TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Exploring the Cultural Gaps in Facial Expression Recognition Systems by Visual Features

Prarinya Siritanawan
School of Information Science
JAIST
Nomi,Ishikawa, Japan
prarinya@jaist.ac.jp

Haruyuki Kojima
Faculty of Human Sciences,
Kanazawa University
Kanazawa, Ishikawa, Japan
hkojima@staff.kanazawa-u.ac.jp

Kazunori Kotani
Division of Transdisciplinary Sciences
JAIST
Nomi, Ishikawa, Japan
ikko@jaist.ac.jp

***Abstract*—This study investigates the cultural dependence of a facial expression recognition (FER) system in an interactive agent by analyzing the performance of several recognition models in different cultural domains. A comprehensive cross-domain classification performance assessment reveals disparities in model performance across different cultural contexts, indicating challenges in cross-cultural FER. To further investigate these characteristics, several public datasets across regions and our cross-cultural dataset of facial expressions derived from Thai and Japanese TV shows are analyzed. By evaluating the capacity of existing FER models to interpret our newly collected data, we found significant variations in emotion interpretation across these cultural contexts, highlighting the necessity for culturally inclusive algorithms. These findings underscore the critical need for more consideration of cultural diversity in FER research, marking a crucial step toward more inclusive and culturally sensitive artificial intelligence technologies.**

***Index Terms*—Facial expression recognition, Affective computing, Multiculture**

I. Introduction

The research and development of Facial Expression Recognition (FER) systems has emerged in an area of affective computing that requires knowledge of computer vision, artificial intelligence, psychology, and social communication. The ability of an interactive computational agent to effectively interpret facial expressions has profound implications for many applications such as human-computer interaction, emotion analysis, mental health monitoring, and social robotics. Rapid advances in deep learning techniques and sophisticated image processing methods have significantly improved the accuracy and reliability of FER systems. These advances are driving the implementation of facial expression recognition systems in the real world.

However, despite significant progress in the research of FER systems, one of the key limitations is the lack of consideration of the cultural diversity of facial expressions across different cultural domains. As a form of nonverbal communication, facial expressions are influenced by numerous factors including geographic location, social dynamics, and cultural background. As a result, the interpretation of facial expressions can exhibit subtle variations across different cultures. Although certain expressions of fundamental emotions may be universally recognizable, the nuanced expressions of these emotions frequently often differ from one culture to another. Therefore, for a FER system to be effective in a global context, it must be able to recognize and adapt to these cultural nuances. Despite these needs, much of the existing research in this field has not accounted for cultural differences. This paper will explore these cultural gaps in FER systems, examining the extent to which they can recognize and adapt to cultural nuances in facial expressions. By addressing this issue, we aim to enhance globally applicable culturally sensitive FER systems, thereby improving their utility in multicultural societies.

II. Related Works

Facial expression recognition (FER) systems have evolved significantly over the past decade with the introduction of various feature extraction methods. The procedure of a FER system typically consists of several steps starting with the preprocessing step that detects and extracts facial regions from images. Depending on the characteristics of an image, different normalization methods are utilized. Then, the process is followed by the feature extraction step, which extracts relevant features from a face image. Given the way image features are extract, features can be an *appearance-based feature* that exploits the texture of an entire face image [1] [2] [3], a *geometric-based feature* that uses more simplified representations of faces such as distances between facial landmark points [4], or a *hybrid feature* that is derived from both types of features [5]. This step is often accompanied by a dimensionality reduction or a feature selection technique, such as Principal Component Analysis or Boosting algorithms, to select the most significant features. Finally, the classification step uses a machine learning technique such as Support Vector Machine or Random Forest, to estimate the emotion class of the previously extracted feature. The composition of these steps generally depicts most of the state-of-the-art FER systems. The representative traditional methods include the work based on the Local Binary Patterns (LBP) which has been widely used due to its simplicity [1] [2], and the significant work of Bartlett et al. [3] which utilized an variety of feature selection techniques and classifiers to discriminate facial expressions from an extensive set of Gabor features.

