symmetry

MDPI

Article

A New Descriptor for Smile Classification Based on Cascade Classifier in Unconstrained Scenarios

Oday A. Hassen [1,*], Nur Azman Abu [1], Zaheera Zainal Abidin [1] and Saad M. Darwish [2]

[1] Department of Information Technology, University Technical Malaysia Melaka, Hang Taya, Melaka 76100, Malaysia; nura@utem.edu.my (N.A.A.); Zaheera@utem.edu.my (Z.Z.A.)

[2] Institute of Graduate Studies and Research, University of Alexandria, 163 Horreya Avenue, El Shatby 21526, Alexandria P.O. Box 832, Egypt; saad.darwish@alexu.edu.eg

* Correspondence: Odayali@uowasit.edu.iq; Tel.: +60-9647827554545

Abstract: In the development of human–machine interfaces, facial expression analysis has attracted considerable attention, as it provides a natural and efficient way of communication. Congruence between facial and behavioral inference in face processing is considered a serious challenge that needs to be solved in the near future. Automatic facial expression is a difficult classification issue because of the high interclass variability caused by the significant interdependence of the environmental conditions on the face appearance caused by head pose, scale, and illumination occlusions from their variances. In this paper, an adaptive model for smile classification is suggested that integrates a row-transform-based feature extraction algorithm and a cascade classifier to increase the precision of facial recognition. We suggest a histogram-based cascade smile classification method utilizing different facial features. The candidate feature set was designed based on the first-order histogram probability, and a cascade classifier with a variety of parameters was used at the classification stage. Row transformation is used to exclude any unnecessary coefficients in a vector, thereby enhancing the discriminatory capacity of the extracted features and reducing the sophistication of the calculations. Cascading gives the opportunity to train an extremely precise classification by taking a weighted average of poor learners' decisions. Through accumulating positive and negative images of a single object, this algorithm can build a complete classifier capable of classifying different smiles in a limited amount of time (near real time) and with a high level of precision (92.2–98.8%) as opposed to other algorithms by large margins (5% compared with traditional neural network and 2% compared with Deep Neural Network based methods).

Keywords: cascade classifier; row transformation; smile detection; features extraction

check for updates

Citation: Hassen, O.A.; Abu, N.A.; Abidin, Z.Z.; Darwish, S.M. A New Descriptor for Smile Classification Based on Cascade Classifier in Unconstrained Scenarios. *Symmetry* **2021**, *13*, 805. https://doi.org/10.3390/sym13050805

Academic Editor: Aviv Gibali

Received: 2 April 2021
Accepted: 29 April 2021
Published: 5 May 2021

1. Introduction

Facial expression is one of the potent and prompt means for humans to communicate their emotions, intentions, cognitive states, and opinions to each other [1]. Facial expression plays an important role in the evolution of complex societies, which help to coordinate social interaction, promote group cohesion, and maintain social affiliations [2]. Potential expression recognition technology applications include tutoring systems that are sensitive to students' expression, computer-assisted deceit detection, clinical disorder diagnosis and monitoring, behavioral and pharmacological treatment assessment, new entertainment system interfaces, smart digital cameras, and social robots. To convey emotions of joy, happiness, and satisfaction, a smile is the most typical facial expression in humans [3]. Modern longitudinal studies have used smile data from images to predict future social and health outcomes [4].

Researchers have made substantial progress in developing automatic facial expression detection systems in the literature [5]. Anger, surprise, disgust, sadness, happiness, and fear are many of the six basic facial expressions and emotions commonly referred to. Among the various facial expressions, happiness, as usually demonstrated in a smile, often occurs in the daily life of a person. Two components of facial muscle movements are included in a smile, namely Cheek Raiser (AU6) and Lip Corner Puller (AU12), as shown in Figure 1 [6]. In computer vision and its fields of operation, the automated facial expression recognition system has become very interesting and difficult [7].

