Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

RESEARCH ARTICLE

A SYSTEMATIC REVIEW OF METHODS OF EMOTION RECOGNITION BY FACIAL EXPRESSIONS

Muazu Abdulwakil Auma, Eric Manzi and Jibril Aminu

…….....

Manuscript Info	*Abstract*
……………………. 	……………………………………………………………… Facial recognition is integral and essential in today's society, and the recognition of emotions based on facial expressions is already becoming more usual. This paper analytically provides an overview of the databases of video data of facial expressions and several approaches to recognizing emotions by facial expressions by including the three main image analysis stages, which are pre-processing, feature extraction, and classification. The paper presents approaches based on deep learning using deep neural networks and traditional means to recognizing human emotions based on visual facial features. The current results of some existing algorithms are presented. When reviewing scientific and technical literature, the focus was mainly on sources containing theoretical and research information of the methods under consideration and comparing traditional techniques and methods based on deep neural networks supported by experimental research. An analysis of scientific and technical literature describing methods and algorithms for analyzing and recognizing facial expressions and world scientific research results has shown that traditional methods of classifying facial expressions are inferior in speed and accuracy to artificial neural networks. This review's main contributions provide a general understanding of modern approaches to facial expression recognition, which will allow new researchers to understand the main components and trends in facial expression recognition. A comparison of world scientific research results has shown that the combination of traditional approaches and approaches based on deep neural networks show better classification accuracy. However, the best classification methods are artificial neural networks.

…….....

Introduction:-

Several facial expressions are displayed on the human face due to various emotions. These emotions are disgust, fear, happiness, surprise, sadness, anger, contempt, and neutral, which are deeply imprinted in our bodies. Our mirror-neuron systems effortlessly recognize these universal emotions in every human interaction we go through. To date, there are a large number of algorithms that can automatically recognize a person's emotions by facial expressions [1, 2, 3, 4]. However, the quality of emotion recognition systems based on facial expressions deteriorates due to the following problems:

1. A small amount of training data.
2. Ethnicity, gender, age.
3. Affectation of emotions.

Corresponding Author:- Muazu Abdulwakil Auma

4. Intra-class difference and inter-class convergence.
5. Occlusion.
6. Different head rotation angle.
7. Illumination level.
8. Differences in facial proportions.

Modern facial expression recognition (FER) systems include the following main stages [5]:
1. Pre-processing of the image consists of finding the face area, cropping and scaling the located area, equalizing the face, and adjusting the contrast.
2. Extraction of visual features.
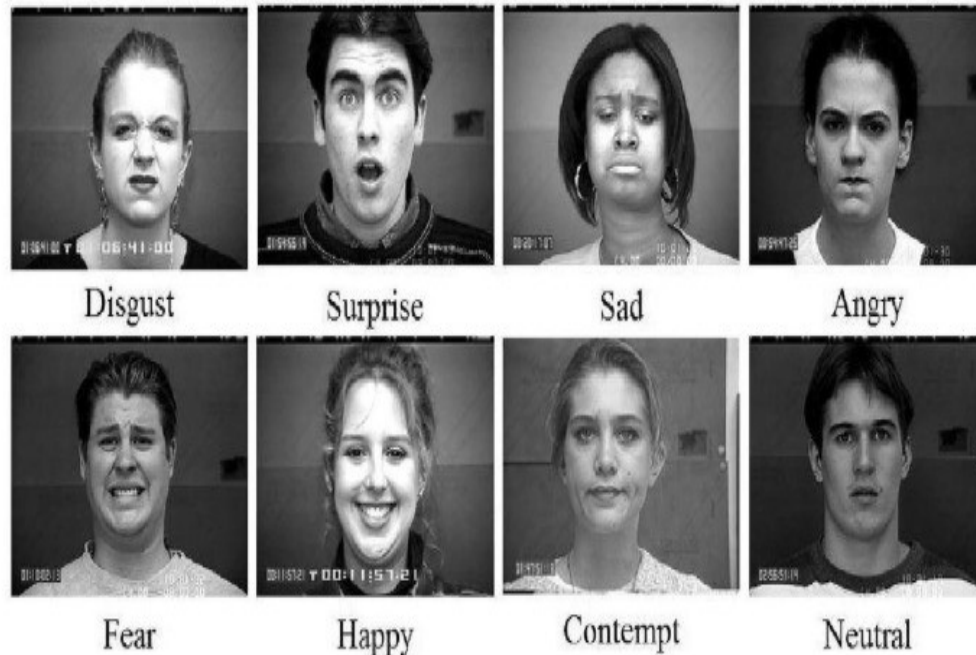3. Classification of emotions.

Figure 1:- Example of various facial expressions.

Figure 1 shows various examples of a person's facial expression.

Pre-processing allows you to cope with the problems described earlier [6, 7]. The face area's localization is performed at this stage, cropping and scaling the found area, changing the image's contrast. Feature extraction is based on the geometry and appearance of the face. By geometry, we mean the following components of the face: (their shape and location on the face), like the eyes, mouth, nose, etc., and under the appearance of the face is the skin's texture.

The classification of features is aimed at the development of an appropriate classification algorithm for identifying facial expressions.

