# MODIFIED LOCAL TERNARY PATTERN WITH CONVOLUTIONAL NEURAL NETWORK FOR FACE EXPRESSION RECOGNITION

**Syavira Tiara Zulkarnain[1] and Nanik Suciati[2]**
[1, 2]Department of Informatics, Faculty of Intelligent Electrical and Information Technology,
Institut Teknologi Sepuluh Nopember
Sukolilo, Surabaya
e-mail: syavira.zulkarnain@gmail.com[1], nanik@if.its.ac.id[2]

**ABSTRACT**

*Facial expression recognition (FER) on images with illumination variation and noises is a challenging problem in the computer vision field. We solve this using deep learning approaches that have been successfully applied in various fields, especially in uncontrolled input conditions. We apply a sequence of processes including face detection, normalization, augmentation, and texture representation, to develop FER based on Convolutional Neural Network (CNN). The combination of TanTriggs normalization technique and Adaptive Gaussian Transformation Method is used to reduce light variation. The number of images is augmented using a geometric augmentation technique to prevent overfitting due to lack of training data. We propose a representation of Modified Local Ternary Pattern (Modified LTP) texture image that is more discriminating and less sensitive to noise by combining the upper and lower parts of the original LTP using the logical AND operation followed by average calculation. The Modified LTP texture images are then used to train a CNN-based classification model. Experiments on the KDEF dataset show that the proposed approach provides a promising result with an accuracy of 81.15%.*

***Keywords**: Facial expression recognition, statistical approach, feature extraction, local ternary pattern, convolution neural network.*

# MODIFIKASI *LOCAL TERNARY PATTERN* DENGAN *CONVOLUTIONAL NEURAL NETWORK* PADA PENGENALAN EKSPRESI WAJAH

**Syavira Tiara Zulkarnain[1] dan Nanik Suciati[2]**
[1, 2]Departemen Informatika, Fakultas Teknik Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember
Sukolilo, Surabaya
e-mail: syavira.zulkarnain@gmail.com[1], nanik@if.its.ac.id[2]

**ABSTRAK**

*Pengenalan ekspresi wajah (FER) pada gambar dengan variasi pencahayaan dan noise merupakan salah satu tantangan dalam bidang visi komputer. Kami menyelesaikannya menggunakan pendekatan deep learning yang telah berhasil diterapkan di berbagai bidang, terutama pada kondisi gambar pada lingkungan tidak terkontrol. Proses yang kami terapkan untuk mengembangkan FER berdasarkan Convolutional Neural Network (CNN) meliputi deteksi wajah, normalisasi, augmentasi, dan representasi tekstur. Kombinasi dari metode normalisasi TanTriggs dan Metode Transformasi Gaussian Adaptif digunakan untuk mengurangi variasi cahaya. Jumlah gambar ditambah menggunakan teknik augmentasi geometris untuk mencegah overfitting karena kurangnya data pelatihan. Kami mengusulkan representasi gambar tekstur Modified Local Ternary Pattern (Modified LTP) yang lebih diskriminan dan tidak sensitif terhadap noise dengan menggabungkan bagian pola atas dan pola bawah dari LTP asli menggunakan operasi AND diikuti dengan perhitungan nilai rata-rata piksel. Citra tekstur LTP yang dimodifikasi kemudian digunakan untuk melatih model klasifikasi berbasis CNN. Eksperimen pada dataset KDEF menunjukkan bahwa pendekatan yang diusulkan memberikan hasil dengan akurasi 81.15%.*

***Kata Kunci**: Pengenalan ekspresi wajah, pendekatan statistik, ekstraksi fitur, local ternary pattern, convolution neural network.*

## I. INTRODUCTION

Facial expression is an essential non-verbal mode to show intentions. Most of the messages relating to emotion are shown in facial expression; the rest are in words spoken and the intonation of the speech. In recent years, the development of Facial Expression Recognition (FER) that can automatically detect expressions from facial images has received huge attention. FER has a wide range of applications, such as human-computer interaction [1], advanced driver assistance systems [2], and education [3].


http://dx.doi.org/10.12962/j24068535.v19i1.a1031

There are two approaches in developing FER, namely conventional and deep learning. The conventional approach consists of three stages: facial data acquisition, feature extraction, and classification [1]. The performance of the conventional approach depends on how robust the feature extraction algorithm is. Deniz et al. observed that the conventional approach had weak performance when handling image datasets with pose and illumination variation created in a laboratory setting [4]. Conversely, the deep learning approach mostly produces high accuracy when applied to image classification in the wild setting but needs massive amounts of data in the learning process.

