

Article

Emotion Recognition in Individuals with Down Syndrome: A Convolutional Neural Network-Based Algorithm Proposal

Nancy Paredes ^{1,2,*}, Eduardo Caicedo-Bravo ^{1,*}  and Bladimir Bacca ^{1,*} 

¹ School of Electrical and Electronics Engineering, Faculty of Engineering, University of Valle, Cali 25360, Colombia

² Department of Electrical, Electronic and Telecommunications, ESPE Armed Forces University, Sangolquí 171103, Ecuador

* Correspondence: nancy.paredes@correounivalle.edu.co or niparedes@espe.edu.ec (N.P.); eduardo.caicedo@correounivalle.edu.co (E.C.-B.); bladimir.bacca@correounivalle.edu.co (B.B.)

Abstract: This research introduces an algorithm that automatically detects five primary emotions in individuals with Down syndrome: happiness, anger, sadness, surprise, and neutrality. The study was conducted in a specialized institution dedicated to caring for individuals with Down syndrome, which allowed for collecting samples in uncontrolled environments and capturing spontaneous emotions. Collecting samples through facial images strictly followed a protocol approved by certified Ethics Committees in Ecuador and Colombia. The proposed system consists of three convolutional neural networks (CNNs). The first network analyzes facial microexpressions by assessing the intensity of action units associated with each emotion. The second network utilizes transfer learning based on the mini-Xception architecture, using the Dataset-DS, comprising images collected from individuals with Down syndrome as the validation dataset. Finally, these two networks are combined in a CNN network to enhance accuracy. The final CNN processes the information, resulting in an accuracy of 85.30% in emotion recognition. In addition, the algorithm was optimized by tuning specific hyperparameters of the network, leading to a 91.48% accuracy in emotion recognition accuracy, specifically for people with Down syndrome.

Keywords: convolutional neural network; microexpressions; Down syndrome; transfer learning



Citation: Paredes, N.; Caicedo-Bravo, E.; Bacca, B. Emotion Recognition in Individuals with Down Syndrome: A Convolutional Neural Network-Based Algorithm Proposal. *Symmetry* **2023**, *15*, 1435. <https://doi.org/10.3390/sym15071435>

Academic Editor: Jeng-Shyang Pan

Received: 10 June 2023

Revised: 2 July 2023

Accepted: 5 July 2023

Published: 17 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human beings need to interact with the world around them through communication, emotions being a crucial point when establishing and maintaining social relationships. Recognizing someone's emotions allows us to understand what has not been said, even leading to adjustments in how we respond or behave. This allows us to support ourselves to infer the next behaviors or actions of another person positively [1].

In the case of people with Down syndrome (DS), communication is a complex process due to linguistic problems derived from cognitive, affective, and social elements [2]. The emotions of this group of people are not controlled and inhibit their behavior, or they can also manifest it effusively due to problems in mental control and self-regulation [3,4].

Advances in artificial intelligence have allowed these techniques to be applied to people with Down syndrome, especially focused on the detection of the syndrome, based on studies carried out on mothers since pregnancy [5]. Where their studies focus on the face of this group of people [6] based on images that allow calculating measurements between specific facial points [5,7,8], in some cases these processes use neural networks [9]. In other cases, authors such as Zhao use a combination of textures and geometric points to identify Down syndrome [10].

Within the field of facial expression recognition (FER), with human–computer interaction and health care [11] where people with Down syndrome are involved, this study is more relevant if it is carried out in real time. FER recognition for emotions focuses on basic

expressions such as anger, disgust, fear, happiness, neutrality, sadness, and surprise [12,13], where one of the main methods used is deep learning (DL). Within this, we have convolution neural networks that currently play a fundamental role in image classification applications [13].

Several investigations were conducted using CNN for emotion recognition with different approaches, such as G. Yang et al. [14] introducing a deep neural network (DNN) model that inputs vectorized facial features. This model demonstrates an accuracy of 84.33% in predicting various emotions. Similarly, Liu et al. [15] employed a two-layer fer2013 and CNN dataset to classify facial emotions. They compared their model with four existing models and achieved a test accuracy of 49.8%. Pranav [16] proposed a two-layer convolutional network model for facial emotion recognition. This model utilizes an Adam optimizer to minimize the loss function and achieves an accuracy of 78.04% when tested.

On the topic of facial analysis, microexpressions have received significant attention. Bishay [17] conducted a study investigating the utilization of various convolutional neural networks to detect action units (AUs). The study compared ten CNNs, including ResNet, VGG, and DenseNet. The findings indicated that the choice of CNN depends on the specific AU under consideration. The research explores the detection of microexpressions while considering the appropriate CNN for each scenario. In a different vein, Hammal [18] focused on action units in infants, which present unique challenges due to reduced facial texture and frequent rapid facial movements. The study employed a multi-label convolutional neural network and tested it on 86 babies during tasks designed to elicit pleasure and frustration. More than 230,000 frames were manually coded using a Baby FACS extension to establish ground reality. The CNN achieved results comparable to the manual coding performed by the FACS validation Kappa system, with Kappa scores ranging from 0.69 to 0.93.

Considering these works previously reviewed, it is observed that the use of convolutional neural networks used for emotion recognition has been widely used for some years. But, one of the challenges of deep learning techniques is the large amount of data [19] necessary to achieve good results in facial recognition, which is one of the main drawbacks in the case of vulnerable groups such as people with Down syndrome. In addition, it is a limitation to being able to analyze the FER of people with DS applying DL techniques. The number of samples is challenging for people from vulnerable groups and people of typical development. In the case of this last group, “learning by transfer” has been used to compensate for this inconvenience [19] caused by the number of samples required to use DL techniques.

According to Yen, transfer learning means that through a previously trained model on a new task, knowledge from the trained model is “transferred” to the pre-trained model, reducing hardware requirements and costs and increasing the accuracy of the system [19,20].

The emotion recognition algorithm in people with Down syndrome presented in this proposal is part of an investigation that works with people with DS who attend special education institutions to support their daily activities in this institution. Emotions being the pillar of this study and constituting the basis for defining the achievements of people with DS within their daily activities, an algorithm proposal designed for people with DS is presented, based on the hand on analysis of facial characteristics of this group of people, which is based on microexpressions through action units (AU) [21,22] through a CNN architecture. On the other hand, “transfer learning” is carried out based on an architecture previously analyzed in a study [22]. In this study, people showed their emotions in real time; as our study presented, the emotions shown by subjects with Down syndrome are spontaneous in their daily activities. Finally, these two architectures converge in a final CNN that makes it possible to obtain the prediction of the emotions of people with Down syndrome.

This work makes several significant contributions. First, it presents an algorithm for emotion recognition explicitly designed for people with Down syndrome, using facial features unique to this population. The algorithm is an integral component of a broader tool designed for implementation in therapeutic processes involving people with Down

syndrome. Its primary goal is to accurately identify the emotions experienced by children participating in these therapy sessions.