This analytical review's primary purpose is to compare the methods of pre-processing facial images, extraction of visual features, and machine classification of emotions. This will allow us to determine the further direction of research for creating a new automatic system for recognizing human emotions by facial expressions. Figure 2 shows a general diagram of the image analysis method for identifying emotions based on a person's facial expression.
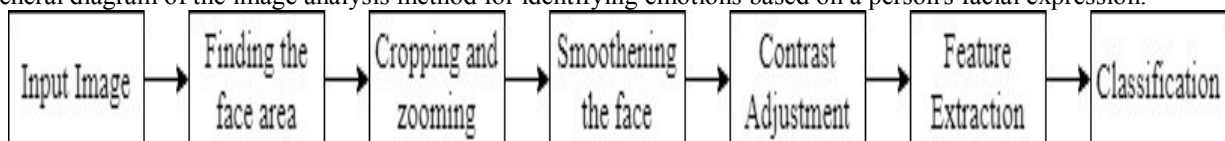

Figure 2:- Diagram of the image analysis method for recognizing emotions from human facial expressions.

The paper presents a brief overview of the databases of facial expressions, the primary processing methods, extraction and classification of features used for the FER task, and the current results of some existing algorithms are presented, as well as a comparison of FER methods.

Overview of facial expression databases
Databases of emotional facial expressions are divided into static and dynamic images in a sequence of frames. Static images capture only the peak level of intensity of the emotion being experienced, while dynamic images capture facial expressions that change over time. To create FER systems, it is promising to use databases containing video sequences. Table 1 provides a summary of some existing databases.

Table 1:- A brief overview of databases containing emotional facial expressions

Database	Number of Samples	Number of Subjects	Resolution, Pixels	Chromaticity	Emotions	Features
KDEF	4900	70	562 ×562	RGB	7	Fixing the face in five angles
JAFFE	213	10	256 ×256	Gray	7	Asian appearance
CK+	72939	123	640 ×490 640 ×480	RGB, Gray	8	Persons of different nationalities
FER2013	35888	-	48 × 48	Gray	7	Images from the Internet
SAVEE	480 audio-video	4	320 × 256	RGB	7	The presence of 60 markers on the face
CREMA-D [8]	7442 audio-video	91	480 × 360	RGB	6	Age from 20 to 74 years, different nationalities, 2443 annotators
RAVDESS [9]	4904 audio-video	24	1280 × 720	RGB	7	Varying degrees of emotion intensity, 247 annotators

Image pre-processing
Image pre-processing allows one to cope with such problems as a lack of data on facial expressions, intra-class differences and inter-class similarities, small changes in the face's appearance, changes in the head position, and illumination improve the accuracy of FER systems. Image pre-processing can include the following steps: localize the face area, crop and scale the found area, align the face and adjust the contrast.
1. Localization of the face area allows one to determine the face's size and location in the image. The most commonly used localization methods:
o Viola-Jones object detection (VJ) [10].
o Single shot multi box detector (SSD) [11].
o Histogram of directional gradients (HOG) [12] .
o Max margin object detection (MMOD) [13].
2. Cropping and scaling the face area's found area is carried out according to the face area's localization methods' coordinates. Since the located areas of the face have different sizes, it is necessary to scale the image, i.e., bring all the images to the same resolution. For these tasks, the following are applicable.
• Bessel's correction [14].
• Gaussian distribution.

For example, the application of these methods is presented in [15, 16].
3. Face alignment reduces intra- class differences. For example, for each facial expression, a reference image is selected and divided by color components or the most informative areas of the face (for example, the forehead, eyes). The remaining images are aligned relative to the reference images. The following methods are used for this task.
    * Scale invariant feature transform (SIFT) [17].
    * Region of interest (ROI) [18].
Examples of the use of these methods are presented in [19, 20].
4. Contrast adjustment allows one to smooth out images, reduce noise, increase the contrast of the face image, and improve saturation, enabling you to cope. For example, with the problem of illumination. Methods of contrast adjustment are:
    * Histogram equalization [21].
    * Linear contrast stretching [22].

Examples of the use of these methods are presented in [23].

Various image manipulations, such as rotation or displacement, allow you to increase the diversity of images and expand databases. Extending the database with variable images is helpful for deep learning-based methods. While for traditional methods, it is more beneficial to use models based on face alignment, which, on the contrary, reduces the variability associated with changes in the head position, which allows you to minimize intra-class differences and increase inter-class similarities. In this case, we selected the following models for each class: it has its standard, according to which each class's images are aligned. Choosing a suitable pre-processing method takes a long time, as facial recognition speed and accuracy depend on it.

Extracting visual features
The next stage of the FER is the extraction of signs. At this stage, the elements that are most informative for further processing are found. Depending on the functions performed, methods for extracting informative visual features are divided into several main types. Each feature extraction method is described in detail below.
1. Methods based on geometric objects allow one to extract information about geometric objects, such as the mouth, nose, eyebrows, and other objects and determine their location. With details about geometric objects and their locations, you can calculate the distance between the objects; the obtained distances and coordinates of the objects' positions are signs for further classification of emotions. Methods based on geometric objects include:
    * The Line Edge Map (LEM) descriptor [24] measure facial expressions' similarity based on objects' boundaries. An example of using the descriptor is presented in [25].
    * Active Shape Model (ASM) [26] detects objects' edges using facial landmarks, representing a chain of sequences of feature points.
    * Active appearance model (AAM) [27] is an extended version of ASM, which also forms the textual features of facial expressions. In [28], examples of using the methods are presented ASM and ASM.
    * HOG allows you to compare facial expressions by the direction of the gradients. An example of using HOG is shown in [6].
    * Curvelet Transform [29] transmits information about the object's location and spatial frequency. An example of using the transformation is shown in [30].
2. Methods based on appearance models will allow you to extract information about the textural features of the face. For example, the different number of wrinkles in the eye area indicates different facial expressions, so using methods based on appearance models is the most informative for classifying emotions. Methods based on appearance models include:
    * The Gabor filter [31] is a classical method of distinguishing facial expression features, allowing you to select different deformation models for each emotion.
    * Local Binary Patterns (LBP) [32] allow you to represent the image pixel areas in binary code containing the distinctive features of both local and global (texture) areas of facial expressions.
    * Local Phase Quantization (LPQ) is resistant to ISO blurring. The method is based on short-term Fourier transformation, which allows us to identify periodic components in images of facial expressions and assess their contribution to the initial data formation. Examples of using the LBP and LPQ methods are presented in [33].

