

# FACIAL EXPRESSION RECOGNITION: A SURVEY AND ITS APPLICATIONS

Vu-Tuan Dang<sup>\*1</sup>, Hong-Quan Do<sup>\*1</sup>, Viet-Vu Vu<sup>\*</sup>, Byeongnam Yoon<sup>\*\*</sup>

<sup>\*</sup>VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam

<sup>\*\*</sup>Computer Science Department, Kyonggi University, Suwon Korea

vuvietvietnam@gmail.com, quandh@vnu.edu.vn, vuvietvu@vnu.edu.vn, tomayoon@kgu.ac.kr

**Abstract**— Automatic facial expression recognition is an important component for efficient human-computer interaction system, and over the past decades, it has become a highly active research area. Numerous algorithms have been proposed in the literature to cope with the problem of face expression recognition (FER). General speaking, current existing FER methods can be categorized into two main groups, i.e., traditional machine learning-based approaches and deep learning-based approaches. Different from other surveys, in this study, we aim to not only comprehensively highlight the differences and similarities of the two approaches above, but also the new trend of hybrid and ensemble learning in FER systems by providing a general framework for each type and review the possible technologies that can be employed in its components. We conduct more specific and detailed competitive performances and experimental comparisons of researches from 2014 to 2020 on widely used datasets. We then extend our survey to our current application scenarios in Vietnam.

**Keywords**— Facial Expression Recognition, FER applications, Machine learning based, Deep learning based, Hybrid FER systems, Ensemble FER systems, Vietnam E-Government

## I. INTRODUCTION

Expression critically influences all aspects of our lives, from how we live, work, and to the decisions that we make. Facial expression plays the major role in non-verbal communication, according to Mehrabian [1], 55% communicative cues can be judged by facial expression. Back to the year 1872, Darwin published “The Expression of the Emotions in Man and Animals”, in which he argued that all humans, and even other animals, show emotion through remarkably similar behaviours. Darwin treated the emotions as separate discrete entities, or modules, such as anger, fear, disgust, etc. Many kinds of research—neuroscience, perception, and cross-cultural evidence—show that Darwin's conceptualization of emotions as separate discrete entities is correct.

In 1971, Paul Ekman et al.'s facial expression work [2] gave the correlation between a person's emotional state & psychological state, which described the FACS. It is based on the muscular contractions which produce our facial expressions. Next in 1992, he defined 7 face expressions [3]:

Happy, Neutral, Angry, Disgust, Fear, Sad, and Surprise. They are named as 7 universal emotions and are used by most researchers (Figure 1).

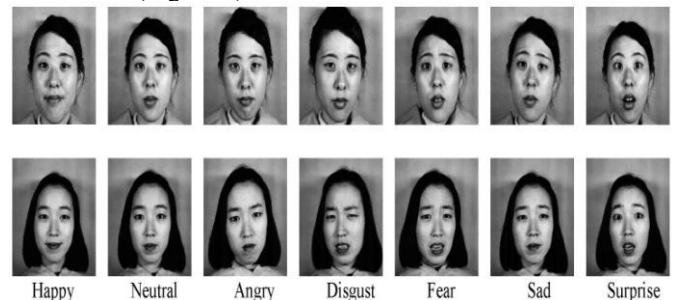


Figure 1. Seven basic expressions (taken from the JAFFE dataset [4])

Facial expression recognition (FER) provides the machine a way of sensing emotions that can be considered one of the mostly used artificial intelligence and pattern analysis applications. This classification problem is handled either by machine learning or deep learning methods. Unlike traditional machine learning and computer vision approaches where features are defined by hand, a deep learning method learns to extract the features itself directly from the training database using iterative algorithms like gradient descent (Figure 2).

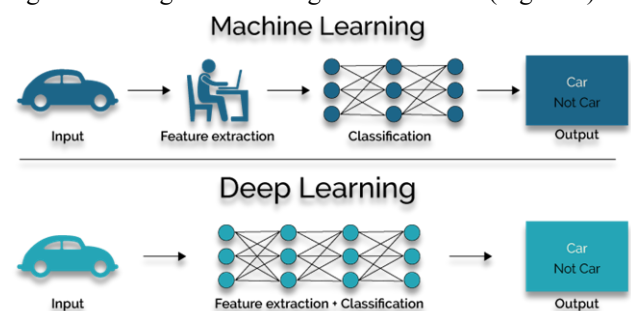


Figure 2. Machine Learning vs. Deep Learning

Several surveys on automatic expression analysis have been published in recent years. However, those study focus either on traditional machine learning-based methods [5] [6], or recently deep learning-based methods [7] [8] [9]. Different from these surveys, in this work, we conduct more specific

<sup>1</sup> Vu-Tuan Dang and Hong-Quan Do are co-first authors.  
Corresponding author: Hong-Quan Do (quandh@vnu.edu.vn).

and detailed research from 2014 to 2020 with an aim to not only comprehensively highlight their differences and similarities between the two types, but also the new trend of hybrid and ensemble learning in FER systems. We then extend our survey to current application scenarios in Vietnam.

The rest of this article is organized as follows. Firstly, we introduce some of the most common public facial expression recognition datasets in Section 2. Next, we discuss various traditional machine learning-based FER systems in Section 3, deep learning-based in Section 4, and lastly the hybrid and ensemble FER systems in Section 5. After that, Section 6 reports the comparison results and our discussions. Section 7 describes our current applications of FER in Vietnam, and finally we draw a conclusion in Section 8.

