Using Spatial Covariance of Geometric and Shape Based Features for Recognition of Basic and Compound Emotions

Syed Farooq Ali[1], Khalid Ijaz[3], Faria Feroz[1], Junaid Jabbar Fiazi[2], Hassan Ali[2], , Nadeem Arif[4], Aadil Zia Khan[2], Raees Hafeez[2]

[1]Department of Software Engineering, SST University of Management and Technology C-II, Johar Town, Lahore, Pakistan

[2]Department of Computer Science, SST University of Management and Technology C-II, Johar Town, Lahore, Pakistan

[3]Department of Artificial Intelligence, SST University of Management and Technology C-II, Johar Town, Lahore, Pakistan

[4]Department of Informatics and Systems, SST University of Management and Technology C-II, Johar Town, Lahore, Pakistan

*Correspondence: khalid.ijaz@umt.edu.pk,

Citation | Ali S. F, Ijaz. K , Feroz. F, Fiazi. J. J, Ali. H, Arif. N, Khan. A. Z, Hafeez. R, "Using Spatial Covariance of Geometric and Shape Based Features for Recognition of Basic and Compound Emotions", IJIST, Vol. 6 Issue. 2 pp 754-771, June 2024

DOI | https://doi.org/10.33411/ijist/202462754771

Received | May 20, 2024, **Revised |** June 16, 2024 **Accepted |** June 18, 2024 **Published |** June 20, 2024.

Introduction / Importance of Study: Compound emotion recognition has been an emerging area of research for the last decade due to its vast applications in surveillance systems, suspicious person detection, detection of mental disorders, pain detection, automated patient observation in hospitals, and driver monitoring.

Objectives: This study focuses on emotions, highlighting the fact that the existing knowledge lacks adequate research on compound emotions. This research work emphasizes compound emotions along with basic emotions.

Novelty Statement: The contribution of this paper is three-fold. The study proposes an approach relying on geometric and shape-based features using SVM and then fusing the obtained geometric and shape-based features for both basic as well as compound emotion recognition.

Materials and Method: This study provides a comparison with six state-of-the-art approaches in terms of percentage accuracy and time.

Dataset: The experiments are performed on a publicly available compound emotion recognition dataset that contains images with facial fiducial points and action units.

Result and Discussion: The results show that the proposed approach outperforms the existing approaches. The best accuracy achieved is 98.57% and 77.33% for basic and compound emotion recognition, respectively. The proposed approach is compared with existing state-of-the-art deep Neural Network architecture. The comparison of the proposed approach has been extended further to various existing classifiers both in terms of percentage accuracy and time.

Concluding Remarks: The extensive experiments reveal that the proposed approach using SVM outperforms the state-of-the-art deep Neural network architecture and existing classifiers including Naive Bayes, AdaBoost, Decision Table, NNge, and J48.

Keywords: Compound Emotions; Fiducial Points; Shape Features; Position Based Features; SVM.

Introduction:

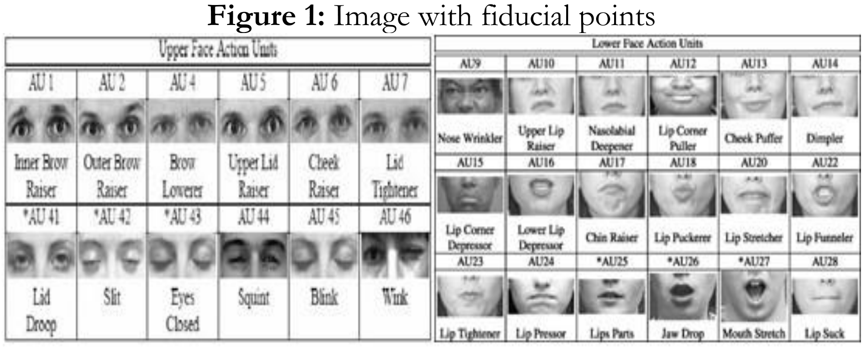
Emotion is typically experienced as a distinct mental state, intermittently often accompanied by physical changes, expressions, or activities [1]. There are seven basic emotions universally recognized: neutral, angry, disgust, sad, happy, surprise, and fear as shown in Figure 1. Du et al. were pioneers in discussing compound emotions, which are combinations of two or more basic emotions [2]. They identified 15 different compound emotions, including happily disgusted, happily surprised, sadly disgusted, sadly surprised, sadly angry, sadly fearful, fearfully disgusted, fearfully surprised, fearfully angry, angrily surprised, angrily disgusted, disgustedly surprised, appalled, hateful, and awed [2] as shown in Figure 2.

Emotions including compound emotions have their significance in many areas including surveillance-related problems, suspicious activity recognition, human behavior understanding, detecting mental disorders, synthetic human expressions, driver monitoring systems for safe driving, and many more. The rise in terrorist activities necessitates biometric systems capable of restricting access to unauthorized individuals and those displaying potentially harmful facial emotions, such as anger and fear. Individuals exhibiting these harmful emotions could be considered threats to sensitive buildings and should be restricted. A biometric system can detect facial emotions and flag suspicious individuals. Emotion analysis can also assist in job interviews, aiding interviewers in assessing a candidate's suitability. Automated emotion analysis may help a medical specialist judge the level of pain in his patients and save them time as well. The interaction between humans and computers will be more natural if computers can recognize emotional inputs. For example, detecting and using facial expressions in video games will provide a more natural interaction, thereby making video games more entertaining [3]. Emotion analysis can also help in criminal investigations.

Angles θ	Points
0	1
45	2
90	3
135	4
180	5
225	6
270	7
315	8

(a): Image (b): Neutral face with 78 fiducial points (c): Slope table of lips

Figure 1: Image with fiducial points



(a): Upper face action units (b): Lower face action units

Figure 2: (a). Upper face action units (b): Lower face action units

Automated emotion recognition can also help in developing sophisticated driver monitoring systems and ensuring reduced number of accidents [4][5][6][7]. High-anger drivers may end up with more collisions and crashes [8][9]. A survey taken from English drivers highlights the fact that drivers having more angry and aggressive behavior violate traffic rules more frequently [10]. To normalize the driver's emotions, a smart music-playing system can be developed based on emotion recognition [11].

The study in this paper proposes the use of novel features with different machine learning techniques for recognizing basic and compound emotions. The main contributions of this paper are as follows:

- The study proposes an approach for basic and compound emotion recognition and compares it with six state-of-the-art approaches - achieving accuracy as high as 98.57% and 77.33% for basic and compound emotion recognition, respectively.
- Our proposed approach is compared with existing state-of-the-art deep architectures, namely; VGG-16, ResNet-50, VGG-19, and DenseNet-121. The proposed approach achieves better percentage accuracy as compared to existing deep architectures.
- We also compare the proposed approach with various existing classifiers and observe that the proposed approach using SVM outperforms the other classifiers including Naive Bayes, Neural Network, AdaBoost, Decision Table, NNge, and J48.
- We also analyzed the results of our approach for combined emotions where we implemented our approach by combining the categories of basic (7) and compound emotions (15), hence, achieving improved accuracy on a total of 22 categories.
- The proposed approach also exhibits better time efficiency for 7 basic categories of emotions, 15 compound categories of emotions, and 22 combined categories of emotions in a dataset.

