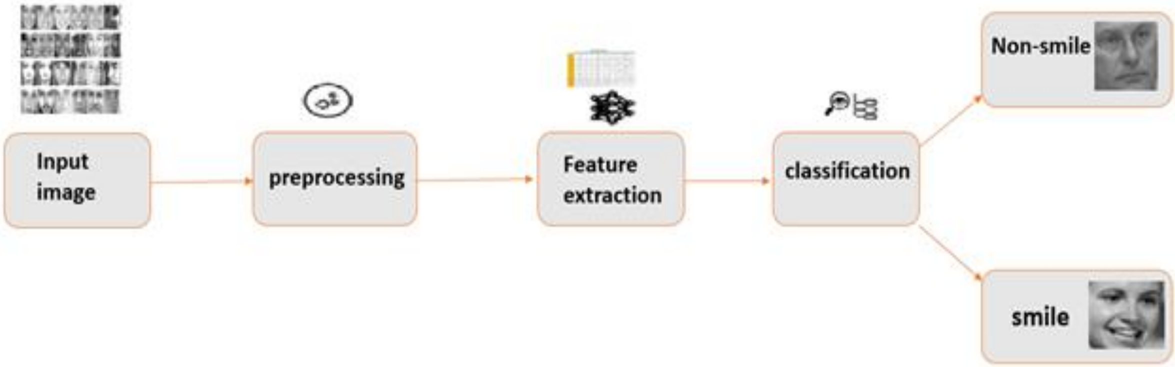

Figure 1. Basic Architecture of smile detection System

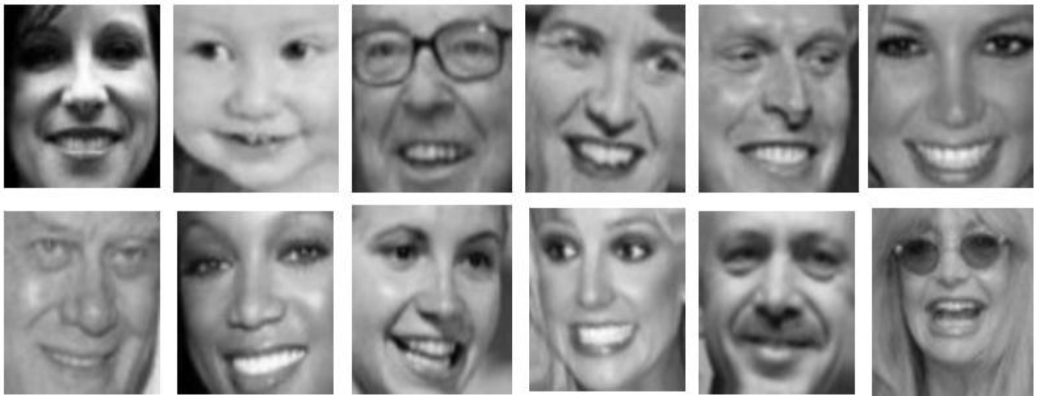

Figure 2. Sample of smile face images from SMILEsmilesD datasets [43]

DOI: 10.21608/ERJENG.2023.194599.1156

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

Figure 3. Sample of face non-smiling images from SMILEsmilesD dataset [43]

The paper is organized as follows: in Section 2, we present the related work and current research in the field. Section 3 contains the description of the model architectures for the smile detection. We describe the experimental environments, datasets used for our experiments, results of the proposed model in Section 4 and we discuss the results of the aforementioned experiments in Section 5. The challenges and future work is given in Section 6 while the collusions are drawn in Section 7.

II. Related work

Face detection, face registration, feature extraction, and recognition are the four steps in the standard smile recognition processing pipeline. Feature extraction and recognition are currently hot research topics. Traditional learning-based methods and deep learning-based methods are the two types of recognition methods. In this section, we will provide a brief overview of the most recent works in these two fields. Whitehill et al. [35] created the GENKI-4 K image database where real-world images were collected to investigate the necessary training dataset characteristics, feature learning, image registration, and algorithms for recognizing smiles in the wild. Shan et al. [27] proposed a practical method based on pixel intensity differences. For recognizing smiles in the real world, AdaBoost [6] selects and combines weak classifiers to form a strong one, but they use a trick that labels eye positions manually, which is not applicable in many situations. Liu et al. [19] improved the recognition performance in the wild by using unlabeled reference data. In [14], Jain et al. used both Gaussian derivatives and LBPs to create a robust descriptor that could effectively extract features from facial images, and these feature descriptors were used for smile recognition with (Support Vector Machine) SVM. To obtain appearance representations from faces, An et al. [31] combined the LBP and HOG. Then, to reduce the dimensionality of the features, they used principal component analysis (PCA) [33]. Finally, an extreme learning machine (ELM) was used as a recognition classifier, outperforming SVM and Linear Discriminant Analysis (LDA). Gao et al. propose a new type of feature, GSS, inspired by CSS in pedestrian detection in [7]. To improve performance, the authors combine multiple features (HOG31 + GSS + Raw pixel) with multiple classifiers (AdaBoost + Linear ELM). However, in the experiments, they remove images with ambiguity or serious lighting issues. The above-mentioned traditional learning methods all relied on handcrafted features, but designing and selecting an optimal feature representation is difficult.