## II. FACIAL EXPRESSION RECOGNITION DATASETS

In this section, we discuss the publicly available datasets that are widely used in our reviewed papers. Table 1 provides an overview of these datasets, including the number of image or video samples, number of subjects, collection environment, and expression distribution. Approach to collect FER-related data was mostly through the images captured in the laboratory, such as JAFFE [4] and CK+ [10], in which volunteers make corresponding expressions under particular instructions. However, since 2013, emotion recognition competitions have collected large-scale and unconstrained datasets, for example, FER2013 [11] queried automatically by the Google image search API. This implicitly promotes the transition of FER from lab-controlled to real-world scenarios.

## III. TRADITIONAL MACHINE LEARNING-BASED APPROACH

In a traditional machine learning-based approach, facial expression recognition can be divided into four major steps: (1) face acquisition stage to automatically find the face region for the input images; (2) normalization of intensity, uniform size and shape; (3) facial data extraction and representation-extracting and representing the information about the encountered facial expression in an automatic way; and (4) facial expression recognition step that classifies the features extracted in the appropriate expressions. Figure 3 illustrates the block diagram of a traditional FER system.

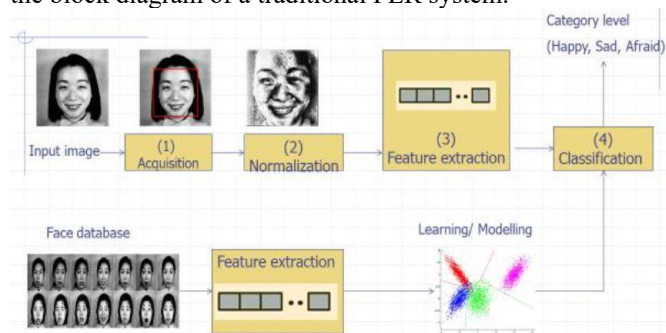


Figure 3. Block diagram of a traditional FER system

Face Acquisitions a process of localizing and extracting the face region from the background. The Viola-Jones algorithm is a widely used mechanism for facial detection. The method

was devised by Viola and Jones, 2001 [12] that allows the detection of image features in real-time. The Viola and Jones algorithm uses Haar-like features to detect faces (Figure 4). Given an image, the algorithm looks at many smaller sub-regions and tries to find a face by looking for specific features in each subregion. It needs to check many different positions and scales because an image can contain many faces of various sizes. Viola-Jones was designed for frontal faces, so it is able to detect frontal the best rather than faces looking sideways, upwards or downwards.

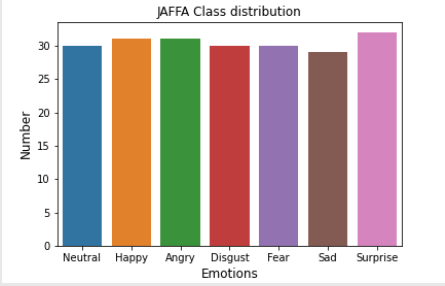
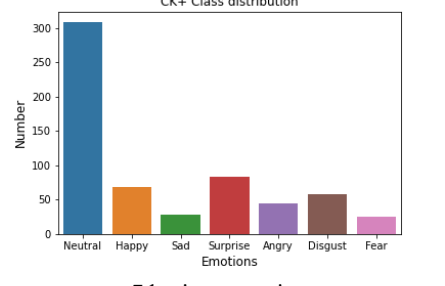
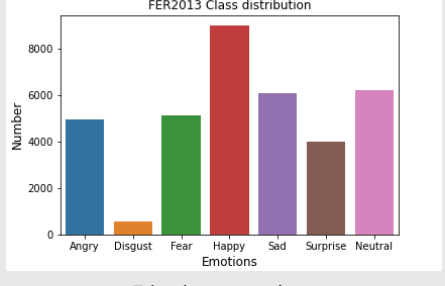


Figure 4. Detect face region in the image with Viola and Jones algorithm using Haar-like features

After the presence of a face has been detected in the observed scene, normalization can be carried out that aims to obtain images in a normalized intensity, uniform size, and shape. Thereby, the performance of the FER system can be improved. This phase includes different tasks such as orientation normalization, image scaling, contrast adjustment, and additional enhancement to improve the expression frames.

The next step is to extract the information about the encountered facial expression in an automatic way. Feature extraction is finding and depicting of positive features of concern within an image. This process mostly requires domain level expertise, that is reason why we call these obtained features by this way are “handcrafted” features. The facial features can be extracted from the entire face or particular facial regions, and categorized into two types, i.e. geometric-based (or shape-based) and appearance-based (Figure 5). Geometric-based representation considers the shape information (in-transient features, e.g., facial points or locations of eyebrows, eyes, the mouth, the nose) explicitly and ignores the texture. Notably, this representation is typically vulnerable to illumination variation. Most geometric feature-based approaches use the active appearance model (AAM) or its variations [13] [14], to track a dense set of facial points. Appearance-based representation, on the other hand, uses the intensity values or the pixels that refer to the textural changes, such as wrinkles and furrows. The techniques adopted as an Appearance-based include Local binary pattern (LBP) [15], scale invariant feature transform (SIFT) [16], histograms of oriented gradients (HOG) [15] [17] [18], Gabor wavelet representation [19] [20] - originally introduced by Dennis Gabor, 1946 [21], and principal component analysis (PCA) [22] [23] - called eigenfaces in [24] and widely used for dimensionality reduction and recorded a great performance in face expression recognition [15] [25]. In addition, researchers have also developed systems for facial expression recognition by utilizing the advantages of both geometric-based and appearance-based features [26] [27].