The organization of this paper is as follows. Section 2 describes the previous work mainly related to the classification of basic emotions and compound emotions. Section 3 explains the data set used in this research. In Section 4, the details of our proposed feature-based approach using different classifiers are presented. Section 5 gives a detailed explanation of the features. In Section 6, results are presented in terms of percentage accuracy and time efficiency while Section 7 provides the conclusion and gives future directions in which open issues and challenges related to basic and compound emotions are presented.

Literature Review:

The problem of emotion recognition dates back to the early 18th century, with significant advancements beginning with Darwin's book "The Expression of the Emotions in Man and Animals (1872/1998)" [12]. In 1938 Woodworth gave an idea of 6 emotions including happiness, sadness, disgust, anger, fear, and surprise by evaluating the responses of observers. Hence, a total of seven basic emotions were identified by adding neutral emotion to the existing six emotions. These laid the foundation for an exposition of compound emotions many years later.

Classification Based on Input Types:

Image and video based:

The development of facial action units marked a major revolution in the area of emotion recognition. Researchers realized that multiple facial muscles worked collectively to create certain emotions, carrying equal importance. Therefore, each facial action unit was assigned a number corresponding to a specific facial muscle. Detecting these facial action units, or their combinations, became integral to identifying emotions[13]. Initially, facial muscle action detection systems were developed only for faces having frontal view but then different algorithms and techniques were established that also deal with the temporal dynamics of facial actions. These algorithms can automatically segment input videos into facial expressions and recognize the temporal segments of 27 action units within the video [14].

Human facial expression analysis traditionally relied on either 2D static images or 2D video sequences, which struggled to handle large pose variations. Although, 3D face recognition techniques were widely used 3D facial expression recognition had rarely been found in literature due to the scarcity and non-availability of 3D facial expression datasets. The first attempt to create 3D facial expression repositories aimed to nurture the research in human facial expression by the authors in [15][16]. Facial expression analysis, specifically for people of Bangladeshi origin, was made and a modified set of action units was proposed for a more accurate classification of their facial expressions [17]. Facial expressions have gained popularity in recent years and various methods of optical flow-based detection of facial expressions have been developed in [18]. The Optical flow-based proposed technique is used to detect basic emotions using feature point tracking on facial image area (mouth, eyebrow, and eyes) yielding an accuracy of 76% (considering only the dataset of classified image sequences).

To solve the exact head, posture estimation, and facial expression tracking problems, two vision-based techniques were proposed in [19]. This work has been carried out in two phases; dynamic head pose estimation, followed by facial expression cloning. The dynamic head pose approach is used to visualize the 3D face pose from video images. The facial region is obtained from an input face image by using the HT skin color model, and the exact face is extracted. PCA is applied to the average image obtained from the set of training images. In the second phase, realistic facial expression is produced. Facial features are extracted and variations are tracked by using optical flow. Gaussian RBF (Radial Basis Function) is then used to produce the facial expressions. The results showed that the facial expression cloning technique automatically extracts the 3D head pose and generates realistic 3D facial expressions in real time. Tony et al. proposed an approach based on the histogram of oriented gradients and canny operator, that can precisely detect the deformation of eyes, eyebrows, and lips for facial expression recognition [20].

Classification Based on Approaches:

Geometric Approaches:

Different applications in Human-Computer Interaction (HCI) can benefit from recognizing expressions from images[21]. Previously developed methods used the 2D distribution of facial features. In Soyel et al., the authors achieved a 91.3% accuracy in facial expression recognition by using 3D facial feature distances [22]. Cohn-Kanade data sets have been widely used since 2000 for the development and evaluation of algorithms. Its refined version, containing a greater number of sequences, more subjects, and revised and validated labels was made available to promote further research in this area [23]. The baseline results using Action Appearance Models (AAMs) and SVM classifiers for both AU and emotion detection for pose data were also made available for the research societies that are working on emotion recognition.

In 2010, facial expression recognition became a significant area of interest in computer science [24]. A 2D appearance-based local approach was used to extract facial features for the recognition of four expressions including happiness, surprise, sadness, and anger, achieving an accuracy of 81% for grayscale images. An application was developed using an improved Active Appearance Model (AAM) in combination with a Probabilistic Neural Network (PNN) for the recognition of seven basic facial expressions [25]. The experimental results of this application achieved an average expression recognition accuracy of 96% on the JAFFE database which outperformed all the existing methods. It further motivated researchers to use artificial intelligence and machine learning algorithms in the future.

An approach using an AdaBoost classifier for face detection followed by token finding and matching with a back-propagation neural network was developed for facial expression recognition [26]. The real-time vision-based facial expression recognition system was developed that detected faces from the video stream and then employed Principal Component Analysis (PCA) to classify the expression into one of the five categories including neutral, happiness, surprise, sadness, and disgust, yielding an average accuracy of 88% [27]. Recognizing a person's feelings through facial expressions has also found applications in 3D gaming, using Active Appearance Models (AAM) to create a natural link between the player and computer [28]. The research in the paper detected facial expressions by monitoring the change of key features in AAM using Fuzzy Logic.

Template Based Approach:

There are two leading models through which humans can identify the classification of facial expressions, i.e., continuous model and categorical model in the field of cognitive science and neuroscience [29]. In different intensities, the continuous model explains how expressions of emotion can be viewed. The categorical model, on the other hand, recognized the image containing only a happy and surprised face. As both of these models had problems in recognizing the combination of emotion categories such as happily surprised versus angrily surprised versus surprise, the researchers worked on a revised model consisting of C distinct continuous spaces. Multiple (Compound) emotion categories can be recognized by linearly combining these -C-face spaces.

Compound emotion is the combination of two or more two basic emotions like happily surprised or happily disgusted [2]. The paper presents 22 distinct emotion categories consisting of both basic emotions having 7 categories and compound emotions having 15 categories. The Facial Action Coding System analysis shows that the production of these compound emotion categories is consistent with the subordinate categories they represent, i.e., happily disgusted expression combines the muscle movement observed in happiness. For the recognition of basic emotions, a database of 1610 images having 230 identities was created. A percentage accuracy of 96.86% was obtained when using both shape and appearance with a 10-fold cross-validation test. For the recognition of both basic and compound emotions, a database of 5,060 images corresponding to the 22 categories for the 230 identities was created that yielded an accuracy of 76.91% when using both shape and appearance features with 10-fold cross-validation.

Machine Learning Approach:

Machine Learning (ML) based techniques enable the system to automatically learn the applications and provide accurate results. These approaches are also used to enhance the performance of algorithms. Different machine learning techniques and algorithms including AdaBoost, support vector machines, and linear discriminant analysis were used to classify these basic emotions into 7 categories [30]. The best result with 93.3% accuracy was obtained by selecting Gabor filters using AdaBoost and support vector machines.

To deal with the challenge of head movement, face deformation, etc. [31] proposed a bi-modality multi-feature framework to detect emotion in the wild. The emotion recognition data is categorized into two components audio and video. To extract emotion recognition information from audio-based data STFT and open SMILE networks are used. Meanwhile, four diverse networks i.e., VGG-Face, Restnet18, Densenet121, and VGG16, for spatial features are used to capture information from video-based data. Each of the scores from image flow and audio flow is weighted with a grid-search optimization. The experimental outcome shows that our proposed framework is more robust for emotion recognition within the wild.

