Development Of A Machine Learning Based Micro Expression Recognition System

D.M.I.S. Dissanayaka
li Dept. of Electrical and Telecommunication Engineering
Faculty of Engineering,
South Eastern University of Sri Lanka
ireshasandamalidm@gmail.com

K.L.I.M Dissanayaka
Dept. of Information Systems Engineering
School of computing
University of Colombo
ireshmdev@gmail.com

M.N.M. Aashiq
Dept. of Computer Science and Engineering,
Faculty of Engineering,
South Eastern University of Sri Lanka
aashiqmnm@seu.ac.lk

***Abstract*—Micro-expressions are brief, spontaneous facial expressions that last for a short time frame (1/25 second to 1/5 second) and are subtle for the naked eye to perceive. These fleeting and spontaneous facial expressions often appear in high-stake situations when people try to suppress or conceal their actual sentiments. Therefore, Automatic micro-expression recognition using machine learning techniques is becoming a growing research area since it has many potential applications such as criminal interrogations, security, or psychological examination. In this study, feature vectors are generated by dividing micro expression videos into sets of frames using Local Binary Pattern on Three Orthogonal Planes (LBP-TOP), a spatiotemporal feature extraction technique, and another feature extraction is performed on the onset, apex and offset micro expression frames taking the difference of landmarks coordinates. For classification and recognition, Support Vector Machine (SVM) was used as the machine learning model which was mostly utilized in previous research with better performance. This research indicates that using both temporal and static features with a machine learning algorithm gives a moderate amount of accuracy for automatic microexpression recognition.**

Keywords—Emotion recognition, Feature extraction, Feature vectors, Micro-expressions, Support vector machine

I. INTRODUCTION

In social interaction, human beings communicate with each other to fulfill day-to-day activities using many physical ways such as speaking, body gesturing, and displaying emotions. When we consider displaying emotions, facial expressions are the most natural and effective way to identify a human's emotions, given that 55% of the human's communication patterns are expressed by facial expressions [1]. Facial expressions can be distinguished as either micro or macro expressions by their duration. Macro expressions are the normal expressions that are fairly easy to notice in day-to-day interaction with people and last between 0.5 seconds and 4 seconds [2]. Micro-expressions are extremely quick facial expressions that a person attempts to either deliberately or unconsciously conceals their genuine emotions [3]. In 1966, Micro-expression was initially detected as micromomentary facial expressions by Haggard and Isaacs and later by the work of Ekman and Friesen in 1969 [4]. Ekman indicated that micro-expressions are important clues for detecting when people were in high-stakes situations in which people have something valuable to gain or lose [5]. He also developed and introduced a Micro-Expression Training Tool (METT) to find concealed emotions that are imperceptible for most people without taking the training sessions [6].

The micro-expression can be outlined by its time evolution, amplitude and symmetry. According to time evolution, onset, apex and offset are the three key moments, described as the moment when the Micro Expression (ME) starts, the moment of maximum amplitude and the moment when it fades out respectively [7]. Ordinary people without training only perform slightly better than chance on average for the micro-expression recognition task since MEs are characterized by short duration and low intensity. So, the computer vision and machine learning methods for an automatic micro-expression recognition system are very much needed. Therefore, Over the previous couple of years, automatic facial micro-expression analysis has acquired increasing attention from experts across different disciplines because of its potential applications in various multidisciplinary domains, ranging from behavioral psychology, forensic investigation,human-computerr interaction, business negotiation, clinical diagnosis, deceit detection and security systems [3].

Furthermore, there is a taxonomy with eight universal micro emotions according to the psychological theory such as disgust, surprise, sadness, anger, happiness, fear, contempt, and neutral [2], [8]. To classify these emotions, different machine learning algorithms have been used in previous studies. Neural networks, Extreme Learning Machine (ELM), k Nearest Neighbors (KNN), random forest classifier, decision trees and Support Vector Machine (SVM) are some of the supervised machine learning algorithms used in the existing studies [9]–[11]. According to the previous studies, SVM seems to be the most utilized machine learning model for the micro emotion recognition purpose due to its good generalization performance irrespective of bias in training sample [2]. In this study SVM machine learning classifier is used for classification in both of the implemented methods.

The organization of this paper is as follows. The literature survey which is related to our study is explained in section II. Proposed method is discussed in section III which includes data collection /preparation, feature extraction and classification. Experimental results and discussion of the model is explained in section IV Conclusion is given in section V .

