UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Tecnura

https://revistas.udistrital.edu.co/ojs/index.php/Tecnura/issue/view/1214

DOI: https://doi.org/10.14483/22487638.22100

INVESTIGACIÓN

Emotion Recognition in the Eye Region Using Textural Features, IBP and HOG

Reconocimiento de Emociones en la Región de los Ojos Utilizando Características Texturales, IBP y HOG

Laura D. Jalili [1], Josué Espejel-Cabrera [2], Jair Cervantes [3] y Farid García-Lamont [4]

Fecha de Recepción: 30 de enero de 2023 **Fecha de Aceptación**: 25 de abril de 2024

Cómo citar: Dominguez-Jalili., L.Y., Espejel-Cabrera J., Cervantes J. y García-Lamont F. (2024). Emotion recognition by eyes region using textural features, lbp and hog. *Tecnura*, **28**(79), 12-33. https://doi.org/10.14483/22487638.22100

Abstract

Objective: Our objective is to develop a robust emotion recognition system based on facial expressions, with a particular emphasis on two key regions: the eyes and the mouth. This paper presents a comprehensive analysis of emotion recognition achieved through the examination of various facial regions. Facial expressions serve as invaluable indicators of human emotions, with the eyes and mouth being particularly expressive areas. By focusing on these regions, we aim to accurately capture the nuances of emotional states.

Methodology: The algorithm we devised not only detects facial features but also autonomously isolates the eyes and mouth regions. To enhance classification accuracy, we utilized various feature extraction and selection techniques. Subsequently, we assessed the performance of multiple classifiers, including Support Vector Machine (SVM), Logistic Regression, Bayesian Regression, and Decision Trees, to identify the most effective approach.

[1] She received her Bachelor's degree in Computer Engineering from the Autonomous University of the State of Mexico, Mexico, in 2014. Subsequently, she pursued both her Master's and Doctorate degrees at the same institution, completing them in 2018 and 2022 respectively. Since 2018, she has served as a professor in the Department of Computing at the Autonomous University of the State of Mexico (CU UAEMex-Texcoco). Her research interests encompass pattern recognition, neural networks, and data science.
Email: lydominguezj@uaemex.mx

[2] Received his Bachelor's degree in computer sciences from C.U UAEM Texoco in 2012, and in 2015 received his master's degree in computer sciences. get into the private sector for a few years, and come back to C.U UAEM Texcoco to study a Ph. D. in computer sciences and obtain his degree in 2022. His research interests include deep learning, computer vision, and pattern recognition.
Email: jespejelc331@alumno.uaemex.mx

[3] Received the B.S. degree in Mechanical Engineering from Orizaba Technologic Institute, Veracruz, Mexico, in 2001 and the M.S degree and Ph. D. from CINVESTAV-IPN, México, in 2005 and 2009. Since 2010, he has been a professor of the Departamento de Computación, Universidad Autónoma del Estado de México (UAEMex-Texcoco). His research interests include support vector machines, pattern classification, neural networks, fuzzy logic and clustering. jcervantesc@uaemex.mx

[4] Farid Garcia-Lamont received a PhD degree in Computer Science from CINVESTAV-IPN, Mexico in 2010. Since 2012 he has been professor at the Autonomous University of State of Mexico. His research interests are pattern recognition, applications of artificial intelligence and image processing.
Email: fgarcial@uaemex.mx

Results: Our experimental methodology involved employing various classification techniques to assess performance across different models. Among these, SVM exhibited exceptional performance, boasting an impressive accuracy rate of 99.2 %. This outstanding result surpassed the performance of all other methods examined in our study. Through meticulous examination and experimentation, we explore the effectiveness of different facial regions in conveying emotions. Our analysis encompasses two datasets and evaluation methodologies to ensure a comprehensive understanding of emotion recognition capabilities.
Conclusions: Our investigation presents compelling evidence that analyzing the eye region using a Support Vector Machine (SVM) along with textural, HoG, and LBP features achieves an outstanding accuracy rate of 99.2 %. This remarkable finding underscores the significant potential of prioritizing the eyes alone for precise emotion recognition. In doing so, it challenges the conventional approach of including the entire facial area for analysis.

Keywords: Emotion recognition, regions, textural features, LBP, HoG

RESUMEN

Objetivo: Nuestro objetivo es desarrollar un sistema robusto de reconocimiento de emociones basado en expresiones faciales, con especial énfasis en dos regiones clave: los ojos y la boca. Este artículo presenta un análisis exhaustivo del reconocimiento de emociones logrado mediante el examen de varias regiones faciales. Las expresiones faciales sirven como indicadores invaluables de las emociones humanas, siendo los ojos y la boca áreas particularmente expresivas. Al centrarnos en estas regiones, nuestro objetivo es capturar con precisión los matices de los estados emocionales.
Metodología: El algoritmo que ideamos no sólo detecta rasgos faciales, sino que también aísla de forma autónoma las regiones de los ojos y la boca. Para aumentar la precisión de la clasificación, utilizamos varias técnicas de extracción y selección de características. Posteriormente, evaluamos el rendimiento de múltiples clasificadores, incluida la máquina de vectores de soporte (SVM), la regresión logística, la regresión bayesiana y los árboles de decisión, para identificar el enfoque más eficaz.
Resultados: Nuestra metodología experimental implicó la utilización de varias técnicas de clasificación para evaluar el rendimiento en diferentes modelos. Entre ellos, la SVM exhibió un rendimiento excepcional, con una impresionante tasa de precisión del 99,2 %. Este resultado sobresaliente superó el rendimiento de todos los demás métodos examinados en nuestro estudio. A través de un examen y una experimentación meticulosos, exploramos la eficacia de diferentes regiones faciales para transmitir emociones. Nuestro análisis abarca dos conjuntos de datos y metodologías de evaluación para garantizar una comprensión integral del reconocimiento de emociones.
Conclusiones: Nuestra investigación presenta evidencia convincente de que cuando se analiza la región del ojo utilizando la máquina de vectores de soporte (SVM) junto con las características de textura, HoG y LBP, se obtiene singularmente una tasa de precisión excepcional del 99,2 %. Este notable hallazgo subraya el importante potencial de priorizar únicamente los ojos para el reconocimiento preciso de las emociones. Al hacerlo, desafía el enfoque convencional de incluir toda el área facial para el análisis.