Facial Animation Parameters (FAPs) and multi-stream Hidden Markov models (HMMs) were used for automatic facial expression identification [32]. FAPs supported by MPEG-4 were used as a feature for facial expression classification. Experiments showed that the multi-stream HMM facial expression system significantly reduced the recognition error compared to the single-stream HMM system. The scarcity of image repositories and databases containing facial emotions created obstacles for researchers to evaluate their ideas, techniques, and algorithms effectively. To overcome this obstacle and fulfill the growing trend toward emotional intelligence, a database of accurate facial expressions was generated which shows that facial expressions depend on the person's emotional state [33]. Furthermore, some machine learning methods including SVMs, Bayesian networks, and decision trees were calculated to identify emotions. The database which is generated for facial expressions in which the HMMs respond to the pure emotional state of the subjects, was one of the main contributions.

Analyzing human emotions involves several aspects; these include facial expressions, posture and movement, language usage, and psychological measurement. Authors in [34] utilized photographs of a person's gait to predict the person's emotions using a Bayesian network. In this study, facial features were extracted using Gabor wavelets [35]. An ensemble machine learning algorithm was used to identify the eye and mouth area. The Ada boost classifier was used to classify these particular features for detecting the emotions. This method was deployed on several databases and they showed significant improvements.

Materials and Methods

A proposed research framework is shown in Figure 3. First of all, the compound emotion recognition dataset has been collected and then converted into grayscale images. Afterward, the shaped features and geometric features have been extracted and then concatenated. After the fusion of the features, the classifier has been implemented to detect the emotions in an image. $Y^{\wedge}$ is the output class of an emotion. In the state-of-the-art algorithm implementation, the Geometric Feature Extraction block, Shaped Feature Extraction block, and Features Fusion block can be replaced by the existing algorithms.

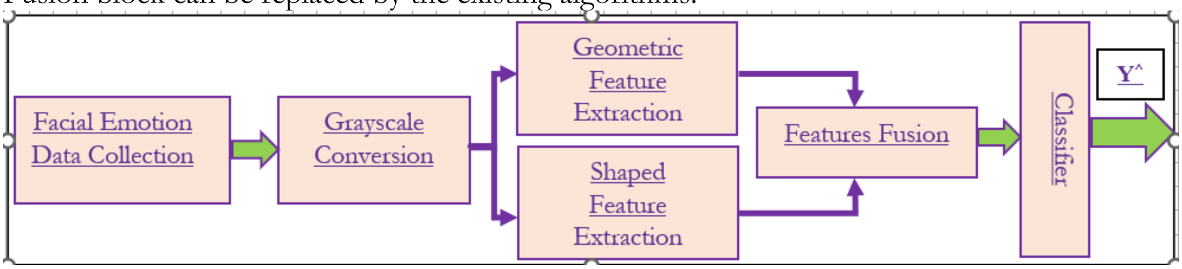


Figure 3: A Proposed Research Framework

Dataset Description:

The compound emotion recognition dataset was collected from the repository of images for both basic and compound emotions presented in [2]. This repository contains a total of 5060 images of 230 persons (subjects), out of which 1610 images contain basic emotions while 3450 images have compound emotions. Each image of a subject has a combination of action units and 78 fiducial facial points as described in Table 1. There is a total of 22 emotions out of which 7 are basic emotions and 15 are compound emotions. Basic emotions include neutral, angry, disgusted, sad, happy, surprised, and fearful as shown in Figure 5.

Table 1: Total number of images of all emotions (basic, compound), total facial points of the individual subject, total action units, total number of persons and face pose from the front

Total No. of Images	5060
Basic emotion images	1610
Compound emotion images	3450
Mixed emotion images	5060
Total facial points of individual subject	78
Total action units	26
No. of persons	230
Face pose	Front

Proposed Approach: The approach, presented in this paper, uses novel features of the slope table along with existing features using a Support Vector Machine (SVM) to achieve the best result in terms of percentage accuracy and execution time among existing approaches. The classifier is trained using features of either existing approaches or proposed approaches), fiducial points and action units of the training images (converted to grayscale). Then the trained classifier is used to categorize test images into two distinct categories, where category 1 corresponds to neutral, and category 2 corresponds to different basic and compound emotions.

Equation (1) shows that the Feature Vector (FV) along with Fiducial Facial Points and Action Units available in the data set of the training images are given to the classifier for training.

Trained Classifier = Classifier (FV (Ii), FP (Ii), AU (Ii)), (1)

Where $1 \leq i \leq N$, N = 5060,

FV (Ii) is the Feature Vector of the ith training image,

FP(Ii) is the Fiducial Facial Points of ith training image,

AU(Ii) is the Action Units of ith training image.

To detect the emotion of a given test image Itest, the FV of this test image is computed and given to the trained classifier along with fiducial feature points and action units as shown in equation (2). The trained classifier categorizes the image into 22 categories of emotions including basic and compound emotions. These categories are as follows: Neutral, Angry, Disgusted, Sad, Happy, Surprised, Fearful, Happily Disgusted, Happily Surprised, Sadly Disgusted, Sadly Surprised, Sadly Angry, Sadly Fearful, Fearfully Disgusted, Fearfully Surprised, Fearfully Angry, Angrily Disgusted, Angrily Surprised, Disgustedly Surprised, Appalled, Hatred and Awed.

$$Category_{1\leq j\leq 22} = \begin{array}{r} TrainedClassifier(FV(I_{test}), \\ FP(I_{test}), AU(I_{test})) \end{array} \quad (2)$$

Where $FV(I_{test})$ is the Feature Vector of the test image,

FP (I_{test}) is the Fiducial Facial Points of the test image,

AU (I_{test}) is the Action Units of the test image.

Fusion of Features:

A novel feature of the slope table is fused with the existing features including action units, length-to-width ratio of right and left eye, angles, area of triangle, ratios of triangles, and histogram of gradients for facial expression recognition. The slope table is a shape-based feature that varies significantly with the change in different emotions of the same subject. The slope table of the left eye, right eye, and lips is used to recognize the facial expression as a feature. The intuition for using this feature comes with the change in slopes (calculated using fiducial facial point and centroid) of different boundary points of the left eye, right eye, and lips with the changing emotions.

The boundary points of the left eye, right eye, and lips are already given as fiducial facial points in the repository of images [2]. The centroid of each facial component i.e., left eye, right eye, and lips is calculated by the formula given in equation (3). Hence, for each face, three centroids are calculated, first for the left eye, second for the right eye, and third for lips.

$$c = (c_x, c_y) = (\sum x_i/n, \sum y_i/n) \qquad (3)$$

Let pi be the distance between the centroid and the ith boundary point that is calculated with the help of the Euclidean Distance formula. The slope of the ith boundary is calculated with the help of equation (4).

m = tan = pix/piy (4)

Where pix is the distance along the x-axis of the boundary point to the centroid and piy is the distance along the y-axis of the boundary point to the centroid.

The calculated slopes of each point are stored in the slope table of that facial component as shown in Figure 1. The slope table for each facial component i.e., left eye, right eye, and lips is calculated separately, hence, three slope tables were used for each face. The eyes and lips have a boundary of 8 points and the nose have 4 points.