II. RELATED WORK

In the literature, most of the previous methods have been developed for micro expressions recognition based on temporal data while few used static data for other methods. In one of the past studies, they conducted their micro expression recognition using Local Binary Patterns (LBP), a static feature extraction technique performed on apex micro expression frames and LBP on Three Orthogonal planes, a temporal feature extraction technique with the ELM and SVM machine learning algorithm to achieve promising results with an efficient and higher learning speed. Their results show that recognition of micro expression using temporal features is more effective than recognition of micro expression using static features [2]. Apart from that, LBP-Six Intersection Points (LBP-SIP) and LBP-Three Mean Orthogonal Planes (LBP-MOP) are the two feature extraction approaches for micro-expression recognition. In LBP-SIP method, neighbor points from the three orthogonal planes were computed by considering only six unique intersection points on the three intersecting lines of three orthogonal planes removing redundancy and conversely in LBP- Mean Orthogonal plane, the mean image of each stack of orthogonal planes were computed. Both LBP-SIP and LBP-MOP approaches, tested on two existing spontaneous micro-expression datasets Chinese Academy of Sciences Micro expression (CASME II) and SMIC consume significantly less time and space in feature extraction tasks [12]. On the other hand, some researchers attempted to improve the method analyzes the movement variations that occur in a given time frame using image differences. Here two algorithms were used and evaluated to determine if a ME occurred at a given frame t and integrated into a full micro expression analysis. In order to ensure the robustness of the algorithm, the raw response of the classifier is further post-processed. This system recorded of TPR (True Positive Rate) of 86.95% [7]. Furthermore, learning a spatio-temporal feature representation with expression state constraints was presented as a method for recognizing micro expressions. As the first step Convolutional Neural Network (CNN) was utilized to encode the spatial characteristics of the facial expression at different states. Then the learned model was transferred to a micro expression temporal feature learning stage which adopts Long Short-Term Memory (LSTM) with expression-state constraints. Consequent LSTM was able to encode the temporal characteristics of the different states of the micro expression. Here experimental results were shown that a 60.98% of accuracy [13]. Moreover, some researchers attempted to encode micro expressions features using LBP-TOP and Bi-Weighted Oriented Optical Flow (Bi-WOOF) techniques by utilizing only the apex frame. This system achieved a 48.15% and 59.26% of accuracy for LBP-TOP and Bi-Woof techniques respectively [14]. The comparative analysis of previous related studies which are linked with our proposed methodology is depicted in TABLE I.

TABLE I. SUMMARY OF LITERATURE REVIEW

Title	Dataset	Features	Algorithm	Performance (%)
Facial micro expression recognition: A machine learning approach	CASME II	LBP	SVM	94..11
			ELM	92.73
		LBP-TOP	SVM	96.26
			ELM	97.65
Detecting micro expressions in real time using high speed video sequence	CASME II	Frame differences	Custom algorithm	86.95
Micro expression recognition with expression-state constrained Spatio-Temporal feature representation	CASME II	Custom spatiotemporal feature	CNN and LSTM	60.98
Micro-expression recognition from video using apex frame	CAS(ME)^2	LBP-TOP	Custom algorithm	48.15
		Bi-Woof		59.26

Note: SVM - Support Vector Machine, ELM - Extreme Learning Machine, CNN – Convolutional Neural Network, LSTM - Long Short-Term Memory, Bi-Woof - Bi-Weighted Oriented Optical Flow

Datasets which contain either genuine or posed micro expressions are a central part in micro expression research. The micro emotion datasets which are publicly available can be categorized into acted or spontaneous samples. All these datasets were collected under stable laboratory conditions such as artificial lighting condition, and the subjects could not move their heads freely and had to keep a near-front pose due to the micro expressions are difficult to evoke. Acted micro expressions samples were collected by asking the subject to perform basic emotions with low amplitude and go back to the neutral state instantly and these are not genuine emotions. On the other hand, spontaneous samples were elicited by asking the subject to watch several sentimental videos and attempt to conceal or suppress all their facial expressions occurring in the experiment. Some of the available spontaneous datasets are SAMM, SMIC, CASME, CASME II, CAS(ME)2 and conversely Polikovsky, USF-HD are the posed (non-spontaneous) datasets [15]. Summary of publicly available datasets containing micro facial expressions are presented in TABLE II.