**TABLE 1.** AN OVERVIEW OF THE FACIAL EXPRESSION DATASETS

Dataset	Samples	Resolution	Subject	Collection Env.	Expression distribution
<b>JAFFE (1998)</b>	213 images (Frontal and 30-degree images)	256x256 pixel grayscale	10	Lab	 <p>7 basic expressions</p>
<b>CK+ (2000)</b>	593 image sequences	640x490 for grayscale, 640x480 for 24-bit colour	123	Lab	 <p>7 basic expressions</p>
<b>FER 2013 (2013)</b>	35.887 images	48x48 pixel grayscale	N/A	Web	 <p>7 basic expressions</p>

Finally, these handcrafted features are fed as an input to the classification. Many classification methods have been employed in the FER systems, such as Linear Discriminant Analysis (LDA) [15], Support vector machine (SVM) [29], Random forest [30], AdaBoost [31], Decision tree [32], Naïve Bayes [33], and Extreme Learning Machine (ELM) that is a type of single hidden-layer feed-forward neural networks (SLFN) [34].



**Figure 5.** (a) Results of Geometric Feature Extraction. (b) Results of Appearance Feature Extraction [28]

#### IV. DEEP LEARNING-BASED FER SYSTEM

One of the key challenges in computer vision is to deal with the variance of data in the real world. It can be said that Deep learning networks are designed to cope with this problem. Since the beginning of the 21<sup>st</sup> century, deep learning is becoming more and more popular as it can highly reduce the reliance on image pre-processing and feature extraction. Unlike a traditional machine learning-based method, a deep learning method makes all these hand-crafted features obsolete, as it defines the features space itself. In terms of FER, a deep learning-based approach may consist of only three major steps as shown in Figure 6.

Deep and large networks have exhibited impressive results when there are large training data sets and computation resources, such as many CPU cores and/or GPU. Several well-known deep learning networks have been existed. Due to the space limitation, we introduce the principles of convolutional neural networks, deep belief networks, and deep auto encoders.



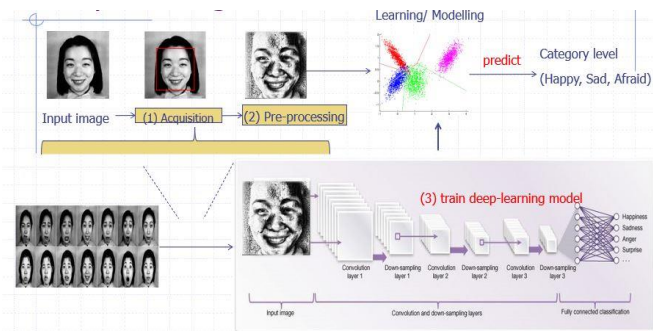


Figure 6. Block diagram of a deep learning-based FER system

### A. Convolutional Neural Network (CNN)

The success of convolutional neural networks (CNNs) [35] in tasks like image classification [36] has been extended to the problem of facial expression recognition. CNN is usually combined with feed-forward neural network classifier which makes the model end-to-end trainable on the dataset. It consists of three types of heterogeneous layers: convolutional layers, pooling layers, and fully connected layers. A visualization of a typical CNN architecture is shown in figure 7. The first part consists of Convolutional and max-pooling layers which act as the feature extractor. The second part consists of the fully connected layer which performs non-linear transformations of the extracted features and acts as the classifier.

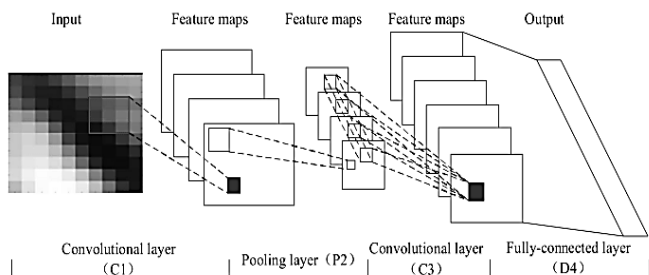


Figure 7. An illustration of a typical CNN architecture showing how they are connected from input to output

Based on the foundation architecture of the CNN, some variations have also been applied to solve the problem of Facial Expression Recognition, including Deep convolutional neural network (DCNN) [37] – a feedforward neural network that applies back propagation algorithms to adjust the parameters (weights and biases), AlexNet [38], Inception [39] [40], Residual Neural Network (Resnet) [40], and 2-Channel CNN [41] where one channel is a standard CNN network and another is trained as a Convolutional Autoencoder, or [42] in which each channel trains different facial parts. However, having a large number of parameters is the main problem of those networks. This makes them impracticable to be trained on small datasets. Besides, they also require long training/test time and large memory cost.

### B. Deep belief network (DBN)

DBN proposed by Hinton et al. [43] is a graphical model that uses probabilities and unsupervised learning to extract a

deep hierarchical representation of the training data. Its network is like a stack of Restricted Boltzmann Machines (RBMs) [44] in the middle, and the last layer is a classifier. An RBM is a stochastic recurrent neural network that consists of a layer of visible units  $v$ , and a layer of binary hidden units  $h$ .

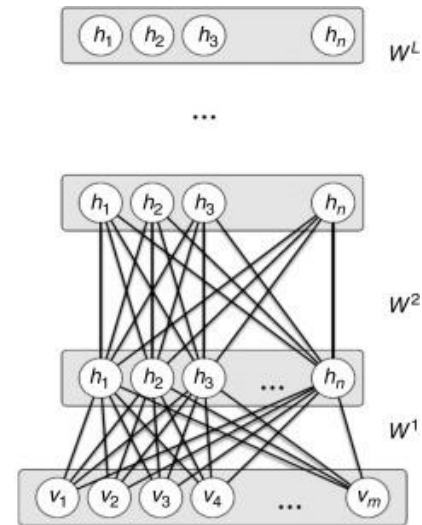


Figure 8. Deep belief network (DBN) architecture composed by stacked restricted Boltzmann machines (RBMs) [45]

The nodes in each layer of a DBN are connected to all the nodes in the previous and subsequent layer (Figure 8). Unlike RBMs, they do not communicate laterally within their layer. Besides, in CNN the first layers only filter the input for basic features, such as edges, and the later layers recombine all the simple patterns found by the previous layers. Deep belief networks, on the other hand, work globally and each layer will learn the entire input and is regulated in order. Greedy learning algorithms are used to pre-train deep belief networks. It makes the optimal choice at each layer in the sequence, eventually finding a global optimum.

