IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, November 2022

Micro Expression Recognition Using Machine Learning Approach

M. Raghava[1], M. Mani Chandan[2], M. Sahithi[3], M. Hemanth Kumar[4], M. Aravind[5]
GMR Institute of Technology, Rajam, Andhra Pradesh, India
20341A05B3@gmrit.edu.in[1], 20341A05B4@gmrit.edu.in[2], 20341A05B6@gmrit.edu.in[3]
20341A05B7@gmrit.edu.in[4], 20341A05B8@gmrit.edu.in[5]

Abstract: *Micro-expressions are characterized by short duration and low intensity, hence, efforts to train humans in recognizing them have resulted in very low performances. Automatic recognition of micro-expressions using machine learning techniques thus promises a more effective result and saves time and resources. In this study, we explore the use of Extreme Learning Machine (ELM) for micro-expression recognition because of its fast learning ability and higher performance when compared with other models. Support Vector Machine (SVM) is used as a baseline model and its recognition performance and its training time compared with ELM training time. Feature extraction is performed on apex micro-expression frames using Local Binary Pattern (LBP) and on micro-expression videos divided into image sequences using a spatiotemporal feature extraction technique called Local Binary Pattern on Three Orthogonal Planes (LBP-TOP). Evaluation of the two models is performed on spontaneous facial micro-expression samples acquired from Chinese Academy of Sciences (CASME II).*

Keywords: Extreme Learning Machine, Support vector Machine, Local Binary Pattern, The apex frame, Feature selection, Micro-Expression

I. INTRODUCTION

In interpersonal communications, facial expressions, body movements, and language constitute the emotional expression system. Among them, facial expressions are the most important way for the communication among people, and they are divided into macro-expression (MaE) and microexpression (ME). MEs are spontaneous expressions, and even when people try to hide their inner emotions, their facial muscles move uncontrollably and produce ME. ME is a physiological stress reaction that occurs naturally and is not controlled by humans. Therefore, it is difficult to fake and suppress. ME reflects the true emotions of people and has practical value in the fields of medical treatment, polygraph detection, and criminal investigation. Currently, the MaE recognition task has been widely studied owing to a large number of MaE data sets available, and the proposed algorithms can reach over 90% accuracy scores on the public data sets. ME can be classified into six typical emotions: happiness, sadness, fear, anger, disgust, and surprise But owing to the small sample size problem, the ME data sets are split into three categories instead: positive, negative and surprise.

A typical ME recognition system can be divided into three stages: image preprocessing, feature extraction, and emotion classification. First, in the image preprocessing stage, image sequences are filtered to select frames that contain MEs, and the redundant and invalid frames are removed.Three ME data sets are used in this study: SMIC, SAMM, and CASME II. In CASME II, happiness is labeled as the positive class, and disgust and depression are labeled as the negative class. Emotions such as anger, disgust, contempt, fear, and sadness in SAMM are labeled as the negative class. In this study, SVM is used as the baseline recognition model, but because of its slow learning speed, ELM which has a faster learning speed is also explored for micro-expression recognition. Features are extracted from apex micro-expression frames using Local Binary Patterns (LBP) and extracted from the entire micro-expression videos using (LBP-TOP). The performance of the two models (SVM and ELM) are compared on both static and temporal features and their overall training time was compared. The models were evaluated using CASME II micro-expression database

II. LITERATURE SURVEY

[1] Zhang, J., Yan, B., Du, X., Guo, Q., Hao, R., Liu, J., ... &Liu,Y. (2022). Motion magnification multi-feature relation network for facial microexpression recognition. *Complex & Intelligent Systems*, 1-14.

- Microexpressions cannot be observed easily due to their short
- duration and small-expression range. These properties pose considerable challenges for the recognition of microexpressions.
- Thus, video motion magnification techniques help us to see small motions previously invisible to the naked eye. This study aimed to enhance the microexpression features with the help of motion amplification technology.

[2] Choi, D. Y., & Song, B. C. (2020). Facial micro-expression recognition using two-dimensional landmark feature maps. *IEEE Access*, *8*, 121549-121563.