In addition to the traditional methods mentioned above, a growing number of CNNs have been used in smile recognition because they can automatically extract the optimal feature from raw data based on the objective function. Glauner et al. [11] used the DISFA [24] database to train separate CNNs for smile recognition using the entire face and mouth regions. Their research, however, focused on smile recognition in a controlled laboratory setting. Bianco et al. [12] proposed a robust processing pipeline for recognizing smiles using a CNN. The pipeline consists of detecting faces using a multi-view face detector, aligning facial images using an eye-based approach, and predicting whether a face is smiling or not using an ad hoc designed CNN. Chen et al. proposed a deep convolutional network called Smile-CNN, which is a CNN that incorporates an SVM and AdaBoost, for performing smile recognition in [13]. However, the above two studies used only conventional CNN to conduct experiments without algorithm optimization, and the results obtained are far from optimal.

Zhang et al. [26] proposed training the CNN model with both recognition and verification signals. The extracted features are then fed into a two-way soft-max classifier for smile recognition in this method. Their competitive approach results in a relatively high recognition rate. However, the convergence of their new loss function was not theoretically demonstrated. A summary of the previous smile detection methods are listed below along with their accuracy in table 1. As seen from the results, the accuracy reached is still low and need to be modified.

III. The Proposed Model

A. Modified LeNet-5 CNN

Yann LeCun et al. [10] proposed Lenet-5 as one of the first

DOI: 10.21608/ERJENG.2023.194599.1156

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

pre-trained models in 1998. This architecture was used to recognize handwritten and machine-printed characters. The popularity of this model was primarily due to its simple and straightforward architecture. It is a multi-layer convolution neural network for image classification. Lenet-5 is the name given to the network because it has five layers with learnable parameters. As input, the model is given a grayscale image. Three convolution layers, two average pooling layers, and two fully connected layers with a SoftMax classifier make up the architecture. There are 60000 trainable parameters. Figure 4 depicts the LeNet-5 architecture. The number of LeNet-5 trainable parameters are inconsiderable comparing with other approaches. This situation ensures that the proposed method requires less training time. The proposed model is an improved version of the LeNet-5, in which we aimed to reduce the trainable parameters with maintaining the detection performance.

Table 1. Pervious methods in smile detection

Method	Dataset	Feature	Classifier	Accuracy
C. Shan [20]	GENKI-4K	Pixel Comparisons	AdaBoost	89.70±0.45%
Y. Zhang et al. [37]	GENKI-4K	Intensity Difference	AdaBoost	88%
H. Yadappa Navar, et al. [15]	GENKI	Machine learning mythology	Haar Classifier	82.2%
C. Shan et al. [40]	GENKI-4K	Pixel Differences	AdaBoost	85% (20 pairs of pixels)
V. Jain et al. [16]	GENKI and CohnKanade	GaborEnergyFilters	Support vector machine (SVM)	90.78% using GEF
George et al. [17]	Random Datasets	K- Nearest Neighbor (KNN)	Haar-cascade classifier	66.6%
I. K. Timotius et al. [29]	VISiO lab lip image dataset	Arithmetic means	Edge Orientation Histograms (EOH)	87.8%
C.Chang et al. [8]	Random Datasets	Mouth Region Segmentation	Mouth Corner Features (MCFs)	87.5% using single level smile measurement and 80% using multilevel smile measurement
Le An et al. [9]	MIX database (FEI, Multi-PIE, CASPEAL, CK+) AND GENKI-4K	Holistic flow-based face registration method	Extreme Learning Machine (ELM)	MIX database: 94.4% and GENKI4K:88.2%
A. Tsai et al. [22]	Random Dataset	OPENCV	Simplified Mouth Corner Features (MCFs)	72.7%
K. Zhang et al. [38]	ChaLearn16	VGGFaces and fine tune model	Two convolutional neural networks (CNNs) GNet and SNet.	88.79%
R. Ranjan et al. [39]	ChaLearn Faces of the World	Multi-task Learning (CNN)	Convolutional neural networks (CNNs) GNet and SNet.	Faces of the World: 90.83%