Action Units:

In this section, geometric features including action units, such as the length-the-to-width ratio of right and left eye, angles, area of triangle, ratios of triangles, and histogram of gradients are trained and tested using various classifiers including Naive Bayes, Neural Networks, and Decision Table on a data set provided by [2]. The combination of these features along with Naive Bayes, Neural Network, and Decision Table laid down the foundation of various approaches named Fab [36], Azm [37], and Jian [38] respectively.

Action units have been used in various existing papers including [36][39][40][41] and [42] whereas the concept of geometric features has been introduced and implemented in [43][44]. The concept of angles computed using facial fiducial points has been introduced in [45]. Ekman and Friesen introduced a Facial Action coding system in which facial expression can be detected with the help of action units corresponding to different facial muscles [12]. The action units, AU4, AU7, AU5, and AU23 are specified universally for anger. The action unit AU4 is for brow lowered, the AU7 tells about lid tightening, the AU5 expresses about upper lid raiser while the action unit AU23 states lip tightened [3] as shown in Figure 2. We got a complete dataset by [2] in which basic and compound emotions were predefined for every individual according to the combination of their action units. Hence, action units can be used as a strong feature for the detection of both basic and compound emotions.

The ratio of Length to Width of Right and Left Eye:

The ratio of length to width of the right and left eye not only serves as a strong feature but also helps in scale-invariant emotion recognition [37][44]. Eyes length and width can be calculated by the Euclidean distance formula as shown in equation (5).

$$Eye\ length\ or\ Eye\ width = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (5)$$

The fiducial facial points (x, y) are given as input along with images and action units.

Area of a Triangle:

The area under the dotted points [46] (as shown in Figure. 1) changes significantly with the change in different emotions, hence it can be used as a strong feature for the detection of both basic and compound emotions. The area under the dotted points is calculated using Equation 6 as written below:

$$Area\ of\ triangle\ = \sqrt{(s \times (s-a)(s-b)(s-c))} \qquad (6)$$

The s can be calculated by equation (7).

$$s = \frac{(a+b+c)}{2} \qquad (7)$$

Where s is the average length of the sides of a triangle. The length of sides of a triangle a, b, and c is calculated using the Euclidean distance formula.

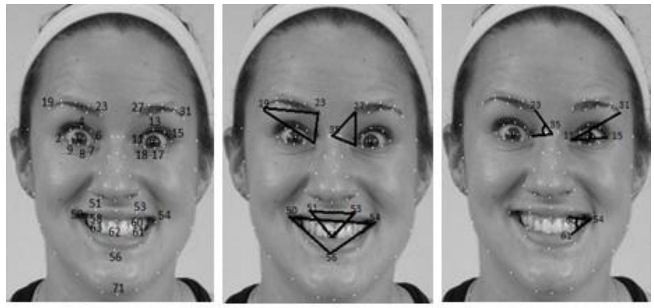
Ratio of Area of Triangles:

The ratio of the area of triangles [46] not only serves as a strong feature but also helps in making the emotion recognition scale invariant.

- The ratio of the following triangles is used in the left eyebrow: (Δ19,23,6) / (Δ19,23,2), (Δ19,23,6)/(Δ19,23,35), (Δ23,6,35)/(Δ19,23,35).
- The ratio of the following triangles is used in the right eyebrow: (Δ11,27,31) / (Δ27,35,11), (Δ11,27,31) / (Δ27,31,15), (Δ27,35,11) / (Δ27,31,35).
- The ratio of the following triangles is used in the left eye: (Δ4,8,6) / (Δ4,2,8), (Δ4,2,8) / (Δ4,8,9), (Δ4,8,7) / (Δ4,8,9).
- The ratio of the following triangles is used in the right eye: (Δ13,11,17) / (Δ13,15,17), (Δ13,11,17) / (Δ13,17,18), (Δ13,11,17) / (Δ13,17,16).

- The ratio of the following triangles is used in lips: (Δ50,62,54) / (Δ51,62,53), (Δ50,62,54) / (Δ51,56,53), (Δ51,56,53) / (Δ50,56,54).

The fiducial facial points along with their numbers, triangles, and angles are marked in Figure 4.



(a): fiducial points (b): triangles (c): angles

Figure 4: (a). fiducial points (b). triangle (c). angles

Angles:

Angles can help in detecting facial expressions as the angles of a face change frequently with the change in emotions [47]. Different angles have been identified in this approach that exhibit change for different emotions (basic and compound) of the same subject, hence making it a strong feature for the detection of both basic and compound emotions. Angles are calculated using the following equations (8-10).

$$\alpha = cos^{-1}\frac{(b^2+c^2-a^2)}{2\times b\times c} \qquad (8)$$

$$\beta = cos^{-1}\frac{(a^2+c^2-b^2)}{2\times a\times c} \qquad (9)$$

$$\gamma = cos^{-1}\frac{(a^2+b^2-c^2)}{2\times a\times b} \qquad (10)$$

Where a, b, c can be calculated using Euclidean distance formula.

The angles used for both left and right eyes and lips are: <23,35,6, <27,35,11, <19,6,2, <60,54,61, <58,50,63, <31,11,15 and <50,71,54.

Histogram of Gradients (HOG):

A histogram of gradient direction [37] is used as a feature for the detection of basic and compound emotions. This choice is motivated by observing how the gradient direction changes across various fiducial points as emotions change. Essentially, this feature aims to assess image slopes by computing their derivatives and subsequently determining the gradient direction, as illustrated in Equation (11).

$$Gradient\ Direction = tan^{-1}(\frac{f_y}{f_x}) \qquad (11)$$

Histogram can be computed from values of gradient directions as shown in equation (12).

$$H = hist\ (\theta) \qquad (12)$$

Local-Based Approach (LB):

In the local-based approach [48], the face image is divided into various local regions that are given to Kernel Subclass Discriminant Analysis (KSDA) and nearest-mean classifier to classify emotions.

Discriminant Analysis Based Approach (DA):

This section contains a discriminant analysis based existing state-of-the-art approach that uses a combination of shape and appearance feature space [2]. The approach achieves an accuracy of 96.86% for basic emotions and 76.91% for both basic and compound emotions using 10-fold cross-validation. However, it is noted that these values are lower than those attained by the proposed approach in this paper.

Support Vector Machine (SVM):

Support Vector Machine is employed in this paper for the detection of human fall events. The diverse features outlined in the "Features Used for Fall Detection" section are fed into the SVM for classifying human falls, with the results presented in the "Experiments on Du Dataset - Comp Em" section. SVM, a pattern classification algorithm pioneered by V. Vapnik and his team at AT&T Bell Labs [49] is utilized for this purpose. This method maps the data into higher dimensional input space and constructs an optimal plane separating the hyper-plane in this space. From a perspective of implementation, Support Vector Machine training is equivalent to solving a Quadratic Programming (QP) problem with linear constrained, in which the number of variables is equal to the number of data points. In [50], Foroughi et al. 2008 implemented a support vector machine for the classification of the event either as a fall or as no fall using a feature-based approach and achieved a reliable recognition rate of 88.08%. Various other human fall detection systems use Support Vector Machines for the classification of fall events (by their features) [51][52][53]. As a result of the extensive utilization of Support Vector Machines (SVMs) within human fall detection systems, we have incorporated them into our study. Leveraging the features outlined in the "Features Used for Fall Detection" section, we implemented SVMs and produced outcomes in terms of both accuracy percentages and computational efficiency. These results serve as a benchmark for comparison against our proposed methodology, which integrates motion, geometric orientation, and geometric location-based features through boosting with J48 (referred to as F M-GO-GL boosting-J48), as delineated in the "Experiments and Results" section.