To achieve this objective, collaboration with a reputable institution specializing in children with disabilities was established. Faiyaz Doctor [23] asserts that incorporating emotions into therapeutic practices enhances outcomes and boosts the immune system. Consequently, this algorithm equips therapists with valuable insights into the emotional states of children with DS, enabling them to develop effective strategies to facilitate the completion of assigned activities.

In summary, the key contributions of this work include the development of an emotion recognition algorithm tailored to individuals with Down syndrome and its integration into a therapeutic tool used during sessions with children. By leveraging this algorithm, therapists can better understand the emotions of their patients, leading to improved therapeutic outcomes and the formulation of targeted strategies.

In this work, Section 1 shows a compilation of research on the importance of emotions in the daily life of people with typical development, and in people with DS, papers are presented where machine learning and deep learning techniques have been used for DS detection. Section 2 shows the methodology applied in this research. Section 3 presents the results obtained, and a sample taken at the institution participating in the research with the proposed algorithm working is presented. Finally, Section 4 presents the conclusions and discussion.

2. Materials and Methods

People with Down syndrome exhibit distinctive facial features. This article introduces an algorithm designed to accurately recognize five primary emotions (anger, happiness, sadness, surprise, and neutrality) in individuals with Down syndrome. The proposed algorithm employs a composition of three convolutional neural networks. The first two networks were previously evaluated in a study conducted by the authors and published in Paredes' research [22]. These networks analyze facial microexpressions, focusing on the intensity of action units. The second network utilizes transfer learning based on the mini-Xception architecture. As part of the ongoing research, the authors propose combining these two networks into a single CNN to enhance emotion recognition accuracy.

2.1. Dataset

The research utilized a dataset compiled from 8- to 12-year-old Down syndrome students at a special education institution. The acquisition of data adhered to protocols approved by the Ethics Committees in Ecuador and Colombia. The dataset includes 1200 images of children displaying spontaneous emotions with their therapist or tutor during daily activities.

Figure 1 presented the dataset that was categorized according to various emotions: 250 were angry, 300 were happy, 200 were sad, 150 were surprised, and 300 were neutral. Horizontal flipping and rotation (ranging from -60 to 60 degrees) were used for data augmentation purposes in order to balance the dataset. This augmented dataset, designated Dataset-DS and depicted in Figure 2, will be utilized in the proposed method described in Sections 2.2 and 2.3.

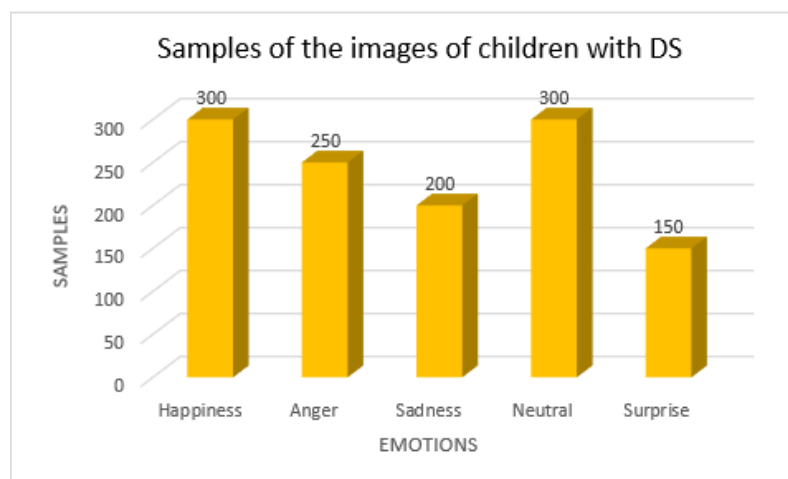


Figure 1. Samples of the images of children with DS (Dataset-DS).

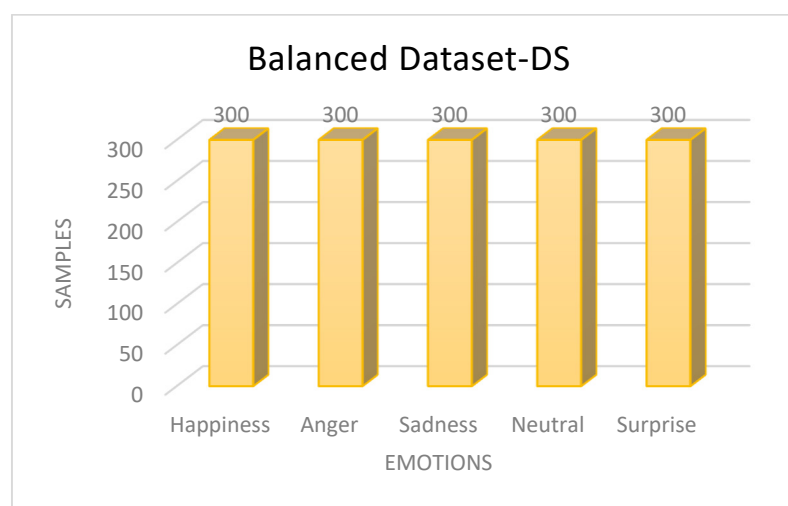


Figure 2. Balanced Dataset-DS.

2.2. Analysis of Techniques for Recognizing Emotions in People with DS

This section summarizes the key findings from the authors' previous research [22]. This work is primarily focused on analyzing the action units (AUs) observed in the facial expressions of individuals with Down syndrome. The intensity levels of AUs in the upper and lower parts of the face were examined using OpenFace (CNN). To detect these intensity levels, the research aimed to identify the most representative AUs for people with Down syndrome across five emotions. Table 1 displays these results, highlighting the significant AUs associated with happiness, sadness, anger, surprise, and neutrality in individuals with Down syndrome.

Table 1. Relevant action units present in the emotions of people with Down syndrome [22].

Expression	AUs
Anger	4, 9, 15, 17
Happiness	6, 7, 10, 12, 14, 20
Sadness	1, 4, 6, 7, 9, 12, 15, 17, 20
Surprise	1, 2, 5, 25, 26
Neutral	2, 5

Multiple machine learning techniques were assessed to identify emotions in individuals with Down syndrome, with the best performance achieved using support vector machines (SVM) with an accuracy of 66.20%. Transfer learning was explored utilizing the mini-Xception architecture (a convolutional neural network) to further improve the outcomes attained through machine learning techniques. In this work, the Dataset-DS was utilized as described in Section 2.1. This dataset was incorporated into the pre-trained mini-Xception architecture as a test data block. The system then determined the recognized emotion by selecting the prediction with the highest value. Consequently, the architecture achieved a 74.8% accuracy in emotion recognition.

Figure 3 shows the diagram of the transfer learning architecture used in [22], using a CNN network (mini-Xception). This trained system effectively transfers the knowledge acquired to recognize emotions, specifically in people with Down syndrome. In other words, the figure shows that both the mini-Xception trained model and the transfer learning system enter an image showing the emotion of happiness, obtaining at the end of the model the emotion “happiness ✓”.