It was demonstrated that expression recognition can benefit from performing feature learning and classifier construction together, back and forth in a Deep Belief Network (DBN) [46] [47]. Motivated by this, Ping Liu et al., 2014 [48] proposed Boosted Deep Belief Network (BDBN) to perform the three stages in a unified loop framework.

### C. Deep Autoencoders (DAE)

A deep autoencoder is trained in the same way as a single-layer neural network. It is composed of two, symmetrical deep-belief networks (Figure 9). One network for encoding and another for decoding. The encoder layer compresses the input data while the decoder layer does the reverse to produce the uncompressed version of the data. The goal of this architecture is to reproduce the input at the output layer as accurately as possible. However, the actual use of autoencoders is for determining a compressed version of the input data with the lowest amount of loss in data. Important data is not lost but the overall size of the data is reduced

significantly. This concept is called Dimensionality Reduction, similarly to what PCA does.

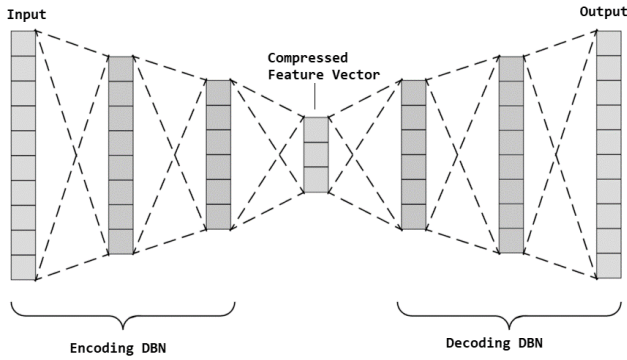


Figure 9. Deep Autoencoder (DAE) architecture

Inspired by the great success of DAE based deep network architecture in facial expression recognition [49], deep sparse autoencoders have been stacked into a deep structure (DSAE) and proposed to solve the problem [50]. Different from autoencoders, sparse autoencoders have a greater number of hidden nodes than the number of input nodes. A sparsity penalty is applied on the hidden layer in addition to the reconstruction error in order to prevent overfitting.

## V. HYBRID AND ENSEMBLE BASED FER SYSTEMS

It should mention that the advances in hardware mostly GPUs is accelerating deep learning. We are currently beholding the trend of hybrid or ensemble learning methods. Both methods make use of the information fusion concept but in slightly different ways.

On the one hand, hybrid models are made through integration of completely different, heterogeneous machine learning approaches, for example traditional machine learning features (or called handcrafted features) and deep learning features. Some of the current hybrid methods proposed to solve the problem of FER include HoG-DAE [49], CNN-SIFT [51], and CNN-HoG [52]. Figure 10 depicts a typical hybrid model.

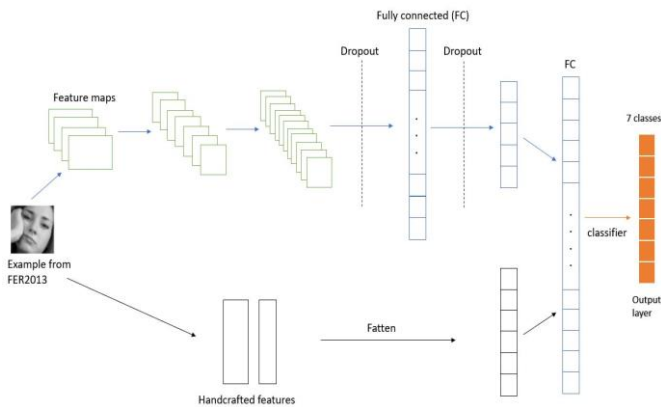


Figure 10. An example of a hybrid model

On the other hand, the ensemble methods are made using various grouping techniques such as bagging (Figure 11) or boosting (Figure 12) to train multiple but normally homogeneous models and combine their predictions. Ensemble voting reportedly improves the test accuracy by 2-3% [67], [68]. In [67], the author minimized the log likelihood loss, and the hinge loss to learn the ensemble weights of the multiple CNN models while in [68], the work coped with the failure in face alignment, and provided a fusion of nine deep convolutional neural networks (DCNN) trained using three different methods for input normalization and weight initialization to estimate aligned states of non-alignable faces. Above all, C. Pramerdorfer et al. in 2016 [59] formed an ensemble of eight modern deep CNNs, and obtained FER2013 test accuracy of 75.2%, outperforming previous works without requiring auxiliary training data or face registration.

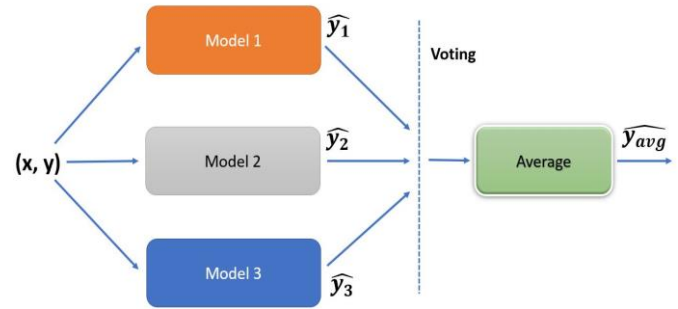


Figure 11. An illustration of a bagging ensemble model

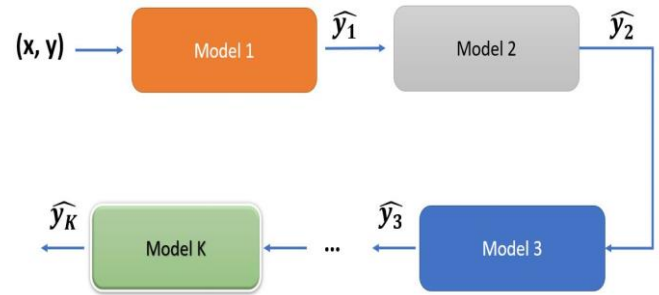


Figure 12. An illustration of a boosting ensemble model

## VI. EVALUATION RESULTS AND DISCUSSIONS

### A. Evaluation Criterion

In this evaluation part, the performance measurement, accuracy, for each referred experiment is presented. It can be defined as the sum of correct decisions divided by the total number of classifications as shown in Equation 5.1.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (\text{Equation 5.1})$$

However, it is worth noting that accuracy alone does not tell the full story when you are working with a class-imbalanced data set.