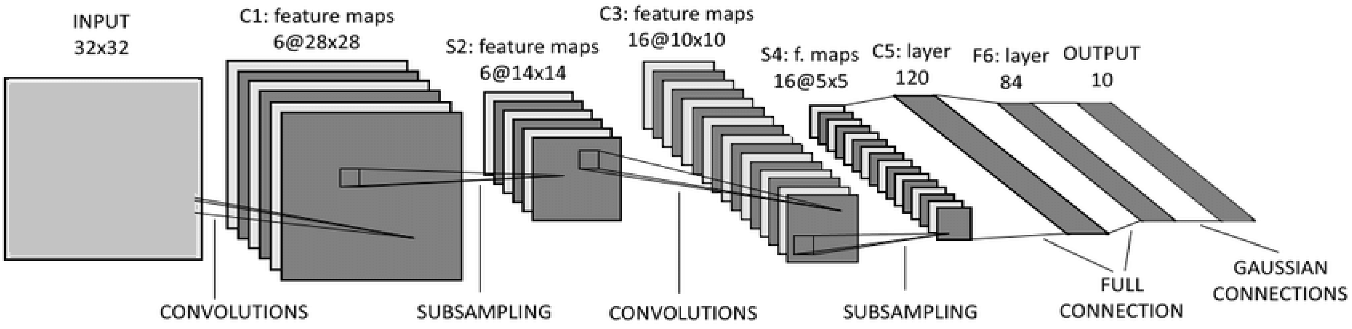

Figure 4. Architecture of LeNet-5 CNN model

DOI: 10.21608/ERJENG.2023.194599.1156

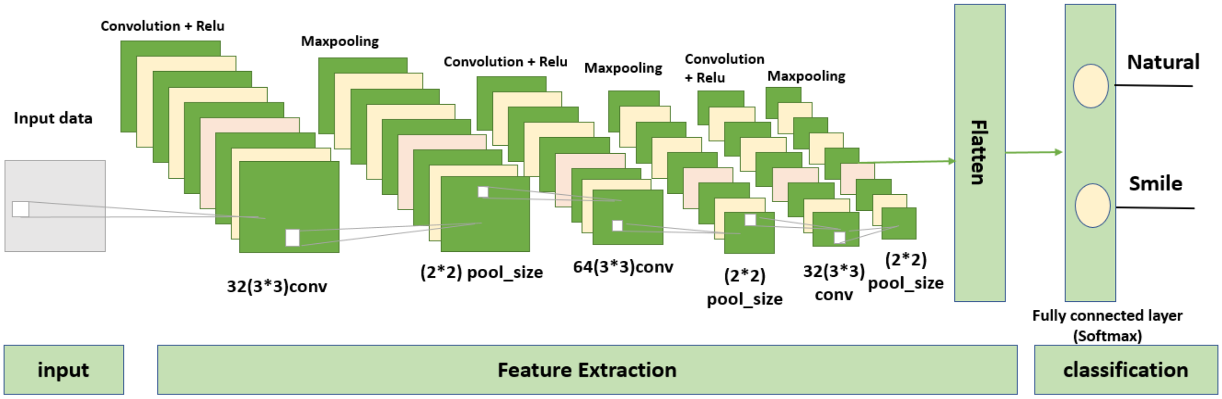

Figure Error! No text of specified style in document.**5. The proposed Model (MLeNet-5)**

In this paper, we suggest an improved version of the LeNet-5 known as the Modified LeNet-5 (MLeNet-5), which is depicted in Figure 5 along with the corresponding output shape for each layer. The input image for this model must be preprocessed; it is normalized to 180 x192, then reduced to 60 x64 and converted to a greyscale image, that further maps to a fully connected layer with two outputs indicating a natural image and a smiling image. The first part of this model consists of convolutional layers with 32 kernels of size 3 x3. Following that, a convolution layer filter with 64 kernels of size 3 x 3 is applied, and finally, to reduce the spatial size of the feature map, the model employs a convolutional layer with 32 kernels of size 3 x 3. Each convolutional layer is followed by a max pooling layer. Following the convolutional layers, the extracted features are reduced by passing them through a fully connected layer, resulting in a two-dimensional one-dimensional array representing a natural image and a smiling image.

The main modification of the proposed MLeNet-5 over the LeNet-5 are:

1. Adding one more layer of convolution network followed by a max pooling layer.
2. Eliminating one fully connected layer.

This modification improves the accuracy of smile detection compared to LeNet-5 and in the same time decreases the number of parameters compared to LeNet-5 (as given in table 3) and table 4 illustrated the comparation between MLeNET-5 and other methods in number of trainable parameters. With fewer parameters, training and inference time are reduced while still achieving satisfied accuracy results for the smile classification task. The dominance of small filters significantly reduces the number of parameters in the MLeNet-5 network which making it much faster than previous CNN generations. Table 2 shows the proposed MLeNet-5 CNN model configuration. In our model, we use a nonlinearity function in the activation layer between the rectified linear unit and the convolutional layers (ReLU). Finally, the root mean squared propagation (RMSProp) is employed to train the proposed MLeNet-5 model.

A.1 Data Preprocessing

We preprocess the face images, before detecting smiles, using two methods:

1) All extracted face images are normalized to the same size of about 180*192 then reduced to 60*64. As a result, the detection rate can be investigated to see if it is related to image resolution.
2) All face images are converted to grayscale to reduce computational complexity.

Table 2. Configuration of the proposed Smile Detection MLeNet-5 model

NO	Layer	Configuration
Smile Detection-CNN Configuration		
1	Input	60 x 64 grayscale image
2	Convolution	3 x 3 — 32 kernel
3	Maxpooling	2 x 2
4	Convolution	3 x 3 — 64 kernel
5	Maxpooling	2 x 2
6	Convolution	3 x 3 — 32 kernel
7	Maxpooling	2 x 2
8	Fully connected	Length: 2

Table 3. Number of parameters in LeNET-5 and MLeNet-5

Method	No. of parameters
LeNet-5	1,420,634
MLeNet-5	**98,914**

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

A.2 Feature extraction

The function of these layers is to extract features in a hierarchy structure. The hidden nodes of each layer are referred to as feature maps or output maps. Convolutional layers obtain features from input images by sliding a series of learnable filters or kernels across the image. In our work, we employ the recently popular rectified linear unit (ReLU) nonlinearity function (f = max (0, x)) [28], which has been demonstrated to fit better than the sigmoid or hyperbolic tangent functions. A pooling layer follows each convolutional layer, which is used to reduce the spatial size of the representation and control over-fitting. To generate a single output from each block, the pooling layer subsamples small square blocks (s x s) from the convolutional layer. The most common pooling method is average or maximum pooling. In our proposed model (MLeNeT-5), we use three convolution layers and one fully connected layer at the end. Following each convolutional layer is a max pooling layer. The extracted features are reduced after the convolutional layers by passing them through a fully connected layer, resulting in a two-dimensional one-dimensional array representing a natural image and a smiling image.

