Enhancing Eye Emotion Recognition with the Haar Classifier Using Co-Evolutionary Hybrid Intelligence

G S Grusha, M Nikhil, Dharani Dharan
A East West Institute of Technology
Bengaluru, India
{grushags, nikhilm, dharanidharan}@ewit.edu.in

Kirill Krinkin, Yulia Shichkina
Saint-Petersburg Electrotechnical University 'LETI'
Saint-Petersburg, Russia
kirill@krinkin.com, shichkina@etu.ai

Prof. T. N Nagabhushana
East West Institute of Technology
Bengaluru, India
ed@ewgi.edu.in

Abstract—**This work explores the challenging task of identifying emotions from the eyes through the analysis of nonverbal cues and facial expressions. The study investigates the various emotions that the eyes can convey and their significance in emotional expression, including happiness, sadness, anger, fear, surprise, and contempt. The practical applications of this research include the potential use of computer vision algorithms to analyze images or videos of the eye region to identify emotions. The coevolution theory suggests that humans and robots can collaborate to achieve a shared goal. Therefore, this study aims to develop algorithms that can identify variations in pupil size, eye movements, and other traits linked to various emotional states, such as arousal, surprise, and disgust. The results of the research show that these algorithms can accurately identify emotions from the eyes, and a human expert can validate their interpretations. The methods used in this study include analyzing facial expressions and nonverbal cues, developing computer vision algorithms, and working with human experts to validate the results. The conclusions drawn from this work suggest that emotional identification from the eyes is a promising area of research that has practical applications in various fields, such as robotics, psychology, and human-computer interaction.**

I Introduction

Complex psychological and physiological states known as emotions are brought on by a variety of internal or external triggers. These stimuli could be people, places, things, or even thoughts. Emotions can be felt in a wide variety of intensities, from moderate to intense, and frequently cause changes in behavior. Emotion classification is typically accomplished by classifying emotions as distinct in nature.

According to the discrete emotion hypothesis, all people are born with a core set of universally recognizable emotions. Because these fundamental emotions may be distinguished by a person's expression and biological processes, they are referred to as distinct emotions.

Emotion classification features regarding the eye: The eyes are an important part of the face that can convey a wide range of emotions. Therefore, eye-related features are commonly used in emotion classification tasks that involve analyzing facial expressions, such as image or video-based emotion recognition. Ultimately, analyzing an individual's emotional state through their eyes can provide important information that can be used in a variety of contexts to foster empathy, understanding, and better communication.

Some of the eye-related features used in emotion classification include -

Eye region shape: The shape of the eye region, including the distance between the eyes, the size of the eyes, and the shape of the eyebrows, can convey different emotions. For example, raised eyebrows and wide-open eyes may indicate surprise or fear.

Eye movements: Different eye movements, such as eye blinks, gaze direction, and the frequency of eye movements, can convey different emotions. For example, avoiding eye contact may indicate shyness or embarrassment, while staring intensely may indicate anger or aggression.

Pupil Dilation: Pupil dilation is a physiological response that is often associated with emotional arousal. Therefore, measuring changes in pupil size can provide valuable information about emotional states.

Eye wrinkles: The presence or absence of wrinkles around the eyes, such as crow's feet or frown lines, can also indicate different emotions. For example, the presence of crow's feet around the eyes may indicate happiness or laughter. Using a classifier and a collection of Haar-like features, a popular computer vision technique known as "Haar-cascades" detects facial features.

The primary goal of utilizing the aforementioned features is to investigate the various emotions that the eyes can convey and their significance in emotional expression.

This research also seeks to develop computer vision algorithms that can accurately identify emotional states through variations in pupil size, eye movements, and other traits linked to emotional expression.

Ultimately, the primary goal of this research is to contribute to the growing body of knowledge on emotional expression and its potential use in practical settings.

II. Related work

There has been a lot of relevant research using eye-tracking technologies to classify emotions

.J.Wang's [2] paper showed that upon the basis of peculiarities in eye movement, the authors suggested a technique for identifying emotions. In order to categorize the feelings, they employed machine learning algorithms to analyze participant eye movement data after seeing emotional imagery.To categorize the emotions, the authors employed decision trees and support vector machine (SVM) methods. The outcomes demonstrated that SVM outperformed the decision tree algorithm in terms of accuracy. Also, the authors discovered that pupil width and saccade velocity were the key indicators of emotion perception. Overall, the study showed that eye movement features might be used to identify emotions.