- Emotion recognition based on facial expressions is very important for effective interaction of humans with artificial intelligence (AI) systems such as social robots.
- On the other hand, in real environment, it is much harder to recognize facial micro-expressions (FMEs) than facial general-expressions having rich emotions. In this paper, They propose a two-dimensional (2D) landmark feature map for effectively recognizing such FMEs.
- The proposed 2D landmark feature map (LFM) is obtained by transforming conventional coordinate-based landmark information into 2D image information. LFM is designed to have an advantageous property independent of the intensity of facial expression change.

[3]**Song, B., Li, K., Zong, Y., Zhu, J., Zheng, W., Shi, J., & Zhao, L. (2020). Recognizing spontaneous micro-expression using a three-stream convolutional neural network. *IEEE Access*, *7*, 184537-184551.**

- In this paper, we propose a three-stream convolutional neural network (TSCNN) to recognize MEs by learning ME-discriminative features in three key frames of ME videos.
- However, it is not sufficient to represent all characteristics of ME videos if only static-spatial components are considered
- TSCNN is used to achieve more promising recognition results when compared with many other methods.
- With the CASME II, SMIC-HS, SAMM, CAS(ME)2 , and CASME databases are 74.05% and 0.7327; 72.74% and 0.7236; 63.53% and 0.6065; 71.62% and 0.7129; and 70.73% and 0.6736, respectively.

[4]Topic, A., & Russo, M. (2021). Emotion recognition based on EEG feature maps through deep learning network. *Engineering Science and Technology, an International Journal*, *24*(6), 1442-1454.

- In this paper Brain-computer interface Electroencephalogram Emotion recognition Valence-arousal model Deep learning are discussed.
- It is a challenging task to develop an intelligent framework that can provide high accuracy for emotion recognition.
- The CNN is a feedforward network and it generally contains convolution layers, activation layers such as ReLU, and pooling layers
- HOLO-FM methods outperformed their approach with a classification accuracy of 88% and 88.45%, respectively.

[5]**Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., &Velastin, S. A. (2021). Facial expression recognition of instructor using deep features and extreme learning machine.** ***Computational Intelligence and Neuroscience*, *2021***

- In cross-sensor and cross-environment testing, the APCER for DenseNet, is 7.47%, and the BPCER is 16.01%. Using the same protocol, the proposed network yields APCER of 3.33% and BPCER of 8.91%.
- These geometry-based and appearance-based methods have the common disadvantage of having to select a good feature to represent facial expression.
- However, it tends to cause overfitting as it is based on the principle of empirical risk minimization (ERM)

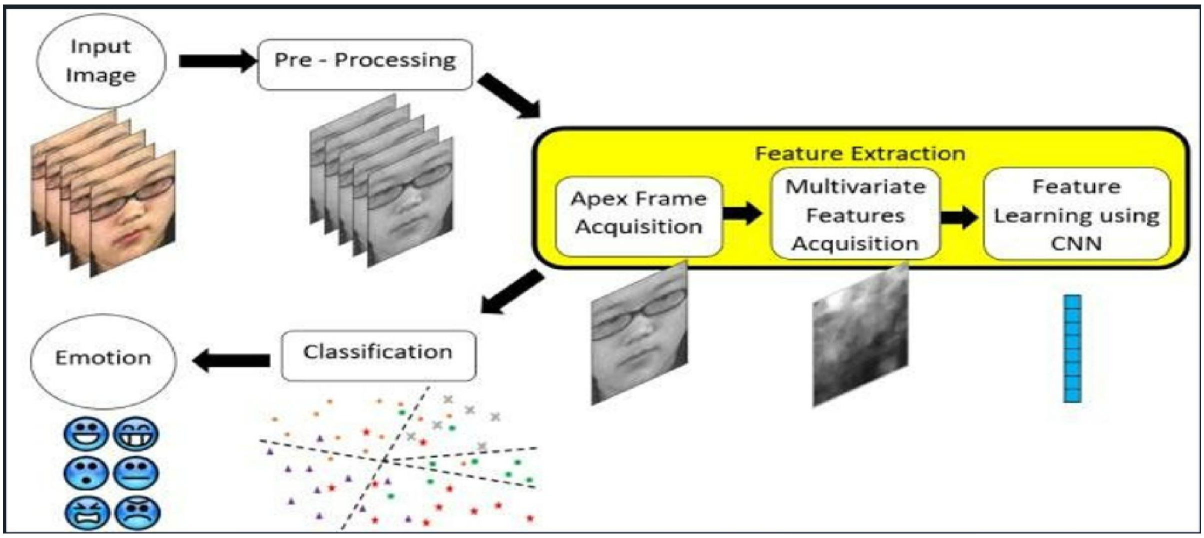
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, November 2022

[61].