The aforementioned traditional methods for training a FER system often struggle with real-world data due to high levels of variance. These methods heavily rely on careful preprocessing and feature extraction, posing significant limitations in the implementation. To address these challenges, deep learning-based approaches have been proposed to learn the features from a large amount of facial expression images. For instance, Mollahosseini et al.'s work proposed a deep learning method using a convolutional neural network

979-8-3503-0219-6/23/$31.00 ©2023 IEEE

(CNN) for automatic feature learning, yielding highly accurate predictions on public datasets [6]. Similarly, Pham et al. presented Residual Masked Network (RMN) which combines CNN with recent concepts such as residual networks, attention models, and masking mechanisms, achieving state-of-the-art performance on public datasets [7]. Recently, some of these cutting-edge techniques for facial expression recognition have been integrated into open-source libraries such as OpenFace [8] or Py-Feat [9]. This integration has facilitated the widespread adoption of FER applications in both academic research and hobbyist endeavors.

Most of the existing FER systems define the recognition as a *classification problem*, where class labels are defined by Ekman's *Basic Emotions* [10]. The theory classifies most fundamental emotions into six categories: *Anger, Disgust, Fear, Sadness, Happiness, Surprise*. Some FER systems may include the less popular *Contempt* class in their implementations [11]. Such a discrete category has been widely adopted more than the dimensional emotion models such as Russell's Circumplex model [12]. This is due to its ease of implementation for understanding facial features. In addition, many of the previous FER systems have used Action Units (AU) [13] and their combinations, which correspond to the basic emotion categories defined by Ekman's conversion [14]. While the basic emotion category itself may not be an issue, the corresponding facial expressions may not follow the universal theory. As the field of psychology has evolved, the consideration of cultural nuances in facial expression recognition has become increasingly important. Barrett et al. [15] posited that cultural factors can profoundly affect how emotions are expressed and interpreted, thereby underscoring the importance of incorporating cultural diversity into FER models. Even among facial expressions of basic emotions, the people in Asian counterparts exhibited different sets of AUs expressing the basic emotions compared to Western counterparts [16].

However, most modern high-performance FER systems are commonly trained on datasets primarily representing Western facial expressions. This limitation raises questions about the applicability of these methods in diverse cultural contexts. In response to these challenges, our research seeks to answer the following key questions:

- Can the current state-of-the-art (SoTA) FER systems effectively comprehend facial expressions from diverse cultural sources?
- How can we quantitatively measure and assess these cultural gaps using computational approaches, and further understand their characteristics through visualization methods?

By exploring these questions, we aim to evaluate the ability of existing FER systems to interpret facial expressions across various cultural contexts and develop numerical methods that enable us to measure and quantify differences in cultural interpretation. In addition, we aim to use computational visualization techniques to gain insight into the specific characteristics of these cultural gaps.

This paper presents a comprehensive analysis of these state-of-the-art FER systems, both open-source and our proprietary. We evaluate their performances on several standard facial expression datasets, and our culturally diversed facial expression dataset. This will highlight the importance of cultural nuances in FER systems and explore their real-world applicability in multicultural societies. This comparative study aims to shed light on the current capabilities of these systems, their limitations, and the steps we need to take to align FER technology with the multicultural global society.

III. Culturally Dependence Analysis of Facial Expression Recognition Systems

In our study, we aim to understand the culturally dependent nuances of FER systems by using several public datasets originating from various cultural backgrounds. This comprehensive evaluation analyzes the diversity of facial expressions expressed in different cultures, and how they are interpreted by FER systems respectively.

A. Performance of Cross-domain Classification by FER Systems on Public Datasets

Cross-domain classification is a task of training a classifier on a source domain and applying it to a target domain with different characteristics. It is a challenging problem as the distribution of data in the target domain may differ significantly from the source domain, leading to a degradation in classification performance. By examining the performance of classification models across multiple domains, we can understand their robustness and adaptability.

1) Public Datasets: The foundation of robust and generalizable FER systems typically lies in the selection and composition of the datasets used for their training and evaluation. The diversity and representativeness of the datasets directly influence the performance of FER systems and their applicability to different contexts. Therefore, we place significant emphasis on identifying the characteristics and potential limitations of the existing datasets commonly employed in FER systems.