## B. Evaluation Results

In this section, we perform some of the aforementioned methods on the benchmark datasets to illustrate the performance of existing state-of-art FER approaches as shown in Table 2, Table 3, and Table 4.

**TABLE 2.** COMPARISON ON JAFEE DATASET

	Method	Features	(Additional) Classifier	Year	Recognition Accuracy
Traditional Machine Learning-based	[18]	Haar	Adaboost	2015	<b>98.90</b>
	[19]	Local Gabor (combine Gabor filters and Local Gradient Code)	Chi-square statistic	2015	93.30
	[15]	PHOG*+LBP+PCA	LDA	2015	87.43
	[23]	PCA	KNN	2015	87.70
	[17]	IVA* + HoG	Adaboost & SVM	2016	88.20
	[20]	ART* + DCT* + Gabor	SVM	2018	97.10
Deep learning-based & Hybrid/ Ensemble Learning	[48]	Boosted Deep Belief Networks (BDBN)		2014	93.00
	[38]	CNN (ImageNet)	SVM	2014	94.40
	[41]	2-Channel CNN (on the whole faces)		2015	94.40
	[37]	DCNN	SVM	2016	<b>98.12</b>
	[42]	2-Channel CNN (on Facial parts)		2017	97.71

\*PHOG: Pyramid Histogram of Orientation Gradients, a spatial shape descriptor

\*IVA: inter vector angle, a geometric-based feature extraction

\*ART: the angular radial transform, a region-based shape descriptor

\*DCT: discrete cosine transforms

**TABLE 3.** COMPARISON ON CK+ DATASET

	Method	Features	(Additional) Classifier	Year	Recognition Accuracy
Traditional Machine Learning-based	[19]	Local Gabor (combine Gabor filters and Local Gradient Code)	Chi-square statistic	2015	90.62
	[18]	HoG	Adaboost	2015	88.90
	[25]	HoG + PCA + LDA	BPNN*	2017	99.51
	[53]	Gabor-mean-DWT	SVM	2018	92.50
	[34]	HoG	ELM	2019	<b>99.79</b>
	[34]	LBP	ELM	2019	98.72
Deep learning-based & Hybrid/ Ensemble Learning	[38]	CNN (AlexNet)	SVM	2014	94.40
	[39]	CNN (Deeper CNN based on the Inception architecture)		2016	93.20
	[54]	CNN (Data Augmentation)		2017	98.62
	[42]	2-Channel CNN (on Facial parts)		2017	95.72
	[40]	Inception-ResNet with CRF*		2017	93.04
	[40]	Inception-ResNet without CRF*		2017	85.77

	[49]	HOG + DAE	SVM	2017	<b>99.60</b>
	[50]	DAE (DSAE)		2018	93.78
	[52]	CNN + HOG	SVM	2020	97.01

\*BPNN: Back-propagation neural network

\*CRF: Conditional Random Field

**TABLE 4.** COMPARISON ON FER2013 DATASET

	Method	Features	(Additional) Classifier	Year	Recognition Accuracy
Traditional Machine Learning-based	[55]		SVM (baseline – with cubic kernel)	2018	57.17
	[34]	HOG	ELM	2019	<b>63.86</b>
	[34]	LBP	ELM	2019	55.11
Deep learning-based & Hybrid/ Ensemble Learning	[56]	VGG CNN		2014	72.70
	[67]	CNN Ensembles (hinge-loss-based weighted fusion)		2015	72.0
	[39]	CNN (Deeper CNN based on the Inception architecture)		2016	61.10
	[57]	CNN	k-NN	2016	71.33
	[58]	ResNet CNN		2016	72.40
	[59]	CNN Ensembles (8 CNN models)		2016	75.20
	[68]	DCNN Ensembles (9 models trained using 3 different methods for input normalization and weight initialization)		2016	73.73
	[54]	CNN		2017	72.10
	[51]	3 CNN models + dense SIFT	SVM	2019	<b>75.42</b>

## C. Discussions

From these results, it can be seen that deep learning-based methods are capable of learning facial characteristics and improving facial expression recognition, but not all the cases. JAFFE and CK+ are small datasets captured in laboratory conditions. Using traditional machine learning-based approach, the problem can be nicely solved, obtaining the accuracy 98.90% (Haar features) and 99.79% (HoG features) corresponding in JAFFE and CK+. Both Haar and HoG are easy to use and fast to train. There is a huge debate always going on which is better in detecting facial expression between these two types. On the one hand, Haar-like features can detect regions brighter or darker than their immediate surrounding region better than HoG features. On the other hand, HoG is purely gradient based and is able to capture the object shape information well, better than Haar features.

However, those evaluations in FER2013 demonstrate that deep learning-based methods are the best candidate to address the problem in real conditions with different illumination, pose and resolution. Deep networks are able to learn invariant features itself well without experts' knowledge. The more data

it has, the better performance it can obtain, that is the reason why data augmentation has been applied. With a classic CNN but using Data Augmentation effectively, Breuer, R. et al. [54] has achieved the second highest accuracy in CK+ compared to other deep learning methods. In addition, data augmentation is applied in most deep learning methods training FER2013.