- RELM is one of the variants of the extreme learning machine (ELM), which is based on a structural risk minimization principle. .e structural risk minimization principle is used to optimize the structure of ELM.
- DenseNet "convolutional layer" and AlexNet "drop7" give the best accuracy of 85% and 83%, respectively.

III. METHODOLOGY/RECENT TECHNOLOGY

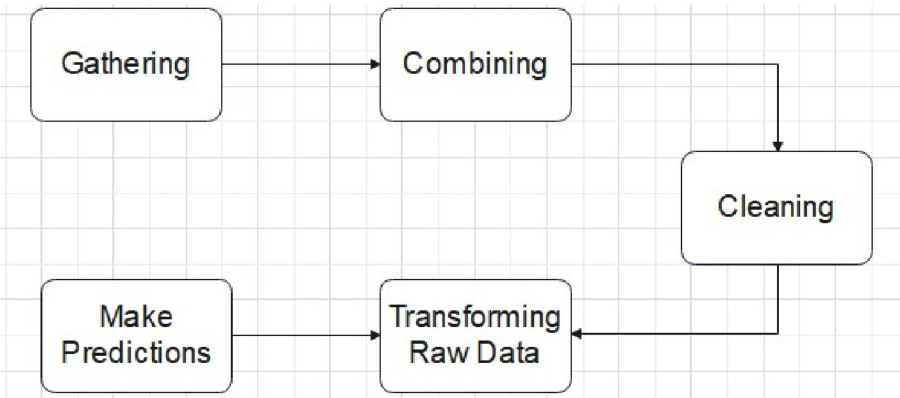
- The process of micro-expression recognition can be divided into three major phases. These include data collection/preparation, feature extraction and classification.



3.1 Data Collection

Data preparation is defined as a gathering, combining, cleaning and transforming raw data to make accurate predictions in Micro-expression Recognition. Micro-expression datasets can be categorized into acted or spontaneous samples. Micro-expression samples elicited spontaneously gives a true picture of what it really is when compared with the acted samples.

CASME II is the most recent, up-to-date and publicly available for research. The dataset consists of 247 facial micro-expression samples retrieved from 26 participants.



Some of the pre-processing carried out by the database owners include face detection and registration, division of recorded videos into frames, labeling of samples with relevant action units and coding with onset, apex and offset frames. Onset frame is the first frame where changes from the neutral expression occurs. Apex frame is the frame where the highest intensity of the expression is reached while offset frame is the last frame before facial expression changes to neutral CASME II database has a total of 247 micro-expression samples with a sampling rate of 200fps. Pre-processing involves conversion of frames from RGB into grey-scale images.

3.2 Feature Extraction

The process of extracting relevant features from data is critical to micro-expression recognition. To recognize micro-expressions accurately and effectively, features (both static and temporal) were extracted from both static and temporal micro-expression samples. These features were extracted using a holistic approach whereby spatio-temporal features are extracted from whole facial images instead of blocks of facial regions. This approach was used so as to reduce information redundancy and avoid complexity during feature extraction.

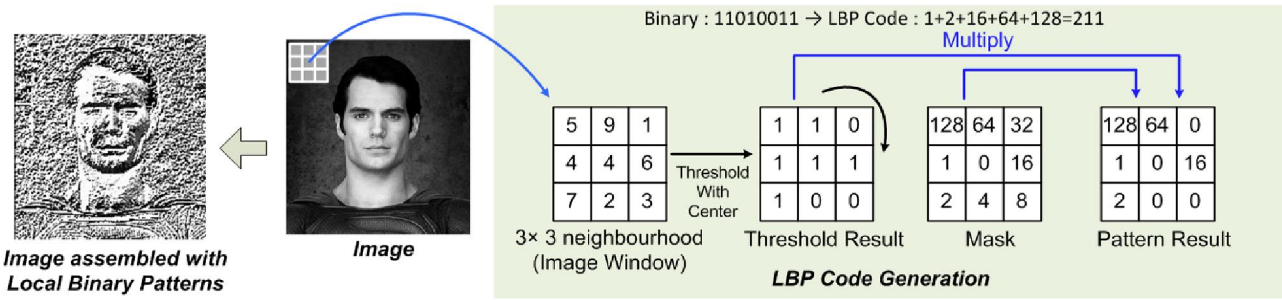
Feature extraction from apex micro-expression frames(static data) using LBP