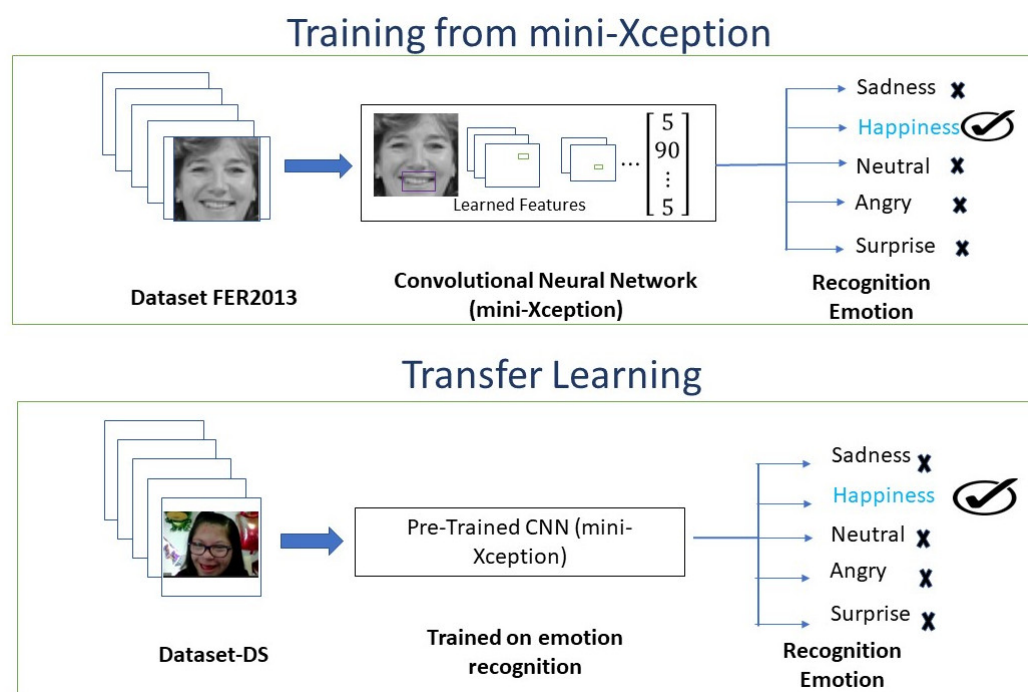


Figure 3. Transfer learning scheme for emotion recognition with SD.

2.3. Improving the Recognition of Emotions for People with DS

In this phase, we present an algorithm proposed for recognizing facial expressions in people with Down syndrome. This section presents an architecture of three convolutional neural networks based on the first part of the analysis made in Section 2.2 of this document.

Our approach involves utilizing a convolutional neural network called OpenFace. This CNN enables us to extract the action units associated with different emotions. With this information, we evaluate the significance of each AU within the specific emotion being analyzed. Our objective is to identify the most pertinent AUs strongly linked to emotions in individuals with Down syndrome. This analysis aims to gain insights into the specific AUs that play a crucial role in expressing emotions among people with DS. According to Baltrušaitis [24], OpenFace is an open source facial recognition system. The architecture of OpenFace combines deep learning and computer vision techniques and is structured as follows:

- **Feature Extraction:** OpenFace employs a convolutional neural network to extract facial features from an image. This CNN is trained explicitly for facial recognition and can accurately detect and locate key facial landmarks such as the eyes, nose, and mouth;

- **Feature Embedding:** Once the facial features are extracted, OpenFace stores them in a high-dimensional vector space. Each face is represented as a numeric vector, commonly called an “inlay,” which captures the unique facial characteristics;
- **Feature Comparison:** To perform facial recognition, OpenFace compares the feature vectors of an individual’s face with the vectors stored in a database. It utilizes similarity measures, such as Euclidean distance or dot product, to determine the degree of similarity between the vectors and whether the faces belong to the same person.

Amos [25] points out that OpenFace, the facial recognition system he developed, operates in real time, providing high accuracy and fast training times. The training and inference processes of neural networks in OpenFace are implemented using Torch, Lua, and LuaJIT. In addition, the Python library is used, with numpy for matrix operations and linear algebra, OpenCV for computer vision primitives, and scikit-learn for classification tasks. OpenFace’s architecture is based on FaceNet, a popular facial recognition model. A dataset of 500,000 images is used to train the system, allowing OpenFace to learn and generalize from a large amount of visual data.

OpenFace employs a modified version of FaceNet’s nn4 network, which effectively reduces the number of parameters, leading to a more efficient and data-friendly model. For a complete overview of the neural network model in OpenFace, see Table 2. Each row in the table corresponds to a specific layer within the neural network. Each of the inception layers is detailed in Santoso [26].

Table 2. Network definition of OpenFace.

Type
conv1 ($7 \times 7 \times 3, 2$)
max pool + norm
inception (2)
norm + max pool
inception (3a)
inception (3b)
inception (3c)
inception (4a)
inception (4e)
inception (5a)
inception (5b)
avg pool
Linear
ℓ_2 normalization

OpenFace is a versatile software that can handle various input types, including real-time video data from web cameras, recorded video files, image sequences, and single images. It also offers the capability to save the processed data outputs, such as facial landmarks, shape parameters, head pose, action units, and gaze vectors, into CSV files. These results are crucial as they serve as inputs for neural networks in subsequent workflow stages.

In our case, we focus on five emotions and utilize the relevant action units identified through analysis. However, specific AUs, namely, AU23, AU28, and AU45, were found to be insignificant in representing the emotions of individuals with Down syndrome based on the analysis conducted in [22], and thus are excluded from the subsequent neural network processing.

The second neural network applies the transfer learning approach using the pre-trained mini-Xception architecture initially trained on the FER2013 dataset. Some authors

mention [27–29] that the mini-Xception architecture is specifically designed for tasks such as image classification, and its structure is as follows:

- **Input Blocks:** The network begins with a convolutional layer that processes the input image and extracts initial features. It utilizes a small filter size and a low number of channels;
- **Parallel Convolution Blocks:** The architecture employs a series of parallel convolution blocks. These blocks consist of depth-separable convolution layers followed by normalization and activation layers. This configuration enables more efficient feature representation and reduces the number of parameters compared to standard convolutions;
- **Reduce Blocks:** Following the parallel convolution blocks, reduced blocks decrease the spatial resolution of features and reduce their dimensionality. These blocks typically consist of a convolutional layer followed by spatial subsampling, such as pooling operations;
- **Global Pooling Layer:** A global pooling layer is applied, further reducing the dimensionality of the features;
- **Fully Connected Layers:** The architecture incorporates fully connected layers specific to the task, such as classification. These layers usually combine convolutional and activation layers, culminating in an output layer with some units corresponding to the output classes or categories.

Mini-Xception architecture belongs to the renowned Xception family of architectures. As described by Arriaga [30], mini-Xception reduces the number of parameters compared to its counterpart, Xception, while effectively performing face detection, gender classification, and emotion classification in a single step.

Figure 4 [29] represents the mini-Xception architecture, showcasing its distinctive features. Notably, mini-Xception eliminates connection layers and incorporates depth-separable convolutions and residual modules, enhancing the model's efficiency and performance. The ADAM optimizer is employed to optimize the training process.