TABLE II. SUMMARY OF PUBLICLY AVAILABLE DATASETS CONTAINING MICRO FACIAL EXPRESSIONS

Dataset	Participants	Resolution	FPS	Samples	Emotion classes
Polikovsky [16]	11	640×480	200	13	7
USF-HD [15]	N/A	720×1280	29.7	100	4
CASME [17]	35	640×480, 1280×720	60	195	7
SMIC [18]	20	640×480	100 and 25	164	3
CASME II [19]	35	640×480	200	247	5
SAMM [20]	32	2040×1088	200	159	7
CAS(ME)^2 [21]	22	640×480	30	250 macro, 53 micro	4

III. PROPOSED METHODS

In this research two methodologies were implemented for micro emotion recognition from static facial images and temporal data. First methodology consists of several steps and feature extraction performed on static facial images such as onset, apex and offset frames taking differences of landmarks coordinates.

In the second methodology, the process of micro expression recognition is made up of three major phases including data collection /preparation, feature extraction and classification. Features were extracted from temporal data using a holistic approach to recognize micro expressions accurately and efficiently [2].

A. Micro emotion recognition from static facial images

Initially the onset, apex and offset frame indexes have been read from a CSV file, provided by the CASME II dataset which contains 247 micro expression samples [19]. Each record in the CSV file represents a single video which contains a single occurrence of a micro emotion. The faces have been detected from the onset, apex and offset frames using a face detector model that has been obtained from an external source. Another pre-trained model for detection of the landmarks within a face has been used. The landmarks detection model detects coordinates of 68 landmark points inside a face in each of those frames. Then difference between the onset, apex frames and also difference between apex, offset frames have been calculated to generate a variation graph by computing the movement of landmark points which translates to 136 coordinate movement values. Then, that variation graph (136 coordinate movement values) was labeled a micro emotion type, also obtained from the corresponding CSV record. An appropriate classifier which is known as SVM has been used for recognizing the emotion type.

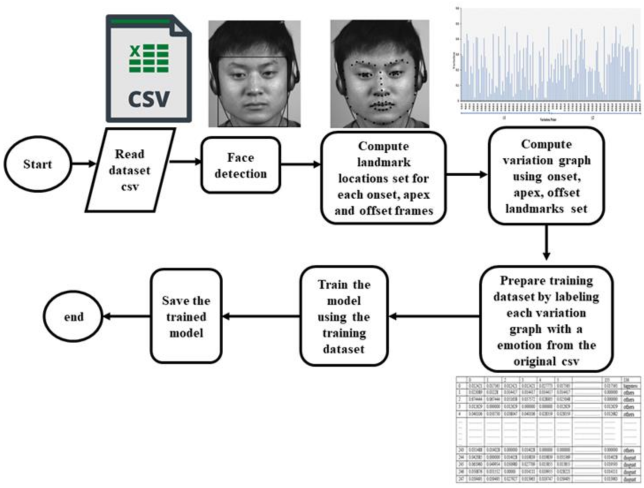

Fig. 1. Proposed model of micro-emotion recognition

B. Micro emotion recognition from temporal facial images

Dataset description and preparation: Micro-expression samples used for this research were acquired from CASME II dataset contains video sequences captured at 200 fps of 26 subjects. In total, 247 micro-expressions were elicited [19]. This dataset can be categorized into spontaneous samples which were retrieved from participants who were made to watch highly emotional video clips. A screen was placed before each of the participants and a high-resolution camera was used to record their emotions while watching the videos. Video recordings from each participant were divided into frames and pre-processed by labelling of samples with onset, apex and offset frames before being made available to the public. Onset, apex and offset frames indicates the three key moments of micro emotion such as the moment when the ME starts, the moment of maximum amplitude and the moment when it fades out. The main drawback of available spontaneous micro- expression database is that all the data were captured in unnatural conditions.

Among a total of 247 data samples from CASME II dataset 230 samples were only used for the experiment involving the use of temporal data (image sequence) since some emotion classes have a number of samples and some classes have a smaller number of samples. 17 samples whose coding were not correctly included were left out. Details on the number of samples for each micro expression class are presented in Table 01. Pre-processing involved conversion of frames from RGB into grey-scale images.