The main idea of LBP is to compare the value of the center pixel C of an image with the value of its neighboring pixel P. If the center pixel value is greater than the neighboring pixel value, then, 0 is assigned, otherwise, 1 is assigned. This results into an 8-digit binary number comprising of 0s and 1s which is converted to decimal number and serves as the LBP value of the center pixel.

Local Binary Pattern is described mathematically as follows:

$$LBP(gp_x, gp_y) = \sum_{P=0}^{P-1} S(gp - gc) \times 2^p$$

where gpx, gpy represents the co-ordinates of the center pixel, gc denotes the intensity value of the center pixel and g is the intensity value of the ith neighboring pixel. 2i is the weight that corresponds to the neighboring pixel locations and f(x) is a sign function



Feature extraction from image sequence (temporal data) using LBP-TOP

LBP-TOP operates on three dimensional spatial-temporal in-formation, the computational cost also sharply increases compared to LBP on two dimensional static images. Suppose there is an image with frame width W and height H, the total number of LBP operations over all the pixels on the image is (W×H) times. Now we have another sequence of dynamic images with the same image size of width W and height H and the number of frames T,LBP-TOP needs to be applied to this three dimensional matrix. Then the total number of LBP operations over all the pixels climbs to (3 ×W×H×T) times.

Moreover, as LBP-TOP traverses all the pixels to compute LBP operations on XY, YT and XT planes respectively, it results in a frequent usage of nested loops in implementation.

Classification (Micro-Expression Recognition)

Five micro-expression classes were used for all experiments (disgust, happiness, repression, surprise and others) via a “one vs all” multi-classification approach. Training was performed using all samples belonging to each class as positive samples labeled as 1 while samples from the remaining four (4) classes were used as negative samples labelled as 0.

Training with Support Vector Machine Model (SVM)

Selection of SVM as baseline model was motivated by its success in past micro-expression studies. SVM uses a hyperplane to separate the group of data into their appropriate classes, considering that we have a dataset with two separate classes that are linearly inseparable. There could be more than a single hyperplane separating the classes and the one with the largest margin is chosen as the best/most correctly classified. Since SVM is a binary classifier, each micro-expression class was trained separately, and the average of their performance was calculated

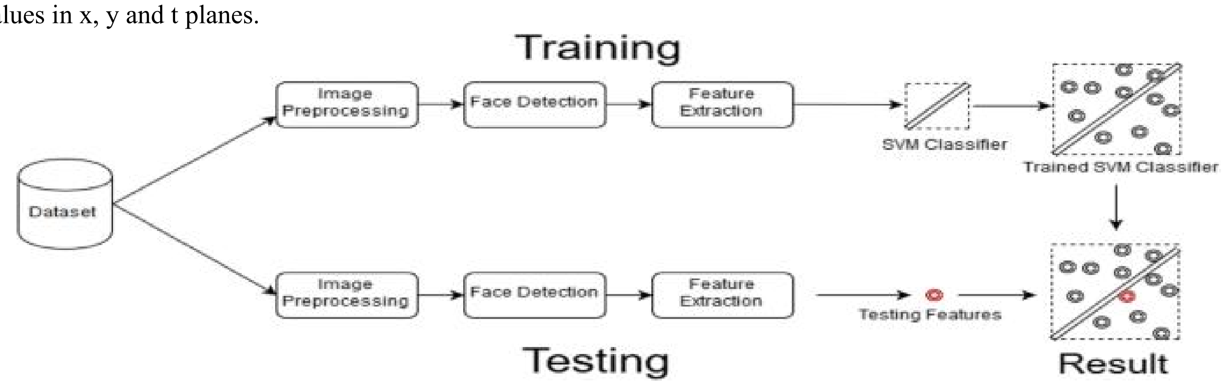
In this study, five-fold cross validation was used to divide all the samples into five subsets. SVM models were built using the two data formats that we had (apex frames and image sequences). A total of 220 samples were used for apex frame experiment while 230 samples were used for experiments performed using image sequences.