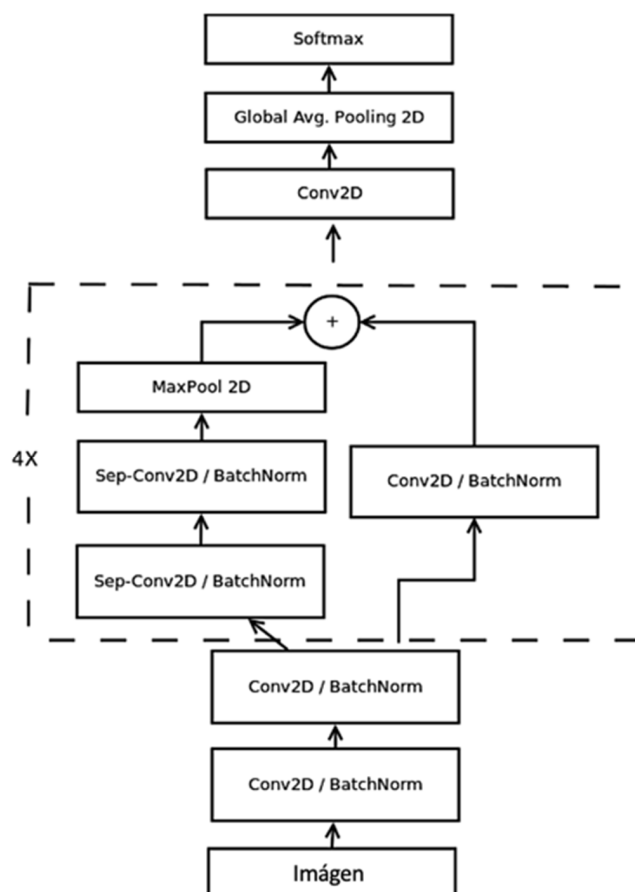


Figure 4. Mini-Xception architecture.

Introducing these advancements in mini-Xception allows for streamlined and simultaneous execution of multiple tasks, making it a valuable architecture for facial analysis applications.

This methodology applies transfer learning to the Dataset-SD, which comprises images of individuals with Down syndrome. By incorporating this dataset as a test data block into the mini-Xception model, we achieved an accuracy of 74.8%. This strategy improves emotion recognition, particularly sadness, and anger, by 11% and 14%, respectively, as mentioned in the previous analysis.

The analysis presented in [22], carried out by the researchers, focused on evaluating the behavior of people with Down syndrome and their emotional expressions. This evaluation led to the creation of Table 1, which highlights this group's most relevant and typical action units. However, when these AUs were analyzed using machine learning techniques, the accuracy of emotion recognition among people with Down syndrome peaked at 66.20%. In particular, the accuracy for identifying sadness and anger was relatively low, at 30.80% and 42.30%, respectively.

To improve emotion identification in people with Down syndrome, transfer learning using the mini-Xception architecture was employed. This architecture was selected based on a study by Paredes [22]. Transfer learning used validation data obtained from samples of people with Down syndrome. This study significantly improved recognition of critical emotions such as sadness and anger, reaching accuracy rates ranging from 41% to 66.7%, respectively.

Considering the results of both analyses, it was decided to combine the intensities obtained from each AU—from the relevant and specific action units of people with Down syndrome with the results obtained through the transfer learning system—which are the values given by the model to each emotion studied, with these values being the ones that enter the final CNN. This approach aimed to improve emotion identification, particularly for the two emotions with the lowest accuracy. A final CNN network was used to achieve this, resulting in an improved accuracy rate of 83.53% for sadness and 87.41% for anger. It is essential to emphasize the focus on these emotions because the existing literature suggests that detecting negative emotions in people with Down syndrome is a highly complex task [31].

A third convolutional neural network is implemented to enhance the accuracy of the previous inference systems. The results obtained from the two CNNs mentioned above are integrated into a final convolutional neural network, enabling a more precise analysis of emotions within the Dataset-DS. This CNN consists of 4 residual blocks, as shown in Figure 5. Each block contains two sets of causal convolution layers dilated with the same dilation factor, followed by spatial normalization layers, ReLU activation, and dropout. The network adds the input of each block to the output of the block (including a 1-by-1 convolution on the input when the number of channels between the input and output does not match) and applies a final activation function. Our final CNN has four of these residual blocks in series, each with twice the dilation factor of the previous layer, starting with a dilation factor of 1. We specify 64 filters for the 1D convolutional layers for the residual blocks, with a filter size of 5 and a dropout factor of 0.005 for the spatial dropout layers.

The proposed architecture scheme, depicted in Figure 6, illustrates the flow of Dataset-DS databases in OpenFace CNN and transfer learning using the mini-Xception architecture. The Dataset-DS is inputted into the latter architecture. The outputs of these two CNNs then enter the final convolutional neural network to obtain the desired results.

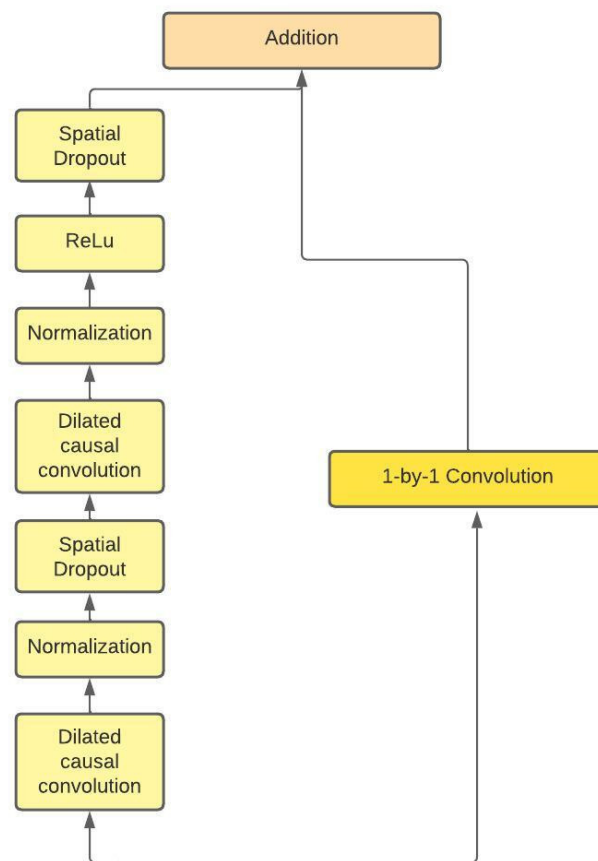


Figure 5. The residual block for third CNN [32].