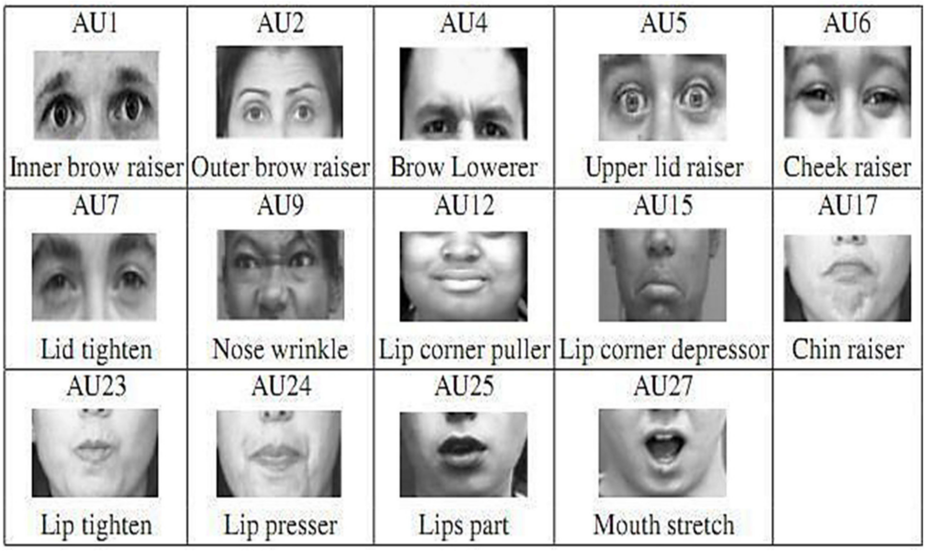

Figure 1. Facial action units (AUs).

In managed settings, current facial expression recognition has promising results, but performance on real-world data sets is still unsatisfactory [8]. This is because there are broad differences in facial appearances through the color of the skin, lighting, posture, expression, orientation, head location, lightening state, and so on. By incorporating deep learning [9], optimization [10], and ensemble classification, automatic methods of identification for facial expression are suggested to deal with existing system difficulties. Five key steps are given for the planned work: preprocessing, a deep evolutionary neural network used for feature extraction, feature selection utilizing swarm optimization, and facial expression classification employing a support vector machine, ensemble classifiers [11], and a neural network [12].

Motivation and Contribution

As an important way to express emotion, facial expression has a vital influence on the communication between people. Currently, in computer vision and pattern recognition, facial expression recognition has become an active research scope. Real-time and effective smile detection can significantly enhance the development of facial expression recognition. The classification of smiles in an unconstrained environment is difficult because of the invertible and wide variety of facial pictures. Faces' extensive optical alterations, such as occlusions, posture transitions, and drastic lightings, make certain functions very difficult in real-world implementations. The majority of current studies deal with smile detection and not smile classification. However, within the current smile classification approaches, their models are not smile attribute specific hence their performance may be limited.

The main goal of this paper is to build an adaptive model for the classification of smiles that incorporates both row-transformation-based features extraction and a cascade classifier to increase the accuracy of classification. In contrast to the current methods of classifying smiles, which rely on deep neural networks to extract features that, in turn, require a large number of samples and more computation, the suggested model relies on row transformation to reduce and improve the discriminatory capability of the extracted features. Furthermore, the suggested model utilizes the cascade classification concept to build an accurate classifier. Cascading classifiers allow the most likely smile pictures to be evaluated for all features that differentiate an individual. The accuracy of a classifier can even be varied. A chain of experiments proves that the suggested model technique is substantially reliable and quicker than other widespread prototypes.

The remainder of the article is organized as follows. Section 2 discusses the current related work. Section 3 presents the proposed model steps in detail. Section 4 explains experimental designs. Section 5 includes the conclusion and future work.

2. Related Work

Several scientific studies have been performed in the field of identification of facial expressions that apply to a range of technologies such as computer vision, image recognition, bioindustry, forensics, authentication of records, etc. [13–15]. In many studies, Principal Component Analysis (PCA) was used to provide a coding framework for facial action that models and recognizes various forms of facial action [16,17]. However, PCA-based solutions are subject to a dilemma in which the projection maximizes variance in all images and negatively affects recognition performance. Independent Component Analysis (ICA) is one of the statistical methods that are adapted to perform expression recognition to elicit statistically independent local face characteristics that proceed better than PCA [18].

Recently, as training and feature extraction are carried out simultaneously, deep learning among the science community has attracted substantial interest in the field of smile detection. The Deep Neural Network (DNN) was the first method of deep study used for the training and classification of models in high-dimensional data [19]. The DNN has one problem: it takes too long to overcome challenges at the preparation stage. The Convolutional Neural Network (CNN) is a deep learning technique that solves DNN problems by reducing preprocessing and thereby enhancing image, audio, and video processing [20,21]. The CNN has great performance while classifying smiles images that are very similar to the data set using a huge computational cost. However, CNNs usually have difficulty in classifying an image that includes some degree of tilt or rotation.