TABLE III. NUMBER OF MICRO EXPRESSION SAMPLE FOR EACH CLASS IN CASME II AND NUMBER OF SELECTED IMAGE SEQUENCE

Class	Original CASME II data	Number of selected image sequences
Disgust	60	59
Happiness	33	30
Repression	27	24
Surprise	25	25
Other	102	92
Total Samples	247	230

Feature extraction: Extracting features from micro expression dataset is critical since encoding quick facial movement is difficult although feature representations are rich. In this methodology, features were extracted from image sequence using Local Binary Pattern on three orthogonal planes.

LBP is a local descriptor of the image based on the neighboring pixel. The main intention of the LBP is to compare the center pixel values of a gray scale image with the value of all the neighboring pixel resulting in an 8-digit binary number converted into decimal.

For a given adjacent pixel **p** in the image has the intensity value $\mathbf{v_p}$ **,** radius **r** and **N** neighboring pixels, a binary label is assigned to each of the neighboring pixel. 1 is assigned to the pixel If the intensity value of the adjacent pixel is equal or greater than that of the center pixel and otherwise 0. Then LBP value is given by the following depicted equation [2].

$$LBP_{x_p,y_p} = \sum_{i=0}^{N-1} (g_i - g_p)2^i \tag{01}$$

where, x_p,y_p indicates the coordinates of the center pixel, g_p represents the intensity value of the center pixel and g_i denotes the value of the ith neighboring pixel. 2^i denotes the weight of corresponding neighboring pixel locations and f(x) is a sign function depicted in (02) [2].

$$f(x) = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{02}$$

LBP technique can only extract features from the images known as static data. Therefore, Zhao and Pietikainen proposed a one of the variants of ordinary LBPs named LBP-TOP to analyze textures that are time dependent.

For a given video with time length T, LBP value is calculated in three planes; XY, XT and YT where XY provides spatial information while XT and YT supplies both spatial and temporal information about space-time transitions [2] . The final feature vector is generated by calculating the LBP value for each of the plane that is concatenated into a single histogram.

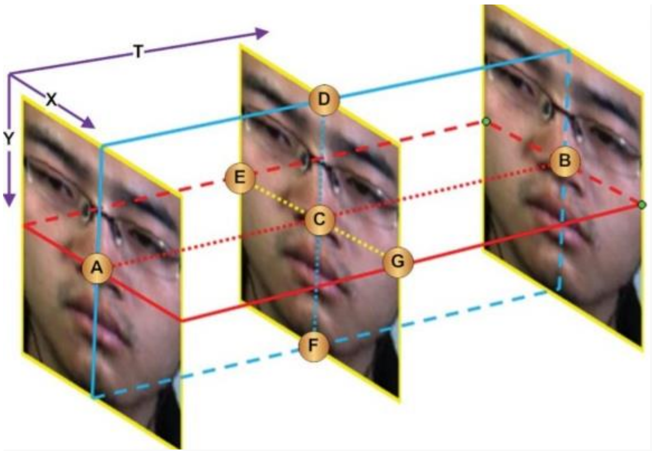

Fig. 2. LBP-TOP illustration [12]

To extract LBP-TOP features, Image sequences which is in CASME II database were read first and converted to grey-scale. Subsequently LBP values were computed in XY, XT and YT planes along with their respective histograms. The histogram corresponds to the three orthogonal planes were concatenated resulting in 3×256 feature vector for each image sequence. Similar to the previous experiment, LBP uniform binning was applied to reduce the length of the feature vector extracting only uniform patterns from the 256 histogram bins. As a consequence, the length of the feature vector was reduced from 256 to 59 [2]. This produced a 3x59 feature vector for each image sequence which was converted into a single row vector of size 1×177. Therefore, for all 230 samples used for these experiments, we had 230×177 features.

Classification: Five micro-expression classes (happiness, disgust, repression, surprise and others) were used for the experiment. With the tuned parameters, LBP-TOP features were extracted for the interval of occurrence of a micro emotion in each video and generated a dataset by labelling those feature vectors with micro emotion names. Then, the SVM model was trained and the accuracy was calculated using the generated dataset. The generated dataset was divided into two sets such as training and testing. The testing data set was kept in memory for another testing phase. At the first phase, we tested the accuracy of the model with the assumption that apex frame and the interval of the micro emotion occurrence were known. At the second phase, feature vectors were generated for every possible interval for each and every video that was used to create the testing data set previously. Second and final phase of testing was done without the assumption that the apex frame and the interval of ME occurrence were known, demonstrating a real-world scenario where the only input is a frame sequence (video).