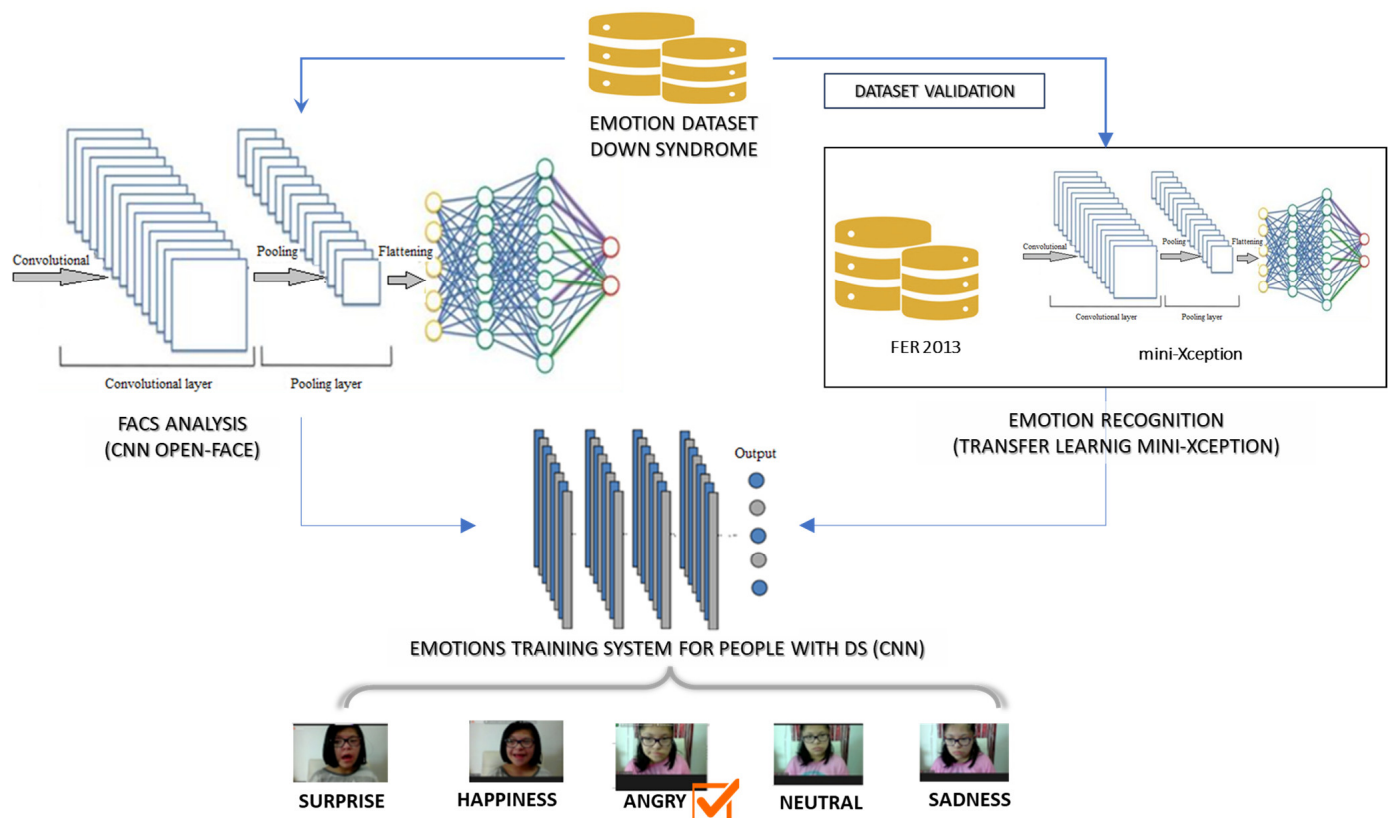


Figure 6. Proposed of architecture scheme for emotion recognition with people with SD.

2.4. Analysis of Ablation in the Proposed Neural Network

In order to understand the individual contributions of each component to the performance of the proposed model, an ablation study was conducted on the neural network described in this research. The study systematically turned off specific system components and evaluated their impact on the overall model performance.

Figure 7 presents the confusion matrix of the architecture when disconnecting the CNN network in charge of identifying the action units (the blue background highlights those results from the rest of the other values. And other background does not represent a particular value). When this network was disconnected, the total accuracy of the model was found to be 79.6%. The results indicated an improvement of 8.49% in recognizing anger and 3.26% in recognizing neutral expressions. However, the accuracy in recognizing happiness, sadness, and surprise decreased by 3.3%, 2%, and 26.31%, respectively.

True Class	Anger	Happiness	Neutral	Sadness	Surprise	Accuracy
	96.00%	4.00%	0.00%	0.00%	0.00%	
	0.00%	88.57%	0.05%	0.00%	11.39%	
	0.00%	12.73%	87.27%	0.00%	0.00%	
	18.77%	0.00%	0.00%	81.23%	0.00%	
	0.00%	0.06%	0.00%	47.94%	52.06%	
Predicted Class						79.6%

Figure 7. Proposed network analysis disconnecting the neural network that analyzes the AU.

In addition, the study examined the effects of disconnecting the learning transfer network, shown in Figure 8 (the blue background highlights those results from the rest of the other values. And other background does not represent a particular value). This resulted in a precision of 79.1% for the model. Disconnecting this network led to a notable improvement of 12.34% in recognizing anger, but it also resulted in decreased precision for happiness (4.28%), sadness (1.72%), and surprise (23.95%) recognition.

True Class	Anger	Happiness	Neutral	Sadness	Surprise	Accuracy
	99.75%	0.25%	0.00%	0.00%	0.00%	
	0.00%	87.59%	0.07%	0.00%	12.34%	
	0.00%	99.81%	83.94%	0.00%	0.03%	
	18.19%	0.00%	0.00%	81.81%	0.00%	
	0.00%	0.00%	0.00%	45.58%	54.42%	
Predicted Class						79.1%

Figure 8. Proposed network analysis disconnecting the neural network of transfer learning.

The results of this ablation study confirm the substantial contributions of each component to the overall performance of the total model. Turning off the CNN network

reduced the model's accuracy but positively impacted the recognition of anger and neutral expressions. On the other hand, disconnecting the learning transfer network improved anger recognition but adversely affected the recognition of happiness, sadness, and surprise. Through this systematic analysis, the researchers gained valuable insights into the importance of each component in achieving optimal performance in the proposed model.

3. Results

The architecture depicted in Figure 6 has enhanced the system's performance based on previous evaluations. This advancement is evident in the achieved system accuracy of 85.30%, as illustrated by the confusion matrix in Figure 9 (the blue background highlights those results from the rest of the other values. And other background does not represent a particular value). The algorithm proposed in this article demonstrates a substantial improvement of 10.50% compared to the authors' previous analysis, where they attained an accuracy of 74.8%.

True Class	Anger	Happiness	Neutral	Sadness	Surprise	Accuracy
	87.41%	1.19%	1.9%	7.84%	1.66%	
	0.00%	91.87%	0.04%	0.18%	7.91%	
	0.00%	9.95%	84.01%	0.54%	5.5%	
	16.39%	0.00%	0.08%	83.53%	0.00%	
	0.00%	0.06%	0.45%	21.11%	78.37%	
Predicted Class						85.30%

Figure 9. Proposal to improve the recognition of emotions for people with DS.

Hyperparameter Tuning

According to Echeverri [33], hyperparameters are predefined values assigned to parameters by the algorithm designer, which define the configuration of the architecture used in research. The selection of appropriate hyperparameters plays a crucial role in CNN's ability to recognize patterns and improve the accuracy and efficiency of the architecture [33,34].