Last but not least, the trend of hybrid and ensemble learning has been attracting more and more researchers since they did help in improving the model accuracy. The work of Georgescu et al., 2019 [51] has achieved the state-of-art result on FER2013, with a test accuracy of 75.42%, in which they fused three different CNN models with SIFT features before the learning stage.

## VII. APPLICATION OF FACIAL EXPRESSION RECOGNITION IN VIETNAM

In recent years, Vietnam has made a progress in building an e-government. According to an e-government survey conducted in 2016 by the United Nations, Vietnam ranked 89<sup>th</sup> of 193 countries and territories on the e-government development index (EGDI), up ten places from 2014. Vietnam ranked 6<sup>th</sup> in ASEAN and 4<sup>th</sup> in Southeast Asia. However, Vietnam E-government has not yet delivered on all its potential opportunities. Currently there are several different definitions on E-Government, but they all share a common characteristic, which is thanks to the transparency and openness, this model will establish a better relationship between the government and its citizens. Public feedback collected through e-government activities will help the government address its shortcomings, enhance transparency and democracy.

Motivated by this objective, we are now working on the direction of how to evaluate citizens' feedbacks. Focusing on users and putting them at the centre of public sector activities then the performance of state agencies in providing services to the public and business enterprises will be improved. In 2019, in realizing the direction of centralized management and administration, Hanoi's authority has installed camera surveillance systems mostly on one-stop shop. It is now able to record photos through camera systems and then extract facial parts to evaluate how the citizens feel about the service.

Since 2017, our own ITI facial data set has been collected with the aim to serve automated facial image analysis and synthesis and for perceptual studies. Till now, one thousand and fifty photos were collected in the real condition with different illumination, pose and resolution, with and without glasses from 150 volunteers living in Hanoi: 70 women and 80 men. The expression in each photo was ratings over 5 subjects, then is given an emotion label. The total images consist of 7 different facial expressions: Angry, Disgust, Happy, Neutral, Sad, Fear, and Surprise. Their class distribution is shown in Figure 13. We also note that getting data in the facial expression task is as difficult as in the other domains, such as annotating gene and disease, speech recognition task, or media labelling since the user is not willing to show their emotions on camera. However, labelling those emotion data is much easier. Also in 2017, we used a

subset of 354 images, then performed our experiment that uses a density-based clustering with side information and active learning as we named MCSSDBS to group facial expression images based on its predicted emotion. The result was published in Intelligent Data Analysis, 2019 [60].

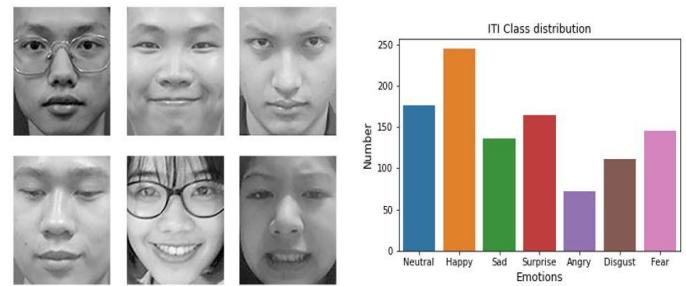


Figure 13. Example of our ITI dataset and its current class distribution

## VIII. CONCLUSIONS

In this study, we showed a comprehensive comparison between the three different approaches of FER systems, i.e., traditional machine learning-based, deep learning-based and the current new trend of hybrid and ensemble learning. The complete results depicted that Haar and HoG are good features in traditional methods while in the real complex environment where data is different in illumination, pose and resolution, deep learning-based methods can solve the FER problem better. Besides, remarkably the current state-of-art result on FER2013 was achieved by a hybrid method. There is no doubt that the trend of hybrid and ensemble learning will attract more and more researchers since they did help in improving the model accuracy. In our future work, we are planning to expand our own dataset, and applying the state-of-art deep-learning models to verify the problem of facial expression recognition in the real applications of our E-Government.

## ACKNOWLEDGMENT

This work was sponsored by The Information Technology Institute, Vietnam National University (VNU) under Project No. QCT.20.03.

## REFERENCES

- [1] M. A., "Communication without Words," *Psychology Today*, vol. 2, no. 4, pp. 53-56, 1968.
- [2] W. V. F. Paul Ekman, "Constants across Cultures in the Face and Emotion," *J. Pers. Psycho. WV*, vol. 17, no. 2, pp. 124-129, 1971.
- [3] P. Ekman, E. Rolls, D. Perrett and H. Ellis, "Facial expressions of emotion: An old controversy and new findings," *Philos. T. Roy. Soc. B.*, vol. 335, no. 1273, p. 63-69, 1992.
- [4] J. Dataset, "Japanese Female FacialExpression Database," [Online]. Available: <https://zenodo.org/record/3451524>.
- [5] B. Fasel and J. Luetin, "Automatic facial expression analysis: a survey," *Pattern recognition*, vol. 36, no. 1, p. 259-275, 2000.
- [6] E. Sariyanidi, H. Gunes and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 6, p. 1113-1133, 2015.