In this study, we selected the datasets from a pool of public sources that have been widely used in previous FER studies [17] [11] [18] [19]. However, these datasets predominantly contain expressions that are representative of certain cultural groups, leading to a potential imbalance in cultural representation. Therefore, our analysis not only extends to the performance of the FER systems trained on these datasets but also examines the datasets themselves to identify potential gaps in cultural diversity.

As shown in the Table I, the details of these datasets are explained as follows:

- *FER2013*: The FER2013 dataset [17] is one of the most widely used datasets in research on the FER system, containing 35,887 grayscale images of faces. Each image has 48×48 pixels, and is labeled under seven emotion categories. The dataset was collected from the search engine and was labeled with associated metadata. This dataset was considered one of the most challenging datasets due to the wide variety of the image conditions such as pose variations or different types of occlusions. Due to this complexity, it is often used to demonstrate the learning capability of deep learning based techniques. However, the dataset lacks of the detailed breakdowns, such as age, gender, demographic

TABLE I
DETAILS OF THE PUBLIC DATASETS FOR FACIAL EXPRESSION RECOGNITION SYSTEMS WITH THE EMOTION LABELS

Dataset	Number of face images	Expression Annotations	Sources	Conditions	Major cultural domains
FER2013	35,887	Basic emotions	Google image search API	Varrying poses, Low res.	US
CK+	593	Basic emotions, AU	Lab setting	Frontal poses, High res.	US
JAFFE	213	Basic emotions	Lab setting	Frontal poses, High res.	Japan
IMFDB	34,512	Basic emotions	Cinematic contents	Varying poses	India

and ethnic representation of the subjects which limits the availability of further use. Furthermore, it can be observed that the data mostly represent Western cultural contexts, which may limit the generalizability of FER systems trained on it.

- *CK+*: The Extended Cohn-Kanade (CK+) dataset [11] is a popular dataset for facial expression studies, including 593 sequences from 123 individuals of various ages and genders. Although it claims to include a variety of nationalities, the dataset primarily represents Western culture. Emotion labels are determined using the Facial Action Coding System (FACS). Each expression sequence starts from a neutral state and culminates at the peak of the emotion. Uniquely, the dataset includes the contempt class and omits the neutral class in its labeling system. This paper utilized the last frame from each sequence for our evaluation of image features, as it represents the peak expression and serves as the labeled reference point.
- *JAFFE*: The Japanese Female Facial Expression (JAFFE) dataset [18] is another widely used standard dataset, containing 213 images posed by 10 Japanese women. It contains seven classes of facial expressions of basic emotions, including neutral. Despite providing an East Asian perspective to the FER problem, its limited size, and the lack of diversity in terms of age, gender, and ethnic representation can constrain the versatility of models trained on it.
- *IMFDB*: The Indian Movie Face Database (IMFDB) [19] provides a diverse range of facial expressions from a South Asian context, consisting of 34,512 images of 100 Indian actors collected from over 100 videos. However, since the expressions in this dataset are extracted from cinematic content from different sources, some images are under challenging conditions such as poor image resolution, noisy data, and varying lighting conditions. The more specific case is the availability of the occlusions, which range from sunglasses to customary clothing and makeup commonly seen on South Asian individuals.

2) Evaluation and Analysis: In this section, we present the experiments conducted on cross-domain classification using FER systems trained from various public datasets as shown in Table II. The performance of these classification models is evaluated using the confusion matrix to assess their ability to generalize across different domains. For brevity, only the best performing methods are shown.

The FER systems in the first and second rows of the Table II are our previous work on facial expression feature extraction using transfer learning [20]. This work employed several standard CNN architectures, such as ResNet [21] or EfficientNet [22] network architectures, combined with data augmentation to optimize the representation of facial expression features corresponding to the source datasets (CK+ and FER2013). In the last row, it is the implementation of RMN method [7] in Py-Feat library [9], which is trained on the several public datasets including FER2013 dataset. To avoid the bias of training, we perform the classifications only on the test sets of each dataset. Note that the IMFDB dataset contains a number of face images with challenging conditions, such as extreme pose variations, and we inevitably had to filter out one third of entire samples. From the results, it is apparent that there is a varying degree of success in cross-domain classification, underscoring the importance of considering cultural differences and dataset composition in FER research.