- The Weber local descriptor [34] extracts features in two stages. The first stage divides the image into local areas (such as mouth and nose) and normalizes the images. The second stage extracts distinctive textural features using the gradient orientation describing facial expressions. An example of using this method is provided in [35].
- Discrete wavelet transform (DWT) [36] extracts texture features by splitting the original image into low-and high-frequency sections. For example, the authors of the article [37] use this descriptor.

3. Methods based on global and local objects are:
- The Principal Component Analysis (PCA) method [38] extracts the distinctive features of facial expressions from the covariance matrix, reducing the vectors' dimension.
- Linear discriminant analysis (LDA) [38] searches for vectors with the most significant differences between classes and groups the same class's features. Examples of using the methods in PCA and LDA are presented in [1].
- The optical flow (OF) [39] assigns a velocity vector to each pixel, so information about the facial muscles' movement is extracted from the sequence of images, which allows us to consider the deformation of the face in dynamics. For example, the authors [40] use this method.

From the presented feature extraction methods, appearance-based methods are the more different methods since they allow extracting textural features of appearance, which are essential parameters for FER, but they are less adapted to occlusions. Methods based on geometric objects are better adapted to occlusion, and the distance between facial landmarks is more characteristic of inter-class differences. The use of hybrid methods to extract features of facial expressions increases the accuracy of classification.
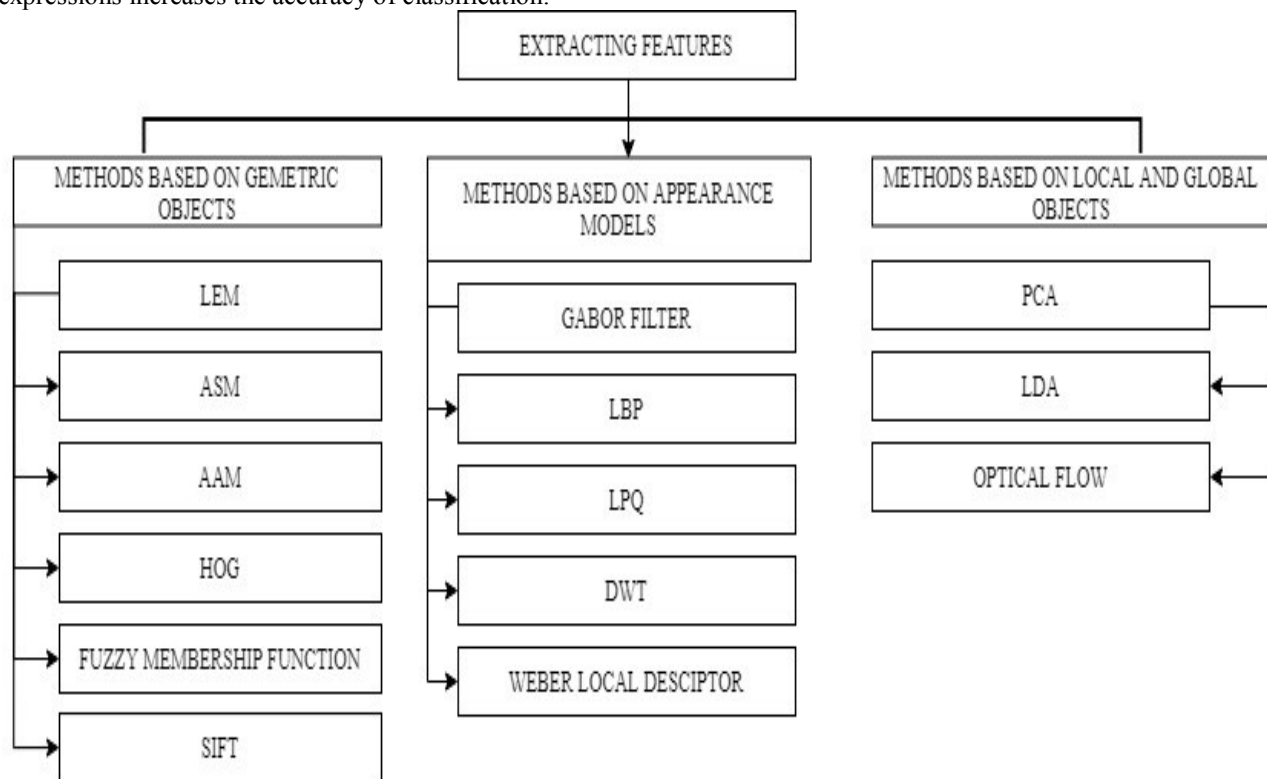

Figure 3:- Systematization of methods for extracting visual features.