As the feature extraction module represents the core module for facial classification, many algorithms inspired by nature are suggested to select the characteristics of the picture [22], among others, primarily in medicinal applications [23]. In order to choose the optimal features in the face, the feature selection strategy is used to classify the smile of a human by excluding unwanted or redundant features. However, traditional optimization solutions do not maximize and converge to the global minimum solution. Through using metaheuristic evolutionary optimization algorithms such as Ant Colony Optimization (ACO) [24], Bee Colony Optimization (BCO) [25], Particle Swarm Optimization (PSO) [26], etc., conventional techniques will minimize drawbacks. Such approaches are inefficient in evaluating the global optimum concerning the pace of convergence, capability for experimentation, and consistency of solution [27]. An updated Cuckoo Search (CS) algorithm is suggested to take several characteristics to perform classification and uses two learning algorithms, namely K-nearest neighbor and Support Vector Machines (SVMs) [28].

In the literature, many other methods for extracting the salient features of an image have been used, such as the chaotic Gray-Wolf Algorithm [29] and Whale Optimization Algorithm (WOA) [30]. Because randomization is so important in exploration and exploitation, using the existing randomization technique in WOA would raise computational time, especially for highly complex problems. The Multiverse Optimization (MVO) algorithm suffers from a low convergence rate and entrapment in local optima. To overcome these problems, a chaotic MVO algorithm (CMVO) is applied that minimizes the slow convergence problem and traps local optima [31].

A graphical model for the extraction and description of functions using a hybrid approach to recognize a person's facial expressions was developed in [32]. However, large memory complexity is the main disadvantage. In this case, matrices can also be a good solution when the graph is roughly complete (every node is connected to almost all of the other nodes). In [33], a Zernike model was developed based on a local moment to classify a person's expressions such as regular, happy, sad, surprise, angry, and fear. Using characteristics for speech recognition and motion change, recognition was done, and SVM was used for the classification. The experiments carried out showed that when compared to the individual descriptor, the integrated system achieves better results. However, this takes a long training time and has a large difficulty to understand and interpret the final model, variable weights, and individual impact.

Recently, several methods for classifying face speech using a neural network approach have been suggested [34,35]. A target-oriented approach using a neural network for facial expression detection was discussed in [34]. There are many limitations of this approach such as stated goals may not be realistic, and unintended outcomes may be ignored. In [36], the detection technique was used to perform automatic recognition of facial expressions using the Elman neural network to recognize feelings such as satisfaction, sadness, frustration, anxiety, disgust, and surprise. The identification rate was analyzed to be lower for pictures of sorrow, anxiety, and disgust. However, neural networks demand processors with parallel processing power by their structure. Furthermore, experience and trial and error are used to achieve the appropriate network structure.

Inspired by the good performance of the CNNs in computer vision tasks, such as image classification and face recognition, several CNN-based smile classification approaches have been proposed in recent years. In [37], a deeper CNN that has a complex CNN network consisting of two convolution layers, each accompanied by a max-pooling and four initiation layers, was suggested for facial expression recognition. It has a single-part architecture that takes face pictures as input and classifies them into one of the seven sentences. Another related work in [38] utilizes deep learning-based facial expression to minimize the dependency on face physics. Herein, the input image is convoluted in the convolution layers with a filter set. To identify the facial expression, the CNN generates a map of functions that are then paired with fully connected networks. In [39], a deep learning approach is introduced to track consumer behavior patterns by measuring customer behavior patterns. The authors in [40] presented a deep region and multilabel learner's scheme for estimation of head poses and study of facial expressions to report the interest of customers. They used a feedback network to isolate vast facial regions.

In general, a deep learning approach gives optimum facial features and classification. However, it is difficult to gather vast amounts of training data for facial expression recognition under different circumstances and more massive calculations are required. Therefore, the calculation time of the deep learning algorithm needs to be reduced. In order to minimize the number of features, a Deep Convolutional Neural Network (DCNN) and Cat Swarm Optimization (CSO) are used for facial expression recognition methods to minimize processing time [10]. Yet, there is no common theory to help choose the best resources for deep learning because they need an understanding of the topology, the process of training, and other parameters; as a consequence, fewer experienced individuals find it impossible to follow.