Palabras clave: Reconocimiento de emociones, regiones, Características texturales, LBP, HoG

INTRODUCTION

The COVID-19 pandemic has profoundly reshaped our interactions, leading to a surge in digital learning, remote work, online shopping, and other virtual activities. Notably, it has also impacted the performance of automated systems, particularly in facial recognition (Manley *et al.*, 2022). With masks becoming an integral part of our daily attire, traditional facial recognition systems are struggling to maintain accuracy and efficiency, especially as concerns emotion detection (Kaur *et al.*, 2022). This paper introduces a novel method designed for emotion recognition under these altered conditions, offering a comparative analysis of our findings.

The advancement of vision systems has found applications across diverse sectors, including agriculture (Cervantes *et al.*, 2017, AlHaditi *et al.*, 2016), medicine (Gonzalez *et al.*, 2017, Villan *et al.*, 2017) and industry (Casas *et al.*, 2017, Sanchez *et al.*, 2017). Emotion recognition has similarly benefited from significant technological progress.

Emotion recognition presents unique challenges due to the intricate and often ambiguous nature of variables influencing emotional states. Emotions manifest differently across individuals, making their recognition a complex endeavour. However, it holds significant importance in various domains, from human health to interpersonal relationships and decision-making processes.

Facial expressions serve as a primary channel of nonverbal communication, conveying our specific moods and offering insights into our interactions. While systems for emotion recognition have explored multiple variables like voice and body movements, the human face remains the most prominent medium for the display of emotion. Moreover, involuntary facial movements often unconsciously reveal our emotional states.

Despite the advancements in emotion recognition technology, most existing systems rely on capturing the entire face to detect emotions. Given the current "new normal"where parts of the face, particularly the mouth area, may be obscured by masks, there is a pressing need for algorithms capable of recognizing emotions from partial facial cues.

In this study, we propose a methodology centred on facial expressions for the recognition of six universally recognized emotions: anger, fear, disgust, happiness, sadness, and surprise, as categorized by Paul Ekman (Ekman, 1993).

RELATED WORK

The study of emotion recognition encompasses several crucial steps, including preprocessing, face detection, and feature extraction techniques, all aimed at enhancing classification accuracy. Preprocessing plays an important role in refining images for optimal results. Notably, histogram equalization emerges as a standout preprocessing technique, facilitating an even intensity distribution across the image, thereby improving tonal consistency for enhanced analysis. Additionally, various filters are employed as needed, with Gaussian filters being a common choice for edge smoothing in most systems (Rao *et al.*, 2019).

In face detection, the widely utilized Viola-Jones technique stands as a cornerstone method (Rao *et al.*, 2019, Xiaohua *et al.*, 2019, Chakraborty *et al.*, 2009, He & Zhang, 2018). However, critiques regarding its efficiency in handling facial occlusion have led researchers to explore alternatives like the Multi-task Cascade Convolutional Neural Network (MTCNN), leveraging its state and regression vector-based approach for improved performance (Hossain & Muhammad, 2019).

On the classification front, Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) are among the most prevalent choices. While SVMs exhibit higher error rates compared to CNNs, the latter's computational intensity and the large amount of data required for CNNs pose challenges for practical implementation (Rao *et al.*, 2019, Hossain & Muhammad, 2019).

Recent years have seen a surge in research efforts aimed at emotion recognition, with diverse approaches leveraging various input modalities such as voice, environmental noise, hand and body movements, and facial expressions. For instance, some studies combine facial and voice features, utilizing algorithms like Mel Frequency Cepstral Coefficients (MFCC) for voice analysis alongside methods like MSER and Viola-Jones for facial feature extraction. However, the ascendancy of CNNs in recent years has reshaped this landscape (Rao *et al.*, 2019, Hossain & Muhammad, 2019).

Alternatively, sensor-based systems offer a novel avenue for emotion detection, albeit with concerns about the invasiveness of their everyday use. Hybrid approaches combining CNNs and recurrent neural networks (RNNs) have gained traction, while electroencephalography (EEG) signals serve as potent indicators of emotional states. Features extracted from EEG signals, including frequency, time, and wavelet characteristics, enable the discernment of emotional responses to audio or visual stimuli (Liang *et al.*, 2019, Kurbalija *et al.*, 2018, Chen *et al.*, 2018).

Facial and bodily movements also play a vital role in emotion recognition, with techniques ranging from 3D sequence analysis employing Riemann's analysis to identify facial expression deformations to mesh-based approaches for facial movement tracking. Body movement tracking involves the identification of interest points and subsequent extraction using algorithms like three-dimensional movement with temporal Hidden Markov Models (HMM) (Santhoshkumar & Geetha, 2019, Amor *et al.*, 2014).

Methodology

Facial expressions serve as a crucial form of non-verbal communication, offering insights into an individual's emotional state. These expressions manifest across various regions of the face, including the eyes, eyebrows, nose, and mouth. For this project, our focus narrows to the intricate dynamics of the mouth and eyes. These areas, known for their significance in conveying emotions, are selected as our primary areas of interest.

The methodology employed in this research endeavor is meticulously outlined and depicted in Figures **1** and **2**, offering a comprehensive overview of our proposed approach.

To begin our analysis, we extract the regions of interest from each image, focusing specifically on the mouth and eyes. The feature extraction process involves isolating the eyes and mouth regions independently, enabling a comprehensive examination of each area.

For feature extraction, we employ three distinct techniques: Haralick Textural features, Histograms of Ordered Gradients (HOG) and Local Binary Patterns (LBP), applied to both the eyes and mouth regions. These techniques enable us to capture intricate details and patterns crucial for emotion classification. Additionally, all color images are converted to grayscale to facilitate feature extraction. Remarkably, due to the grayscale nature of the original images, preprocessing steps were deemed unnecessary for training our final model.