Challenges in developing FER using a deep learning approach are caused by a low number of training data and variation in illumination, noise, and head pose. A FER research [4] that used the state-of-the-art deep learning approach and experimented with five datasets containing only frontal face images achieved the highest accuracy of 83.5%. Other research used a hybrid approach [5] that combined conventional and deep learning using handcrafted methods like Local Binary Pattern (LBP) for feature representation and Convolutional Neural Network (CNN) to handle variation in illumination produced an accuracy of 61.29%. Such results are not optimal because LBP is sensitive to random noise [6], and can't tolerate variations that commonly occur in unconstrained natural images caused by illumination, pose, occlusions [7]. In order to solve this problem, variations of the LBP method have been introduced for handling different lighting conditions to increase robustness against noise. One of the methods is called Local Ternary Pattern (LTP), which provide a more stable image representation in the presence of noise by using t parameter to split the pixels neighbors into upper and lower parts [6][7]. State-the-art research using LTP method for FER such as research [8] shows that LTP achieves an accuracy of 88,9% on CK+ dataset (using 7 class). This accuracy is higher compared to LBP method that reach an accuracy of 83.3% and Local Directional Pattern (LDP) method that reach an accuracy of 88.4%. Although the performance of state-the-art of deep learning and LTP methods is quite good, further research in implementing deep learning to FER, especially for handling variation in illumination, head pose, and noise, is still open to obtain better improvement.

In this paper we proposed representation of Modified Local Ternary Pattern (Modified LTP) texture image for FER using Convolutional Neural Network (CNN)-based approach. This research was developed to handle illumination variation and noise in image. The Modified LTP provides a more discriminant method and less sensitivity to noise by combining the upper and lower parts of the original LTP using the logical AND operation followed by average calculation. The combination of the TanTriggs normalization method and Adaptive Gaussian Transformation was applied to reduce illumination variation. The CNN-based approach unites the learning process of feature extraction and classification.

This article is organized into five sections. Section 1 contains the background and research objectives, Section 2 describes the literature study, Section 3 provides an overview of the proposed method, Section 4 explains the results and discussion, and Section 5 contains conclusions.

## II. Literature Study

According to the research of Huang et al. [9] and Tian et al. [10] we know the important phase for facial expression recognition consists of pre-processing, feature extraction and classification. Pre-processing phase intended to detect essential regions (facial region) of input images. Generally, pre-processing consists of face detection and face enhancement or normalization. Face detection intended to reduce cost calculations at next phase, by removing irrelevant information in input images. Face enhancement or normalization aims to process objective distraction factors such as light intensity, occlusion, size of input images and other interference factors [9]. Research for object detection and face recognition is experiencing rapid advances in method used, ranging from conventional method through deep learning method. Notable research for conventional methods are Haar feature [11], Viola Jones [12], Local Assembled Binary (LAB) [13], Seeta Face [14]. And notable research for deep learning method which is low computation is Dlib [15].

Face extraction phase intended to decompose information of input image in some values of vector features. Usually, the cause of facial expression recognition gets low results due to poor feature extraction. Therefore, it is crucial to consider selecting suitable feature extraction methods manually. A research [10] argued that features can be extracted either based on geometrical relation or presence region of input images. Presence region – also known as appearance features – are frequently used by researchers. Notable research for appearance feature is Local Binary Pattern (LBP), proposed by Ojala et al [16]. Over the past few years, LBP have increased concern in image processing and computer vision [17]. As a non-parametric method, LBP efficiently encapsulates the local image structure by comparing each pixel with its neighboring pixels. The most important properties of LBP are its tolerance to monotonous illumination adjustments. LBP was originally proposed for texture analysis [18] and has proven a simple but powerful approach to describing local structures. LBP descriptors have been widely exploited in many applications, such as, medical image analysis, blood vessels recognition [19] facial image recognition, facial expression recognition [18] [20] [21] and even for clothing pattern recognition [22]. Several variations of the LBP designed to increase robustness in noise include the Local Ternary Pattern (LTP) [6]. LTP reduces the bad effects of noise as well as resolves one of the challenges in LBP by introducing the threshold parameter to suppress noise and be more discriminant.