In contrast to the previous methods, which rely on a deep learning concept for smile classification, and in order to solve the problem facing this type of learning in terms of its difficulty to gather vast amounts of training data for facial expression recognition under different circumstances, the suggested approach utilizes both the row transformation technique and the cascade classifier in a unified framework. The cascade classifier can process a large number of features. Even so, the effectiveness of this method is fundamentally dependent on the extracted features, which may indeed not require much time to realize its purpose algorithm. In this case, row transformation is used to exclude any redundant coefficients from a vector of features, thus increasing the discriminatory capacity of the derived features and reducing computational complexity.

3. Methodology

3.1. Face Detection

The first step in the identification of a smile is to locate the face in the picture. For this function, the Viola–Jones method was used [41]. The face identified represents a Region of Interest (ROI) in the picture of a smile. The Viola–Jones method was also used to locate the eyes and mouth. The area of the eyebrow was determined from the position of the eye region. After the identification of facial regions, different techniques of image processing were used in each of the detected ROIs to remove the eyes and mouth. A search was then carried out on each of the extracted components to identify facial expressions [42]. Figure 2 shows the proposed system block diagram, and Figure 3 shows an example of the facial regions.

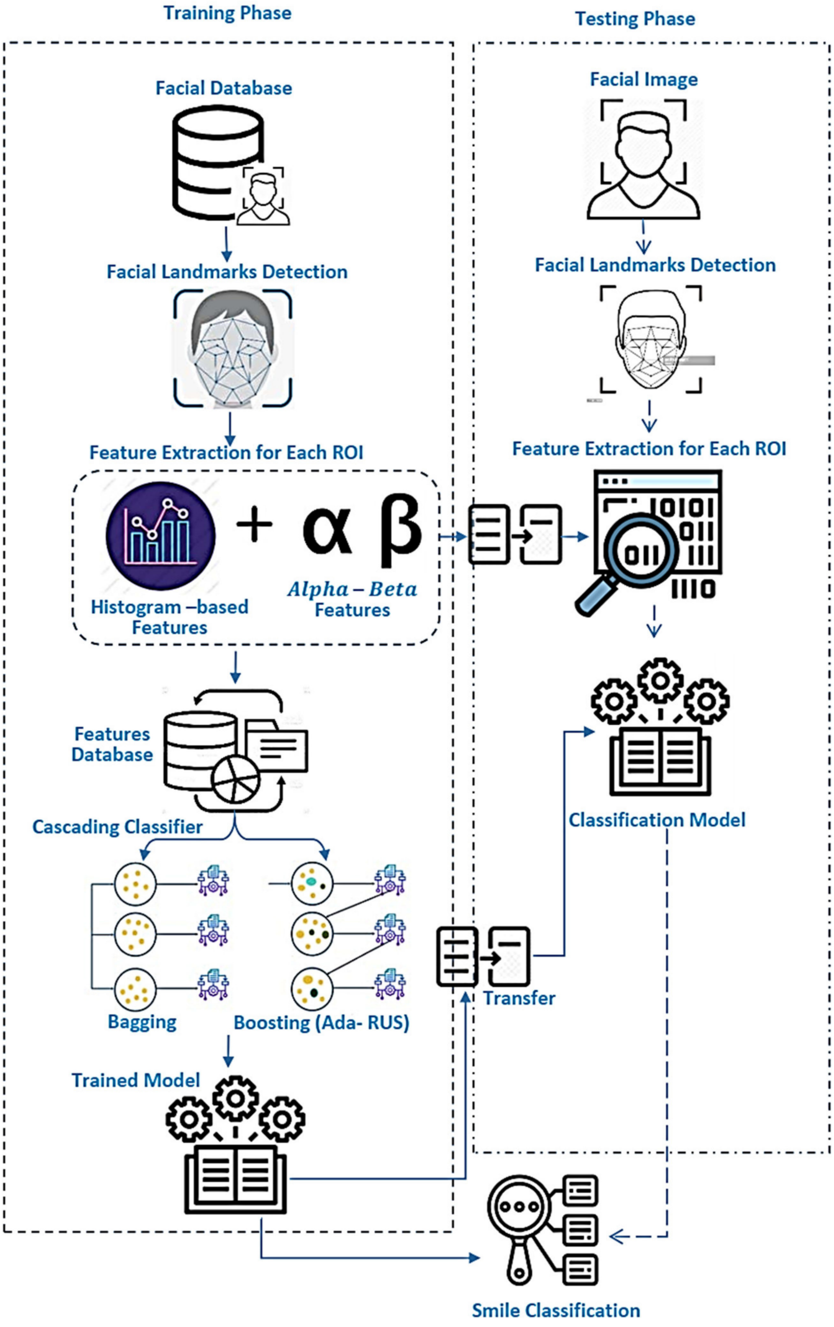

Figure 2. System flowchart.