The selection of appropriate hyperparameters is crucial for optimizing the performance of a CNN, as it can enhance pattern recognition, accuracy, and efficiency. We adjusted the hyperparameters, evaluating the model's performance with all possible combinations. The objective was to find the optimal configuration that maximizes the CNN's performance. In this study, we analyzed the following hyperparameters:

- **Filter size:** The filter size is an essential parameter in a one-dimensional convolutional neural network. It determines the width of the filter used for convolutions and is represented as a positive integer. The standard options for the filter size are two, three, and five. In Figure 6, where the non-optimized proposal is shown, a value of five was used for this hyperparameter, whereas in our optimized system, a filter of size two was used. Choosing a filter size of two allows us to capture local features and short patterns in the input data. This is advantageous as it allows the detection of relevant patterns at a more localized level. By using smaller filters, we can extract more precise information than larger ones, especially those with more than three. By taking advantage of these smaller filters, our system benefits from greater sensitivity to local variations and can capture finer details in the data. This granularity can be crucial for tasks where the shortest patterns or specific local features are important for accurate predictions or analysis. In general, using a filter size of two in our optimized

system improves our ability to extract and interpret local features effectively, leading to better performance in capturing relevant patterns in the input data [35].

- **Number of filters:** This hyperparameter, known as the number of filters, determines the number of filters utilized in the convolutional layer. Each filter detects specific patterns in the input data, generating an output pipeline through the convolution operation. The available options for the number of filters typically include values like 8, 16, 32, 64, or 128. In Figure 6, where the non-optimized proposal is shown, a value of 64 was used for this hyperparameter, while in our optimized system, a filter of size 128 was used. The network can capture more patterns and details within the input data by selecting more filters. This increased capacity allows for improved differentiation among different classes of emotions. Additionally, a more significant number of filters facilitates the network's adaptation to each specific dataset's unique characteristics [35].
- **The number of input channels:** In our approach, we employ the 'automatic' configuration for the optimized model, which automatically determines the number of filters based on the architecture and specific characteristics of the model or data. This parameter provides the network with flexibility as it can seamlessly adapt to various datasets without adjusting the number of input channels. Consequently, this approach enhances the performance and efficiency of the network by automatically adapting to different datasets [35].
- **Dilation factor:** This parameter controls the spacing between elements the filter considers during the convolution operation. Expanding the filters and inserting zeros between each filter element can detect patterns on a larger scale. We experimented with one, two, four, and eight dilation factors. A factor of one indicates no spacing, while factors of two, four, and eight skip one, three, and seven elements, enlarging the filter's receptive field [35].
- **Optimizer:** The optimizer refers to the algorithm that adjusts the weights during training. Standard optimization algorithms include stochastic gradient descent (SGD), Adam, RMSprop, and Adagrad. In our study, we employed the Adam optimizer due to its efficiency in optimization, adaptability to learning rates, and less sensitivity to the selection of hyperparameters. Adam simplifies the process of tuning the chosen hyperparameters [34]. By carefully selecting and tuning these hyperparameters, we aim to enhance the performance of the CNN architecture in terms of pattern recognition, accuracy, and efficiency.

As part of the analysis aimed at improving the accuracy of the emotion recognition algorithm for individuals with Down syndrome, the hyperparameters of the CNN were analyzed. This analysis resulted in an efficiency of 91.4%, enhancing the algorithm's performance, as demonstrated by the confusion matrix in Figure 10 (the blue background highlights those results from the rest of the other values. And other background does not represent a particular value). The critical configuration parameters considered included filter size, number of filters, expansion, buffer value, and optimizer. Table 3 presents an example of the optimal hyperparameter values determined through the algorithm's optimization process and the corresponding mean precision values achieved for each interaction.

This analysis involved evaluating all hyperparameters to determine the optimal system precision. Each hyperparameter modification was tested through 100 iterations. An example of the analysis conducted with modified hyperparameter values for the CNN proposed in this section is presented in Table 3. The rows represent the specific hyperparameter values or parsed parameters, while the columns represent the top hyperparameters. The last column indicates the average precision achieved with each set of hyperparameters parsed for each row. The following hyperparameter values were optimizing the proposed system: filter size = "2"; number of filters = "128"; expansion = "1, 2, 4, 8"; and optimizer = "ADAM".

Finally, the results obtained through the proposed architecture and the hyperparameter analysis, which enabled system optimization, are depicted in Figure 10 using a

confusion matrix. The optimized system achieved a precision of 91.4%, representing a 6.1% improvement compared to the non-optimized system.

True Class	Anger	99.50%	0.00%	0.00%	0.50%	0.00%	
	Happiness	0.00%	93.50%	3.02%	0.05%	3.43%	
	Neutral	0.00%	6.42%	93.45%	0.12%	0.00%	
	Sadness	6.23%	0.00%	3.29%	90.48%	0.00%	
	Surprise	0.00%	0.00%	5.77%	9.03%	85.19%	
							Accuracy
							91.48%
		Anger	Happiness	Neutral	Sadness	Surprise	
Predicted Class							

Figure 10. Classification confusion matrix using proposed for emotion recognition.

Table 3. Sampling of the values of hyperparameters and accuracy.

	Hyperparameters					Mean Accuracy
	Filter Size	Num Channels	Num Filters	Dilation Factor	Optimizer	
Set values	3	auto	64	1, 2, 4, 8	ADAM	0.89
	3	auto	16	1, 2, 4, 8	ADAM	0.77
	3	auto	8	1, 2, 4, 8	ADAM	0.61
	3	auto	64	1	ADAM	0.87
	3	auto	128	2	ADAM	0.88
	2	auto	128	1, 2, 4, 8	ADAM	0.91
	3	auto	128	1, 2, 4, 8	RMSPROP	0.89
	2	auto	128	1, 2, 4, 8	RMSPROP	0.90
	2	auto	128	1, 2, 4, 8	SGDM	0.37

Based on the information obtained from previous analyses conducted by the authors, as mentioned in Section 2.2, specific facial characteristics of individuals with Down syndrome were defined by examining their microexpressions in various emotions. The activation and intensity of these specific microexpressions were considered relevant for this particular population group. The analysis focused on observing the intensity behavior of the action units. To perform this analysis, the authors utilized the Dataset-DS, which consists of samples collected from individuals with Down syndrome engaging in daily activities within an institution. These samples contained spontaneous emotions that were subsequently analyzed using a transfer learning system based on the mini-Xception architecture. The Dataset-DS served as the validation dataset for this architecture.

These analyses have led to the proposal of an algorithm to improve emotion recognition accuracy in individuals with Down syndrome. This proposal involves incorporating the values obtained from the abovementioned analyses into a convolutional neural network, resulting in improved precision values, particularly for the emotions of sadness, surprise, and anger. Despite achieving better results through transfer learning, these emotions still exhibited a true positive rate (TPR) of less than 67% [17].