Another phase beside the feature extraction method is classification or selecting a suitable classifier. There are two kinds of classifiers in the facial expression recognition approach, both have different advantages and challenges.

## III. Overview of The Proposed Method

Figure 2 shows the overall flow of the proposed facial expression recognition method. The method begins with input image obtained from laboratory-controlled dataset, continued with pre-processing, then feature extraction using our proposed method (Modified LTP), then we split the dataset for classification purposes, the output of the system is label of the expression. Each stage of the overall flow is given, respectively.

### *A. Dataset*

We use 4900 images of human facial expression in Karolinska Directed Emotional Faces (KDEF) dataset from the Karolinska Institute, Stockholm, Sweden [23]. These images consist of 70 participants and were photographed separated by lighting condition into 2 sessions. In each session, each participant would have 7 expressions; where each expression would have 5 different angles (-90◦, -45◦, 0◦, 45◦, 90◦) as shown in Figure 1.

### *B. Pre-processing*

We applied the pre-processing stage to the dataset to subtract irrelevant information, localize the face area, and normalize the illumination. The preprocessing stage consists of face detection, normalization, and augmentation, and is aiming to prepare more pertinent data. The number of images is augmented using a geometric augmentation technique to prevent overfitting due to lack of training data.

#### *1) Face Detection*

All image in the dataset contains a face and a background. We use face detector Dlib [15] to subtract irrelevant information and provide only the face area. DLib is an open-source library of machine learning toolkit; therefore, some researchers try to explore DLib, one of which is by using the Convolutional Neural Network (CNN) approach called Max-Margin Object Detection (MMOD) [24]. All face area produced by the face detector then cut from the images with various size, resized into 128× 128 images, and then converted into grayscale.

#### *2) Face Normalization AGT-Me*

Lee et al., proposed a method that reconsiders human eyes has non-linear response to brightness, called Adaptive Gamma Transformation Method (AGT-ME) [25]. The theoretical basis of that method is maximizing differential entropy to minimize gamma distortion, applied for automatic gamma adjustment, contrast enhancement. We applied face normalization using (AGT-Me) to reduce light variations. The AGT-Me algorithm is performed in several steps. The first step is checking whether the input image is grayscale or RGB. If the image is RGB, convert


Figure 1. Sample expression of KDEF dataset.

them into HSV and then take the V component - because the V component represents the amount of illumination. Since the images that we used are greyscale, the first step is skipped. In the second step, each pixel in the image is normalized. Then we transform the data within the range of [0,1]. Then we assign the masking area (whether to mask the whole image or parts of the image) and calculate the optimal gamma of the masking area. After obtaining the optimal gamma transform (power law) are performed. The final step is transforming the pixel value within its original range of grayscale images.

*3) Data Augmentation*

Frequently, deep learning techniques give high recognition rates, but require large amounts of training data. Data augmentation is a vital process to increase the quantity of training data. This creates new images that represent a comprehensive set of possible images. This is done by applying geometric augmentation techniques like flipping horizontally and zooming images. After we applied the augmentation process, the total images increased from 4900 to 14700.

*4) Face Normalization TanTriggs*

Beside applying illumination normalization using AGT-Me, we also applied face normalization of lighting such as the proposed method of Tan, Triggs [6]. The advantage of this normalization method is robustness to lighting changes and other image quality degradations such as blurring. The process includes gamma correction, which is a process to improve the results of the power law transformation, which is an enhancement method to improve the range of the input image in dark areas as well as compress bright and highlight areas with gamma calculations. Then proceed by applying Difference of Gaussian (DoG) filtering which is a band-pass filter that removes high frequency components containing noise and also some low frequency components that represent homogeneous areas. Furthermore, basic normalization is carried out to adjust the interval range up to 8 bits. And the last step of preprocessing is contrast equalization to rescale the image intensity to standardize the overall contrast / intensity variation because after normalizing the image with variations in lighting it will cause contrast variations.