SVM Training with LBP Features

Training SVM on apex micro-expression frames was performed by loading (1 X 177) feature vector for 220 apex samples acquired after extraction of LBP features. Thereafter, five-fold cross validation was performed to divide the samples into five independent subsets. Linear SVM kernel was used for training. An average training accuracy of 96.30% was achieved with LBP features obtained from apex frames.

SVM Training with LBP-TOP Features

Similar procedure used for apex frame experiments was followed to train SVM on micro-expression image sequences. Feature vector of size 1 × 177 for 230 image sequences acquired from LBP-TOP feature extraction were loaded into each classifier. Optimization of the models was performed by recording training accuracy at varying LBP-TOP radii values in x, y and t planes.

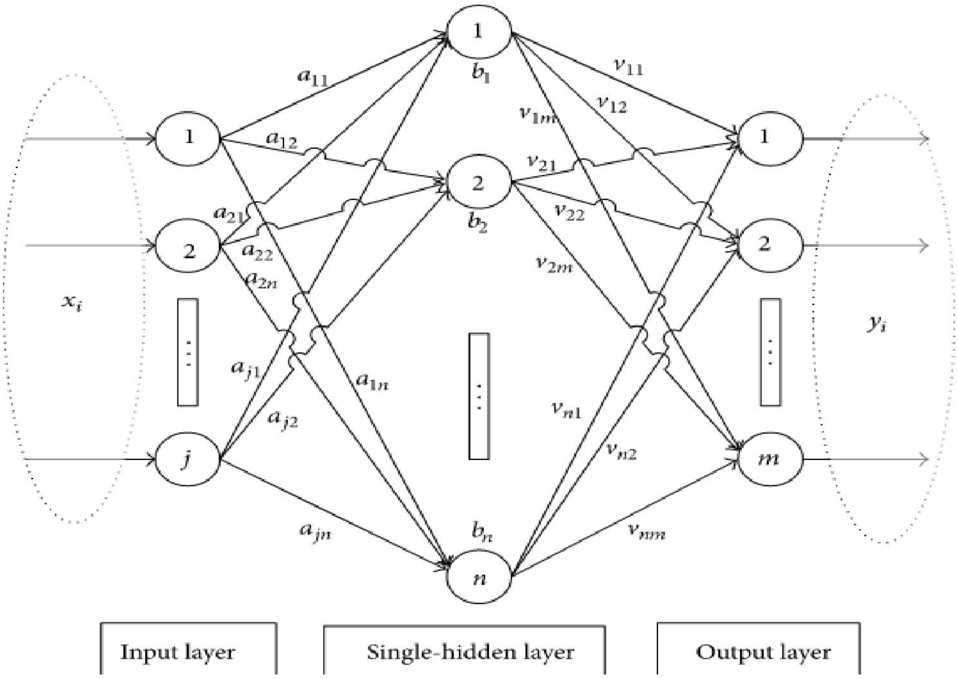


Training with Extreme Learning Machine Model

Extreme Learning Machine is a learning algorithm for the single hidden layer feedforward neural networks (SLFN) proposed by Huang . This learning algorithm was proposed to overcome the drawbacks of traditional feed-forward neural networks. According to Huang one of the major drawbacks of traditional feed-forward neural networks is their slow learning speed. Some of the advantages of ELM over other traditional learning algorithms of SLFN are highlighted below:

- ELM does not require parameter tuning
- ELM has an extremely fast learning speed as compared with other learning algorithms such as back propagation (BP) algorithm
- ELM is very useful in training SLFNs with many non-differentiable activation functions
- Extreme learning machines are feedforward neural networks for classification, regression , clustering with a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes (not just the weights connecting inputs to hidden nodes) need not be tuned.
- Unlike traditional feedforward network learning algorithms like backpropogation algorithm, the ELM does not use a gradient-based technique. With this method, all the parameters are tuned once. This algorithm does NOT need iterative training.
- Extreme Learning Machine algorithm is one of the most efficient machine learning algorithms in the neural networks. Because of the non-iterative training, all the parameters are tuned once. This results in a high training speed. Its implementation is easy to understand, and it can be used to solve complex problems.
- The feedforward neural network was the first and simplest type of artificial neural network. In this network, the information moves in only one direction—forward—from the input nodes, through the hidden nodes (if any) and

to the output nodes. There are no cycles or loops in the network



ELM Training with LBP Features

Training accuracy was recorded for varying number of hidden neurons (between 50 and 400) at an interval of 50 using LBP features. It was discovered that training accuracy increased with increasing number of hidden neurons. This informed the decision to select 300 as number of hidden neurons during validation of the model.