In this research, we employed the RMN method [7] implemented in Py-Feat, which is used as the representative method that used several datasets for its training. Despite using FER2013 as one of many datasets to train the model, it is interesting to see that the performance of classification on FER 2013 can archived only 53% in our experiment which is contradicted to the original report presenting more than 70% of accuracy rate on FER2013 dataset. Despite the poor performance of the RMN method, we found an interesting trends from the recognition using the RMN on the IMFDB (Indian) and JAFFE (Japan) datasets. As shown in the confusion matrices Fig. 1 and 2, we found that the RMN method trained from various datasets in Py-Feat tends to estimate most of facial expressions as angry, while the same model estimated the samples in JAFFE as the sad expression. These results demonstrated the bias of FER models that has been trained from the public datasets which is mostly made from western samples.

B. Performance of FER Systems on Unknown Data of Different Cultural Groups

To make our investigation more exhaustive, we also evaluate the FER systems on our additional data collected

TABLE II
PERFORMANCE OF CROSS-DOMAIN CLASSIFICATION OF FER SYSTEMS TRAINED ON DIFFERENT PUBLIC DATASETS.

Source Training Data	Method	Target Tested Data			
		FER2013	CK+	JAFFE	IMFDB
CK+ [11]	Optimized CNN [20]	0.31	0.92	0.31	0.16
FER2013 [17]	Optimized CNN [20]	0.61	0.47	0.45	0.21
FER2013 + Others	RMN [7]	0.53	0.60	0.36	0.25

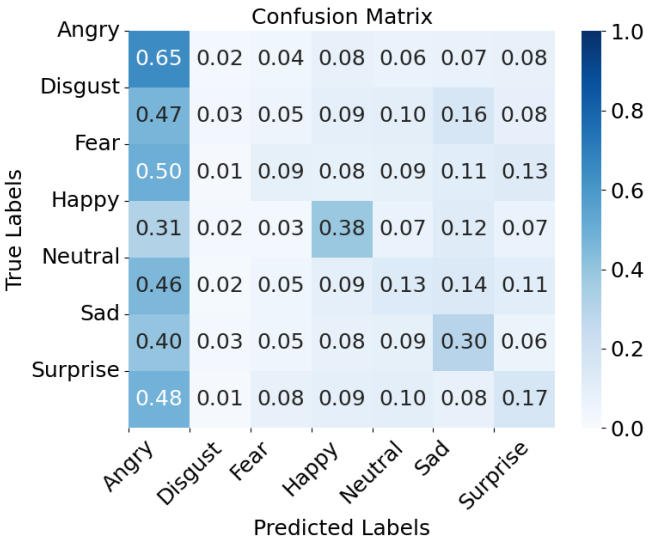

Fig. 1. Confusion matrix showing the estimation of emotion classes in IMFDB samples by using RMN method trained on various datasets

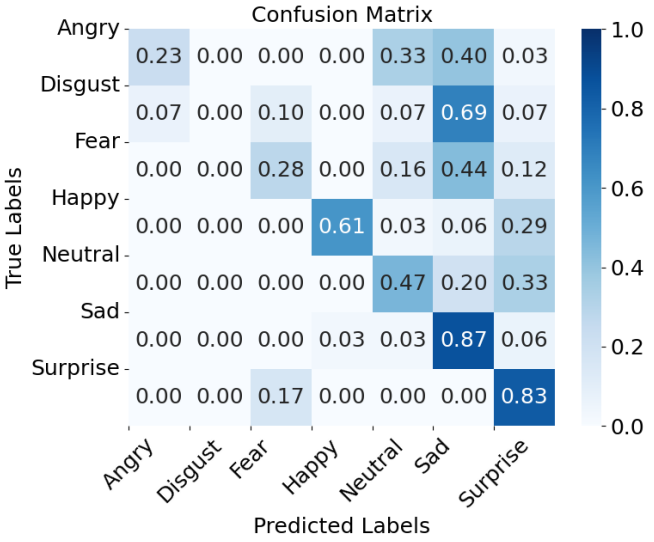

Fig. 2. Confusion matrix showing the estimation of emotion classes in JAFFE samples by using RMN method trained on various datasets

from Japanese and Thai cultural domains in addition to the public datasets. The additional data allows us to evaluate the performance of FER systems on unknown data with a broader cultural spectrum of facial expressions from media in Japan and Thailand. The overall diagram of the performance evaluation and analysis of the FER System on cross-cultural facial expression images is shown in the Fig. 3.