Implementation of proposed method with a GUI: Out of all GUI (Graphical User Interface) methods in python, tkinter is the fastest and easiest way to create the GUI applications. It is a standard Python interface to the Tk GUI toolkit shipped with Python. The main elements of the GUI design for this system are Button, Entry, Label, BooleanVar, Radio button, Scale, String Var, Toplevel, Canvas and Option Menu. The Entry widget is a basic widget of Tkinter used to get single line of text input from the user of an application. In this GUI, five entry widgets were created to input radius, number of neighbors within a given radius and number of frames in a one-time interval when extracting features using LBP and LBP-TOP. Fourteen buttons, standard Tkinter widget were configure with functions and methods to select a video from a list, establish play/pause state, generate LBP and LBP-TOP features and histograms, display the plotted histograms and generate LBP and LBP-TOP histogram for entire videos in CASME II dataset, generate and store LBP and LBP-TOP feature vectors for entire dataset, display training window, load testing data, generate test results and display timeline plot window. Six tkinter Label widgets were used to implement display boxes that can use only one font at a time to display text. The six tkinter string variables were applied to manage the value of Entry widget and Label widget effectively. Progressbar widget was used to reassure some functions in program executing whereas Scale widget provides a sliding bar to select the index of frame while video is playing. The ListBox widget is used to display the name of the video files in CASME II dataset and according to the requirement, items can be selected from the given list. OptionMenu widget was a dropdown menu that displays a group of LBP methods. Radio button was used to implement one-of-two selections from LBP-TOP pattern. Finally, the Canvas widget was used to display the selected video file from the list. GUI model was designed to display the extracted features as histograms, the frame sequence, and to display the occurrence of a micro emotion in the selected video.

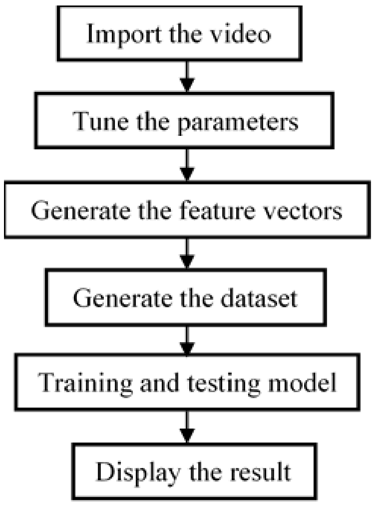

Fig. 3. Operational Flow chart of GUI

Fig.3 illustrates operational flow chart of the GUI. Firstly, we imported a video, selected from video list box. Then we experimented by tuning parameters of LBP-TOP pattern extraction and plotting them to find a better correlation between the extracted pattern/features and the occurrence of a micro emotion in a given video. With the tuned parameters, we extracted LBP-TOP features for the interval of occurrence of a ME in each video and generated the dataset by labelling those feature vectors with micro emotion names. Finally, the results which consist of accuracy and occurrence of a micro emotion in a selected video were displayed after completing training and testing of the model.

IV. EXPERIMENTAL RESUL AND DISCUSSION

In this section, all the results obtained from both feature extraction techniques and performance of models are presented and discussed.

A. *Classification with differences of landmarks coordinates*

In the first experiment, variation graph was generated calculating landmarks coordinates differences between onset frames and the apex frames and also differences between the apex frames and the offset frames. The differences were calculated for each landmark point inside the face. For each emotion this type of variation graph was generated and it was the input to the emotion recognition model.