Region of interes

Initially, attempts were made to extract the areas of interest using a cascade classification model trained specifically for detecting mouth and eye objects. However, these efforts yielded suboptimal

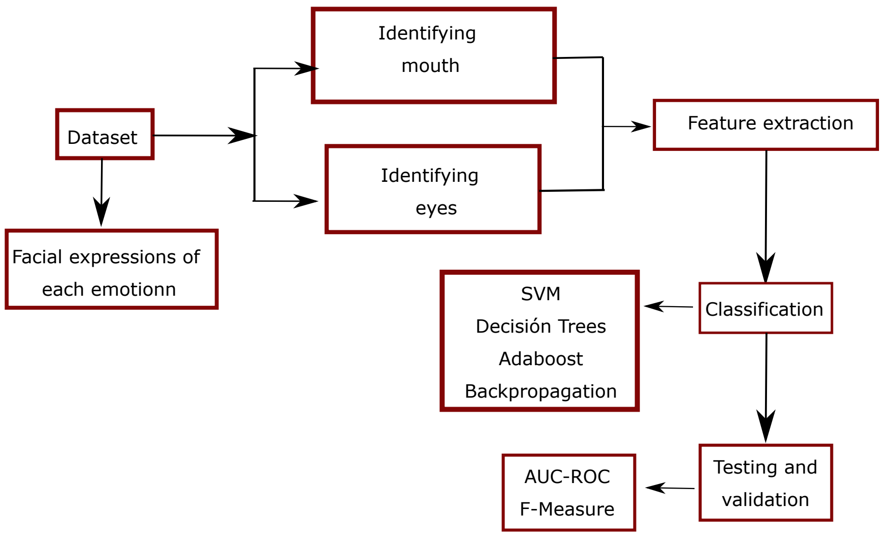

Figure 1. Methodology

Source: own work.

results, as the algorithm frequently misidentified the regions during testing. Consequently, a two-step approach was formulated to accurately identify the regions of interest.

In the first step, the Viola-Jones algorithm is employed to detect faces using cascade classification, providing a foundational framework for subsequent analyses. This algorithm is well-known for its effectiveness in detecting facial features and serves as a reliable starting point for our image analysis.

Then, in the second step, the areas encompassing the eyes and mouth are calculated. This process involves identifying specific points within each region of interest, including the upper-left corner (P_1), upper-right corner (P_2), lower-right corner (P_3), and lower-left corner (P_4). To ensure robustness and accuracy in identifying these key points, an average of 20 images is utilized. The calculation of each Pi point in the new images is performed as follows:

For each point P_i , the coordinates are determined based on the average position of the respective feature across the 20 images. This averaging technique helps to minimize errors and enhances the precision of the feature localization process.

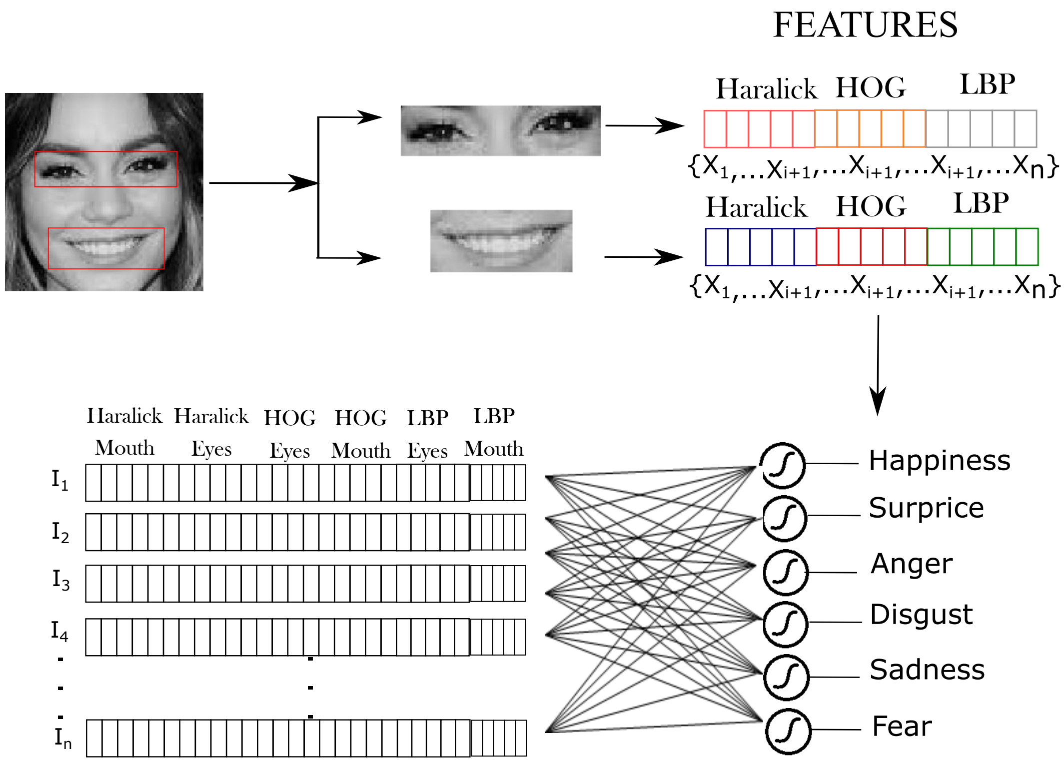

Figure 2. Feature extraction of the proposed methodology

Source: own work.

The improved two-step approach not only enhances the accuracy of the identification of regions of interest but also provides a more reliable foundation for subsequent image analysis tasks. By combining the strengths of the Viola-Jones algorithm with meticulous point calculation, the proposed method aims to overcome the challenges encountered with the initial single-step cascade classification model.

$$P_{ix} = FP_{ix} * NewRes_x \tag{1}$$

$$P_{iy} = FP_{iy} * NewRes_y \tag{2}$$

Here, $NewRes_x$ denotes the resolution along the x axis of the image acquired through the Viola-Jones face recognition algorithm, while $NewRes_y$ denotes the resolution along the y axis of the image, and FP_{ix} and FP_{iy} represent the average factor calculated with 20 images. It is calculated as follows:

$$FP_{ix} = \frac{1}{n} \sum_{k=1}^{n} \frac{x_k}{Res(x_k)} \tag{3}$$

$$FP_{iy} = \frac{1}{n} \sum_{k=1}^{n} \frac{y_k}{Res(y_k)} \tag{4}$$

Where $Res(x_k)$ and $Res(y_k)$ denote the resolution of the face in the x and y dimensions respectively, with n representing the number of figures, which in our case is set to 20.

Using this approach, the regions of interest for the eyes and mouth were determined, and from each region, a segment was extracted from the original image, resulting in two distinct sets. Subsequently, Haralick textural characteristics were extracted from each set, along with the Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) features, which pertain to shape and texture exclusively.

The extraction of Haralick textural features involved obtaining values for the second angular momentum, contrast, correlation, sum of squares, inverse difference moment, sum average, entropy, variance difference, entropy differences, correlation measures, and correction coefficient. Gradients of ordered histograms were derived using nine bins of size 20×20 with a displacement of 10. HOG proved instrumental in capturing facial expression shapes by leveraging directional gradients and magnitudes, thereby accommodating variations in lighting across images. Additionally, the LBP algorithm facilitated the extraction of shape features by comparing neighboring pixels, complementing the capabilities of HOG and contributing valuable features to the project. The three types of features extracted from the data are illustrated in Figure **2**.