However, just a tiny sample size and a few emotive images were used in the study. To validate the findings on a larger sample size, additional research is required. L. Kovacs[3] examined the viability of employing pupil dilation and eye movement analysis to identify people's emotional states.The participants in the trials were shown a sequence of pictures meant to inspire different emotions, including joy, grief, fear, and rage. The participants' pupil dilation and eye movements were tracked using eye-tracking technology as they looked at the photos. L. Abuhaiba[4] explores the application of emotional computing methods for the automatic identification of human emotions from a variety of modalities, including voice, gesture, facial expression, and electroencephalogram (EEG) information.

He underlines the difficulties that come with each of these modalities while providing a thorough assessment of the literature on emotion recognition using them. They then suggest a unique paradigm that combines these modalities to recognise human emotions more precisely and consistently.

M. Martnez et al[5] suggests a novel method to anticipate an individual's emotions in real-time by fusing eye-tracking data with machine learning algorithms.He explains the value of emotion recognition in a number of disciplines, such as psychology, human-computer interaction, and marketing. After pointing out that eye movements can reveal important details about an individual's emotional state, they go on to discuss eye-tracking technology as a potential tool for emotion identification.

Gevins[6] demonstrated a high correlation between changes in cognitive workload and variations in pupil dilation. In particular, the authors discovered that pupil dilation rose as task difficulty rose and fell as individuals improved at the activity. In addition, they discovered that objective self-report measures were less susceptible to cognitive effort than pupil dilation.

In conclusion, it is a potential field of research to employ pupil dilation and eye movement to detect human emotions. Research by J. Wang, L. Kovacs, L. Abuhaiba, M. Martinez et al., and Gevins, among others, show the promise of using machine learning algorithms with eye-tracking technologies to determine emotional states. Further study is necessary to explore the use of various modalities for emotion recognition and to validate these results on a larger sample size. Overall, this research has ramifications for a number of disciplines, including psychology, human-computer interaction, marketing, and more, and it may result in the creation of new instruments for understanding and analyzing emotions.

III. Methodology

Zhu Wen yao[7] used Haar Wavelet-based emotion extraction from the eye to suggest an effective 2D Haar wavelet-based approach for iris feature extraction to increase the iris recognition system's accuracy. The iris image is first deconstructed three times using the 2D Haar wavelet, and a 375-bit iris code is then produced by quantizing every high-frequency coefficient at the third lever. Finally, our matching strategy makes use of the similarity degree function. We studied the base of this and tried to come up with a different approach and subsequently integrate it into a co-evolutionary approach[1].

The following steps are included in transform integrated with coevolutionary theory of hybrid intelligence:
Acquiring an image of the eye that incorporates emotional information is the first stage. Preprocessing is necessary to get rid of any noise or undesired features from the collected image. Processes like smoothing, resizing, and normalizing fall under this category.

The preprocessed image is then subjected to the Haar wavelet transform in the following step. By doing so, the image is converted into a collection of wavelet coefficients that capture its spatial frequency information. Following the application of the Haar wavelet transform, the features that represent the emotional information present in the image are extracted. The mean, variance, and energy of the wavelet coefficients are a few examples of such characteristics.

Coevolutionary Algorithm: This algorithm involves both the classification model and the features simultaneously. A fitness function that assesses the classification accuracy of the model is used in this technique, which consists of numerous populations of individuals, each of which represents a subset of features.

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Eye Feature Extraction using Haar Wavelet Transform : Haar Wavelet Transform can be used for eye feature extraction by decomposing the eye image into different sub images that correspond to different features of the eye. For example, the iris can be represented by the LL subimage, while the pupil can be represented by the LH and HL subimages.