The proposed architecture, depicted in Figure 6, demonstrated enhanced accuracy in recognizing the emotions of anger, sadness, and surprise, with increases of 87.41%, 83.53%, and 78.37%, respectively, as observed in the confusion matrix depicted in Figure 9. These improvements represent significant advancements of 21%, 42%, and 16%, respectively,

compared to the transfer analysis mentioned in Section 2.2. It is worth noting that although the accuracy of happiness and neutrality in Figure 9 decreased by 8% and 4% in the proposed algorithm, the accuracy rates for these emotions remained high at 91.87% and 84.01%, respectively. Overall, the system's accuracy improved to 85.30%, representing a notable increase of 10.1%.

Finally, the results obtained from the proposed architecture and the hyperparameter analysis aimed at optimizing the algorithm are presented in Figure 10 through the confusion matrix. This analysis showcases improvements in the precision of four emotions analyzed in this research: anger, happiness, neutrality, sadness, and surprise, with accuracy rates of 99.50%, 93.50%, 93.45%, 90.48%, and 85.19%, respectively. These improvements represent significant gains of 12.09%, 1.63%, 9.44%, 6.95%, and 6.82% compared to the non-optimized analysis. Overall, the optimized system achieved an enhanced accuracy of 6.18% compared to the system described in Section 2.3. This analysis yields excellent results for analyzing emotions in individuals with Down syndrome during daily activities.

The sensitivity analysis aimed to evaluate the impact of variations in filter size, number of filters, dilation factor, number of channels, and hyperparameters on the architecture's performance. Specifically, the analysis focused on the newer 1D convolutional layer structure, which plays a crucial role in connecting the action unit values of the OpenFace structure with the prediction values generated by the mini-Xception structure.

To perform the analysis, different combinations of hyperparameter values were tested, including defining filter size = "2", a number of filters = "128", expansion = "1, 2, 4, 8", and optimizer = "ADAM", as specified in Table 3. The goal was to understand how changes in these parameters affected the architecture's ability to classify emotions accurately.

The evaluation metric used to assess performance was classification accuracy. Following the methodology described in Section 2.3, the accuracy of the architecture in emotion recognition was measured. The precision reported in Table 3 represents the average performance over 100 iterations of the testing process.

Furthermore, Table 4 provides a comparative overview of the analyzed systems in this section, showcasing the techniques studied, starting from machine learning, moving on to transfer learning analysis, and concluding with the proposed approach. The proposed approach achieves the highest precision in emotion recognition for individuals with Down syndrome.

Table 4. Summary of results for emotion classification of people with DS.

Techniques Applied		True Positive Rates (%)					
		Anger	Happiness	Neutral	Sadness	Surprise	Accuracy
Machine Learning [36,37]	KNN	53.8	91.3	76.1	13.8	60.3	64.9
	Ensemble Subspace Discriminant	46.2	89.4	71.6	26.2	60.3	64.7
	SVM	42.3	82.7	78.9	30.8	64.1	66.2
Transfer Learning	Mini-Xception	66.7	99	88.4	41	62.2	74.8
Proposed	CNN (OpenFace) CNN (Transfer Learning) CNN	99.5	93.5	93.5	90.5	85.2	91.4

4. Conclusions

This work presents a novel algorithm for recognizing the emotions of people with Down syndrome based on facial features. The algorithm builds upon the authors' previous research, which involved analyzing the facial characteristics of individuals with Down syndrome using OpenFace to capture their microexpressions. This analysis identified the primary action units associated with each emotion within this specific population, considering the intensity of their microexpressions.

Various artificial intelligence techniques were explored to enhance recognition accuracy, and transfer learning with the mini-Xception architecture proved remarkably effective. The

algorithm's performance was evaluated using Dataset-SD as the test data set. Additionally, the features extracted from these convolutional neural networks were incorporated into a final convolutional network, further improving the accuracy of the results obtained thus far.

The research presents an optimized algorithm that leverages the distinctive facial features of individuals with DS and deep learning techniques to achieve accurate recognition of people with Down syndrome. The proposed model demonstrates robustness and efficient resource utilization by combining microexpression analysis, transfer learning, and the final CNN. The hyperparameter analysis conducted in this research has led to an optimized system with an impressive accuracy rate of 91.4% in recognizing the emotions of individuals with DS. These results highlight the algorithm's effectiveness, positioning it as a promising solution for Down syndrome recognition based on facial features.

Among the emotions analyzed, happiness consistently exhibited the highest accuracy, with values consistently surpassing 82% for individuals with Down syndrome. However, anger and sadness posed challenges, with precision values ranging from 42% to 30%. This research has achieved highly accurate results, with recognition rates of up to 99% for anger and 90% for sadness. It is important to note that this algorithm was primarily developed to support therapeutic processes for individuals with DS.

This research used an independent test data set comprising samples from people with Down syndrome. Since no architectures designed explicitly for spontaneous emotion recognition in people with Down syndrome existed, this computational model is a pioneering effort in this field. The objective is to recognize emotions during the daily activities carried out by people with Down syndrome, thus serving as a starting point for future studies in this area, which will allow the proposal made to be improved and will contribute to continuous improvement, not only in terms of computer vision and computational performance but also from a specific psychological perspective for this group of individuals.

These findings put us at the forefront of discussions about the importance of conducting research that focuses on the unique needs of this specific population. Given people with Down syndrome's distinctive physical and cognitive characteristics, developing support tools is crucial to improve their daily activities. The proposed approach can be further refined by exploring alternative deep learning techniques and using new architectures. It is essential to make concerted efforts to translate scientific advances into tangible benefits for vulnerable groups, thus bridging the gap between scientific progress and real-world application in their daily lives.

Author Contributions: Conceptualization, N.P., E.C.-B. and B.B.; methodology, B.B. and N.P.; writing—original draft preparation, N.P.; writing—review and editing, N.P., E.C.-B. and B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carvalho, P.; Menezes, P. Classification of FACS-Action Units with CNN Trained from Emotion Labelled Data Sets. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 3766–3770.
2. Matsumoto, D.; Hwang, H.S. Lectura de la Expresión Facial de las Emociones: Investigación Básica en la Mejora del Reconocimiento de Emociones. *Ansiedad Estrés* **2013**, *19*, 121–129.
3. Ruiz, E. Temas de Interés Evaluación de la Capacidad Intelectual en Personas Con Síndrome de Down. 2012. Available online: <http://www.centrodocumentaciondown.com/uploads/documentos/27dcb0a3430e95ea8358a7baca4b423404c386e2.pdf> (accessed on 3 July 2023).
4. Ruiz, E.; Álvarez, R.; Arce, A.; Palazuelos, I.; Schelstraete, G. Programa de educación emocional. Aplicación práctica en niños con síndrome de Down. *Rev. Sindr. De Down* **2009**, *103*, 126–139.
5. Soler Ruiz, V. Lógica Difusa Aplicada a Conjuntos Imbalanceados: Aplicación a la Detección del Síndrome de Down. 2007. Available online: <https://www.tesisenred.net/handle/10803/5777?locale-attribute=ca> (accessed on 1 July 2023).