To construct the training model, feature vectors from all extracted characteristics were consolidated. Various classifiers, including a Support Vector Machine (SVM), Adaboost, decision trees, backpropagation, and regression, were employed.

For the SVM classification model, a Kernel rbf was utilized with a complexity parameter set to $C = 0{,}0001$ and gamma set to 1000. This configuration is chosen to achieve optimal classification performance while maintaining computational efficiency.

The comprehensive feature extraction and utilization of multiple classifiers aim to enhance the robustness and accuracy of the proposed model. By combining a variety of features and employing diverse classification techniques, the model is designed to effectively capture and classify intricate details and variations in facial expressions and features.

RESULTS

Datasets

Two datasets containing spontaneous facial micro-expressions were utilized in the experiments:

SMIC (Spontaneous Micro-Expressions Database):

The SMIC database comprises 164 spontaneous micro-expressions elicited from 16 participants through emotionally evocative movie clips. Data was captured using a high-speed camera at 100 frames per second, resulting in 1909 images with a resolution of 186 x 227 pixels. For this study, five emotions were considered: Joy, Surprise, Disgust, Sadness, and Anger (Li *et al.*, 2018, Li *et al.*, 2013, Pfister *et al.*, 2011).

SAMM (Spontaneous Actions and Micro-Movements): The SAMM dataset consists of 159 spontaneous micro-facial movements elicited through emotional induction from 32 participants representing diverse demographics. The dataset includes individuals from 13 ethnicities, with a mean age between 24 and 33 years (17 men and 16 women). The SAMM dataset contains 3634 images with a resolution of 960 x 650 pixels. Six emotions were analyzed in this study: Joy, Surprise, Anger, Disgust, Sadness, and Fear (Davison *et al.*, 2018, Yap *et al.*, 2020).

Results with both regions

Emotion recognition systems often rely on facial images as input data. While the entire face provides valuable information, research suggests that specific regions, particularly those surrounding the eyes and mouth, hold the most salient features for discerning emotional expressions. This observation stems from the fact that these areas are highly mobile and exhibit distinctive muscle activations associated with various emotions. For instance, the eyes convey emotions through changes in gaze direction, pupil dilation, and the formation of wrinkles around the corners (çrow's feet"). Similarly, the mouth plays a crucial role in expressing happiness through smiles, sadness through frowns, and surprise through open mouth expressions.

To investigate the relative importance of these facial regions in emotion recognition, we conducted experiments focusing on two specific approaches: **a) Combined Eye and Mouth Regions**: This approach leverages information from both the eye and mouth regions, capitalizing on the combined expressive power of these areas. By analyzing features extracted from both regions, we aim to assess the accuracy and effectiveness of emotion recognition when utilizing a comprehensive facial representation.

b) Eye Region Only: This approach focuses solely on the eye region, exploring the hypothesis that the eyes alone may provide sufficient information for accurate emotion recognition. This investigation seeks to determine the extent to which emotions can be effectively discerned without considering features from the mouth region.

By comparing the performance of these two approaches, we can gain valuable insights into the significance of the eye region in emotion recognition. If the eye-only approach achieves comparable accuracy to the combined approach, it suggests that the eyes indeed hold crucial information for discerning emotional states. This finding could have implications for the development of more efficient and targeted emotion recognition systems, potentially requiring less computational resources while maintaining accuracy. Additionally, understanding the specific contributions of the eye region could pave the way for advanced techniques that focus on subtle eye movements and expressions, further enhancing the capabilities of emotion recognition technology.

This section presents a comprehensive evaluation of the proposed method's performance on two distinct datasets, comparing its efficacy with various established classifiers using previously defined performance metrics.

Initially, we present the outcomes achieved using diverse metrics for the classifier exhibiting the highest efficacy. This encompasses the display of confusion matrices generated by the four assessed classifiers.

To ensure consistency and reproducibility, we employed a standardized experimental setup across all evaluations. Data features were normalized to achieve a mean of zero and a standard deviation of one, mitigating the potential influence of scale differences on classifier performance. A 10-fold cross-validation approach was implemented to assess the generalizability of the results and reduce the risk of overfitting. For each classifier, optimal hyperparameters were determined through an exhaustive grid search, ensuring that each model was fine-tuned for the specific datasets and task.

We present the results obtained using the most effective classifier first, showcasing its performance across various metrics. Additionally, confusion matrices are provided for all four evaluated classifiers, offering deeper insights into their classification behavior and potential misclassifications.

Table 1 displays the performance metrics achieved using an SVM classifier with a Gaussian kernel. The hyperparameters were optimized through cross-validation and grid search, resulting in C=1000 and G=0.005. The results demonstrate high precision across all emotion classes, indicating the model's ability to accurately identify emotions with minimal false positives.

Figure **3** illustrates the confusion matrices generated by the four classifiers. Along the diagonal, both the number and percentages of accurately classified instances are depicted. Notably, the SVM classifier exhibits the highest performance, closely followed by Logistic Regression. Across all classes, precision values hover near 1. Remarkably, the experiments revealed only 12 errors out of a dataset comprising 3634 instances.

Table 2 presents the performance metrics achieved using an SVM classifier with a Gaussian kernel for emotion recognition. The hyperparameters, C and G, were optimized through a rigorous process of cross-validation and grid search, resulting in values of 1000 and 0.005, respectively. This optimization ensures that the SVM model is fine-tuned to effectively capture the underlying patterns within the data and maximize its discriminative power for emotion classification.

The results displayed in Table 2 highlight the remarkable precision achieved by the SVM clas-

Tabla 1. SVM classification results with the SAMM dataset

Class	TPR	Accuracy	Recall	F-measure	MCC	AUC-ROC
1	0.993	0.994	0.994	0.992	0.998	0.993
2	0.997	0.995	0.996	0.994	0.998	0.993
3	1.000	0.999	0.999	0.999	1.000	0.999
4	0.998	1.000	0.999	0.999	1.000	0.999
5	1.000	1.000	1.000	1.000	1.000	1.000
6	0.993	0.995	0.994	0.994	0.999	0.992

Source: own work.