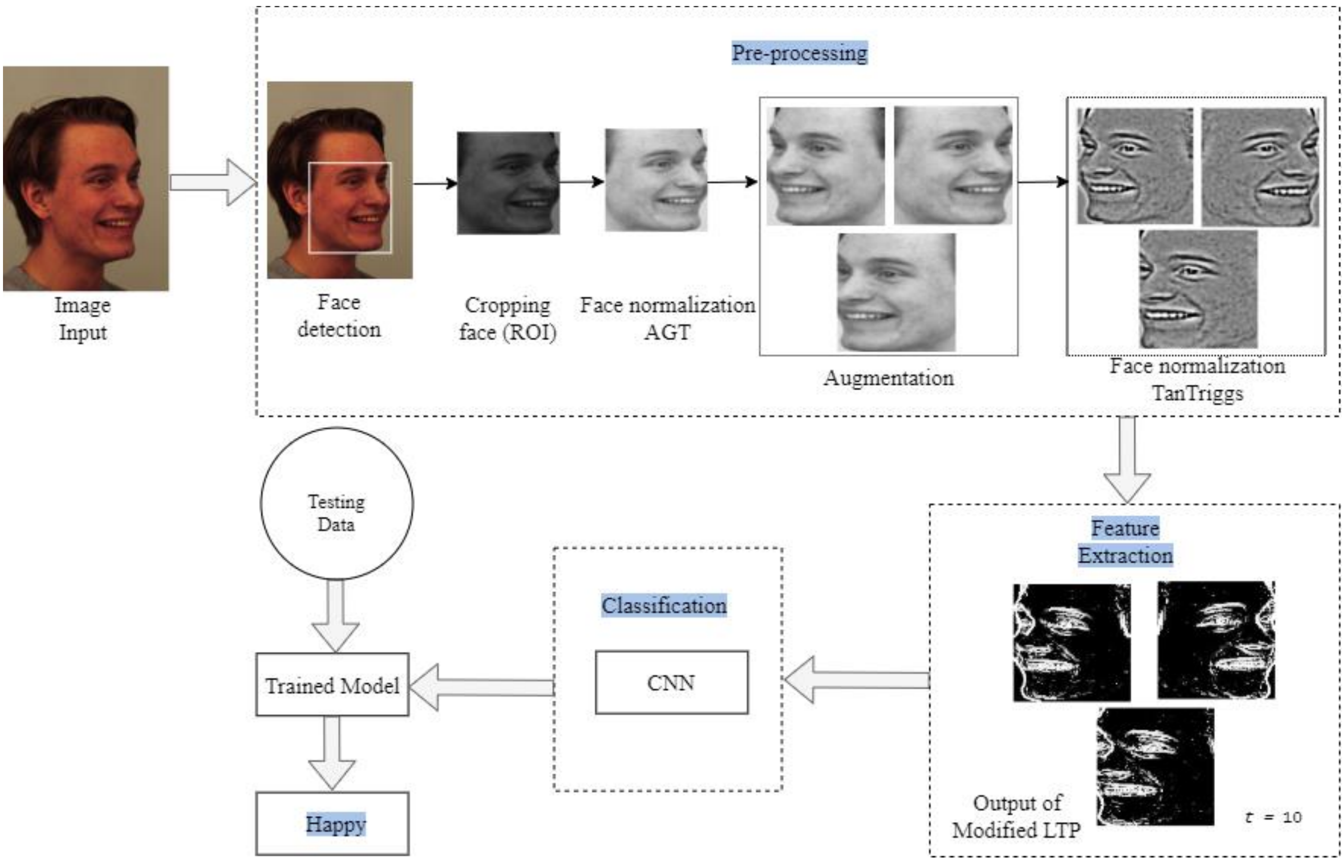


Figure 2. Flow of the proposed method.