1) Japan-Thai Drama Dataset: The dataset of cross-cultural facial expression images was collected from

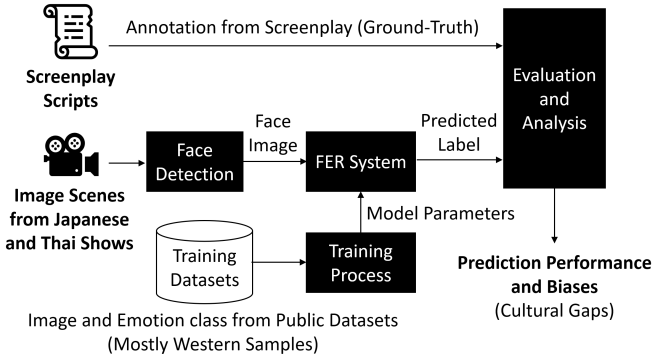

Fig. 3. Overview diagram of the performance evaluation and analysis of the FER System on cross-cutural Japan-Thai facial expression images

Japanese and Thai media, including TV series and movies [23]. The examples of these images are shown in the Fig. 4. Specifically, we collected facial expression images from

Fig. 4. The examples of cross-cultural facial expression images utilized in this experiment[1,2,3,4,5] [23]

the Thai TV series *Rissaya; Jealousy Is A Curse (2016)*[1] and *Hua Jai Ruk See Duang Dao; F4 Thailand: Boys Over Flowers (2021)*[2]. For the Japanese samples, we collected the facial expression images from the movie *Orange (2015)*[3], the TV series *Watashitachi wa Douka Shiteiru; Cursed In Love (2020)*[4] and *Hana yori Dango; Boys Over Flowers (2005)*[5]. Note that, two TV series (*Boys Over Flowers* and *F4 Thailand*) based on the famous Japanese manga *Boys Over Flowers* were chosen to emphasize cross-cultural components in this research. For the data collection process, the annotators selected the video sequences from each show, focusing on scenes that exhibited intense emotions. Occlusions were applied to cover the faces of other non-target actors in the scenes with more than one character. Finally, we collected 715 images with 403 images from Thai media and 312 images from Japanese media, following this procedure:

[1]©KANTANA, Rissaya (2016)
[2]©GMMTV, Hua Jai Ruk See Duang Dao (2021)
[3]©TOHO, Orange (2015)
[4]©NTV, Watashitachi wa Douka Shiteiru (2020)
[5]©TBS, Hana yori Dango (2020)

- Each image in the dataset was annotated with labels corresponding to Ekman's *basic emotions* categories, including *Angry*, *Disgust*, *Fear*, *Happy*, *Sad*, and *Surprise* [10]. For clarification, all data collection and annotation was done by Thai nationals. To ensure that the annotations are independent of the cultural biases of the human annotators, each facial image was labeled with the designated emotion category provided by the screenplay script of that particular scene, rather than the opinions of the annotators, which may differ for the people in different cultures.
- Furthermore, we categorized the facial expressions of the characters in the shows into distinct emotional classes based on their responses to specific situations. The *Angry* class includes the expressions from the characters who are treated unfairly, mocked, annoyed, or harmed. The *Disgust* class is expressed by the characters who encounter creepy things, look down on others, or find an individual action is hateful. The *Fear* class is expressed by the characters who feel threatened, bullied, scared of supernatural situations, or have a personal phobia. The *Happy* class includes the expression of the characters who find solutions to problems, are in joyful situations, or experience a rush of excitement. The *Sad* class is captured from the characters who encounter sorrowful situations or are physically hurt. Lastly, the *Surprise* class contains the facial expressions of the characters who are surprised by unexpected scenarios or experience a jump-scare scene.