- [7] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.2981446, 2020.
- [8] T. Zhang, "Facial expression recognition based on deep learning: A survey," *International Conference on Intelligent and Interactive Systems and Applications*, Springer, p. 345–352, 2017.
- [9] C. Pramerdorfer and M. Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art," *arXiv.1612.02903*, 2016.
- [10] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *Computer Vision and Pattern Recognition Workshops (CVPR)*, 2010.
- [11] D. E. P. L. C. e. a. I. J. Goodfellow, "Challenges in representation learning: a report on three machine learning contests," *Neural Information Processing*, Springer, Berlin, Germany, 2013.
- [12] P. Viola and M. Jones, "Rapid object detection using a Boosted Cascade of Simple features," *Conference on Computer Vision and Pattern Recognition*, 2001.
- [13] A. A., S. J., W. M. and G. R., "Evaluating AAM Fitting Methods for Facial Expression Recognition," *Proceedings of the International Conference on Affective Computing and Intelligent Interaction; Amsterdam, The Netherlands*, pp. 1-8, September 2009.
- [14] S. J. and K. D., "Real-time facial expression recognition using STAAM and layered GDA classifiers," *Image Vision Comput.*, vol. 27, p. 1313–1325, 2009.
- [15] S.L.Happy and A. Routray, "Robust Facial Expression Classification Using Shape and Appearance Features," *Proceedings of 8th International Conference of Advances in Pattern Recognition*, 2015.
- [16] Y. Liu, J. Wang and P. Li, "A Feature Point Tracking Method Based on The Combination of SIFT Algorithm and KLT Matching Algorithm," *J. Astronaut*, vol. 7, 2011.
- [17] R. I. e. al., "SenTion: A framework for Sensing Facial Expressions," *arXiv preprint, arXiv:1608.04489*, 2016.
- [18] C. Liew and T. Yairi, "Facial expression recognition and analysis: A comparison study of feature descriptors," *IPSJ Trans. Comput. Vis. Appl.*, vol. 7, p. 104–120, 2015.
- [19] S. Al-Sumaidae, "Facial Expression Recognition Using Local Gabor Gradient Code-Horizontal Diagonal Descriptor," *School of Electrical and Electronic Engineering, Newcastle University, England, UK*, 2015.
- [20] H. Tsai and Y. Chang, "Facial expression recognition using a combination of multiple facial features and support vector machine," *Soft Comput.*, vol. 22, p. 4389–4405, 2018.
- [21] D. Gabor, "Theory of communication," *Journal of the Institute of Electrical Engineers*, vol. 93, p. 429–457, 1946.
- [22] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Faces," *J. Opt. Soc. Am.*, vol. A4(3), p. 519–524, 1987.
- [23] X. Wang, A. Liu and S. Zhang, "New facial expression recognition based on FSVM and KNN," *Optik2015*, vol. 126, p. 3132–3134, 2015.
- [24] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, p. 71–86, 1991.
- [25] N. B. Kar, K. S. Babu and S. K. Jena, "Face expression recognition using histograms of oriented gradients with reduced features," *Proceedings of International Conference on Computer Vision and Image Processing*, Springer, pp. 209-219, 2017.
- [26] L. M. and S. S., "Coding Facial Expressions with Gabor Wavelets," *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition; Nara, Japan*, p. 200–205, April 1998.
- [27] H. X., Z. G., P. M. and Z. W., "Dynamic Facial Expression Recognition Using Boosted Component-Based Spatiotemporal Features and Multi-Class Classifier Fusion," *Proceedings of Advanced Concepts for Intelligent Vision Systems; Sydney, Australia*, p. 312–322, December 2010.
- [28] Y.-L. Tian, T. Kanade and J. Cohn., "Recognizing action units for facial expression analysis," *IEEE Trans. on Pattern Analysis and Machine Intell.*, vol. 23, no. 2, p. 1–19, 2001.
- [29] N. S. and J. R.P.K., "GCV-Based Regularized Extreme Learning Machine for Facial Expression Recognition," *Advances in Machine Learning and Data Science*, Springer, p. 129–138, 2018.
- [30] B. S., K. K., L. R., M. M. and M. P., "Face analysis through semantic face segmentation," *Signal Process. Image Commun.*, vol. 74, no. doi: 10.1016/j.image.2019.01.005, p. 21–31, 2019.
- [31] V. V.K., S. S., J. T. and J. A., "Local Invariant Feature-Based Gender Recognition from Facial Images," *Soft Computing for Problem Solving*, Springer; Berlin/Heidelberg, Germany., p. 869–878, 2019.
- [32] M. A., H. A., K. S.A., N. M. and A. S., "Illumination invariant facial expression recognition using selected merged binary patterns for real world images," *Optik*, vol. 158, no. doi: 10.1016/j.ijleo.2018.01.003, p. 1016–1025, 2018.
- [33] K. S.A., H. A. and U. M., "Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features," *Multimed. Tools Appl.*, vol. 77, p. 1133–1165, 2018.
- [34] S. S. Shafira, N. Ulfa, H. A. Wibawa and Rismiyati, "Facial Expression Recognition Using Extreme Learning Machine," *3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia*, doi: 10.1109/ICICoS48119.2019.8982443, pp. 1-6, 2019.
- [35] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
- [36] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
- [37] V. M. e. al., "Automatic Facial Expression Recognition Using DCNN," *Procedia Computer Science*93, p. 453–461, 2016.
- [38] S. Ouellet, "Real-time emotion recognition for gaming using deep convolutional network features," *arXiv preprint arXiv:1408.3750*, 2014.
- [39] A. Mollahosseini, D. Chan and M. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA*, pp. 1-10, March 2016.
- [40] B. Hasani and M. Mahoor, "Spatio-temporal facial expression recognition using convolutional neural networks and conditional random field," *IEEE International Conference on Automatic Face & Gesture Recognition. IEEE.*, p. 790–795, 2017.
- [41] D. H. e. al., "Face Expression Recognition with a 2-Channel Convolutional Neural Network," *International Joint Conference on Neural Networks (IJCNN)*, 2015.
- [42] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang and T. Zhang, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts," *IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intell*, 2017.
- [43] G. E. Hinton, S. Osindero and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, p. 1527–1554, 2006.
- [44] G. E. Hinton and T. J. Sejnowski, "Learning and relearning in boltzmann machines," *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, no. 2, pp. 282-317, 1986.
- [45] D. Rodrigues, X.-S. Yang and J.P. Papa, "Chapter 3 - Fine-tuning deep belief networks using cuckoo search," in *Bio-Inspired Computation and Applications in Image Processing*, Academic Press, 2016, pp. 47-59.