ELM Training with LBP-TOP Features

Optimization of parameters was performed by recording training accuracy at varying LBP-TOP radii values in x, y and t planes. These values were varied between 1 and 4 for x and y planes while variation for t plane.Training results showed that the highest average training accuracy (97.57%) using ELM was achieved at Rx=1, Ry=1 and Rt=3. To select optimal number of hidden neurons, average training accuracy was recorded between 50 and 500 at an interval of 50 for each class of micro-expression. The highest average training accuracy (97.54%) was achieved at 200 hidden neurons.

Classification with LBP Features from Apex Frames

Classification was performed for each of the five micro-expression classes, which produced five classifiers. For this experiment, a total of 220 samples (57 disgust, 25 happiness, 27 repression, 25 surprise and 88 others) were divided into training and test sets using five-fold cross validation. The performance of the classifiers was recorded five times (five-fold) and their mean calculated. Number of training samples used for each classifier was within the range of 174 and 176 while number of test samples was within the range of 44 and 46 samples.

Classification with LBP-TOP Features from Videos (Image Sequences)

For this experiment, a total of 230 image sequence samples (59 disgust, 30 happiness, 24 repression, 25 surprise and 92 others) were divided into training and test sets via a five-fold cross validation. Since one versus all classification was used, we had a total of five classifiers. Number of training samples used for each classifier was within the range of 184 and 186 while number of test samples was within the range of 44 and 46 samples. After LBP-TOP features were extracted and training performed, classification was performed on the test subsets using feature vectors from LBP-TOP and trained SVM model for each of the five classifiers.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT

Impact Factor: 6.252

Volume 2, Issue 1, November 2022

IV. RESULTS AND DISCUSSION

In this section, results obtained from all experiments are presented, analyzed and discussed. These include test performance for SVM and ELM models using LBP (static) features from apex micro-expression frames. It also includes test performance for SVM and ELM models using LBP-TOP (temporal) features from microexpression samples. Comparative analysis was carried out to show which of the feature extraction and classification models performed better.

Accuracy, precision, recall and F1 measures were used to evaluate the models.

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1 = (2∗Precision ∗ Recall) / (Precision + Recall)

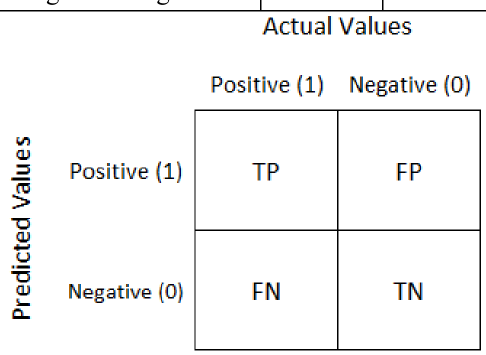
- The accuracy metric is one of the simplest Classification metrics to implement, and it can be determined as the number of correct predictions to the total number of predictions.
- The precision metric is used to overcome the limitation of Accuracy. The precision determines the proportion of positive prediction that was actually correct. It can be calculated as the True Positive or predictions that are actually true to the total positive predictions (True Positive and False Positive).
- Recall and Sensitivity is also similar to the Precision metric; however, it aims to calculate the proportion of actual positive that was identified incorrectly. It can be calculated as True Positive or predictions that are actually true to the total number of positives, either correctly predicted as positive or incorrectly predicted as negative (true Positive and false negative).
- F-score or F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class. It is calculated with the help of Precision and Recall. It is a type of single score that represents both Precision and Recall. So, the F1 Score can be calculated as the harmonic mean of both precision and Recall, assigning equal weight to each of them.