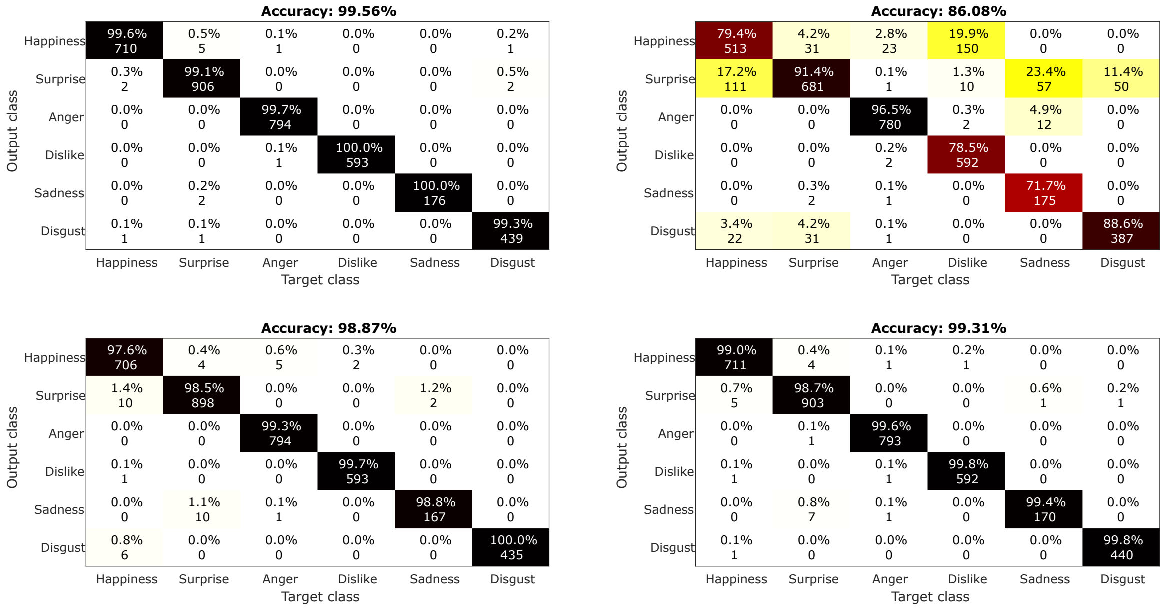

Figure 3. Confusion matrices obtained for the SAMM dataset with eyes and mouth.

Source: Own work.

sifier across all identified emotion classes. This indicates that the model exhibits a low rate of false positive predictions, meaning it rarely misclassifies other emotions as a specific target emotion. This high precision is crucial for real-world applications of emotion recognition, where minimizing false positives is often essential.

Figure **4** presents the confusion matrices for the four evaluated classifiers on Dataset 2, which comprises five distinct emotion classes as previously described. The diagonal elements of each ma-

Tabla 2. SVM classification results with the SMIC dataset

Class	TPR	Accuracy	Recall	F-measure	MCC	AUC-ROC
1	0.980	0.972	0.976	0.968	0.990	0.963
2	0.981	0.984	0.982	0.975	0.994	0.977
4	0.971	0.993	0.982	0.979	0.996	0.978
5	0.991	0.996	0.994	0.993	0.998	0.991
6	0.978	0.965	0.972	0.965	0.990	0.953

Source: own work.

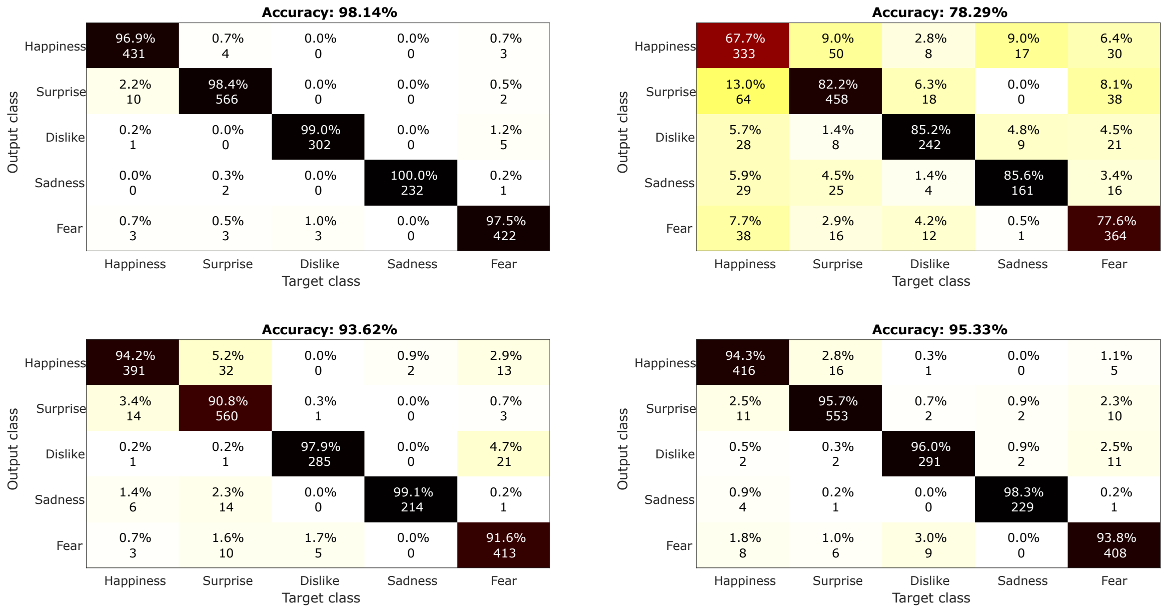

Figure 4. Confusion matrices obtained for the SMIC dataset with eyes and mouth.

Source: own work.

trix represent the number and percentage of correctly classified instances for each emotion category. By analyzing these matrices, we gain valuable insights into the strengths and weaknesses of each classifier and their ability to differentiate between various emotional expressions.

A comparative analysis of the confusion matrices clearly demonstrates that the SVM classifier outperforms the other models, achieving the highest overall accuracy and exhibiting remarkable precision across all emotion classes. This is evident from the diagonal dominance observed in the SVM's confusion matrix, indicating a low rate of misclassifications. Notably, the SVM achieved near-perfect precision, with only 38 errors identified out of a total of 1909 data points in the experiments.

Results with the eye region alone

This section delves into the effectiveness of emotion recognition utilizing solely the eye region as input data. We compare the performance of various classifiers using established metrics, aiming to assess the extent to which emotions can be accurately discerned without information from other facial areas.

Similarly to the previous evaluation, we employed a 10-fold cross-validation approach and grid search for hyperparameter optimization to ensure robust and reliable results. We present the performance metrics for the best-performing classifier first, followed by a detailed analysis of confusion matrices and box plots to provide a comprehensive understanding of the results.