Classification of emotions

Classification is the last stage in FER. The extracted features are classified into facial expressions: happiness, surprise, anger, fear, disgust, sadness, and neutrality. Methods of machine classification of emotions are divided into traditional methods and artificial neural networks. In Figure 4, you can see the methods of machine classification of emotions. Each method of machine classification of emotions is described in detail below.

1. Traditional methods:

- Hausdorff Distance [41] allows you to measure the distance between areas of interest and compare the resulting distances between classes. An example of using the method is presented in [7].
    - Minimum distance classifier (MDC) minimizes the distance between classes. The smaller the distance, the greater the similarity between classes. An example of using the method is presented in [2].
    - The k-nearest neighbor algorithm (KNN) [42] classifies facial emotions by the first k-elements whose distance is minimal. An example of using the method is presented in [3].
    - LDA presents matrix data in the form of a scatter plot, which allows you to separate the features of classes visually. For example, the authors [43] use the LDA method.
    - The Support vector machine (SVM) method [44] builds a hyperplane separating the sample objects. The greater the distance between the separating hyperplane and the shared classes' objects, the lower the classifier's average error. For example, the authors in [1, 3, 4] use the SVM method.
    - The Hidden Markov model (HMM) [45] extracts pixels with a scanning window, converting them into observation vectors, facial expressions classify the resulting vectors. For example, the authors of the article [1] use the HMM method.
    - The Decision tree [46] assigns the value of emotion to the image according to predefined rules, depending on what features the image contains. An example of using the method is presented in [4].
2. Artificial neural networks:
    - The multilayer perceptron (MLP) [47] has three layers: an input layer, a hidden layer, and a processing layer. Each layer contains neurons that have their unique weight value. This perceptron type is called multilayer because its hidden (trainable) layer can consist of layers. Examples of the use of MLP are presented by the authors of the articles [3, 4].
    - Deep neural networks (DNN):
        o The multilayer direct neural network (Multilayer feed forward neural network (MLFFNN) [48] is a relationship of per-centers, where the number of layers in the neural network is the number of perceptrons. The authors in [5] present an example of using MLFFNN.
        o Convolutional neural network (CNN) is a classification method with three layers: a convolution layer, a subsample layer, and a fully connected layer. CNN captures textures on small areas of images. CNN is trained on still images. Examples of using CNN are presented in [5, 6].
        o Recurrent neural network (Recurrent neural network (RNN) uses contextual information about previous images. So, one training set contains a sequence of images, the class-fixing of the entire set will correspond to the classification of the last image.
        o The combination of convolutional and recurrent neural networks can extract local data and use temporal information for more accurate image classification. For example, the authors in [49] use a combination of convolutional and recurrent neural networks.

Other DNNs are modifications of the presented neural networks. The following metrics are used to analyze classification methods' effectiveness: completeness, accuracy, and F-measure. With a good set of training data, the best classification method is deep neural networks. They automatically study and extract features from the input images and detect emotions on the face with higher accuracy and speed than other methods.
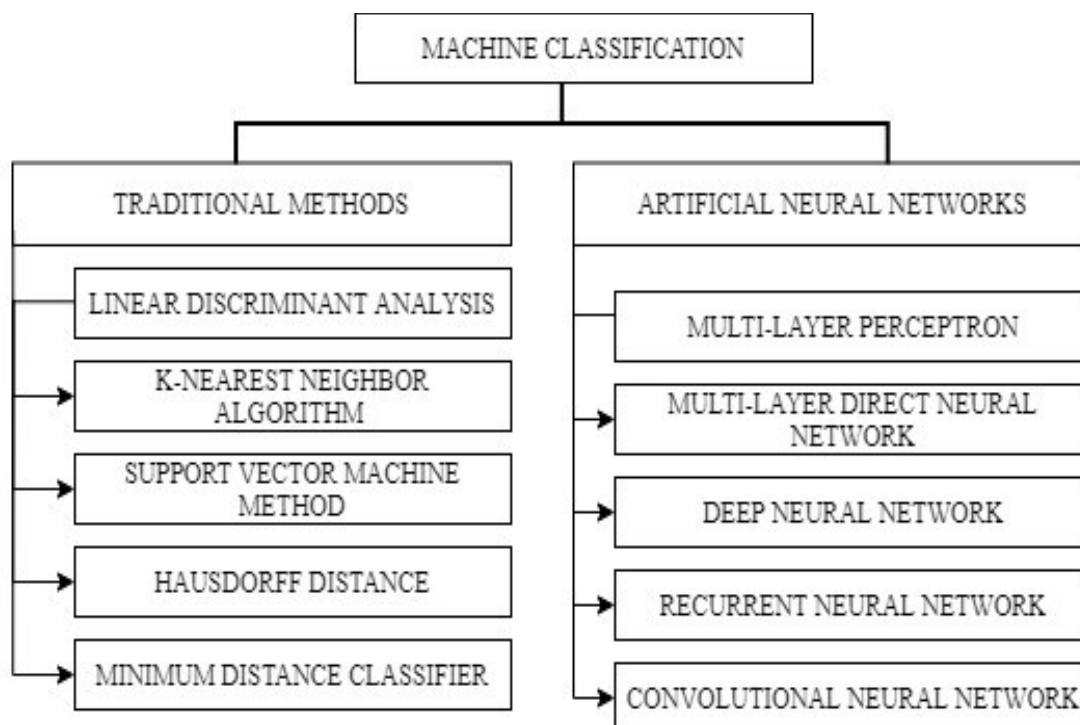
Figure 4:- Systematization of methods of machine classification of emotions.