TABLE IV. EXTRACTED FEATURES FROM ONSET, APEX AND OFFSET FRAMES

	0	**1**	**2**	**3**	**4**	**5**	**…………………**	**135**	**136**
0	0.012421	0.017565	0.012421	0.012421	0.027773	0.017565	………………….	0.017565	happiness
1	0.023089	0.032281	0.014417	0.014417	0.014417	0.014417	………………….	0.000000	others
2	0.674444	0.067444	0.051638	0.037572	0.028005	0.025048	………………….	0.000000	others
3	0.012629	0.000000	0.012629	0.000000	0.000000	0.012629	………………….	0.012629	others
4	0.040106	0.050730	0.038047	0.040106	0.028359	0.028359	………………….	0.012682	others
… … ………	……… ………… ………… … ………	……… ………… ………… … ………	……… ………… ………… … ………	……… ………… ………… … ………	……… ……… ………… …… ………	……… ……… ………… ………… …	…………………. …………………. …………………. …………………. ………………….	……… ……… ……… ……… ………	……… ……… ……… ……… ………
243	0.031488	0.014028	0.000000	0.014028	0.000000	0.000000	………………….	0.000000	others
244	0.042085	0.000000	0.014028	0.019839	0.019839	0.031369	………………….	0.014028	disgust
245	0.061960	0.049954	0.030980	0.027709	0.013855	0.013855	………………….	0.019593	disgust
246	0.050876	0.031552	0.00000	0.014111	0.019955	0.028221	………………….	0.014111	disgust
247	0.039495	0.039495	0.027927	0.013963	0.019747	0.039495	………………….	0.013963	disgust

The data frame which is depicted in Table 4 contains the extracted features from onset, apex and offset frames for each emotion and each of the rows in the data frame corresponds to a single variation graph. One hundred thirty-six (136) coordinates differences in the data frame were taken as X value and emotion label was taken as Y value to feed the emotion recognition model. Then feeding data were split into two sets such as training and testing to calculate the accuracy of making predictions which was the emotion representing from onset, apex and offset frames.

To analyze the data, it is very important to select machine learning algorithm considering accuracy, performance and nature of the input data. Therefore SVM (Support Vector Machine) algorithm was used for classification purpose. When calculating the accuracy, it was very important tuning the SVM hyper-parameters, C which is known as regularization parameter controls the tradeoff between smooth decision boundary and classifying training point correctly and gamma defines how far the influences of a single training example reaches. Finding optimal hyper-parameter is a very hard task to solve but it can be approximated by trying a limited number of combinations and see which values work best.

TABLE V. ACCURACY VARIATION WITH HYPER-PARAMETER TUNNING OF SVM

Gamma (γ)	**C (Regularziation Parameter)**	**Accuracy (%)**
0.1	1-54	61.54
	55-97	65.38
	98-99	57.69
0.2	1-24	61.54
	25-44	65.38
	45	61.54
	46-76	57.69
	77-92	53.85
0.3	1-14	61.54
	15-29	65.38
	30-38	57.69
	39-43	53.85
0.4	1-10	61.54
	11-20	65.38
	21	61.54
	22-26	57.69
	27	53.85
0.5	1-7	61.54
	8-14	65.38
	15-16	61.54
	17-19	57.69
	20	53.85

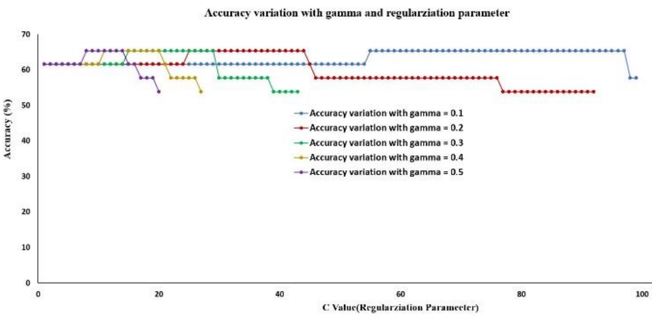

Fig. 4. Accuracy variation with gamma and regularization parameter

The results of the accuracy which vary with SVM hyper-parameters is plotted in Figure 4. According to Table 5, when gamma equals to 0.1, the regularization parameter is able to vary from 1 to 99 with a higher accuracy which is greater than 50%. The peak accuracy was 65.38% when the gamma value is 0.1 and the regularization parameter was between 55 to 97. Therefore, the gamma value equal to 0.1 and the regularization parameter equals to 80 was selected as hyper parameter for the emotion recognition model.

B. *Classification with LBP-TOP features*

This section contains results of training and testing performance for SVM model using LBP-TOP features from micro expression samples.

A total of 230 image sequence samples which consist of 30 happiness, 59 disgust, 24 repression, 25 surprise and 92 others were used for the experiment. Size of each feature vector obtained using LBP-TOP feature extraction was 1x177. A feature vector was generated for 30 frames window such that the apex frame sits in the middle of the window. There were two main parameters that needed tuning for generating a better version of a feature vector, namely the radius and the number of neighboring points.