A.3 Classification

After passing this vector through the final fully connected layer, we will end up with a 2-dimensional vector expressing the scores for the two labels smile and non-smile. This vector is then activated using the SoftMax function. Finally, we have a two-element vector denoting the probability distribution across two classes: smile and non-smile. As the final answer for the input image, the label with the highest probability is chosen. After collecting all of the characteristics, the data is matched from various datasets available, such as GENKI-4K and SMILEsmilesD, and after configuring, it gives the output of whether is smiling or non-smiling.

IV. Experimental Results

In this section, we present the experimental results of the proposed MLeNet-5 based smile detection and its comparison to state-of-the-art methods. But first, we will define the databases we will work on. The model was implemented using Python 3.9.7 (‘base’: conda) involving the Keras framework running Visual Studio Code using a GPU of Nvidia Quadro K5100M on processor: (Intel(R) Core (TM) i7-4930 MX CPU @ 3.00GHz) with 16 GB RAM.

A. The databases

We used two different databases which have different environments and challenges: GENKI-4K [32] and SMILEsmilesD [43]. Their description is given below.

- **SMILEsmilesD database**

SMILEsmilesD [43] is Multisource dataset because it combines more than one data set such as: LFW and Genki. The samples are collected from the Internet. The SMILEsmilesD can be used in the learning transfer technique which facilitating (or even fully automate) the labelling of new images, thereby extending the original dataset with new images. In addition, the images are tightly cropped around the face that making the training easier. The input images can directly be used without any additional processing. This database contains 13165 images of smiling and non-smiling faces. 10532 images are used for training and 2633 for testing. Figure 2 and Figure 3 show samples of the face images available in this database.

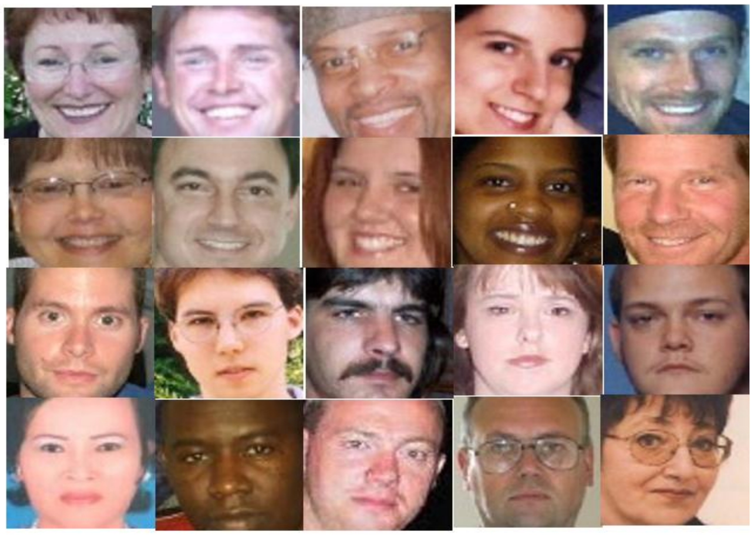

Figure 6. Examples of face images from GENKI-4 K. The first two rows contain smile face images, while the remaining rows contain natural face images.

DOI: 10.21608/ERJENG.2023.194599.1156

- **GENKI-4K database**

In this model, we used GENKI-4K [32] as first database. The GENKI database, which is a growing image database, contains a wide range of illumination conditions, personal identity, poses, geographical locations, camera models, and so on. The GENKI-4K database, a subset of the GENKI database, consists of 4000 real-life face images, downloaded from publicly available Internet repositories for smiling and non-smiling faces, where 2800 instances are used for training and 1200 instances are used for testing. The first two rows of images in Figure 6 represents smile expressions, while the second two row represents non-smile expressions.

To evaluate our model, we have used various performance metrics namely; Accuracy, F1_scorecore, Precision, and Recall which are calculated as below [44]:

$$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{F}_1\text{\_score} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where true negative (TN), true positive (TP), false negative (FN), and false-positive (FP).

B. Results and Comparisons

To train the proposed model, the images dataset are split into 80% of training and 20% for testing. The model is trained for 10 epochs using the root mean squared propagation (RMSProp) optimizer. The results will be presented in two subsections, one for each database.

B.1 Results and comparisons on SMILEsmilesD database

The training loss and accuracy of our MLeNet-5 model for smile detection in 10 epochs on SMILEsmilesD dataset [43] is depicted in Figure 7. The classification accuracy attained by the proposed MLeNet-5 is 92.8%, which is more than the accuracy of LeNet-5 (90%) when applied to SMILEsmilesD database [43] as illustrated table 5 attaining the best accuracy of 92.8 ± 0.66Meanwhile, the rate of smile recognition was increased.