Comparison of facial expression recognition algorithms
Since the FER algorithms' accuracy depends on the database used for training and testing, it is correct to compare different methods on the same data sets. So, to compare the methods, the results obtained from the FER2013 database were studied on static images. The SAVEE database was used to evaluate the recognition algorithms for dynamic changes in facial expressions. To compare the research results, the databases widely used by researchers for this task were selected.

Comparison of results on the FER2013 dataset
The authors in [6] compared the FER algorithms with feature extraction using the HOG and CNN methods. They found that with a small 48 × 48 image expansion and low image quality, the HOG and CNN methods' results are not adequate compared to CNN. The authors in their work [5] experimented with CNN parameters. As a pre-processing method, the authors used the Viola-Jones method. As a result, the localized areas of the face had a resolution of 48 × 48 pixels. Figure 4 shows the CNN architecture proposed by the authors in [5].

The authors of the article [49] proposed to use a multilayer maxout activation function (MAP), which allows coping with such a problem in deep learning as a gradient explosion. To extract visual features, the authors used CN and RN with long short-term memory (LSTM). The combination of models also allows us to obtain information about the dynamic characteristics of facial expressions. The authors used SVM to classify emotions. All images were randomly cropped to 24 × 24 pixels.

The authors in [50] used the JAFFE data set as a set of k-means training data. This was necessary to obtain a set of cluster centers. The k-means method is a clustering method used to divide n observations into k-clusters. Furthermore, the cluster centers' obtained values were fed as the convolution kernel's initial values in the CNN. The classification in the algorithm proposed by the authors was performed by SVM. For pre-processing, the authors used Viola-Jones, resulting in localized areas the faces had a resolution of 48 × 48 pixels.

Table 2 summarizes the FER methods used and the classification accuracy on the FER2013 dataset, presented in [5, 6, 49, 50].

According to the final (Table. 2), it can be noted that the authors used the Viola-Jones pre-processing method. We used methods based on geometric objects (ASM, HOG, SIFT), global and local objects (PCA) to extract features. In some of the papers presented, CNN and the CNN + LSTM combination were the preferred feature extraction methods. The authors used SVM, KNN, and CNN as classifiers. However, the authors showed the highest result [5] using the Viola-Jones method and CNN, experimenting with CNN parameters. The accuracy of the analysis was 89.78 %. From this, we can conclude that CNN copes with classification more accurately than traditional classification methods, such as SVM and KNN.

Table 2:- Results of facial emotion recognition based on the FER2013 database.

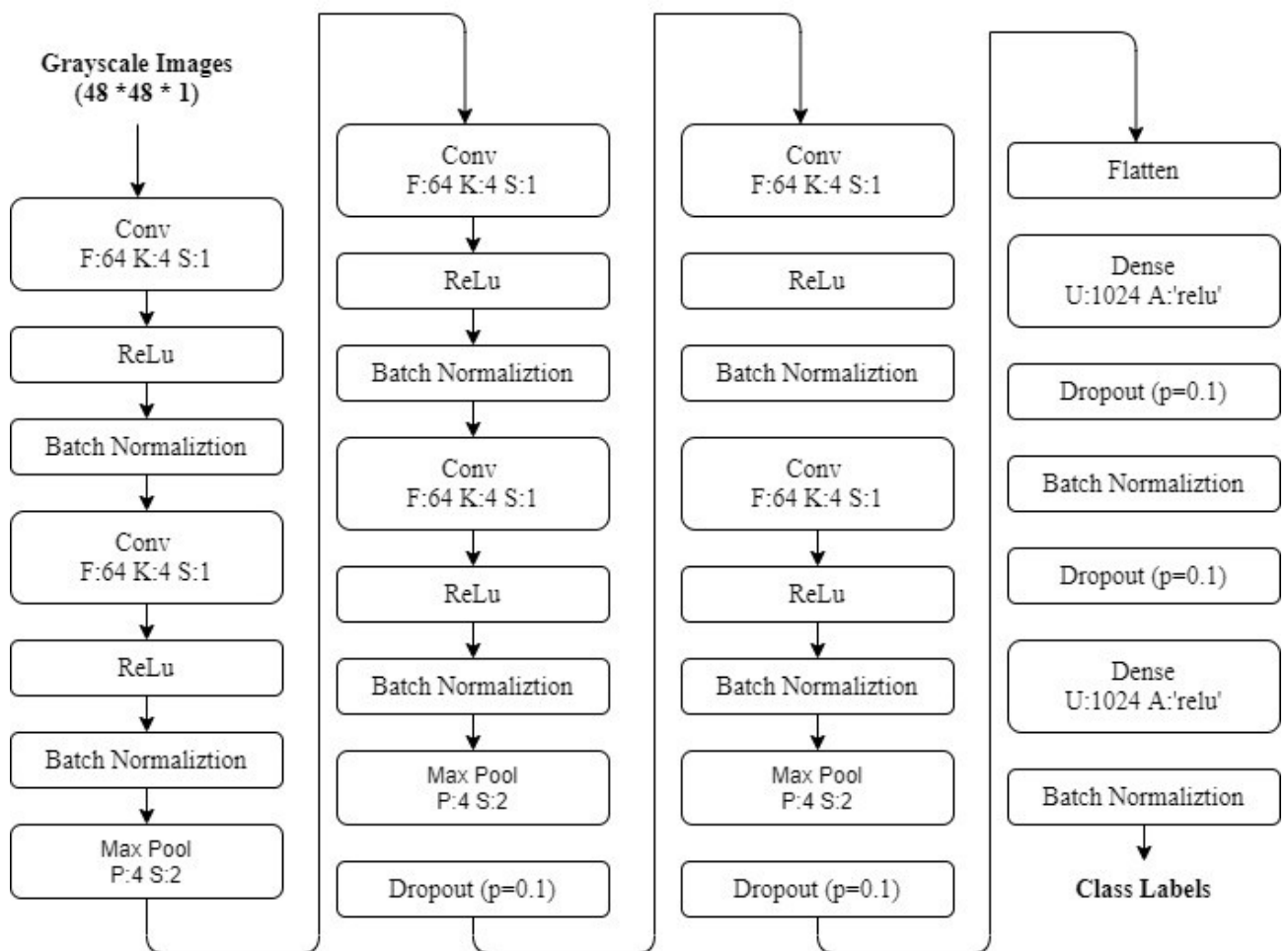| Author, Year	Pre-processing Method	Feature Extraction Method	Classifier	Accuracy %
Sahar et al., 2019 [6]	-	HOG	CNN	70.00%
		-	CNN	72.00%
Cao et al., 2019 [50]	VJ	CNN	KNN	72.86%
		-	CNN	77.43%
		CNN	SVM	78.86%
		K-means	CNN + SVM	80.29%
An et al., [49]	-	MMAF	CNN	84.50%
		MMAF + CNN + LSTM	SVM	86.60%
Isha et al., 2019 [5]	VJ	-	CNN	89.78%