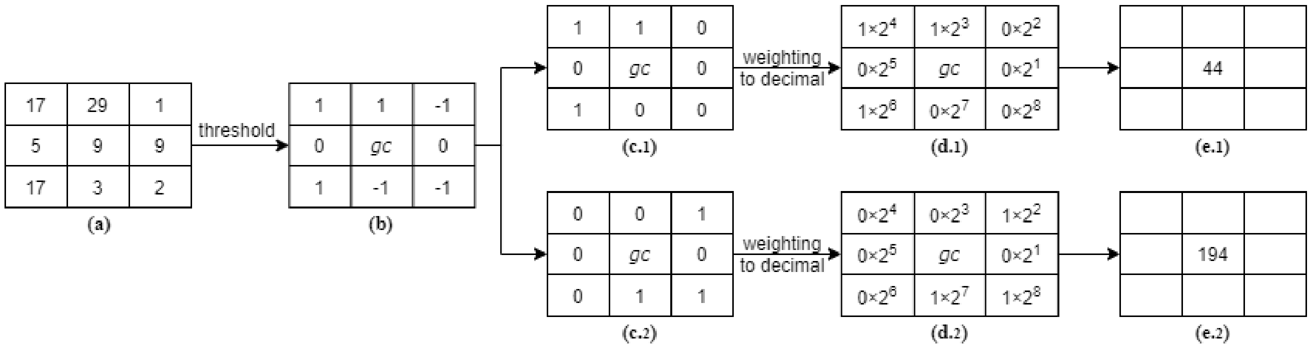


Figure 3. (a) The original gray-scale image on 3x3 neighborhood window, (b) The result of thresholding called Ternary Pattern by using subtract intensity of the center pixel and threshold (e.g., threshold = 5), (c) The result consists of two patterns: (c1) Upper Pattern, (c2) Lower Pattern, (d) Conversion into decimal (d1) Upper Pattern, (d2) Lower Pattern, (e) The result of each pattern.

### C. *Feature Extraction*

Local Ternary Pattern (LTP) was introduced by Tan, Triggs which is extended from Local Binary Pattern (LBP) [7]. LBP labels the neighboring pixels to 0 and 1 by using the center pixel as a threshold which is therefore sensitive to noise, especially for noise that appears in the image, namely gaussian, quantization, or poisson noise illustration shown in Figure 3. Meanwhile, LTP reduces the bad effects of noise as well as resolves one of the challenges in LBP by introducing the $t$ parameter to suppress noise. After obtaining or define the threshold or the $t$ parameter, a comparison of neighboring and center pixels is carried out, in this case the center pixel has been carried out with subtraction and addition operations with a threshold and obtaining ternary pattern. As a result of these subtraction and addition operations, the center pixel becomes 2 binary coded patterns, resulting in lower pattern pixel and upper pattern pixels. In a study [26] of the proposed method and LTP, both managed to survive in the conditions of the image with a gaussian noise. In LTP, the measurement of neighboring pixels becomes 3 quantization, namely 0, -1 and 1 [27]. The final step for LTP original is unite the histogram of two pattern. The quantization equation $f_i$ with $x$ neighboring pixel, $gc$ center pixel, and $t$ threshold is defined as in (1).

$$f_i(x, gc, t) = \begin{cases} +1, & x \geq gc + t \\ 0\,, & |\, x - gc| < t \\ -1\,, & x \leq gc - t \end{cases} \tag{1}$$

The major proposed method in this paper is the modification of LTP method which selects the most important feature using statistical approach. In most cases, the variations classified as texture actually represent changes in the statistical average or mean brightness level. The variation of brightness level which has the highest and lowest scores are discarded in favour of average. By discarding the extreme values, noise is rejected, so the average of the remaining pixel values are used as the new brightness [28]. Those are the reasons for choosing average, a simple statistical method, to reduce noise.

The general concept of LTP is the same as LBP, there is a difference in the threshold that needs to be defined first, so we get two patterns, each capturing different features, upper pattern for feature with positive value and lower pattern for feature with negative value. The modified LTP generally has the same concept with LTP original. First we defined the $t$ parameter = 10 based on experiment. The modification in the final result of the calculation after the pixel converted into decimal value which consists of two patterns, namely lower and upper. The LTP operator is the concatenation of the code of the upper pattern and the lower pattern and we applied the calculation of the average to get the most important feature. At the end, output of modified LTP is an image.

### D. *Classification*

For classification, we use a deep learning approach called Convolutional Neural Network (CNN). We construct a CNN architecture shown in TABLE 1, which consists of five convolution layers, four batch normalization, four down-samplings with kernel size 2 x 2, two dropouts, and a fully connected layer. The input is a feature of a grayscale channel with the shape 128x128. Detail layers of the architectures shown in TABLE 1.