Results and Discussion:

Experiment 1: Comparison with Existing Approaches:

The section presents the comparison of the proposed approach PA SVM with six state-of-the-art existing approaches including Fab [36], Azm [37], Jain [38], LB [48], and DA [2]. The proposed approach PA SVM uses a fusion of geometric, holistic, and shape-based features with a supervised learning algorithm of Support Vector Machine (SVM).

Two experiments were conducted on the dataset [2] with 10-fold cross-validation. As the dataset [2] has 7 basic and 15 compound emotions as shown in Figure 6, hence, it is a multiclass problem with 7, 15, and 22 categories for basic, compound, and combined (basic and compound) emotions, respectively. Table 2 and Figure 7 show that the proposed approach PA SVM outperforms the existing approaches in terms of percentage accuracy. The reason may be the use of novel features of slope tables in our proposed approach. As the number of categories (classes) increases from basic to combined, the percentage accuracy of existing approaches decreases more rapidly than the proposed approach. This happens due to the stable performance of the Support Vector Machine over the increase in classes as compared to other classifiers including Naive Bayes, Neural Network, and Decision Table. It is also observed that the proposed approach PA SVM has better time efficiency i.e., less execution time as compared to the existing approaches of Jain and Azm. The increase in time efficiency can be due to the fast training and testing time of SVM [54].

Figure 5: Basic emotions of the subject [2]

Experiment 2: Comparison with Deep Neural Network (NN) Architecture:

The study compares our proposed approach with state-of-the-art deep NN architecture as shown in Figure 8. Deep NN architecture is computationally more expensive and requires more training time and data. In comparison, our approach is much simpler, yet gives us classification accuracy comparable to the deep learning-based approaches [55].

Experiment 3: Comparison with Existing Classifiers:

The comparison of the proposed approach PA SVM has also been provided with existing classifiers including Naive Bayes, Neural Network, AdaBoost, Decision Table, Nearest-Neighbor- like Algorithm using Non-Nested Generalized Exemplars (NNge), and J48 using the same set of proposed features.

Table 3 and Figure 7 show that the percentage accuracy of the Support Vector Machine of the proposed approach is significantly higher than the existing classifiers including Naive Bayes, Neural Network, AdaBoost, Decision Table, NNge, and J48 using the same combination of proposed features. It can also be observed that as the number of classes increases from basic to combined, the percentage accuracy of each existing classifier decreases more rapidly as compared to the proposed classifier i.e., SVM. The SVM also shows increased time efficiency as compared to Neural Network, Decision Table, NNge, and J48.

System Specifications:

The system uses Windows 8.1 Pro 64-bit Operating System. The processor specifications of a system are Intel (R) Core (TM) i3-3110M CPU @2.40GHz while RAM size is 4096 MB. The classifiers were executed on Weka 3.8.2.

Conclusion and Future Work:

Facial emotion recognition has been an area of high interest in the past. In the last three to five years, the focus has been extended to compound emotions and researchers are coming up with novel ideas and approaches for their recognition. Our proposed approach uses the feature-based system for compound emotion recognition using a Support Vector Machine (SVM). These features include action units, angles, triangles, ratios of length to width of eyes and lips, ratios of angles, ratios of triangles, regional properties, histogram of gradients, and slope table. It can be concluded that the proposed approach outperforms the existing approaches of Fab, Azm, Jain, LB, and DA both in terms of percentage accuracy and time efficiency. Moreover, the Support Vector Machine of the proposed approach outperforms the existing classifiers both in terms of percentage accuracy and time efficiency.

It can also be safely concluded that feature-based approaches using machine learning techniques seem to provide promising results in the future. The work can be extended to further improve the percentage accuracy and time efficiency of compound emotion recognition. Besides, proposing more novel and robust features, the development of large image repositories is also important in improving the percentage accuracy of emotions. Initially, the facial emotion categories were 7, i.e., only basic emotions were discovered. However, with the addition of compound emotions, the total number of categories increased to 22. As the number of categories increases, the data requirements also increase exponentially. Hence, the availability of large data sets will be a significant requirement in the future.

The focus of research should also be extended to compound emotion recognition of different regions including Asian, Sub-continent, Chinese, American, and African American. Besides, working on frontal faces, the algorithms and techniques should also be developed for non-frontal faces. The system should further be tested on different fiducial facial point algorithms including Active Shape Model (ASM), BoRMaN, and Local Evidence Aggregation for Regression-based facial point detection (LEAR). The addition of night vision functionality and robustness against light variation to the proposed system can be an important extension of the system.