- [46] M. Ranzato, J. Susskind, V. Mnih and G. Hinton, "On deep generative models with applications to recognition," *CVPR*, p. 2857–2864, 2011.
- [47] S. Rifai, Y. Bengio, A. Courville, P. Vincent and M. Mirza, "Disentangling factors of variation for facial expression recognition," *ECCV, Springer*, p. 808–822, 2012.
- [48] P. L. e. al., "Facial Expression Recognition via a Boosted Deep Belief Network," *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, p. 1805–1812, 2014.
- [49] M. Usman, S. Latif and J. Qadir, "Using deep autoencoders for facial expression recognition," *13th International Conference on Emerging Technologies (ICET), Islamabad*, doi: 10.1109/ICET.2017.8281753, pp. 1–6, 2017.
- [50] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, p. 643–649, 2018.
- [51] M.-I. Georgescu, R. T. Ionescu and M. Popescu, "Local Learning With Deep and Handcrafted Features for Facial Expression Recognition," *IEEE Access*, vol. 7, p. 64827–64836, 2019.
- [52] X. Pan, "Fusing HOG and convolutional neural network spatial-temporal features for video-based facial expression recognition," *IET Image Processing*, vol. 14, no. 1, p. 176–182, 2020.
- [53] G. Mattela and S. Gupta, "Facial Expression Recognition Using Gabor-Mean-DWT Feature Extraction Technique.," *Proceedings of the 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India*, pp. 575–580, February 2018.
- [54] R. Breuer and R. Kimmel, "A deep learning perspective on the origin of facial expressions," *arXiv2017, arXiv:1705.01842*, 2017.
- [55] S. Saeed, J. Baber, M. Bakhtyar, I. Ullah, N. Sheikh, I. Dad and A. Ali, "Empirical evaluation of svm for facial expression recognition," *International Journal of Advanced Computer Science and Applications*, vol. 9, 2018.
- [56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [57] Y. Guo, D. Tao, J. Yu, H. Xiong, Y. Li and D. Tao, "Deep neural networks with relativity learning for facial expression recognition," *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on. IEEE*, p. 1–6, 2016.
- [58] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 770–778, 2016.
- [59] C. Pramerdorfer and M. Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art," *arXiv:1612.02903*, 2016.
- [60] V.-V. Vu, H.-Q. Do, V.-T. Dang and D. Toan, "An efficient density-based clustering with side information and active learning: A case study for facial expression recognition task.," *Intelligent Data Analysis*, doi: 10.3233/IDA-173781, vol. 23, pp. 227–240, 2019.
- [61] W. Zhao, R. Chellappa and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," *Proc. Computer Vision and Pattern Recognition, Santa Barbara, CA*, p. 164–169, 1998.
- [62] C. K. and V. S., "A survey on facial recognition based on local directional and local binary patterns," *Proceedings of the 2018 Conference on Information Communications Technology and Society (ICTAS)*, p. 1–6, March 2018.
- [63] H. Yang, U. Ciftci and L. Yin, "Facial expression recognition by de-expression residue learning," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 2168–2177, 2018.
- [64] Y. Huang, F. Chen, S. Lv and X. Wang, "Facial Expression Recognition: A Survey," *Symmetry*, 11.1189. 10.3390/sym11101189, 2019.
- [65] C. S. e. al., "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, 2009.
- [66] P. Belhumeur, J. Hespanha and D. Kriegman, "Using discriminant eigenfeatures for image retrieval," *PAMI*, vol. 19(7), p. 711–720, 1997.
- [67] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," *ACM International Conference on Multimodal Interaction (MMI)*, pp. 435–442, 2015.
- [68] B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, and S.-Y. Lee, "Fusing Aligned and Non-Aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 48–57, 2016.



**MSc. Vu-Tuan Dang** is working at Hanoi Department of Information and Communication, and currently he is also a PhD candidate at the Information Technology Institute, Vietnam National University, Hanoi. His research focuses mainly on applications in E-government.



**MSc. Hong-Quan Do** received a double M.S. degree in Information and Communication Technology from University of Science and Technology of Hanoi, Vietnam and The University of Rennes 1, France in 2015. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research concentrates primarily on Clustering, Semi-supervised clustering, and Image processing. At the present, he has been involved in many projects related to E-government, and E-Commerce Recommendation applications.



**Dr. Viet-Vu Vu** received the B.S. degree in Computer Science from Ha Noi University of Education in 2000, a M.S. degree in Computer Science from Hanoi University of Technology in 2004, and a Doctor Degree in Computer Science from Paris 6 University in 2011. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research interests include clustering, active learning, semi-supervised clustering, and E-government applications.



**Byeongnam Yoon (M'97)** He became a Member of IEEE in 1997. He was born in Seoul Korea 1949. He got the PhD in computer science, Chungnam National University, Korea, 1997. He worked for the Sperryrand UNIVAC as a Computer Specialist 1974 -1978, Samsung as a Manager of Telecommunications Section 1978 -1982, Electronics & Telecommunications Research Institute (ETRI) as a Principal Researcher 1982 - 1999, National Information society Agency (NIA) as a Senior Executive Director General 1999 - 2010, Kyonggi University (KGU) as an Associate Professor, Faculty of Computer Science 2010 - 2016. Global IT Research Institute GIRI as a President 1999 - current. His research area includes a Telecommunications, Internet, Software, Web programming & security, e-Government, Enterprise Architecture, Work Flow, Information Control Nets. System Work Method. SPICE, CMMI, BPM, etc.