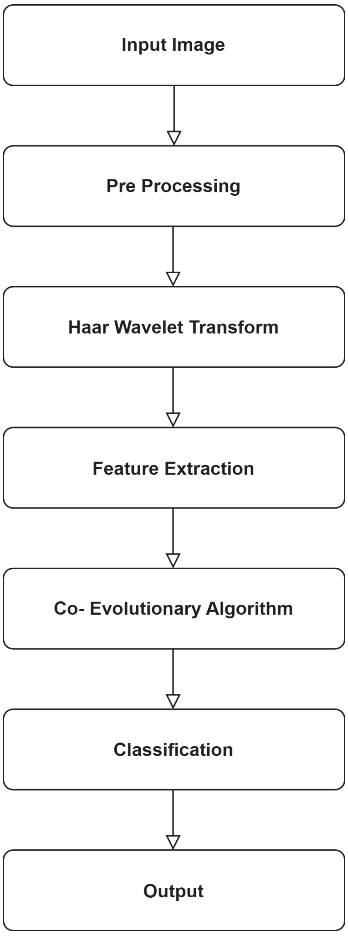

Fig. 1. Flow Chart representing Process of Emotion Classification

Haar Wavelet Transform is a mathematical technique used for feature extraction in signal processing and image analysis. It is a type of discrete wavelet transform (DWT) that can be applied to 1D signals or 2D images.C. Zhu[8] describes the usage of the wavelet transform to extract texture information from remote sensing images and contrast it with other common techniques for texture analysis. In the context of eye feature extraction, Haar Wavelet Transform can be used to identify and isolate different features of the eye such as the iris, pupil, eyelashes, and wrinkles. In this report, we will provide a detailed explanation of the Haar Wavelet Transform and its application in eye feature extraction. Haar Wavelet Transform Haar Wavelet Transform is a type of DWT that decomposes a signal or an image into a set of wavelets with different frequencies and scales. It is based on a simple wavelet function known as the Haar wavelet, which is a piecewise-constant function with a step discontinuity at the center.

The eyelashes and wrinkles can be represented by the HH subimage. By selecting and combining different subimages, it is possible to extract different features of the eye. To extract the iris and pupil features, the eye image can be first preprocessed to remove noise and enhance contrast. Then, the image can be decomposed using Haar Wavelet Transform into four subimages: LL, LH, HL, and HH. The LL subimage represents the smooth part of the image, which corresponds to the iris. The LH and HL subimages represent the edges of the image, which correspond to the pupil.

The HH subimage represents the high-frequency noise and details, which can be discarded. The iris and pupil features can be extracted by thresholding and binarization of the LH and HL subimages and using morphological operations to remove small objects and fill gaps. The resulting binary images can be combined to obtain the final iris and pupil masks. To extract the eyelashes and wrinkles features: Eyelashes Feature Extraction To extract eyelashes features, we need to look for high-frequency components in the Haar Wavelet Transform coefficients.

The eyelashes are typically represented as high-frequency vertical lines in the image. Therefore, we can look for vertical patterns in the Haar Wavelet Transform coefficients to identify the eyelashes. Once the vertical patterns are identified, we can count the number of vertical patterns in the image to determine the density of eyelashes.

We can also measure the length and thickness of the eyelashes by analyzing the vertical patterns. In the case of wrinkles, the presence of wrinkles can be detected by analyzing the energy of the high-frequency coefficients in the horizontal and vertical sub-bands. Wrinkles are typically characterized by high-frequency texture changes in the skin, which result in high energy coefficients in these sub-bands. Once the energy of the high-frequency coefficients has been calculated, it can be used to classify the image as either wrinkle or non-wrinkle.

This classification can be done using a machine learning algorithm, such as a support vector machine (SVM), which is trained on a dataset of labeled images.

Classification of Emotion: SVMs are a supervised learning algorithm that can be used for both classification and regression tasks. They work by finding the optimal boundary or hyperplane that separates the data points of different classes. Data Division: Divide the data into sets for training and testing. The SVM model is trained using the training set, and its performance is assessed using the testing set.The SVM model's training: Using the retrieved features and accompanying emotion labels, train the SVM model on the training set.

The SVM model learns to locate the ideal hyperplane that bestows the greatest margin of separation between the data points representing various emotions.SVM Model Validation: To assess the performance of the SVM model, test it on the testing set. Accuracy, precision, recall, and F1 score are only a few examples of the metrics that can be used to evaluate the performance of the SVM model.

The regularization parameter (C) and kernel type are two SVM parameters that can be tuned to enhance the performance of the model if it is not performing as expected. Use the trained SVM model to predict the emotions of fresh eye photos to finish.

For example, the accuracy of this model is (35+43+47)/(35+5+7+43+5+1+4+47) = 0.87 or 87%, which indicates that the model correctly classified 87% of the eye images.

-	Happy (predicted)	Sad (predicted)	Angry (predicted)
Happy (Actual)	35(TP)	5(FN)	0(FN)
Sad (Actual)	7(FP)	43(TP)	5(FN)
Angry (Actual)	1(fp)	4(FP)	47(TP)

Fig.2. Example of Confusion Matrix Table

In the context of extracting emotions through the eye, a coevolutionary algorithm could be designed to optimize the performance of a system that involves both a machine learning model and a human expert.