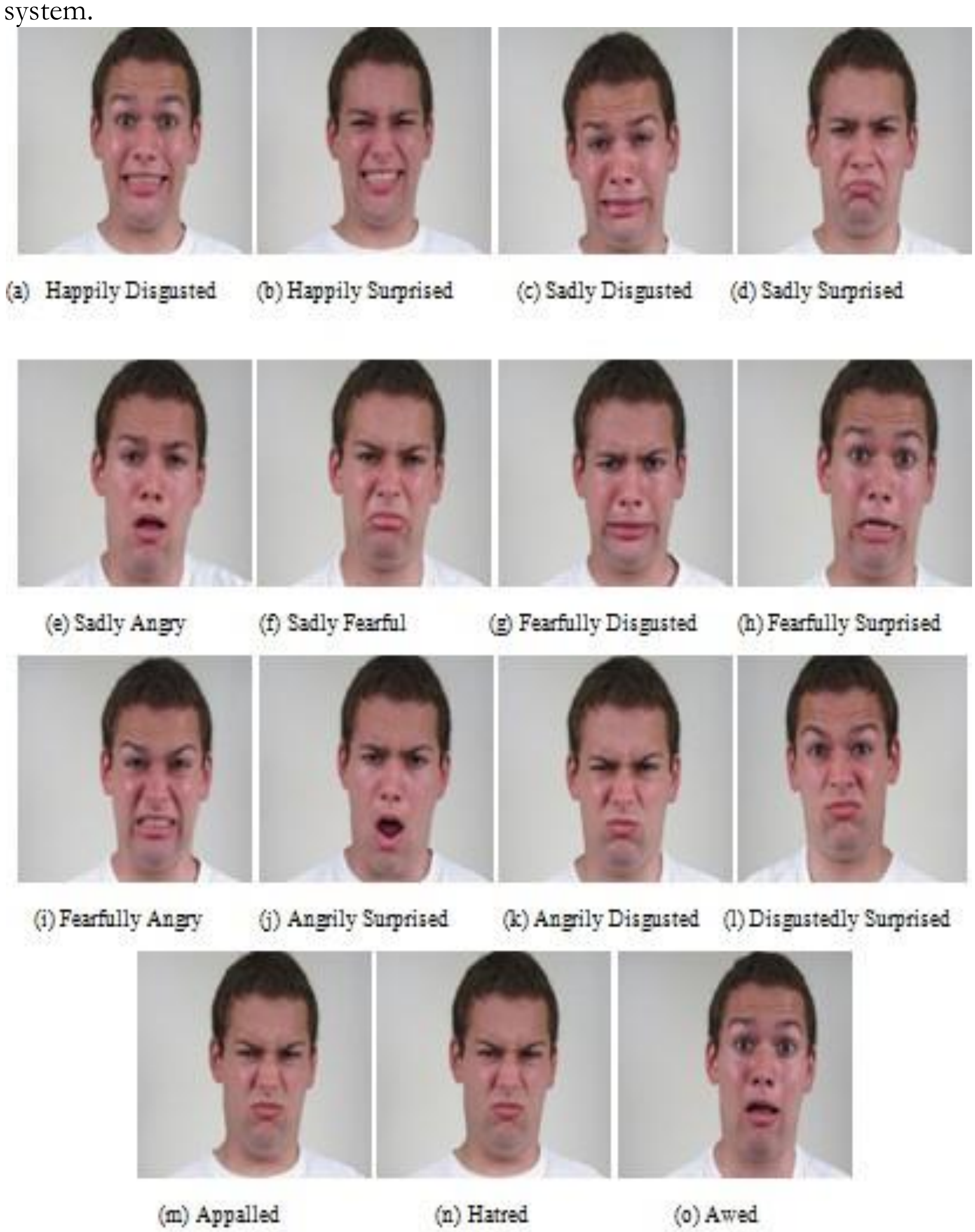
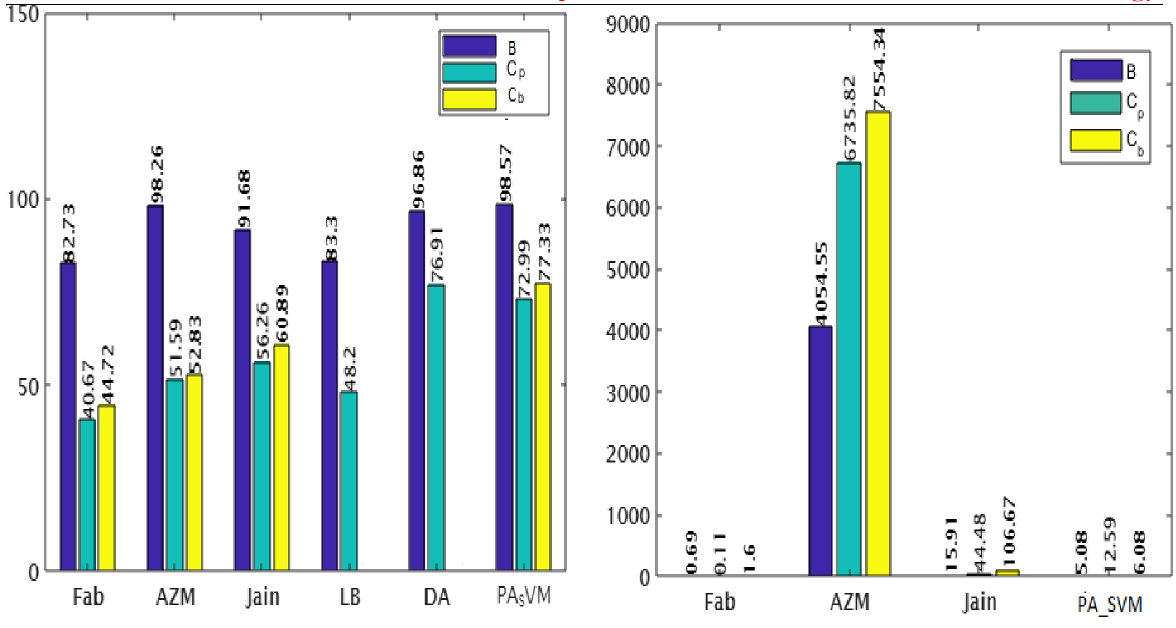
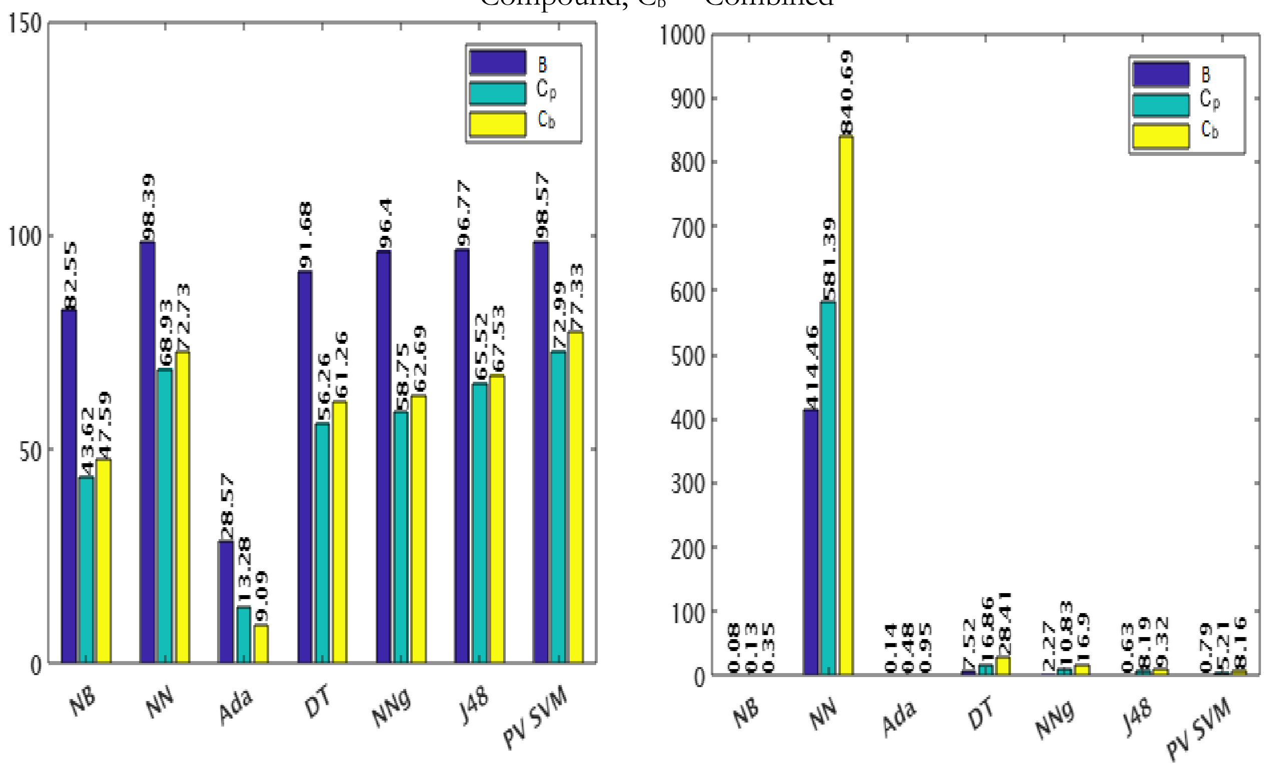


Figure 6: Compound emotions of the subject [2]



(a) (b)

Figure 7: Figure showing a comparison of existing approaches including Fab, Azm, Jain, LB, DA, and proposed approach PA SVM with 10-fold cross-validation. a) Comparison in terms of Percentage Accuracy b) Comparison in terms of Time (in sec). B = Basic, C_p = Compound, C_b = Combined



(a) (b)