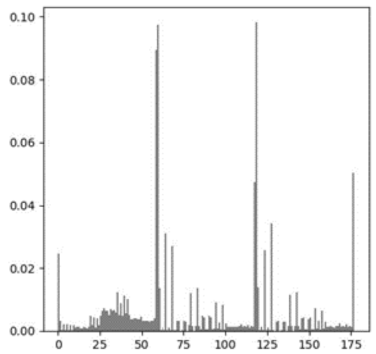

Fig. 5. Histogram of LBP-TOP patterns for 30 frames window

A graphical user interface was developed to visualize and find better correlations between the generated feature vectors and the different micro emotion classes. That made it easier to change different parameters and visualize the results to identity changes. Finally, the classification results for a given video are displayed in GUI. Graphical User Interface is depicted in Fig.6.

Fig. 6. Graphical User Interface

The performance of the model can be evaluated via accuracy, precision, recall and F1. Accuracy is the most intuitive performance measure defined as the ratio of correctly predicted observation to the total observation. Precision is interpreted as the ratio of correctly predicted positive observations to the total positive observations predicted. Recall calculates the ratio of correctly predicted positives to the all observations in actual class. F1 score is the weight average of recall and precision [2]. The equations for calculation of accuracy, precision, recall and F1 are presented in equations (3), (4), (5) and (6).

$$\text{Accuracy } = \frac{TP+TN}{TP+TN+FP+FN} \tag{03}$$

$$\text{Precision } = \frac{TP}{TP+FP} \tag{04}$$

$$\text{Reall} = \frac{TP}{TP+FN} \tag{05}$$

$$\text{F1} = \frac{2\times Precision \times Recall}{Precision + Recall} \tag{06}$$

For a given class TP stands for true positives which are the correct predictions for the class. TN is the true negatives and they are the predictions for the classes except the class considered whose true values are not equal to the considered class. FP represents the false positives which are the wrong predictions for the class. FN stands for false negatives and they are the predictions for the classes except the considered class whose true values are equal to the considered class.Confusion matrix is described in TABLE VI.

TABLE VI. CONFUSION MATRIX

		Predicted Values				
		happiness	disgust	surprise	repression	other
Actual Values	happiness	1	0	0	0	1
	disgust	1	5	0	2	0
	surprise	0	1	0	0	1
	repression	0	0	1	4	1
	other	0	2	1	0	2

Resulting confusion matrix where the model was evaluated against the test data set. The test dataset consist of 10% of the samples from the complete dataset (23 samples).

TABLE VII. PERFORMANCE OF SVM MODEL USING LBP-TOP FEATURES

Class	happiness	disgust	repression	surprise	other
Accuracy (%)	52.17	52.17	52.17	52.17	52.17
Precision (%)	50.00	62.00	67.00	0.00	40.00
Recall (%)	50.00	62.00	67.00	0.00	40.00
F1 (%)	50.00	62.00	67.00	0.00	40.00

Generally, temporal feature extraction methods had resulted in higher accuracy than static feature extraction methods but this research suggested otherwise. It should be noted that might have been due to an error in practical model evaluation.

V. CONCLUSION

This research reveals that automatic micro-expression recognition can be performed using temporal feature extraction techniques and static feature extraction techniques with a machine learning algorithm known as SVM. We have implemented two different micro-expression recognition schemes separately by utilizing the two different feature extraction methods, such as LBP-TOP and a customized approach. Features were extracted from onset, apex and offset frames calculating landmarks coordinates differences in the customized approach. In this study, the CASME II dataset which includes both temporal and static data samples, was used. Accuracies given by trained models using two feature extraction techniques are 65.38% and 52.17% for static only and the temporal method respectively. This implementation can be further improved with higher accuracy in future studies and it can assist in identifying criminals with bad intentions who are trying to suppress their emotions.

VI. REFERENCES

[1] S. Ramakrishnan, H. Ravishankar, P. Das, and S. Palaniswamy, "Emotion Recognition from Facial Expressions Using Prototypical Networks for Arbitrary Pose Images," p. 14.

[2] I. P. Adegun and H. B. Vadapalli, "Facial micro-expression recognition: A machine learning approach," Sci. Afr., vol. 8, p. e00465, Jul. 2020, doi: 10.1016/j.sciaf.2020.e00465.