Figure 5:- Neural network architecture in the emotion recognition system [5].

Comparison of results on the SAVEE dataset
As mentioned earlier, the FER2013 database consists of static images, but since facial expressions change dynamically, it is necessary to compare the results obtained on a database containing video recordings. A database will be used to compare the FER algorithms SAVEE.

The authors in [51] use a new hybrid structure called Tandem modeling (TM), which consists of two hierarchically connected neural networks with a direct connection of the same type. The first network is a Not Fully Connected Neural Network (NFCN), the second is a standard fully connected neural network (FCN). Both networks use a hidden bottleneck connection layer that allows all outputs to be combined. As a preliminary treatment, the authors used the following methods: Viola-Jones localization, as a result of which the localized areas of the face had a resolution of 128×128 pixels, LBP was used on three orthogonal planes (Local binary patterns on three orthogonal planes, LBP-TOP). Other LBPs matrices of each class were independently distributed through 5-layer MLP and fed to the neural network.

Table 3 summarizes the FER methods used and the classification accuracy on the SAVEE dataset, as presented by the authors in [33, 40, 51].

The authors [40] used the following approach in their work. Rough and precise networks with an automatic encoder were responsible for pre-processing (coarse-to-fine auto-encoder networks, CFAN), allowing them to gradually optimize and align facial expression images. As a result, the localized areas of the face had a resolution of 64×64 pixels. Two sets of functions were responsible for extracting global features: CNN and Gist, which use Gabor filters. Local features were extracted by LBP and LPQ-TOP. The authors used a discriminative multiple canonical correlation analysis (DMCA) to determine these indicators' characteristics to combine local and global. To reduce the functions' dimension, the Kernel entropy component analysis (KECA) algorithm was used. The SVM classifier terminates the algorithm.

The authors [40] used the methods OF and Cumulative OF to extract the features. optical flow, AOF). 3D CNN was chosen by the authors because it allows us to extract features from spatial and temporal dimensions. In 3D CNN, the core moves in three directions, and the input and output data in this network are 4-dimensional. To select a suitable set of hyperparameters, the authors chose a Bayesian optimizer. The Viola-Jones methods and histogram equalization (HE) were applied to pre-process the image, resulting in those localized areas of the face had a resolution of 64×64 pixels.

Total, based on the results obtained from the database SAVEE (table. 3), we can conclude. The following methods were used at the preliminary image processing stage: Viola-Jones, histogram alignment, and CFAN. The following methods were used at the feature extraction stage: LBP, LPS, OF, CNN, and Gist. The following methods are selected to classify facial expressions: SVM, MLP, TM, 3D CNN. Thus, the highest accuracy was obtained by the authors of the work [51], the recognition accuracy was 100 %. The authors used the Viola-Jones method for image pre-processing. The method is a local binary pattern on three orthographic surface planes and was selected at the feature extraction stage. Next, tandem modelling was used.

Table 3:- Results of facial emotion recognition on the SAVEE database.

Author, Year	Pre-Processing Method	Feature Extraction Method	Classifier	Accuracy %
Fan et al., 2018 [34]	CFAN	CNN	SVM	62.3%
		Gist	SVM	62.7%
		LPQ-TOP	SVM	64%
		LBP	SVM	68.5%
		LBP+LPQ-TOP	SVM	73.33%
		CNN+Gist	SVM	75%
		All methods	SVM	85.83%
Zhao et al., 2018 [40]	VJ, HE	OF	3D CNN	86%
		AOF	3D CNN	90.8%
		-	3D CNN	97.92%
Salma et al., 2019	VJ	LBP-TOP	MLP	95.08%

[51]		LBP-TOP	TM	100%

As a result of comparing the FER algorithms on the SAVEE and FER2013 databases, we can conclude:
1. At the stage of preliminary processing, the Viola-Jones localization method is preferred.
2. At the stage of feature extraction, the following methods were used:
    - Based on geometric objects (ASM, HOG, SIFT).
    - Based on appearance (Gabor filter, LBP).
    - Based on global and local objects (PCA, OF).
    - CNN.
3. At the classification stage, the authors used SVM, MLP, CNN, and FCN.