Figure 8: The comparison of Proposed Classifier SVM with the existing classifiers using the same combination of proposed features with 10-fold cross-validation. a) Comparison in terms of Percentage Accuracy b) Comparison in terms of Time (in sec). B = Basic, C_p = Compound, C_b = Combined

Table 2: Results of existing approaches including Fab, Azm, Jain, LB, DA, and proposed approach PA SVM in terms of percentage accuracy and time with 10-fold cross-validation. B = Basic, Cp = Compound, Cb = Combined, NB = Naive Bayes, NN = Neural Network, DT = Decision Table, A = Percentage Accuracy, T = Time in seconds

	Existing Approach															Proposed Approach		
	Fab			Azm			Jain			LB			DA			PA SVM		
	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb
A	82.73	40.67	44.72	98.26	51.59	52.83	91.68	56.26	60.89	83.3	48.2		96.86	76.91		98.57	72.99	77.33
T	0.69	0.11	1.6	4054.55	6735.82	7554.34	15.91	44.48	106.67							5.08	12.59	6.08

Table 3: Comparison of SVM of Proposed Approach with the existing classifiers using the same combination of proposed features in terms of percentage accuracy and time with 10-fold cross-validation. B = Basic, Cp = Compound, Cb = Combined, NB = Naive Bayes, NN = Neural Network, DT = Decision Table, A = Percentage Accuracy, T = Time in seconds

	Existing Approach																		Proposed Approach		
	NB			NN			Ada			DT			NNg			J48			PV SVM		
	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb	B	Cp	Cb
A	82.55	43.62	47.59	98.39	68.93	72.73	28.57	13.28	9.09	91.68	56.26	61.26	96.4	58.75	62.69	96.77	65.52	67.53	98.57	72.99	77.33
T	0.08	0.13	0.35	414.46	581.39	840.69	0.14	0.48	0.95	7.52	16.86	28.41	2.27	10.83	16.9	0.63	8.19	9.32	0.79	5.21	8.16

Acknowledgments:

We are thankful to Du et al. for the dataset provided by them. We really appreciate Muhammad Abdul Rehman, Ali Ammar, Fahad Zafar, and Tahir Khalil for their support in this work.

References:

[1] J. Mesken, "Determinants and consequences of drivers' emotions", [Online]. Available: https://pure.rug.nl/ws/portalfiles/portal/9807378/thesis.pdf

[2] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," Proc. Natl. Acad. Sci. U. S. A., vol. 111, no. 15, pp. E1454–E1462, Apr. 2014, doi: 10.1073/PNAS.1322355111/ASSET/C2BBF816-8771-46B5-9871-0AA84A513697/ASSETS/GRAPHIC/PNAS.1322355111I91.GIF.

[3] "Intensity Estimation of Unknown Expression Based on a Study of facial Permanent Features Deformations." Accessed: Jun. 04, 2024. [Online]. Available: https://www.researchgate.net/publication/235344027_Intensity_Estimation_of_Unknown_Expression_Based_on_a_Study_of_facial_Permanent_Features_Deformations

[4] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: A review," Accid. Anal. Prev., vol. 38, no. 2, pp. 215–224, Mar. 2006, doi: 10.1016/J.AAP.2005.07.004.

[5] "TRAFFIC SAFETY FACTS 2011", [Online]. Available: https://www.cs.cmu.edu/smartheadlight/media_v3/etc/NHTSA_safety_stats.pdf

[6] M. Chan and A. Singhal, "The emotional side of cognitive distraction: Implications for road safety," Accid. Anal. Prev., vol. 50, pp. 147–154, Jan. 2013, doi: 10.1016/J.AAP.2012.04.004.

[7] G. Rebolledo-Mendez, A. Reyes, S. Paszkowicz, M. C. Domingo, and L. Skrypchuk, "Developing a body sensor network to detect emotions during driving," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 4, pp. 1850–1854, 2014, doi: 10.1109/TITS.2014.2335151.

[8] J. L. Deffenbacher, D. M. Deffenbacher, R. S. Lynch, and T. L. Richards, "Anger, aggression, and risky behavior: a comparison of high and low anger drivers," Behav. Res. Ther., vol. 41, no. 6, pp. 701–718, Jun. 2003, doi: 10.1016/S0005-7967(02)00046-3.

[9] G. Underwood, P. Chapman, S. Wright, and D. Crundall, "Anger while driving," Transp. Res. Part F Traffic Psychol. Behav., vol. 2, no. 1, pp. 55–68, Mar. 1999, doi: 10.1016/S1369-8478(99)00006-6.

[10] Y. King and D. Parker, "Driving violations, aggression and perceived consensus," Eur. Rev. Appl. Psychol., vol. 58, no. 1, pp. 43–49, Mar. 2008, doi: 10.1016/J.ERAP.2006.05.001.

[11] N. Dibben and V. J. Williamson, "An exploratory survey of in-vehicle music listening," http://dx.doi.org/10.1177/0305735607079725, vol. 35, no. 4, pp. 571–589, Oct. 2007, doi: 10.1177/0305735607079725.

[12] "Facial expression of emotion", [Online]. Available: https://www.paulekman.com/wp-content/uploads/2013/07/Facial-Expression-Of-Emotion1.pdf

[13] P. Ekman and W. V. Friesen, "Measuring facial movement," Environ. Psychol. Nonverbal Behav., vol. 1, no. 1, pp. 56–75, Sep. 1976, doi: 10.1007/BF01115465/METRICS.

[14] "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences", [Online]. Available: https://ieeexplore.ieee.org/document/1605389

[15] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D facial expression recognition based on primitive surface feature distribution," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2, pp. 1399–1406, 2006, doi: 10.1109/CVPR.2006.14.