When we increase the number of epochs to 50, the accuracy was increased to 92.8%. So, increasing the number of epochs enhances the performance of the proposed model.

B.2 Experimental results and comparisons on GENKI-4 K database

Experimental results on the GENKI-4K database with other methods for smile detection are listed in Table 3. The recognition rate obtained by our method is listed at the last row in Table 6. The table shows that the proposed method does not achieve the highest accuracy in the GENKI-4K database, because the images in the GENKI-4K dataset were acquired from the internet in a variety of real-world contexts (unlike other face data sets, which are commonly acquired in the same scene), making detection more complicated. Moreover, some of the images in the database are unclear (i.e. not clear whether the person is smiling or not). Therefore, we need to improve the accuracy of our model based on GENKI-4K database in the future work by introducing solutions for these problems. Most of the related work algorithms depend on a complex architecture of CNN with high depth of layers. This leads to high performance but with high dimensional features. In addition, their complex architecture increases the computational overhead in terms of the used memory and processing resources. Sometimes, it influences the efficiency of dealing with pattern recognition problems. The key aim of recent researches is to reduce the feature dimension using the best mapping function with maintaining the nature of the original data.

Table 4. Number of parameters in MLeNet-5 and other methods

Method	No. of parameters
[45]	695,472
Face-CNN [2]	733,952
Mouth-CNN [2]	242,756
BKNet [52]	2,418,722
Our method	**98,914**

Table 5. Accuracy of LeNet-5 and MLeNet-5 on SMILEsmilesD database

Method	Feature	Classifier	Accuracy
LeNet-5	LeNet-5	SoftMax	90%
Proposed MLeNet-5 method	**MLeNet-5**	**SoftMax**	**92.8 ±0.66%**

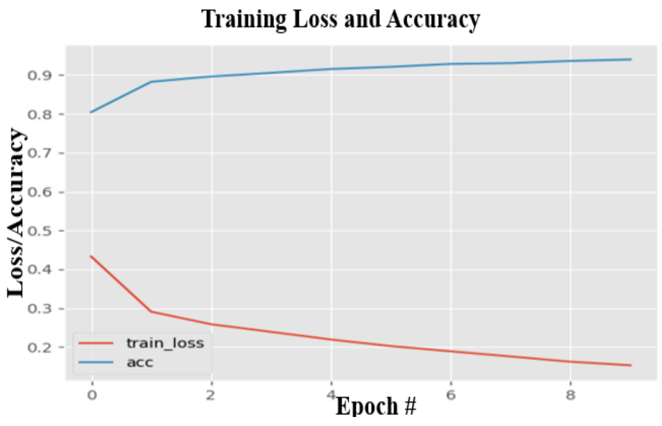

Figure 7. Training loss and accuracy of our MLeNet-5 model for smile detection in 10 epochs on SMILEsmilesD dataset [43]

DOI: 10.21608/ERJENG.2023.194599.1156

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

The proposed MLeNet-5 model achieved a classification accuracy of 84.8% which was more than accuracy of LeNet-5 applying to GENKI-4K as illustrated in Figure 8. According to table 6, we compare our proposed model smile detection results to other method of smile detection like RPN + VGG-16, and RPN + Resnet50.

We also compare the results with a smile detector and classifier based on Extreme Learning Machine [50] (ELM) with pixel values (rather than extracted handcrafted features) and RPN + BKNet (based on Faster R-CNN) [51, Finally, we include reference results from using Support Vector Machine (SVM) with pixel values [50] as a classifier for non-neural network smile detection.

According to the confusion matrix of binary classifier with smile and non-smile based on SMILEsmilesD in Figure 9, the label 'smile' refers to smile face images, and the label 'non-smile' refers to non-smile face images. A total of 2633 images were tested. The classifier predicted as 'positive' 1919 times and as 'negative' 714 times out of the 2633 images (regardless of whether the predictions were correct or not). In actual fact, 1898 of the images in the test have a smile face, while 735 have a non-smile face.

Confusion matrix does not only give your insight into the value of classifier' errors, but also the error types. In unbalanced dataset, this helps overcoming the limitation of the classification accuracy which can be biased for one classifier output. The accuracy represents the number of correctly classified output over the total number of testing samples regardless the classifier types. The confusion matrix gives a comprehensive analysis of the classifier decision in all cases whether correct or incorrect. The confusion matrix of binary classifier with smile and non-smile based on GENKI-4k is shown in figure 10. This means that our proposed model achieves the result more closed to actual fact.