2) Evaluation and Analysis: To evaluate the efficacy of the facial expression recognition (FER) system, we cross-examine its recognition outcomes with annotations derived from screenplays. This comparison aims to understand how the present-day FER system can interpret unknown data originating from distinct cultural contexts, as the dataset currently in use contains facial images from Japan and Thailand. The condition of this data differs significantly from the public data used to train state-of-the-art FER systems. To further investigate the performance of such systems under these unfamiliar circumstances, we evaluate the performance of the classification prediction on facial expression images from Japanese shows and Thai shows (Table III and IV). This matrix provides a visualization of the recognition results obtained by the FER system as opposed to the annotation of the *ground truth* labels taken from the screenplay scripts.

The experimental result of the FER system on Japanese facial images in Table III represents the performance of the FER system on data from Japanese screenplays showing several insights. The FER system demonstrated an effective capability to accurately recognize *Happy* expressions, with a recognition rate of 69%. However, it struggles significantly with the *Disgust* expression, incorrectly predicting it as *Sad* around 50%. The FER system also seems to misinterpret *Fear* as *Sad* for 41%. Moreover, *Angry* and *Surprise* expressions are often misconstrued as *Neutral*, with rates of 36% and 26% respectively. This indicates the bias of the FER system on Japanese facial images towards *Sad* emotion expressions, which is partly similar to the findings in the previous test on the JAFFE dataset shown in Fig. 2.

Furthermore, Table IV demonstrates the performance of the FER system on Thai facial images. Similarly, the system is highly successful in recognizing *Happy* expressions, with a rate of 70%. However, it struggles with *Angry*, *Disgust*, *Sad*, and *Surprise* expressions, often predicting them as *Neutral* at rates of 39%, 50%, 47% and 30% respectively. *Fear* expressions show a strong bias being identified as *Sad* for 43%. This indicates the bias of the FER system on Thai facial images, perceiving most of the samples as the *Neutral* emotion.

Finally, *Happy* expressions seem to be recognized relatively accurately in both cultural contexts. However, there are notable struggles in the recognition of *Disgust*, *Fear*, *Angry*, and *Surprise* expressions, which often get misclassified as *Sad* or *Neutral*. This might be due to the complexity and subtlety of these expressions, which can vary across different cultural contexts. By evaluating and analyzing the prediction performance and biases as the cultural gaps between the FER system and the unknown cross-cultural data, these results will be instrumental in refining the FER system for better performance and more accurate cross-cultural FER system.

IV. Conclusion

In this study, we investigated the cultural dependency of facial expression recognition systems, addressing a critical gap in current research on FER systems. By leveraging several publicly available datasets and implementing cross-domain classification performance evaluations, we gained important insights into the varying degrees of recognition efficiency in different cultural contexts.

Our research demonstrated the disparity in the performance of FER systems when tested on data from different cultural domains, highlighting the challenges and potential points for improvement in cross-cultural facial expression recognition. These findings indicate that there is a significant need to improve the robustness and generalization of FER systems to better address cultural diversity.

In addition, the experiment of the FER system on the culturally specific dataset (Thai and Japanese TV shows) revealed the challenges of interpreting emotions from unknown data from various cultural contexts. This emphasizes the importance of developing culturally sensitive algorithms in FER research.

Although there are critical issues that we have raised in this paper that have not been fully explored, these issues are important to address in light of the future development of a culturally inclusive FER system. We plan to extend our study by including more public datasets and culturally diverse facial expression samples. This is expected to further explore the influence of culture on facial expression recognition and provide valuable insights to advance the field of FER system study. We believe that the results of our research will contribute to the ongoing discussion on cultural diversity in facial expression recognition and pave the way for more inclusive and culturally sensitive technologies in the future.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 23K16925. In addition, the authors extend their gratitude to A. Chaiyaroj, P. Tantawanich, and K. Sirinaksomboon for their efforts in data collection and annotation of the Japan-Thai drama dataset utilized in this research.