[16] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," FGR 2006 Proc. 7th Int. Conf. Autom. Face Gesture Recognit., vol. 2006, pp. 211–216, 2006, doi: 10.1109/FGR.2006.6.
[17] P. N. Tumpi, W. R. Rahman, and M. A. Haider, "An efficient facial expression detection system," Mach. Graph. Vis., vol. Vol. 16, No. 3/4, pp. 377–399, 2007.
[18] A. Besinger et al., "Optical flow based analyses to detect emotion from human facial image data," Expert Syst. Appl., vol. 37, no. 12, pp. 8897–8902, Dec. 2010, doi: 10.1016/J.ESWA.2010.05.063.
[19] J. Chun and O. Kwon, "A vision-based approach for facial expression cloning by facial motion tracking," KSII Trans. Internet Inf. Syst., vol. 2, no. 2, pp. 120–133, Apr. 2008, doi: 10.3837/TIIS.2008.02.004.
[20] Y. Tong, Y. Shen, B. Gao, F. Sun, R. Chen, and Y. Xu, "A noisy-robust approach for facial expression recognition," KSII Trans. Internet Inf. Syst., vol. 11, no. 4, pp. 2124–2148, 2017, doi: 10.3837/TIIS.2017.04.017.
[21] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," Image Vis. Comput., vol. 27, no. 6, pp. 803–816, May 2009, doi: 10.1016/J.IMAVIS.2008.08.005.
[22] H. Soyel and H. Demirel, "Facial Expression Recognition Using 3D Facial Feature Distances," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4633 LNCS, pp. 831–838, 2007, doi: 10.1007/978-3-540-74260-9_74.
[23] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010, pp. 94–101, 2010, doi: 10.1109/CVPRW.2010.5543262.
[24] S. B. Sarode Neeta, "Facial expression recognition", [Online]. Available: https://www.researchgate.net/publication/49965368_Facial_Expression_Recognition
[25] N. Neggaz, M. Besnassi, and A. Benyettou, "Application of improved AAM and probabilistic neural network to facial expression recognition," J. Appl. Sci., vol. 10, no. 15, pp. 1572–1579, 2010, doi: 10.3923/JAS.2010.1572.1579.
[26] U. K. Jagdish Lal Raheja, "Hum an facial expression detection from detected in captured im age using back propagation neural netw ork", [Online]. Available: https://airccse.org/journal/jcsit/0210ijcsit10.pdf
[27] N. A. Tawhid, N. U. Laskar, and H. Ali, "A Vision-based Facial Expression Recognition and Adaptation System from Video Stream," Int. J. Mach. Learn. Comput., pp. 535–539, 2011, doi: 10.7763/IJMLC.2012.V2.183.
[28] Sujono and A. A. S. Gunawan, "Face Expression Detection on Kinect Using Active Appearance Model and Fuzzy Logic," Procedia Comput. Sci., vol. 59, pp. 268–274, Jan. 2015, doi: 10.1016/J.PROCS.2015.07.558.
[29] "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives", [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3742375/
[30] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," Image Vis. Comput., vol. 24, no. 6, pp. 615–625, Jun. 2006, doi: 10.1016/J.IMAVIS.2005.09.011.
[31] S. Li et al., "Bi-modality fusion for emotion recognition in the wild," ICMI 2019 - Proc. 2019 Int. Conf. Multimodal Interact., pp. 589–594, Oct. 2019, doi: 10.1145/3340555.3355719.
[32] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," IEEE Trans. Inf. Forensics Secur., vol.

1, no. 1, pp. 3–11, Mar. 2006, doi: 10.1109/TIFS.2005.863510.
[33] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, "Authentic facial expression analysis," Image Vis. Comput., vol. 25, no. 12, pp. 1856–1863, Dec. 2007, doi: 10.1016/J.IMAVIS.2005.12.021.
[34] O. J. Urizar, E. I. Barakova, L. Marcenaro, C. S. Regazzoni, and M. Rauterberg, "Emotion estimation in crowds: The interplay of motivations and expectations in individual emotions," Eur. Signal Process. Conf., vol. 2018-September, pp. 1092–1096, Nov. 2018, doi: 10.23919/EUSIPCO.2018.8553370.
[35] K. M. P ITHAYA RANI, "Emotion recognition based on facial components", [Online]. Available: https://link.springer.com/article/10.1007/s12046-018-0801-6
[36] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 5562–5570, Dec. 2016, doi: 10.1109/CVPR.2016.600.
[37] A. Azman, Q. Meng, and E. Edirisinghe, "Non intrusive physiological measurement for driver cognitive distraction detection: Eye and mouth movements," ICACTE 2010 - 2010 3rd Int. Conf. Adv. Comput. Theory Eng. Proc., vol. 3, 2010, doi: 10.1109/ICACTE.2010.5579547.
[38] X. X. Yu-Gang Jiang, Baohan Xu, "Predicting emotions in user-generated videos", [Online]. Available: https://cdn.aaai.org/ojs/8724/8724-13-12252-1-2-20201228.pdf
[39] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," Proc. - 4th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2000, pp. 46–53, 2000, doi: 10.1109/AFGR.2000.840611.
[40] A. Savran et al., "Biometrics and Identity Management," Biometrics Identity Manag., vol. 5372, pp. 47–56, 2008, doi: 10.1007/978-3-540-89991-4_6.
[41] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," IEEE Trans. Syst. Man, Cybern. Part B Cybern., vol. 42, no. 1, pp. 28–43, Feb. 2012, doi: 10.1109/TSMCB.2011.2163710.
[42] K. Zhao, W. S. Chu, F. De La Torre, J. F. Cohn, and H. Zhang, "Joint patch and multi-label learning for facial action unit detection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07-12-June-2015, pp. 2207–2216, Oct. 2015, doi: 10.1109/CVPR.2015.7298833.
[43] S. Achiche and S. Ahmed, "Mapping Shape Geometry and Emotions Using Fuzzy Logic," Proc. ASME Des. Eng. Tech. Conf., vol. 4, pp. 387–395, Jul. 2009, doi: 10.1115/DETC2008-49290.
[44] F. Tang and B. Deng, "Facial expression recognition using AAM and local facial features," Proc. - Third Int. Conf. Nat. Comput. ICNC 2007, vol. 2, pp. 632–635, 2007, doi: 10.1109/ICNC.2007.373.
[45] E. Cambria, I. Hupont, A. Hussain, E. Cerezo, and S. Baldassarri, "Sentic Avatar: Multimodal Affective Conversational Agent with Common Sense," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6456 LNCS, pp. 81–95, 2011, doi: 10.1007/978-3-642-18184-9_8.
[46] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," IEEE Trans. Intell. Transp. Syst., vol. 7, no. 1, pp. 63–77, Mar. 2006, doi: 10.1109/TITS.2006.869598.
[47] K. P. Yao, W. H. Lin, C. Y. Fang, J. M. Wang, S. L. Chang, and S. W. Chen, "Real-time vision-based driver drowsiness/fatigue detection system," IEEE Veh. Technol. Conf., 2010, doi: 10.1109/VETECS.2010.5493972.
[48] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," IEEE Trans. Pattern Anal. Mach. Intell.,

vol. 24, no. 6, pp. 748–763, Jun. 2002, doi: 10.1109/TPAMI.2002.1008382.
[49] "An improved training algorithm for support vector machines", [Online]. Available: https://ieeexplore.ieee.org/document/622408
[50] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," Proc. 11th Int. Conf. Comput. Inf. Technol. ICCIT 2008, pp. 219–224, 2008, doi: 10.1109/ICCITECHN.2008.4803020.
[51] M. Bosch-Jorge, A. J. Sánchez-Salmerón, Á. Valera, and C. Ricolfe-Viala, "Fall detection based on the gravity vector using a wide-angle camera," Expert Syst. Appl., vol. 41, no. 17, pp. 7980–7986, Dec. 2014, doi: 10.1016/J.ESWA.2014.06.045.
[52] B. V. Glen Debard, Peter Karsmakers, Mieke Deschodt, Ellen Vlaeyen, Jonas Van Den Bergh, Eddy Dejaeger, Koen Milisen, Toon Goedemé, Tinne Tuytelaars, "Camera based fall detection using multiple features validated with real life video".
[53] M. Yu, Y. Yu, A. Rhuma, S. M. R. Naqvi, L. Wang, and J. A. Chambers, "An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment," IEEE J. Biomed. Heal. Informatics, vol. 17, no. 6, pp. 1002–1014, 2013, doi: 10.1109/JBHI.2013.2274479.
[54] S. Dumais, J. Piatt, D. Heckerman, and M. Sahami, "Inductive Learning Algorithms and Representations for Text Categorization," Int. Conf. Inf. Knowl. Manag. Proc., vol. 1998-January, pp. 148–155, 1998, doi: 10.1145/288627.288651.
[55] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †," Energies 2018, Vol. 11, Page 1636, vol. 11, no. 7, p. 1636, Jun. 2018, doi: 10.3390/EN11071636.