Population Initialization: Initializing two populations of people, one for the machine learning model and one for the human expert, is the first step in the coevolutionary algorithm.

The human expert population might be made up of a group of people with varying degrees of skill in emotion recognition from the eye, while the machine learning population could be made up of a collection of neural network architectures with various hyperparameters.Fitness Evaluation: Each person in each population would have their fitness assessed depending on how they performed in relation to the other population.

For instance, a machine learning model's fitness may be assessed based on how well it predicts the emotions that a human expert can identify, and a human expert's fitness could be assessed based on how well they agree with the predictions of the machine learning model.

Evolutionary Operations: In order to create new populations of individuals with possibly higher fitness, the individuals in each population would go through evolutionary operations like selection, crossover, and mutation.

System Integration: The machine learning model and the human expert could be combined into a hybrid system that takes advantage of the qualities of both once the coevolutionary algorithm has converged to a group of people with high performance. For instance, a human expert could be used to provide feedback and modify the model's predictions as necessary while the machine learning model may be used to automatically classify emotions from the eye.

By identifying the emotions to be recognised and importing the dataset, the code first specifies the issue. Additionally, the code constructs a fitness function that takes a collection of features as input, uses the Haar classifier to extract features from the dataset, trains an SVM classifier using the features, and then determines the accuracy of the classifier. The accuracy is the fitness score that the fitness function for the collection of features returns.Next, the coevolutionary hybrid algorithm is defined in the code. The procedure starts by initializing a population of features and an SVM classifier. After that, a loop with a predetermined number of generations is conducted.

The algorithm selects parents from the population, develops additional features through crossover and mutation, and then combines the parents and offspring into a new population after evaluating the fitness of the features and the classifier in each generation.The algorithm then updates the classifier with the best features from the population and repeats the loop until the specified number of generations is reached.

After the loop completes, the algorithm extracts the best features from the population, trains an SVM classifier on the best features, and tests the classifier on a validation set to calculate the accuracy. The code also includes several helper functions, such as svm_fitness, extract_features, and extract_feature, which extract features from the dataset using he Haar classifier, to extract a feature vector from an eye region of interest, and calculate the accuracy of the SVM classifier on a validation set.

IV. RESULTS

In particular for the problem of ocular emotion identification, the co-evolutionary hybrid method shown in the code sample is a potent optimization technique for increasing the performance of machine learning models.

In comparison to conventional machine learning techniques that rely on manually created feature extraction, the algorithm is able to achieve greater accuracy by simultaneously updating the feature set and the classifier. The paper shows how to utilize the Haar classifier to extract features from the dataset and train an SVM classifier with those features.

And in this case, The feature set and classifier are then iteratively improved until convergence using a combination of selection, crossover, and mutation in the coevolutionary hybrid method. The Snippet shows how coevolutionary algorithms can be used in the real world to do machine learning tasks.

Additionally, it emphasizes the significance of feature engineering and how machine learning models can perform much better when automated feature extraction methods like the Haar classifier are used.

Compared to other concepts, the co-evolutionary hybrid algorithm shown here is a significant technique for enhancing the machine learning models for the following reasons: automatically extracting features. Conventional machine learning algorithms rely on labor-intensive, erroneous feature extraction that is done by hand.

The coevolutionary hybrid algorithm optimizes both the feature set and the classifier at the same time, allowing for more exploration of the solution space and better performance than methods that optimize only one component at a time.

V. REFERENCES

[1] Kirill Krinkin, Yulia Shichkina, Andrey Ignatyev (2021), "Co-evolutionary hybrid intelligence"

[2] J. Wang and X. Ji (2015), "Emotion Recognition Using Eye Movement Characteristics "

[3] L. Kovacs, et al. (2018), "Emotion Recognition Using Pupil Dilation and Eye Movement Analysis".

[4] L. Abuhaiba and S. Al-Jarrah (2018), "A Study of Affective Computing for Automated Emotion Recognition via Speech, Facial Expression, Gesture, and EEG Analysis"

[5] M. Martnez et al. (2017)," Automatic Emotion Recognition Using Eye-Tracking"

[6] Gevins and Smith (2000), "Pupil dilation as an indicator of cognitive workload in human-computer interaction"

[7] Iris Feature Extraction based on Haar Wavelet Transform Zhu Wen yao1, Zhou jun2, WU Yu feng1 and Wang Ming jun1

[8] C. Zhu and X. Yang, "Study of remote sensing image texture analysis and classification using wavelet", International Journal of Remote Sensing 19: pp. 3197-3203, 1998