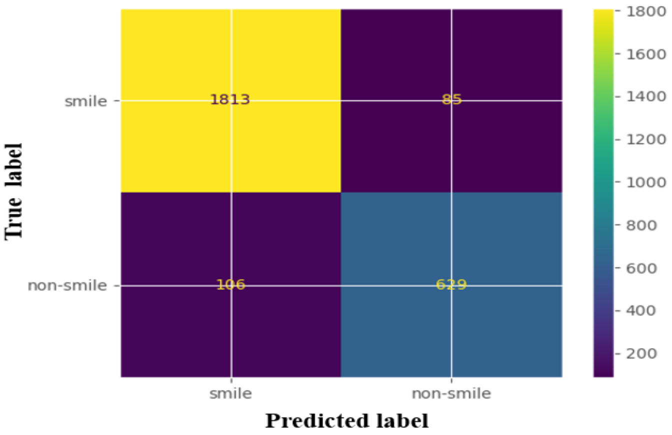

Figure 8. Confusion matrix of binary classifier with smile and non-smile based on SMILEsmilesD

Table 6. Experimental results on the GENKI-4 K database [32]

Method	Feature	Classifier	Accuracy
An et al [31]	HOG	ELM	88.5%
Zhang et al. [41]	MFs	AdaBoost	89.21%
Shan et al [42]	Pixel comparison	AdaBoost	89.7 ± 0.45%
Chen et al [9]	CNN	AdaBoost	91.8 ± 0.95%
Liu et al. [46]	HOG	SVM	92.29±0.81%
Jain et al. [47]	Guassian	SVM	92.97%
Kahou et al. [48]	LBP	SVM	93.2±0.92%
[45]	CNN-Basic CNN-2Loss	SoftMax SoftMax	93.6±0.47% 94.6±0.29%
[49]	HOG31 + GSS + Raw pixel HOG31 + GSS + Raw pixel Linear HOG31 + GSS + Raw pixel Linear HOG31 + GSS + Raw pixel HOG31 + GSS + Raw pixel	AdaBoost SVM ELM Adaboost + Linear SVM Adaboost + Linear ELM	92.51±0.40 94.28±0.60 94.21±0.35 94.56±0. 62 94.61±0.53
Proposed MLeNet-5 method	**MLeNet-5**	**SoftMax**	**87.8%**

DOI: 10.21608/ERJENG.2023.194599.1156

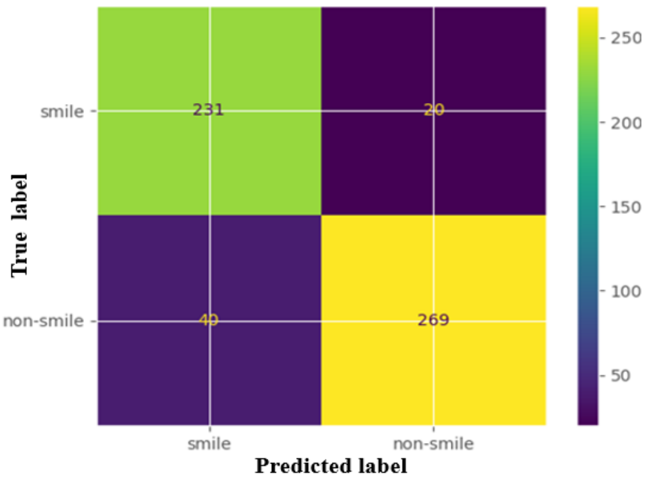

Figure 9. Confusion matrix of binary classifier with smile and non-smile based on GENKI-4 K

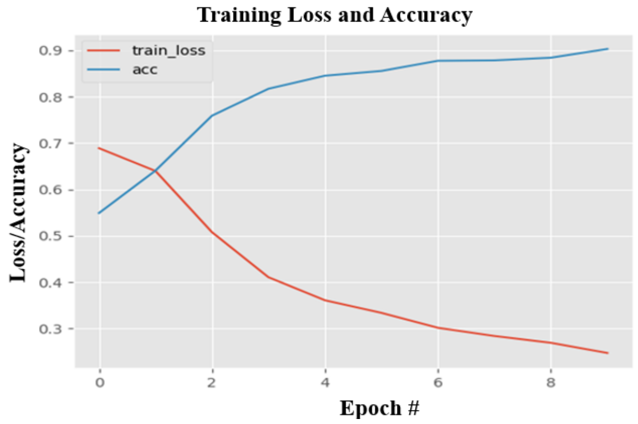

Figure 10. Training loss and accuracy of our CNN- smile detection with 10 epochs on GENKI-4k dataset

The results of precision, recall, F1_score and accuracy based on SMILEsmilesD and GENKI-4 K databases according to MLeNet-5 model illustrated below in table 7. Table 8 compares our propose MLent-5 model to existing smile detection models GENKI-4 K database. As seen from the results, our model accuracy is still beyond the existing models, however, it has lower number of parameters which means less time for training and response and less memory size which facilitate using it in real-life applications on smaller devices.

Table 7. Precision, Recall, F1-score and Accuracy for SMILEsmilesD and GENKI-4 K databases

Database	Precision	Recall	F1-Score	Accuracy
SMILEsmilesD	0.94	0.96	0.95	92.7%
GENKI-4 K	0.85	0.92	0.89	89%

Table 8. Comparison of other method in smile detection for GENKI-4 K database

Method	Accuracy
RPN + VGG-16	92.5%
RPN + ResNet50	93.2%
ELM (pixel)	79.3%
SVM	80.6%
RPN + BKNet	84.5%
Our MLent-5 model	**87.8%**

V. CHALLENGES AND FUTURE WORK

The scientific community in smile detection has turned its attention to natural smile expression identification. Identifying natural and fake smiles in video and images constitutes one of the most difficult challenges. People will make up fake expressions if they are aware that they are being photographed or videotaped, according to Sebe et al [36]. They installed a hidden camera to capture photos and videos with natural expressions to address this issue. After that is resolved, the next challenge is to improve natural expression in various lighting and obstacle situations. The hidden camera planting techniques will not work in this situation. When people wear specs and mufflers (to provide an eye and mouth obstruction), it is difficult to detect their expressions. Another significant challenge is labelling available data. Unlabeled data is easily obtained in large quantities. However, labelling that unlabeled data is a tedious process with a high risk of error.