Comparative Result for SVM and ELM Training Time

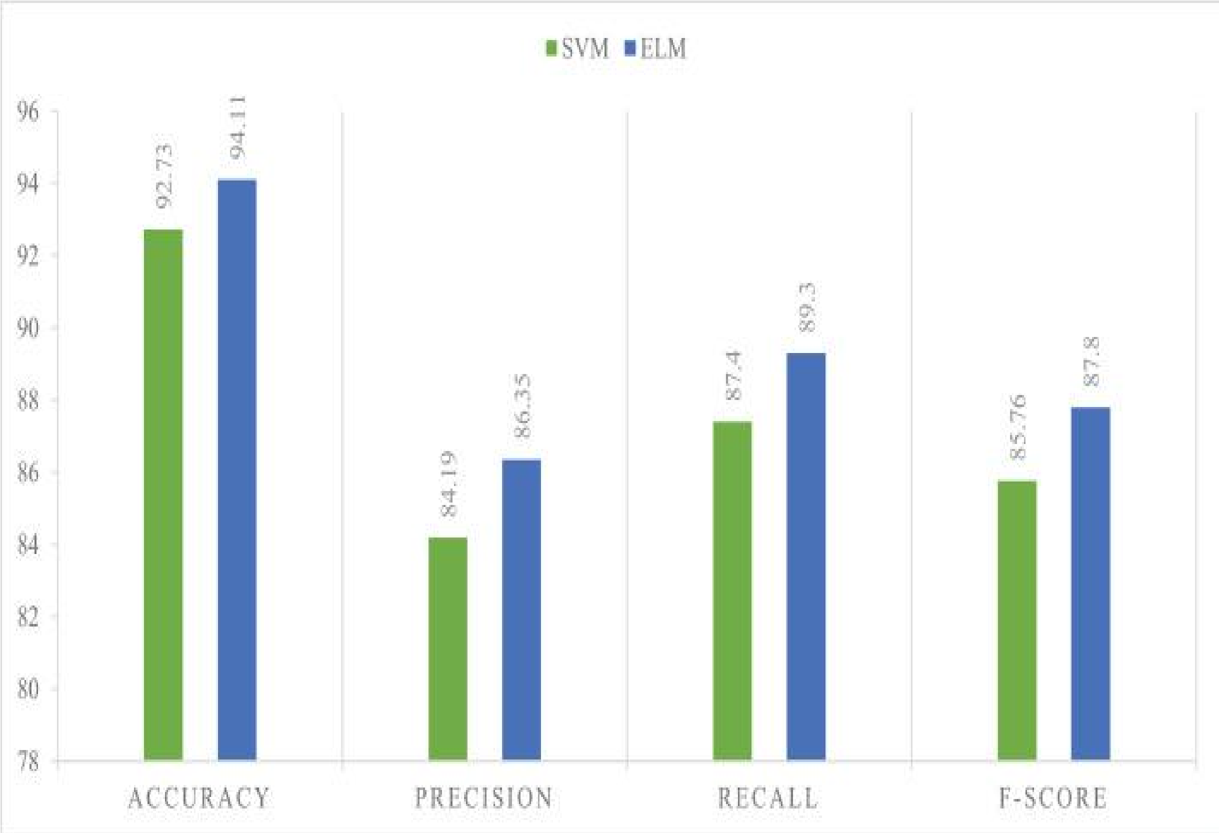
Training time for each class of micro-expression was recorded and the average calculated. Results were recorded using both SVM and ELM. Average training time using SVM was 0.3405 seconds while ELM had an average training time of 0.0499 seconds. Hence, we can deduce that ELM learns faster than SVM

Training time (in seconds) of SVM and ELM using LBP-TOP features.