TABLE III
PERFORMANCE OF THE EMOTION ESTIMATION ON FACIAL EXPRESSION IMAGES USING THE FER SYSTEM OVER SCREENPLAY ANNOTATION IN JAPANESE SHOWS

		Predicted Labels from FER system						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Annotation from Screenplay (Ground-truth)	Angry	0.11	0.04	0.11	0.04	**0.36**	0.19	0.15
	Disgust	0.08	0.00	0.08	0.08	0.17	**0.50**	0.08
	Fear	0.00	0.00	0.12	0.06	0.29	**0.41**	0.12
	Happy	0.00	0.04	0.01	**0.69**	0.17	0.07	0.01
	Sad	0.01	0.03	0.06	0.03	0.21	**0.66**	0.00
	Surprise	0.01	0.00	0.20	0.07	**0.26**	0.22	0.24

TABLE IV
PERFORMANCE OF THE EMOTION ESTIMATION ON FACIAL EXPRESSION IMAGES USING THE FER SYSTEM OVER SCREENPLAY ANNOTATION IN THAI SHOWS

		Predicted Labels from FER system						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Annotation from Screenplay (Ground-truth)	Angry	0.07	0.05	0.19	0.07	**0.39**	0.14	0.10
	Disgust	0.02	0.02	0.07	0.07	**0.50**	0.23	0.09
	Fear	0.00	0.02	0.21	0.00	0.30	**0.43**	0.04
	Happy	0.00	0.00	0.03	**0.70**	0.17	0.09	0.00
	Sad	0.00	0.04	0.12	0.02	**0.47**	0.33	0.02
	Surprise	0.03	0.01	0.24	0.01	**0.30**	0.16	0.25

REFERENCES

[1] Y. Hu, Z. Zeng, L. Yin, X. Wei, X. Zhou, and T. S. Huang, "Multi-view facial expression recognition," in *8th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–6, 2008.

[2] C. Shan, S. Gong, and P. W. McOwan, "Appearance manifold of facial expression," in *Computer Vision in Human-Computer Interaction* (N. Sebe, M. Lew, and T. S. Huang, eds.), pp. 221–230, Springer Berlin Heidelberg, 2005.

[3] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2005*, vol. 2, pp. 568–573 vol. 2, 2005.

[4] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[5] L. Zhang, D. Tjondronegoro, and V. Chandran, "Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition," in *ICME*, pp. 1027–1032, 2012.

[6] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, 2016.

[7] L. Pham, T. H. Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in *25th International Conference on Pattern Recognition (ICPR)*, pp. 4513–4519, 2021.

[8] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: An open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, 2016.

[9] J. H. Cheong, T. Xie, S. Byrne, and L. J. Chang, "Py-feat: Python facial expression analysis toolbox," *CoRR*, vol. abs/2104.03509, 2021.

[10] P. Ekman, "Universals and Cultural Differences in Facial Expressions of Emotions," in *Nebraska Symposium on Motivation*, pp. 207–283, 1972.

[11] P. Lucey, J. F. 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," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 94–101, June 2010. ISSN: 2160-7516.

[12] J. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[13] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto: Consulting Psychologists Press, 1978.

[14] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial action coding system: Investigator Guide*. Salt Lake City: A Human Face, 2002.

[15] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019. PMID: 31313636.

[16] W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa, "Facial expressions of basic emotions in japanese laypeople," *Frontiers in Psychology*, vol. 10, 2019.

[17] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," in *Neural Information Processing* (M. Lee, A. Hirose, Z.-G. Hou, and R. M. Kil, eds.), pp. 117–124, Springer Berlin Heidelberg, 2013.

[18] M. J. Lyons, M. Kamachi, and J. Gyoba, "Japanese female facial expressions (jaffe)," 1997.

[19] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy, R. Vaddi, V. Hemadri, J. C. Karure, R. Raju, B. Rajan, V. Kumar, and C. V. Jawahar, "Indian movie face database: A benchmark for face recognition under wide variations," in *4th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 1–5, 2013.

[20] W. S. S. Khine, P. Siritanawan, and K. Kotani, "Facial expression features analysis with transfer learning," in *14th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1–6, 2022.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[22] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning*, vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114, PMLR, 09–15 Jun 2019.

[23] P. Siritanawan, A. Chaiyaroj, P. Tantawanich, K. Sirinaksomboon, and K. Kotani, "Facial expression analysis interpreting emotion in multicultural settings," in *62th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2023.