The accuracy of the FER is high and is mainly achieved by artificial neural networks. However, the articles' authors did not conduct experiments of the trained models they obtained with other datasets, indicating that the trained models are not universal.

Conclusion:-
Automatic facial expression recognition is an essential component of multimodal interfaces of human-computer interaction and computer paralinguistics systems. For accurate recognition of emotions from facial expressions, it is necessary to choose suitable methods for pre-processing the image, extracting visual features of facial expressions, and classifying emotions. Today, traditional classification methods are inferior in speed and accuracy to artificial neural networks. However, despite the large number of experiments carried out, the accuracy of the algorithm's facial expression recognition is still not high enough for various input parameters, so the task of creating a universal algorithm remains relevant.

For further research, the authors plan to combine the SAVEE databases, CREMA-D, and RAVDESS, which will solve the problem of lack of data for training models. In addition, it will train the classifier to automatically recognize emotions regardless of ethnicity, age, or gender. To localize the face area, it is assumed to use an active shape model, which will allow you to divide the edges of objects using facial landmarks. This method shows high speed and accuracy face area detection. The images will be scaled with a resolution of 48×48, 64×64, 128×128, 224×224 pixels, which will allow us to conclude about the effect of resolution on the accuracy of facial expression recognition. Also, to adjust the contrast, the histogram equalization method will be used, which will reduce the noise and the dependence of the facial expression recognition algorithm on the illumination. Using the active shape model will also allow you to extract features about the distance between the face landmarks and the center of mass, and the angle turning the head. It is planned to use CNN to classify emotions based on the obtained pixel arrays with different resolutions.

It is planned to use FCN and SVM methods to classify emotions based on the extracted features (the coordinates of the facial orientations, the distance to the center of mass, and the angle of rotation of the head).

References:-
[1]   S. Varma, M. Shinde and S. S. Chavan, "Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers, Techno-Societal 2018," in Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications - Volume 1, 2020.
[2]   D. B. M. Yin, A. A. Mukhlas, R. Z. W. Chik, A. T. Othman and S. Omar, "A Proposed Approach for Biometric-Based Authentication Using of Face and Facial Expression Recognition," in 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS), Singapore, 2018.
[3]   H. I. Dino and M. B. Abdulrazzaq, "Facial Expression Classification Based on SVM, KNN and MLP Classifiers," in 2019 International Conference on Advanced Science and Engineering (ICOASE), Zakho - Duhok, Iraq, 2019.
[4]   A. Tripathi, S. Pandey, and H. Jangir, "Efficient Facial Expression Recognition System Based on Geometric Features Using Neural Network," in Information and Communication Technology for Sustainable Development, Singapore, 2018.
[5]   T. Isha, J. Kalyani, V. Shreya, K. Rucha and K. Anagha, "Real Time Facial Expression Recognition using Deep Learning," in Proceedings of International Conference on Communication and Information Processing (ICCIP), 2019.

[6]   Z. Sahar, A. Fayyaz, G. Subhash, A. Irfan, K. Asif and Z. Adnan, "Facial Expression Recognition with Histogram of Oriented Gradients using CNN,"Indian Journal of Science and Technology, vol. 12, no. 24, pp. 1-8, 2019.

[7]   B. D. Ravi, S. R. Shiva, M. Gadiraju and M. K.V.S.S, "Facial expression recognition using bezier curves with hausdorff distance," in International Conference on IoT and Application (ICIOT), Nagapattinam, 2017.

[8]   H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova and R. Verma, "CREMA-D: Crowd-Sourced Emotional Multimodal Actors Dataset,"IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 377-390, 2014.

[9]   S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English,"PLoS ONE, vol. 13, no. 5, 2018.

[10]   P. Viola and M. J. Jones, "Robust Real-Time Face Detection,"International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.

[11]   W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in Computer Vision – ECCV 2016, vol. 9905, Springer, Cham, 2016, pp. 21-37.

[12]   O. Deniz, G. Bueno, J. Salido, and F. De La Torre, "Face recognition using Histograms of Oriented Gradients,"Pattern Recognition Letters, vol. 32, no. 12, pp. 1598-1603, 2011. [13]   D. E. King, "Max-Margin Object Detection."

[14]   G. P. Mohan, C. Prakash, and S. V. Gangashetty, "Bessel transform for image resizing," in 18th International Conference on Systems, Signals and Image Processing, Sarajevo, Bosnia, and Herzegovina, 2011.

[15]   E. Owusu, J.-D. Abdulai and Y. Zhan, "Face detection based on multilayer feedforward neural network and Haar features,"Software: Practice and Experience.

[16]   J. Su, L. Gao, W. Li, Y. Xia, N. Cao, and R. Wang, "Fast Face Tracking-by-Detection Algorithm for Secure Monitoring,"Applied Sciences, vol. 9, no. 18, p. 3774, 2019.

[17]   D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints,"International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[18]   A. Hernandez-Matamorosa, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake and H. Perez-Meana, "Facial expression recognition with automatic segmentation of face regions using a fuzzy-based classification approach,"Communications in Computer and Information Science, vol. 532, pp. 529-540, 2015.