## IV. Result and Discussion

### A. *Experimental Configuration*

In this section contains setting data and specification training for classification purposes, the data will be divided into 2, namely training data and testing data, with total of training data containing 13230 images and 1470 testing images. Because the images in each person appears 2 times in the dataset, division is done by entering all the data

TABLE 1
DETAIL ARCHITECTURE CNN LAYERS.

| Layer | Input | Output | Spesification |
|---|---|---|---|
| Input layer | 128,128,1 | | - |
| Convolution Layer 1 | 128, 128,1 | 124, 124, 32 | Filter: 5 × 5 , 32<br>Stride: 1 × 1<br>ReLU |
| Max Pooling Layer 1 | 124, 124, 32 | 62, 62, 32 | Kernel: 2 × 2 |
| Convolution Layer 2 | 62, 62, 32 | 58, 58, 32 | Filter: 5 × 5 , 32<br>Stride: 1 × 1<br>ReLU |
| Max Pooling Layer 2 | 58, 58, 32 | 29, 29, 32 | Kernel: 2 × 2 |
| Convolution Layer 3 | 29, 29, 32 | 25, 25, 32 | Filter: 5 × 5 , 32<br>Stride: 1 × 1<br>ReLU |
| Max Pooling Layer 3 | 25, 25, 32 | 12, 12, 32 | Kernel: 2 × 2 |
| Dropout Layer 1 | 12, 12, 32 | 12, 12, 32 | Probability: 0.4 |
| Convolution Layer 4 | 12, 12, 32 | 8, 8, 64 | Filter: 5 × 5 , 64<br>Stride: 1 × 1<br>ReLU |
| Max Pooling Layer 4 | 8, 8, 64 | 4, 4, 64 | Kernel: 2 × 2 |
| Dropout Layer 2 | 4, 4, 64 | 4, 4, 64 | Probability: 0.1 |
| Convolution Layer 5 | 4, 4, 64 | 1, 1, 7 | Filter: 4 × 4 , 7<br>Stride: 1 × 1<br>ReLU |
| Fully Connected Layer | 1, 1, 7 | 7 | Softmax |

in the first session into the training data, then images in the second session will be shuffled using k-fold cross validation with k=10 to be included in the testing data and another training data. This division is done to ensure that all individuals and all angle/head pose in the dataset have entered the training model at least once so that the data is divide balanced.

In each experiment scenario using architecture with specification of parameters epoch 35, batch size 100, optimizer using Adam with a learning rate 0.01 and using early stopping with loss monitor validation to avoid overfitting or underfitting.

### *B. Comparison with different combinations of the proposed method*

TABLE 2 shows the experiment result using a dataset that we explained at the experimental setting part. The first experiment combination used all of the images in dataset and augmentation images as an input image, with total images 14700, and followed by normalization using Tantriggs and feature extraction LTP with upper pattern images. The result of this combination is an accuracy of 79.18%. The second combination, still using the same data, pre-processing method, and normalization, but we try using lower Pattern as type for feature extraction. This combined result increased with an accuracy of 77.89%. For the next combination, we tried to combine the lower and upper pattern using AND operator, for this combination result decreased with an accuracy 71.02%. For the fourth experiment, we used a statistical approach such as computing maximum value beside merging two patterns. The result of this combination is an accuracy of 70.47%. The sixth experiment used two normalization AGT-Me and TanTriggs and followed by upper pattern feature extraction. This combination achieved performance 78.97%.

The next experiment we used a lower pattern as a type of feature extraction, and the result little bit increased with an accuracy 79.79%. But in the next experiment, the result achieved an increase with an accuracy of 80.20%. We found the combination method to be suitable to increase performance of FER. The highest accuracy results from combination corresponding proposed method achieved an accuracy 81.15% used AGT-Me and TanTriggs as normalization method, and concatenate following by compute the average value as feature extraction. This FER testing is using KDEF datasets that consists of Caucasian ethnicity images. This method should work in recognizing facial expressions from ethnicity such as Asian. The performance of facial expression recognition of another ethnicity is able to be improved by training the model using a multiethnic dataset. The performance of this method's recognition for each expression are shown in TABLE 3.