Table 3 presents the performance metrics achieved by the SVM classifier with a Gaussian kernel (C=1000, G=0.005). As observed in the previous experiment, the SVM exhibits high precision across all emotion classes, indicating its ability to accurately identify emotions based solely on features extracted from the eye region.

Figure **5** shows the confusion matrices of the four classifiers. On the diagonal, the number and percentages of well-classified classes are shown. As can be seen, the classifier with the best performance is SVM followed very closely by Logistic Regression. For all classes, the precision is very close to 1. In the experiments, only 39 errors were found out of a set of 3634 data. The results show that with the eye region alone, the accuracy of the classifiers falls somewhat. However, this decline is only significant– dropping from 83 % to 76 % accuracy– with the Bayesian classifier, with no significant effect in the three other 3 classifiers, for which the drop is minimal. With this we can deduce that the recognition of emotions using only the eye region with the proposed method is not severely affected.

Figure **5** displays the confusion matrices for all four evaluated classifiers. The diagonal elements represent the number and percentage of correctly classified instances for each emotion category. Consistent with the overall performance metrics, the SVM demonstrates superior accuracy, followed closely by Logistic Regression. Both classifiers achieve near-perfect precision, with only 39 errors out of

Tabla 3. SVM classification results with the SAMM dataset and with only the eyes region

Class	TPR	Precision	Recall	F-measure	MCC	AUC-ROC
1	0.999	0.991	0.999	0.995	0.993	0.998
2	0.983	0.983	0.983	0.983	0.978	0.993
3	0.998	0.999	0.998	0.998	0.998	0.999
4	0.998	0.998	0.998	0.998	0.998	1.000
5	0.909	0.944	0.909	0.926	0.923	0.991
6	1.000	0.998	1.000	0.999	0.999	1.000

Source: own work.

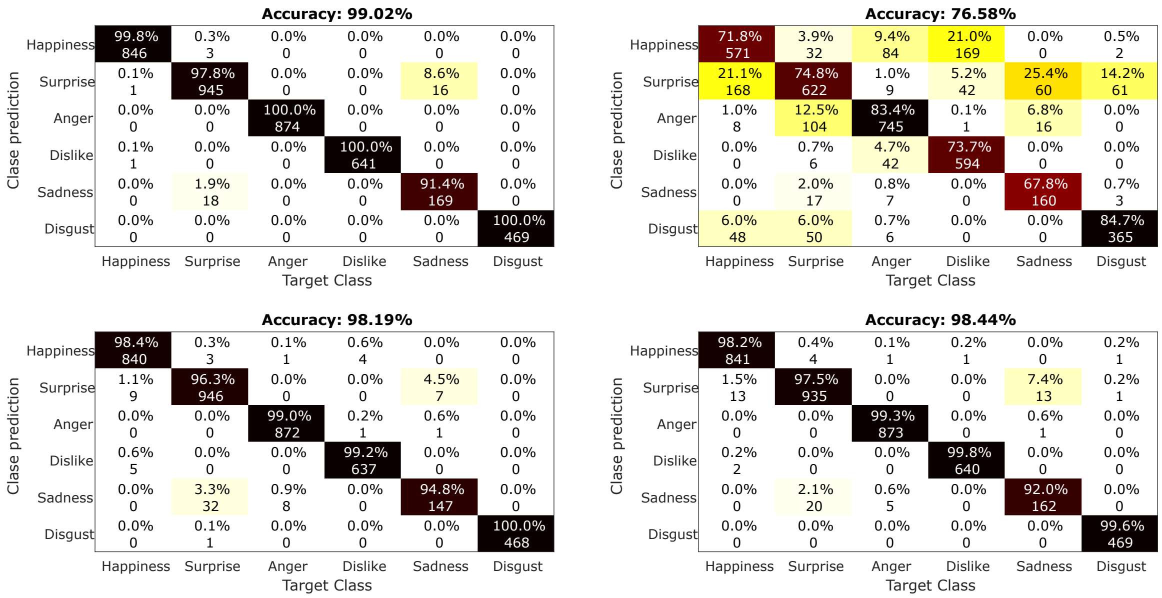

Figure 5. Confusion matrices obtained for the SAMM dataset with eyes and mouth.

Source: own work

3634 data points.

Comparing these results with the previous experiment utilizing both eye and mouth regions, we observe a slight decrease in overall accuracy when relying solely on the eye region. However, this decrease is minimal for the SVM, Logistic Regression, and K-Nearest Neighbors classifiers, suggesting that the eye region provides sufficient information for effective emotion recognition. Notably, the Bayesian classifier experiences a more significant drop in accuracy, highlighting its potential sensitivity to the reduced feature set.

Our findings suggest that the eye region plays a crucial role in emotion recognition, enabling accurate classification even without information from the mouth region. This has significant implications for developing efficient and targeted emotion recognition systems, potentially allowing for reduced computational complexity while maintaining high performance.

Finally, we present the results obtained by applying the proposed method to the SMIC dataset, focusing solely on the eye region for emotion recognition. Table 4 displays the performance metrics achieved using an SVM classifier with a Gaussian kernel, employing the same hyperparameter optimization process as with the SAMM dataset.

The results presented in Table 4 are striking, revealing that the performance achieved using the