[19]   S. Naz, S. Ziauddin and A. R. Shahid, "Driver Fatigue Detection using Mean Intensity, SVM, and SIFT,"International Journal of Interactive Multimedia and Artificial, vol. 5, no. 4, pp. 86-93, 2017.

[20]   R. V. Priya, "Emotion recognition from geometric fuzzy membership functions,"Multimedia Tools and Applications, vol. 78, no. 13, p. 17847–17878, 2019.

[21]   X. Wang and L. Chen, "Contrast enhancement using feature-preserving bi-histogram equalization,"Signal, Image, and Video Processing, vol. 12, no. 4, p. 685–692, 2017.

[22]   A. Mustapha, A. Oulefki, M. Bengherabi, E. Boutellaa and M. A. Algaet, "Towards nonuniform illumination face enhancement via adaptive contrast stretching,"Multimedia Tools and Applications, vol. 76, no. 21, p. 21961–21999., 2017.

[23]   M. Oloyede, G. Hancke, H. Myburgh, and A. Onumanyi, "A new evaluation function for face image enhancement in unconstrained environments using metaheuristic algorithms,"EURASIP Journal on Image and Video Processing, no. 1, p. 27, 2019.

[24]   Y. Gao and M. K. Leung, "Face recognition using line edge map,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, p. 764–779.

[25]   F. M. Hussain, H. Wang, and K. C. Santosh, "Gray level face recognition using spatial features,"Communications in Computer and Information Science, vol. 1035, p. 216–229.

[26]   T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application,"Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, 1995.

[27]   T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, 2001.

[28]   M. Iqtait, F. S. Mohamad and M. Mamat, "Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM)," in IOP Conference Series: Materials Science and Engineering.

[29]   E. Candès, L. Demanet, D. Donoho and L. Ying, "Fast Discrete Curvelet Transforms,"Multiscale Modeling & Simulation, vol. 5, no. 3, p. 861–899.

[30]   X. Fu, K. Fu, Y. Zhang, Q. Zhou, and X. Fu, "Facial Expression Recognition Based on Curvelet Transform and Sparse Representation," in 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018.

[31]   A. Tanveer, T. Jabid and U.-P. Chong, "Facial Expression Recognition Using Local Transitional Pattern on Gabor Filtered Facial Images,"IETE Technical Review, vol. 30, no. 1, p. 47–52, 2013.

[32]   C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study,"Image and Vision Computing, vol. 27, no. 6, p. 803–816.

[33]   J. Fan, Y. Tie and L. Qi, "Facial Expression Recognition Based on Multiple Feature Fusion in Video," in International Conference on Computing and Pattern Recognition (ICCPR), Shenzhen, China, 2018.

[34]   S. Li, D. Gong and Y. Yuan, "Face recognition using Weber local descriptors,"Neurocomputing, vol. 122, p. 272–283, 2013.

[35]   M. I. Revina and W. Emmanuel, "Face Expression Recognition Using Weber Local Descriptor and F-RBFNN," Madurai, India.

[36]   P. S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, Second Edition, CRC Press, 2017, p. 464.

[37]   S. Nigam, R. Singh and A. K. Misra, "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain,"Multimedia Tools and Applications, vol. 77, no. 21, p. 28725–28747, 2018.

[38]   A. Martinez and A. C. Kak, "PCA versus LDA,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233.

[39]   S. Negahdaripour, "Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 9, p. 961–979., 1998.

[40]   J. Zhao, X. Mao, and J. Zhang, "Learning deep facial expression features from image and optical flow sequences using 3D CNN,"The Visual Computer, vol. 34, no. 10, p. 1461–1475, 2018.

[41]   B. Guo, K.-M. Lam, K.-H. Lin and W.-C. Siu, "Human face recognition based on spatially weighted Hausdorff distance,"IEEE International Symposium on Circuits and Systems (ISCAS), vol. 2, p. 145–148, 2001.

[42]   I. Tayari-Meftah, N. L. Thanh and C. Ben-Amar, "Emotion Recognition Using KNN Classification for User Modeling and Sharing of Affect States,"Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7663, p. 234–242, 2012.

[43]   L. Greche, M. Akil, R. Kachouri and N. Es-Sbai, "A new pipeline for the recognition of universal expressions of multiple faces in a video sequence,"Journal of Real-Time Image Processing, vol. 17, p. 1389–1402, 2020.

[44]   M. AbdulRahman and A. Eleyan, "Facial expression recognition using Support Vector Machines," in 2015 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 2015.

[45]   P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs,"IEEE Transactions on Information Forensics and Security, vol. 1, no. 1, pp. 3-11, 2006.

[46]   S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology,"IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, 1991.

[47]   P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression Recognition."

[48]   D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multilayer feedforward neural networks,"Chemometrics and Intelligent Laboratory Systems, vol. 39, no. 1, pp. 43-62, 1997.

[49]   F. An and Z. Liu, "Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM,"The Visual Computer, vol. 36, no. 3, p. 483–498, 2020.

[50]   T. Cao and M. Li, "Facial Expression Recognition Algorithm Based on the Combination of CNN and K-Means," in 11th International Conference on Machine Learning and Computing (ICMLC), 2019.

[51]   S. Kasraoui, Z. Lachiri, and K. Madani, "Tandem Modelling Based Emotion Recognition in Videos," in Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science, vol. 11507, 2019, Springer.