Class	SVM	ELM
Disgust	0.2808	0.0468
Happiness	0.3806	0.0499
Repression	0.2434	0.0593
Surprise	0.2995	0.0530
Others	0.3182	0.0406
Average Training Time	0.3405	0.0499



Comparison of performance metrics between SVM and ELM using LBP-TOP Features



V. CONCLUSION

This study reveals that it is possible to recognize micro-expressions automatically and achieve promising results using temporal feature extraction technique (LBP-TOP) and a machine learning algorithm with an efficient and very fast learning speed (ELM). Two data formats were used for the experiments. The first format includes apex frames extracted from CASME II micro-expression samples. The second data format includes temporal image sequences from CASME II micro-expression samples. The first experiment was conducted by extracting LBP features from apex frame samples using SVM and ELM for classification. The second experiment was conducted by extracting LBP-TOP features from micro-expression image sequence samples using SVM and ELM for classification. Results show that supervised machine learning algorithms (SVM and ELM) can successfully classify micro-expressions into their appropriate classes. It also shows that recognition of micro-expression using temporal features is more effective than recognition of micro-expression using static features. However, this depends on the classification model used and the number of samples. Comparison between SVM and ELM training time also shows that ELM learns faster than SVM. An average training time of 0.3405 seconds is achieved for SVM while an average training time of 0.0409 seconds is achieved for ELM for the five selected micro-expression.

On a more general note, recognition of micro-expressions can assist in identifying criminals with bad intentions and are trying to suppress their emotions. This can act as a useful tool in achieving the 16th Sustainable Development Goal (Peace, Justice and Strong Institutions).

REFERENCES

[1]. Adegun, I. P., &Vadapalli, H. B. (2020). Facial micro-expression recognition: A machine learning approach. Scientific African, 8, e00465.

[2]. Zhang, J., Yan, B., Du, X., Guo, Q., Hao, R., Liu, J., ... & Liu, Y. (2022). Motion magnification multi-feature relation network for facial microexpression recognition. Complex & Intelligent Systems, 1-14.

[3]. Allaert, B., Bilasco, I. M., &Djeraba, C. (2020). Micro and macro facial expression recognition using advanced Local Motion Patterns. IEEE Transactions on Affective Computing, 13(1), 147-158.

[4]. Li, Y., Huang, X., & Zhao, G. (2020). Joint local and global information learning with single apex frame detection for micro-expression recognition. IEEE Transactions on Image Processing, 30, 249-263.

[5]. Sergeeva, A. D., Savin, A. V., Sablina, V. A., &Melnik, O. V. (2019, June). Emotion recognition from micro-expressions: search for the face and eyes. In 2019 8th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE.).

[6]. Choi, D. Y., & Song, B. C. (2020). Facial micro-expression recognition using two-dimensional landmark feature maps. IEEE Access, 8, 121549-121563.

[7]. Zhang, Y., Jiang, H., Li, X., Lu, B., Rabie, K. M., & Rehman, A. U. (2020). A new framework combining local-region division and feature selection for micro-expressions recognition. IEEE Access, 8, 94499-94509.

[8]. Xie, H. X., Lo, L., Shuai, H. H., & Cheng, W. H. (2020). An overview of facial micro-expression analysis: Data, methodology and challenge. arXiv preprint arXiv:2012.11307.

[9]. Song, B., Li, K., Zong, Y., Zhu, J., Zheng, W., Shi, J., & Zhao, L. (2020). Recognizing spontaneous micro-expression using a three-stream convolutional neural network. IEEE Access, 7, 184537-184551.

[10]. Lu, S., Xue, L., & Gu, X. (2021). Sparse Representation Classifier Embedding Subspace Mapping and Support Vector for Facial Expression Recognition. Wireless Communications and Mobile Computing, 2021.

[11]. Guo, C., Liang, J., Zhan, G., Liu, Z., Pietikäinen, M., & Liu, L. (2020). Extended local binary patterns for efficient and robust spontaneous facial micro-expression recognition. IEEE Access, 7, 174517-174530.

[12]. Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M., & Tao, J. (2021). Review of micro-expression spotting and recognition in video sequences. Virtual Reality & Intelligent Hardware, 3(1), 1-17

[13]. Saeed, S., Shah, A. A., Ehsan, M. K., Amirzada, M. R., Mahmood, A., &Mezgebo, T. (2022). Automated Facial Expression Recognition Framework Using Deep Learning. Journal of Healthcare Engineering, 2022.

[14]. Niu, B., Gao, Z., & Guo, B. (2021). Facial expression recognition with LBP and ORB features. Computational Intelligence and Neuroscience, 2021.

[15]. Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., &Velastin, S. A. (2021). Facial expression recognition of instructor using deep features and extreme learning machine. Computational Intelligence and Neuroscience, 2021