6. Agbolade, O.; Nazri, A.; Yaakob, R.; Ghani, A.A.; Cheah, Y.K. Down syndrome face recognition: A review. *Symmetry* **2020**, *12*, 1182. [\[CrossRef\]](#)
7. Cornejo, J.Y.R.; Pedrini, H.; Lima, A.M.; Nunes, F.D.L.D.S. Down syndrome detection based on facial features using a geometric descriptor. *J. Med. Imaging* **2017**, *4*, 044008. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101. [\[CrossRef\]](#)
9. Eroglu, O.; Sipahi, M.E.; Tunca, Y.; Vurucu, S. Recognition of Down syndromes using image analysis. In Proceedings of the 14th National Biomedical Engineering Meeting, Izmir, Turkey, 20–22 May 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–4. [\[CrossRef\]](#)
10. Zhao, Q.; Okada, K.; Rosenbaum, K.; Kehoe, L.; Zand, D.J.; Sze, R.; Summar, M.; Linguraru, M.G. Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA. *Med. Image Anal.* **2014**, *18*, 699–710. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Alsharekh, M.F. Facial Emotion Recognition in Verbal Communication Based on Deep Learning. *Sensors* **2022**, *22*, 6105. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Atabansi, C.C.; Chen, T.; Cao, R.; Xu, X. Transfer Learning Technique with VGG-16 for Near-Infrared Facial Expression Recognition. *J. Phys. Conf. Ser.* **2021**, *1873*, 012033. [\[CrossRef\]](#)
13. Bodapati, J.D.; Naik, D.S.B.; Suvama, B.; Naralasetti, V. A Deep Learning Framework with Cross Pooled Soft Attention for Facial Expression Recognition. *J. Inst. Eng. Ser. B* **2022**, *103*, 1395–1405. [\[CrossRef\]](#)
14. Yang, G.; Ortoneda, J.S.Y.; Saniie, J. Emotion Recognition Using Deep Neural Network with Vectorized Facial Features. In Proceedings of the 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 3–5 May 2018; pp. 318–322. [\[CrossRef\]](#)
15. Liu, L. Human face expression recognition based on deep learning-deep convolutional neural network. In Proceedings of the Proceedings—2019 International Conference on Smart Grid and Electrical Automation, ICSGEA 2019, Xiangtan, China, 10–11 August 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 221–224. [\[CrossRef\]](#)
16. Pranav, E.; Suraj, K.; Satheesh, C.; Supriya, M.H. Facial Emotion Recognition Using Deep Convolutional Neural Network. In Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 April 2020; pp. 317–320. [\[CrossRef\]](#)
17. Bishay, M.; Ghoneim, A.; Ashraf, M.; Mavadati, M. Which CNNs and Training Settings to Choose for Action Unit Detection? A Study Based on a Large-Scale Dataset. In Proceedings of the Proceedings—2021 16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, 15–18 December 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021. [\[CrossRef\]](#)
18. Hammal, Z.; Chu, W.S.; Cohn, J.F.; Heike, C.; Speltz, M.L. Automatic Action Unit Detection in Infants Using Convolutional Neural Network. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 216–221. [\[CrossRef\]](#)
19. Yen, C.T.; Li, K.H. Discussions of Different Deep Transfer Learning Models for Emotion Recognitions. *IEEE Access* **2022**, *10*, 102860–102875. [\[CrossRef\]](#)
20. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. 2018. Available online: <http://arxiv.org/abs/1808.01974> (accessed on 29 June 2023).
21. Paredes, N.; Bravo, E.C.; Cortes, B.B. Experimental Analysis Using Action Units as Feature Descriptor for Emotion in People with down Syndrome. *Lect. Notes Electr. Eng.* **2021**, *762*, 253–265.
22. Paredes, N.; Caicedo-Bravo, E.F.; Bacca, B.; Olmedo, G. Emotion Recognition of Down Syndrome People Based on the Evaluation of Artificial Intelligence and Statistical Analysis Methods. *Symmetry* **2022**, *14*, 2492. [\[CrossRef\]](#)
23. Doctor, F.; Karyotis, C.; Iqbal, R.; James, A. An intelligent framework for emotion aware e-healthcare support systems. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016, Athens, Greece, 6–9 December 2016. [\[CrossRef\]](#)
24. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial behavior analysis toolkit. In Proceedings of the Proceedings—13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Xi'an, China, 15–19 May 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 59–66. [\[CrossRef\]](#)
25. Amos, B.; Bartosz, L.; Mahadev, S. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications*; Carnegie Mellon University: Pittsburgh, PA, USA, 2016.
26. Santoso, K.; Kusuma, G.P. Face Recognition Using Modified OpenFace. *Procedia Comput. Sci.* **2018**, *135*, 510–517. [\[CrossRef\]](#)
27. Fatima, S.A.; Kumar, A.; Raoof, S.S. Real Time Emotion Detection of Humans Using Mini-Xception Algorithm. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1042*, 012027. [\[CrossRef\]](#)
28. Sun, L.; Ge, C.; Zhong, Y. Design and implementation of face emotion recognition system based on CNN Mini_Xception frameworks. *J. Phys. Conf. Ser.* **2021**, *2010*, 012123. [\[CrossRef\]](#)
29. Behera, B.; Prakash, A.; Gupta, U.; Semwal, V.B.; Chauhan, A. *Statistical Prediction of Facial Emotions Using Mini Xception CNN and Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 397–410. [\[CrossRef\]](#)
30. Arriaga, O.; Plöger, P.G.; Valdenegro, M. Real-time Convolutional Neural Networks for Emotion and Gender Classification. *arXiv* **2017**, arXiv:1710.07557.

31. Williams, K.R.; Wishart, J.G.; Pitcairn, T.K.; Willis, D.S.; Williams, J.G.W.K.R.; Dykens, E.; Channell, M.M.; Conners, F.A.; Barth, J.M.; Virji-Babul, N.; et al. Emotion Recognition by Children with Down Syndrome: Investigation of Specific Impairments and Error Patterns. *Am. J. Ment. Retard.* **2005**, *110*, 378. [CrossRef] [PubMed]
32. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. 2018. Available online: <http://github.com/locuslab/TCN> (accessed on 29 June 2023).
33. Echeverri, C.J.O.; Cordoba, D.A.L.; Quintero, E.A. Ajuste de hiperparámetros de una red neuronal convolucional para el reconocimiento de lengua de señas. *Con-Cienc. Técnica* **2021**, *5*, 48–55. Available online: <https://revistas.sena.edu.co/index.php/conciencia/article/view/3926> (accessed on 9 April 2023).
34. Zhou, S.; Song, W. Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection. *Autom. Constr.* **2020**, *114*, 103171. [CrossRef]
35. 1-D Convolutional layer-MATLAB-MathWorks América Latina. Available online: <https://la.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.convolution1dlayer.html> (accessed on 4 June 2023).
36. Batta, M. Machine Learning Algorithms—A Review. *Int. J. Sci. Res. IJSR* **2019**, *9*, 381–386. [CrossRef]
37. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 3. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.