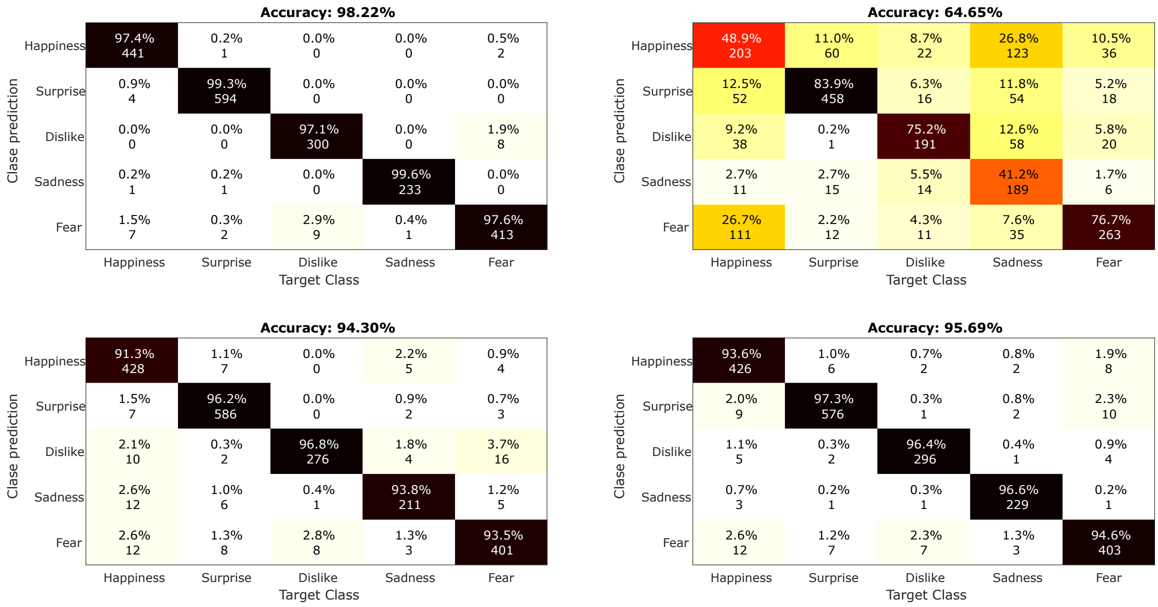

Figure 6. Confusion matrices of different classifiers with SMIC Dataset

Source: own work.

eye region alone surpasses the results obtained when utilizing both the eye and mouth regions in the previous experiment. This finding is particularly significant as the number of extracted features is reduced by half when focusing solely on the eyes, leading to a substantial decrease in computational complexity.

Figure **6** presents the confusion matrices for the four evaluated classifiers on the SMIC dataset, focusing solely on the eye region for emotion recognition. As with previous analyses, the diagonal

Tabla 4. SVM classification results with the SMIC dataset and with only the eyes region

Class	TPR	Accuracy	Recall	F-Measure	MCC	AUCROC
1	0.993	0.974	0.993	0.983	0.979	0.995
2	0.993	0.993	0.993	0.993	0.990	0.996
4	0.974	0.971	0.974	0.972	0.967	0.994
5	0.991	0.996	0.991	0.994	0.993	0.998
6	0.956	0.976	0.956	0.966	0.957	0.985

Source: own work.

elements of each matrix represent the number and percentage of correctly classified instances for each emotion category, providing valuable insights into the classifiers' ability to distinguish between different emotional expressions.

Consistent with the overall performance metrics, the confusion matrices demonstrate that SVM and Logistic Regression achieve the highest accuracy, with near-perfect precision across all emotion classes. Only 36 errors were identified out of a total of 2017 images, highlighting the effectiveness of these classifiers in recognizing emotions based solely on eye region features.

Interestingly, comparing these results with the previous experiment that utilized both eye and mouth regions reveals a noteworthy consideration: focusing solely on the eye region leads to a slight improvement in overall accuracy for the SMIC dataset. This suggests that, for this particular dataset, the eye region provides more discriminative and informative features for emotion recognition than the combined eye and mouth regions.

The observed advantage of the eye-only approach might be attributed to specific characteristics of the SMIC dataset. For instance, the dataset might contain subtle eye expressions that are highly indicative of emotional states, while mouth expressions may be less pronounced or consistent.

The effectiveness of the eye-focused approach could also depend on the specific feature extraction techniques employed. It is possible that the extracted eye features capture more nuanced and relevant information compared to the mouth features, leading to improved performance.

Individual variations in expressiveness and facial muscle activation patterns could also contribute to the observed results. For example, some individuals might exhibit more pronounced emotional cues around the eyes, making the eye region a more reliable source of information for emotion recognition.

The ability to achieve high accuracy with only the eye region demonstrates the potential for developing computationally efficient emotion recognition systems without compromising performance. This has significant implications for real-world applications where computational resources may be limited, such as embedded systems or mobile devices.

The results reinforce the notion that the eye region contains highly discriminative features for emotion recognition, potentially even more so than the mouth region. This highlights the need for further investigation into the specific eye features that contribute most significantly to accurate emotion classification.

The success of the eye-only approach suggests the possibility of developing specialized emotion recognition systems tailored to specific contexts. For instance, systems designed for video conferencing or human-computer interaction could prioritize the eye region, optimizing processing speed and resource allocation.

CONCLUSIONS

This paper introduces a novel algorithm for emotion recognition that leverages advanced feature extraction techniques and autonomous region-of-interest identification. Building upon the foundational face detection capabilities of the Viola-Jones algorithm, our approach strategically targets specific facial regions to optimize the accuracy of emotion classification.

The research presented herein focuses on a comparative analysis of emotion recognition performance when utilizing two distinct facial regions: the eye area and the entire face. By systematically evaluating the efficacy of our algorithm across these regions, we aim to shed light on the relative importance of targeted feature extraction in emotion detection systems.

Our experimental results demonstrate that focusing on specific facial regions, particularly the eye area, can lead to improved emotion recognition accuracy compared to utilizing the entire face. This finding highlights the potential for developing more efficient and targeted emotion detection systems that prioritize informative facial features.

By extracting features from a smaller region of interest, our algorithm reduces computational demands without compromising performance. This has significant implications for real-world applications where computational resources may be limited, such as in embedded systems or mobile devices.

The comparative analysis provides valuable insights into the role of different facial regions in conveying emotional states. Our findings suggest that the eye area plays a crucial role in emotion recognition, potentially even more so than the mouth region for specific datasets.

While our study focused on the eyes and the entire face, future research could explore the role of additional facial areas, such as the forehead and nose, in emotion recognition. Investigating the contributions of these regions could further enhance the accuracy and robustness of emotion detection systems.

A deeper analysis of individual features within each region, such as eyebrow position, eyelid movement, and lip curvature, is crucial for understanding their specific contributions to the recognition process. This could involve techniques like feature selection or visualization methods to identify the most discriminative features for each emotion.

The expression of emotions exhibits cultural variations, which can influence the effectiveness of emotion recognition systems. Future research should investigate how cultural differences in facial expressions impact recognition accuracy and explore strategies to develop culturally-aware or adaptable models.