The study of obelic faces with micro expressions could be a future extension of smile detection research. At the moment, there are only a few methods for dealing with obelic face with micro expressions. Another significant challenge is detecting a smile in people who have lost their spontaneous expressions due to medical issues like; facial weakness, facial paralysis, Asperger syndrome, depressive disorders, hepatolenticular degeneration, depression, Parkinson's disease, autistic disorder, major depressive disorder, Wilson's Disease, scleroderma, and Bell's palsy.

VI. CONCLUSIONS

The research on smile detection is still in its early stages, and a more powerful and stable smile detection system is required. Deep learning, as opposed to previous research that performed feature extraction and classification separately, can effectively combine the two steps into a single trainable model. In this paper, we present an efficient deep learning-based approach to

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

improve smile detection system and also decrease the number of parameters while maintaining the robustness and effectiveness of the results. In our study, we build a deep convolutional network called Modified Lenet-5 (MLenet-5) to detect smiles and deal with "big data". The outputs from the final hidden layer serve as the learned features, which are used to train the SoftMax classifiers for comparison, in part because a deep convolutional network can extract features hierarchically and higher-level representations are more abstract and condensed. The experimental results demonstrate the impressive discriminative power of these features; MLent-5 outperforms GENKI-4K on the public SMILEsmilesD database, which we aim to enhance in future work. Research on smile recognition is still currently in progress in hopes of improving precision and tackle various kinds of detection difficulties (lighting, pose, oblique face, etc.).

Funding: This research has not been conducted under any fund.

Conflicts of Interest: The authors declare that there is no conflict of interest.

REFERENCES

1. Ramakrishna, A. G. (2021). Review on Smile Detection. International Journal of Scientific Research in Computer Science Engineering and Information Technology.
2. Chen, J. C. (2019). Novel multi-convolutional neural network fusion approach. Springer Science+ Business Media, LLC, part of Springer Nature.
3. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: A comprehensive study. Image Vis Comput 27(6):803–816.
4. Sikka K, Wu T, Susskind J, Bartlett M (2012) Exploring bag of words architectures in the facial expression domain. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, pp. 250–259.
5. Dahmane M, Meunier J (2011) Emotion recognition using dynamic grid-based HoG features. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pp. 884–888.
6. Freund Y, Schapire R E (1995) A desicion-theoretic generalization of on-line learning and an application to boosting. In: European conference on computational learning theory, pp. 23–37.
7. Gao Y, Liu H, Wu P, Wang C (2016) A new descriptor of gradients self-similarity for smile detection in unconstrained scenarios. Neurocomputing 174:1077–1086.
8. C. Chang et al., "Multi-level smile intensity measuring based on mouth corner features for happiness detection," 2014 International Conference on Orange Technologies, Xian, 2014, pp. 181-184.
9. L. An, S. Yang, and B. Bhanu, “Efficient smile detection by Extreme Learning Machine,” Neurocomputing, vol. 149, pp. 354–363, 2015.
10. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. (1998). "Gradient-based learning applied to document recognition" (PDF). Proceedings of the IEEE. 86 (11): 2278–2324. doi:10.1109/5.726791.
11. Glauner P O (2015) Deep convolutional neural networks for smile recognition. arXiv:1508.06535
12. Bianco S, Celona L, Schettini R (2016) Robust smile detection using convolutional neural networks. Journal of Electronic Imaging 25(6):063002–063002.
13. Chen J, Ou Q, Chi Z, Fu H (2017) Smile detection in the wild with deep convolutional neural networks.Mach Vis Appl 28(1–2):173–183.
14. Jain V, Crowley J L, Lux A (2014) Local binary patterns calculated over Gaussian derivative images. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, pp. 3987–3992.
15. H. Yadappanavar and S. S. S, “Machine Learning Approach for Smile Detection in Real Time Images,” International Journal of Image Processing and Vision Sciences (IJIPVS), vol. 1, no. 1, 2012.
16. V. Jain and James L, “Smile Detection Using Multi-scale Gaussian Derivatives,” 12th WSEAS International Conference on Signal Processing, Robotics and Automation, 2013.
17. T. George, S. P. Potty and S. Jose, "Smile detection from still images using KNN algorithm," 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, 2014, pp. 461-465.
18. LeCun Y, Boser B, Denker JS et al (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–551.
19. Liu M, Li S, Shan S, Chen X (2012) Enhancing expression recognition in the wild with unlabeled reference data. In: Asian Conference on Computer Vision, pp. 577–588.
20. C. Shan, "An efficient approach to smile detection," Face and Gesture2011, Santa Barbara, CA, 2011, pp. 759-764.
21. Liu C, Xu W, Wu Q et al (2016) Learning motion and content-dependent features with convolutions for action recognition. Multimedia Tools and Applications 75(21):13023–13039. https://doi.org/10.1007/s11042-015-2550-4.
22. A. Tsai, T. Lin, T. Kuan, K. Bharanitharan, J. Chang and J. Wang, "An efficient smile and frown detection algorithm," 2015 International Conference on Orange Technologies (ICOT), Hong Kong, 2015, pp. 139-143.
23. Lucey P, Cohn JF, Matthews I et al (2011) Automatically detecting pain in video through facial action units.IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 41(3):664–674.
24. Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF (2013) Disfa: A spontaneous facial action intensity database. IEEE Trans Affect Comput 4(2):151–160.
25. Qu T, Zhang Q, Sun S (2017) Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. Multimedia Tools and Applications 76(20):21651–21663. https://doi.org/10.1007/s11042-016-4043-5.
26. Zhang K, Huang Y, Wu H, Wang L (2015) Facial smile detection based on deep learning features. In:Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on. IEEE, pp: 534–538.
27. Shan C (2012) Smile detection by boosting pixel differences. IEEE Trans Image Process 21(1):431–436.
28. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011).
29. I. K. Timotius and I. Setyawan, "Evaluation of Edge Orientation Histograms in smile detection," 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2014, pp. 1-5.
30. Singh R, Om H (2017) Newborn face recognition using deep convolutional neural network. Multimedia Tools and Applications:1–11. https://doi.org/10.1007/s11042-016-4342-x.
31. An L, Yang S, Bhanu B (2015) Efficient smile detection by extreme learning machine. Neurocomputing 149:354–363.
32. The MPLab GENKI-4K Database (2018). http://mplab.ucsd.edu/.
33. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cogn Neurosci 3(1):71–86.
34. Valstar MF, Pantic M (2012) Fully automatic recognition of the temporal phases of facial actions. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42(1):28–43.