[3] Y.-H. Oh, J. See, A. C. Le Ngo, R. C.-W. Phan, and V. M. Baskaran, "A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges," Front. Psychol., vol. 9, p. 1128, Jul. 2018, doi: 10.3389/fpsyg.2018.01128.

[4] G. Zhao, "Automatic Micro-Expression Analysis: Open Challenges," Front. Psychol., vol. 10, p. 4, 2019.

[5] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," in 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, Apr. 2013, pp. 1–6. doi: 10.1109/FG.2013.6553717.

[6] X. Shen, Q. Wu, and X. Fu, "Effects of the duration of expressions on the recognition of microexpressions," J. Zhejiang Univ. Sci. B, vol. 13, no. 3, pp. 221–230, Mar. 2012, doi: 10.1631/jzus.B1100063.

[7] R. Danescu, D. Borza, and R. Itu, "Detecting Micro-Expressions in Real Time Using High-Speed Video Sequences," in Intelligent Video Surveillance, A. J. R. Neves, Ed. IntechOpen, 2019. doi: 10.5772/intechopen.76155.

[8] M. Kowalska and M. Wróbel, "Basic Emotions," in Encyclopedia of Personality and Individual Differences, V. Zeigler-Hill and T. K. Shackelford, Eds. Cham: Springer International Publishing, 2017, pp. 1–6. doi: 10.1007/978-3-319-28099-8_495-1.

[9] Goutami Panchal, K N Pushpalatha, and Dayananda Sagar College of Engineering, Bangalore, "A Local Binary Pattern Based Facial Expression Recognition using K- Nearest Neighbor (KNN) Search," Int. J. Eng. Res., vol. V6, no. 05, p. IJERTV6IS050284, May 2017, doi: 10.17577/IJERTV6IS050284.

[10] S. P. T. Reddy, S. T. Karri, S. R. Dubey, and S. Mukherjee, "Spontaneous Facial Micro-Expression Recognition using 3D Spatiotemporal Convolutional Neural Networks," ArXiv190401390 Cs, Mar. 2019, Accessed: Jun. 26, 2021. [Online]. Available: http://arxiv.org/abs/1904.01390

[11] I. P. Adegun and H. B. Vadapalli, "Automatic recognition of micro-expressions using local binary patterns on three orthogonal planes and extreme learning machine," in 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), Stellenbosch, South Africa, Nov. 2016, pp. 1–5. doi: 10.1109/RoboMech.2016.7813187.

[12] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition," PLOS ONE, vol. 10, no. 5, p. e0124674, May 2015, doi: 10.1371/journal.pone.0124674.

[13] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations," in Proceedings of the 24th ACM international conference on Multimedia, Amsterdam The Netherlands, Oct. 2016, pp. 382–386. doi: 10.1145/2964284.2967247.

[14] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame," Signal Process. Image Commun., vol. 62, pp. 82–92, Mar. 2018, doi: 10.1016/j.image.2017.11.006.

[15] W. Merghani, A. K. Davison, and M. H. Yap, "A Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics," ArXiv180502397 Cs, May 2018, Accessed: Jun. 27, 2021. [Online]. Available: http://arxiv.org/abs/1805.02397

[16] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, UK, 2009, pp. P16–P16. doi: 10.1049/ic.2009.0244.

[17] Wen-Jing Yan, Q. Wu, Yong-Jin Liu, Su-Jing Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, Apr. 2013, pp. 1–7. doi: 10.1109/FG.2013.6553799.

[18] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," in 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, Apr. 2013, pp. 1–6. doi: 10.1109/FG.2013.6553717.

[19] W.-J. Yan et al., "CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation," PLoS ONE, vol. 9, no. 1, p. e86041, Jan. 2014, doi: 10.1371/journal.pone.0086041.

[20] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A Spontaneous Micro-Facial Movement Dataset," IEEE Trans. Affect. Comput., vol. 9, no. 1, pp. 116–129, Jan. 2018, doi: 10.1109/TAFFC.2016.2573832.

[21] F. Qu, S.-J. Wang, W.-J. Yan, H. Li, S. Wu, and X. Fu, "CAS(ME): A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition," IEEE Trans. Affect. Comput., vol. 9, no. 4, pp. 424–436, Oct. 2018, doi: 10.1109/TAFFC.2017.2654440.