References

Al-Hadithi, B. M.; Cena, C. E. G.; León, R. C. & Loor, C. L. (2016). Desarrollo de un Sistema de Iluminación Artificial Inteligente para Cultivos Protegidos. *Revista Iberoamericana de Automática e Informática Industrial RIAI, 13*(4), 421-429. https://doi.org/10.1016/j.riai.2016.07.005

Ben Amor, B.; Drira, H.; Berretti, S.; Daoudi, M. & Srivastava, A. (2014). 4-D Facial Expression Recognition by Learning Geometric Deformations. *IEEE Transactions on Cybernetics, 44*(7-8), 2443-2457. https://doi.org/10.1109/tcyb.2014.2308091

Casas, S.; Portalés, C.; Rueda, S. & Fernández, M. (2017). Simulación de Plataformas Robóticas de Movimiento para Aplicaciones de Realidad Virtual Mediante Filtros Digitales. *Revista Iberoamericana de Automática e Informática Industrial RIAI, 14*(4), 455-466. https://doi.org/10.1016/j.riai.2017.07.001

Cervantes, J.; Taltempa, J.; García-Lamont, F.; Castilla, J. S. R.; Rendon, A. Y. & Jalili, L. D. (2017). Análisis Comparativo de las técnicas utilizadas en un Sistema de Reconocimiento de Hojas de Planta. *Revista Iberoamericana de Automática e Informática Industrial RIAI, 14*(1), 104-114. https://doi.org/10.1016/j.riai.2016.09.005

Chakraborty, A.; Konar, A.; Chakraborty, U.& Chatterjee, A. (2009). Emotion Recognition From Facial Expressions and Its Control Using Fuzzy Logic. *IEEE Transactions on Systems, Man, and Cybernetics*

- *Part A: Systems and Humans, Institute of Electrical and Electronics Engineers (IEEE)*, *39*(4), 726-743. https://doi.org/10.1109/tsmca.2009.2014645

Chen, T.; Ju, S.; Yuan, X.; Elhoseny, M.; Ren, F.; Fan, M. & Chen, Z. (2018). Emotion recognition using empirical mode decomposition and approximation entropy. *Computers amp; Electrical Engineering*, *72*(7), 383-392. https://doi.org/10.1016/j.compeleceng.2018.09.022

Davison, A.; Merghani, W. & Yap, M. (2018). Objective Classes for Micro-Facial Expression Recognition. *Journal of Imaging, MDPI AG*, *4*(10), 119-121. https://doi.org/10.3390/jimaging4100119

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, American Psychological Association (APA)*, *48*(4), 384-392. https://doi.org/10.1037/0003-066x.48.4.384

González, R.; Barrientos, A.; Toapanta, M. & del Cerro, J. (2017). Aplicación de las Máquinas de Soporte Vectorial (SVM) al diagnóstico clínico de la Enfermedad de Párkinson y el Temblor Esencial. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, *14*(4), 394-405. https://doi.org/10.1016/j.riai.2017.07.005

He, X. & Zhang, W. (2018). Emotion recognition by assisted learning with convolutional neural networks. *Neurocomputing*, *291*(1), 187-194. https://doi.org/10.1016/j.neucom.2018.02.073

Hossain, M. S. & Muhammad, G. (2019). Emotion recognition using secure edge and cloud computing. *Information Sciences*, *504*(1), 589-601. https://doi.org/10.1016/j.ins.2019.07.040

Kaur, G.; Sinha, R.; Tiwari, P. K.; Yadav, S. K.; Pandey, P.; Raj, R.; Vashisth, A. & Rakhra, M. (2022). Face mask recognition system using CNN model. *Neuroscience Informatics*, *2*(3), 100035. https://doi.org/10.1016/j.neuri.2021.100035

Kurbalija, V.; Ivanović, M.; Radovanović, M.; Geler, Z.; Dai, W. & Zhao, W. (2018). Emotion perception and recognition: An exploration of cultural differences and similarities. *Cognitive Systems Research*, *52*(1), 103-116. https://doi.org/10.1016/j.cogsys.2018.06.009

Liang, Z.; Oba, S. & Ishii, S. (2019). An unsupervised EEG decoding system for human emotion recognition. *Neural Networks, 116*(1), 257-268. https://doi.org/10.1016/j.neunet.2019.04.003

Li, X.; Hong, X.; Moilanen, A.; Huang, X.; Pfister, T.; Zhao, G. & Pietikainen, M. (2018). Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods. *IEEE Transactions on Affective Computing, Institute of Electrical and Electronics Engineers (IEEE), 9*(4), 563-577. https://doi.org/10.1109/taffc.2017.2667642

Li, X.; Pfister, T.; Huang, X.; Zhao, G. & Pietikainen, M. (2013). A Spontaneous Micro-expression Database: Inducement, collection and baseline. *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 1*(1), 1-6. https://doi.org/10.1109/fg.2013.6553717

Pfister, T.; Li, X.; Zhao, G. & Pietikainen, M. (2011). Recognising spontaneous facial micro-expressions. *2011 International Conference on Computer Vision, IEEE, 1*(1), 1449-1456. https://doi.org/10.1109/iccv.2011.6126401

Prasada Rao, K.; Chandra Sekhara Rao, M. & Hemanth Chowdary, N. (2019). An integrated approach to emotion recognition and gender classification. *Journal of Visual Communication and Image Representation, 60*(1), 339-345. https://doi.org/10.1016/j.jvcir.2019.03.002

Sánchez-Alonso, R. E.; Ortega-Moody, J.; González-Barbosa, J.-J. & Reyes-Morales, G. (2017). Uso de Plataformas para el Desarrollo de Aplicaciones Virtuales en el Modelado de Robot Manipuladores. *Revista Iberoamericana de Automática e Informática Industrial RIAI, 14*(3), 279-287. https://doi.org/10.1016/j.riai.2017.04.001

Santhoshkumar, R. & Geetha, M. K. (2019). Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks. *Procedia Computer Science, 152*(1), 158-165. https://doi.org/10.1016/j.procs.2019.05.038

Villán, A. F.; Fernández, R. U. & Tejedor, R. C. (2017). Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas. *Revista*

Iberoamericana de Automática e Informática Industrial RIAI, *14*(3), 307-328. https://doi.org/10.1016/j.riai.2017.05.001

Xiaohua, W.; Muzi, P.; Lijuan, P.; Min, H.; Chunhua, J. & Fuji, R. (2019). Two-level attention with two-stage multi-task learning for facial emotion recognition. *Journal of Visual Communication and Image Representation*, *62*(1), 217-225. https://doi.org/10.1016/j.jvcir.2019.05.009

Yap, C. H.; Kendrick, C. & Yap, M. H. (2020). SAMM Long Videos: A Spontaneous Facial Micro- and Macro-Expressions Dataset. *Automatic Face and Gesture Recognition*, *1*(1), 771-776. https://doi.org/10.1109/fg47880.2020.00029