DOI: 10.21608/ERJENG.2023.194599.1156

Journal of Engineering Research (ERJ)
Vol. 7 – No. 2, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

35. Whitehill J, Littlewort G, Fasel I, Bartlett M, Movellan J (2009) Toward practical smile detection. IEEE Trans Pattern Anal Mach Intell 31(11):2106–2111.
36. Nicu Sebe, M. S. Lew, Ira Cohen, Yafei Sun, T. Grevers, T. S. Huang, "Authentic Facial Expression Analysis," Image and Vision Computing, Vol. 25, pp. 1856-1863, 2007.
37. Zhang Y, Zhou L, Sun T (2012) A novel approach to detect smile expression. In: Machine Learning and Applications (ICMLA), 2012 11th International Conference on. IEEE 1:482–487.
38. K. Zhang, L. Tan, Z. Li, and Y. Qiao, "Gender and Smile Classification Using Deep Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),2016.
39. R. Ranjan, S. Sankaranarayanan, C. D. Castillo and R. Chellappa, "An All-In-One Convolutional Neural Network for Face Analysis," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, 2017, pp. 17-24.
40. C. Shan, "Smile detection by boosting pixel differences," in IEEEs Transactions on Image Processing, vol. 21, no. 1, pp. 431-436, Jan.2012.
41. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456
42. Cui D, Huang G B, Liu T (2016) Smile detection using Pair-wise Distance Vector and Extreme Learning Machine. In: Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE, pp. 2298–2305
43. The SMILEsmilesD database: https://github.com/hromi/SMILEsmileD.git
44. Mahmoud M Fahmy Amin. "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial". Journal of Engineering Research, vol. 6, no. 5, 2022.
45. K. Zhang, Y. Huang, H. Wu and L. Wang, "Facial smile detection based on deep learning features," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, Malaysia, 2015, pp. 534-538, doi: 10.1109/ACPR.2015.7486560.
46. M. Liu, S. Li, S. Shan, and X. Chen. Enhancing expression recognition in the wild with unlabeled reference data. In Computer Vision–ACCV 2012. 2013
47. V. Jain and J. Crowley. Smile detection using multi-scale gaussian derivatives. In 12th WSEAS International Conference on Signal Processing, Robotics and Automation, 2013. 2, 4
48. S. E. Kahou, P. Froumenty, and C. Pal. Facial expression analysis based on high dimensional binary features. In Computer Vision-ECCV 2014 Workshops, 2014.
49. Yuan Gao, Hong Liu, Pingping Wu, and Can Wang, "A new descriptor of gradients self-similarity for smile detection in unconstrained scenarios," Neurocomputing, vol. 174, pp. 1077–1086, 2016.
50. L. An, S. Yang, and B. Bhanu, "Efficient smile detection by extreme learning machine," Neurocomputing, vol. 149, pp. 354–363, 2015
51. Nguyen, C. C., Tran, G. S., Nghiem, T. P., Doan, N. Q., Gratadour, D., Burie, J. C., & Luong, C. M. (2018, April). Towards real-time smile detection based on faster region convolutional neural network. In *2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR)* (pp. 1-6). IEEE
52. Dinh Viet Sang; Le Tran Bao Cuong; Do Phan Thuan *Facial smile detection using convolutional neural networks* In: 2017 9th International Conference on Knowledge and Systems Engineering (KSE)

.

DOI: 10.21608/ERJENG.2023.194599.1156