As the TABLE 3 shows expressions that have a higher recognition rate based on precision and recall is happy. Because visually, a happy expression does not resemble any other class expression and hence can be easily differentiated. On the other hand, the expression class that most frequently labeled misclassification based on lowest precision is afraid and based on lowest recall is sad. This is happening because visually two classes are similar.

## V. Conclusion

This paper presents a facial expression recognition with KDEF dataset using deep learning technique that aims to obtain more discriminant and less sensitivity to noise using a statistical approach by combining the upper and lower pattern of the original LTP using the logical AND operation followed by average calculation. The proposed system was tested and capable of handling not only frontal facial images and has achieved the best result with an accuracy of 81.15%. The extension of facial expression recognition can be used for emotional analysis for mental health. We will explore different methods such as Facial Action Unit or Facial Landmark to enrich feature vector representation for further research.

TABLE 2
PERFORMANCE COMPARISON OF DIFFERENT COMBINATIONS OF THE PROPOSED METHOD.

| **Augmentation** | **Normalization** | **Type of LTP** | **Accuracy (%)** |
|---|---|---|---|
| Yes | TanTriggs | Upper | 79.18 |
| Yes | TanTriggs | Lower | 77.89 |
| Yes | TanTriggs | Concatenation with Logical Operator (AND) | 71.02 |
| Yes | TanTriggs | Concatenation with Logical Operator (AND), Maximum value | 70.47 |
| Yes | AGT-Me + TanTriggs | Concatenation with Logical Operator (AND), Average value | 75.10 |
| Yes | AGT-Me + TanTriggs | Upper | 78.97 |
| Yes | AGT-Me + TanTriggs | Lower | 79.79 |
| Yes | AGT-Me + TanTriggs | Concatenation with Logical Operator (AND) | 79.86 |
| Yes | AGT-Me + TanTriggs | Concatenation with Logical Operator (AND), Maximum value | 80.20 |
| Yes | AGT-Me + TanTriggs | Concatenation with Logical Operator (AND), Average value | 81.15 |

TABLE 3
PERFORMANCE EACH FEATURE EXPRESSION KDEF WITH OUR PROPOSED METHOD.

| Feature | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| Afraid | 72 | 72 | 72 |
| Angry | 84 | 65 | 73 |
| Disgusted | 80 | 86 | 83 |
| Happy | 93 | 94 | 94 |
| Neutral | 74 | 90 | 81 |
| Sad | 82 | 62 | 71 |
| Surprise | 80 | 94 | 86 |

## REFERENCES


[1] Katsaggelos, Aggelos K, "IEEE Signal Processing Magazine: Farewell," *IEEE Signal Processing Magazine*, pp. 2-4, 2002.

[2] Assari, Mohammad Amin, and M. Rahmati, "Driver drowsiness detection using face expression recognition," in *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 337-341, 2011. https://doi.org/10.1109/ICSIPA.2011.6144162.

[3] A. Kapoor, W. Burleson, R.W. Picard, "Automatic prediction of frustration," *Int. J. Hum. Comput. Stud*, pp. 724–736, 2007. https://doi.org/10.1016/j.ijhcs.2007.02.003.

[4] Fei, Zixiang, Erfu Yang, David Day-Uei Li, Stephen Butler, Winifred Ijomah, Xia Li, and Huiyu Zhou, "Deep convolution network based emotion analysis towards mental health care," *Neurocomputing* 388, pp. 212-227, 2020. https://doi.org/10.1016/j.neucom.2020.01.034.

[5] Levi, Gil, and Tal Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 503-510. 2015. https://doi.org/10.1145/2818346.2830587.

[6] Tan, Xiaoyang, and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing* 19, no. 6, pp. 1635-1650, 2010. https://doi.org/10.1109/TIP.2010.2042645.

[7] Murala, Subrahmanyam, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: a new feature descriptor for content-based image retrieval," *IEEE Transactions on Image Processing* 21, no. 5, pp. 2874-2886, 2012. https://doi.org/10.1109/TIP.2012.2188809.

[8] R.P. Holder, J.R. Tapamo, "Improved gradient local ternary patterns for facial expression recognition," *Eurasip J. Image Video Process*, 2017. https://doi.org/10.1186/s13640-017-0190-5.

[9] Y. Huang, F. Chen, S. Lv, X. Wang, "Facial expression recognition: A survey," *Symmetry (Basel)*, 11, 2019. https://doi.org/10.3390/sym11101189.

[10] Y.-L. Tian, T. Kanade, J.F. Cohn, "Facial Expression Analysis," in *Handbook of Face Recognition*, 2005. https://doi.org/10.1007/0-387-27257-7_12.

[11] C.P. Papageorgiou, M. Oren, T. Poggio, "General framework for object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 555–562, 1998 https://doi.org/10.1109/iccv.1998.710772.

[12] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, no. 1, 2001. https://doi.org/10.1109/cvpr.2001.990517.

[13] S. Yan, S. Shan, X. Chen, W. Gao, "Locally Assembled Binary (LAB) feature with feature-centric cascade for fast and accurate face detection," in *Proc. 26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2008. https://doi.org/10.1109/CVPR.2008.4587802.

[14] S. Wu, M. Kan, Z. He, S. Shan, X. Chen, "Funnel-structured cascade for multi-view face detection with alignment-awareness," *Neurocomputing* no. 221, pp. 138–145, 2017. https://doi.org/10.1016/j.neucom.2016.09.072.

[15] D.E. King, "Dlib-ml: A machine learning toolkit", *J. Mach. Learn. Res. 10*, pp. 1755-1758, 2009.

[16] T. Ojala, M. Pietikäinen, D. Harwood, "A comparative study of texture measures with classification based on featured distributions", *Pattern Recognit. 29*, pp. 51–59, 1996. https://doi.org/10.1016/0031-3203(95)00067-4.

[17] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen, D. Hu, "Extended local binary patterns for face recognition", *Inf. Sci. (Ny).* no. 358–359, pp. 56-72, 2016. https://doi.org/10.1016/j.ins.2016.04.021.

[18] W. Huang, H. Yin, "Robust face recognition with structural binary gradient patterns", *Pattern Recognit* no. 68, pp. 126–140, 2017. https://doi.org/10.1016/j.patcog.2017.03.010.

[19] Z. Li, N. Yang, B. Xie, J. Zhang, "A two-phase face recognition method in frequency domain", *Optik (Stuttg).* no. 124, pp. 6333–6337, 2013. https://doi.org/10.1016/j.ijleo.2013.05.057.

[20] E. khadiri I, C. A, E. merabet Y, R. Y, T. R, "Local directional ternary pattern: A New texture descriptor for texture classification", *Comput. Vis. Image Underst.* no. 169, pp. 14–27, 2018. https://doi.org/https://dx.doi.org/10.1016/j.cviu.2018.01.004.

[21] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, "Local binary patterns and its application to facial image analysis: A survey", *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* no. 41, pp. 765–781, 2011. https://doi.org/10.1109/TSMCC.2011.2118750.

[22] A. Kurniawardhani, N. Suciati, I. Arieshanti, "Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri yang Invariant terhadap Rotasi", *JUTI J. Ilm. Teknol. Inf.* no. 12, pp. 48, 2014. https://doi.org/10.12962/j24068535.v12i2.a322.

[23] D. E.Lundqvist, A. Flykt, A. Öhman, "The Karolinska Directed Emotional Faces - KDEF", *CD ROM from Department of Clinical Neuroscience, in: Psychol. Sect., Psychology*, Karolinska Institutet, 1998.

[24] D.E. King, "Max-Margin Object Detection", *Comput. Vis. Pattern Recognit.* 2015. http://arxiv.org/abs/1502.00046.

[25] Y.-H. Lee, S. Zhang, M. Li, X. He, "Blind Inverse Gamma Correction with Maximized Differential Entropy", *ArXiv.* abs/2007.0, pp. 1-12, 2020.

[26] M.A. Farooque, J. S.Rohankar, "Survey on Various Noises and Techniques for Denoising the Color Image", *Int. J. Appl. or Innov. Eng. Manag.* no. 2, pp. 217-221, 2013.

[27] W. Yang, X. Zhang, J. Li, "A local multiple patterns feature descriptor for face recognition", *Neurocomputing*. no. 373, pp. 109-122, 2020. https://doi.org/10.1016/j.neucom.2019.09.102.

[28] J. C. Russ, "The Image Processing Handbook, Fifth Edition (Image Processing Handbook, Fifth Edit", *CRC Press, Inc*., United State, 2006.