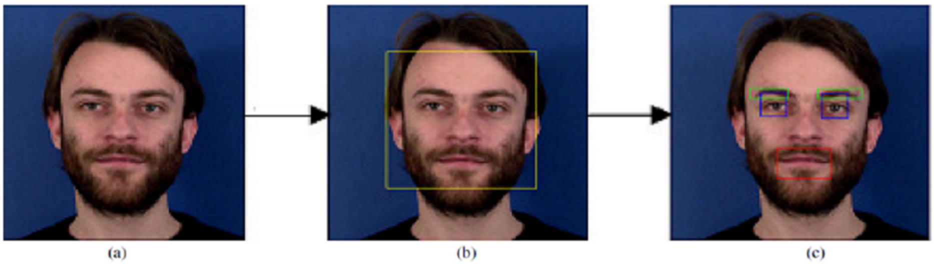

Figure 3. Facial regions. (**a**) Input, (**b**) face detection (Step 1), and (**c**) region detection (Step 2).

To soften the image, minor noises such as defects in the image and scarcely visible lines were discarded. In order to locate points of interest on the face, it was initially important to enhance and extract the relevant information from the image. For this reason, different techniques of image processing were used in this work such as contrast correction, thresholding, context subtraction, contour detection, and Laplacian and Gaussian filters for the extraction of points of interest. To segment the image into regions (set of pixels), two methods were used in segmentation: thresholding and morphological operations. To remove the edges of the eyes and mouth, a canny detector was used, and a search was carried out on each of the resulting edges to detect facial landmarks. Figure 4 illustrates the output facial landmarks, which are detected from the image processing techniques in each of the ROIs (see [42] for in-depth details).

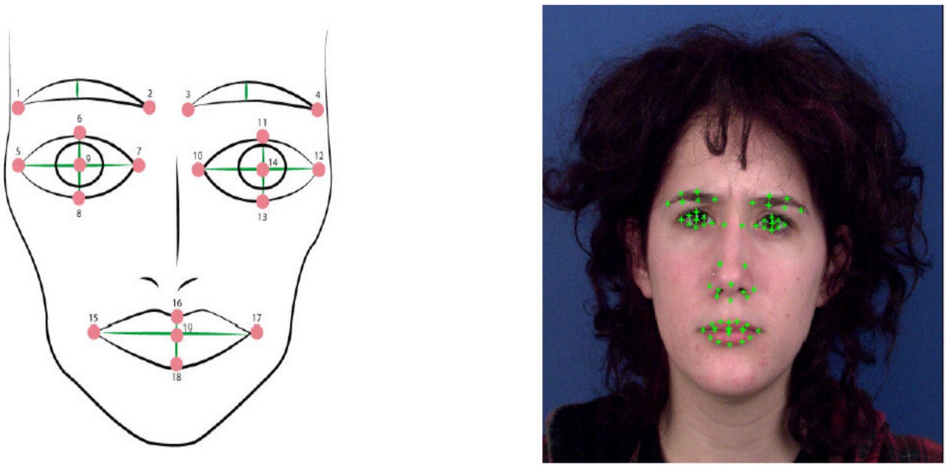

Figure 4. Facial landmarks.

3.2. Feature Extraction

After the preprocessing stage, feature extraction is performed in a facial expression recognition system. The most important knowledge present in the original ROI is a kind of dimensional decrease technique. This is the knowledge gathered in a small space from the photo. The main goal of the extraction function is to minimize the initial ROI size into a manageable processing vector that has histogram and alpha and beta features.

3.2.1. Histogram Feature Extraction

Herein, six parameters of histograms are calculated for each ROI. The histogram features are statically based features as a model of the probability distribution of the gray levels. We define the first-order histogram probability as [43]:

$$P(g) = \frac{N(g)}{M} \tag{1}$$

$N(g)$ is the number of pixels at gray level value g, and M is the number of pixels in the ROI. $P(g)$ has all values less than or equal to 1. The total number of gray levels available will be L, so the gray levels range from 0 to $L - 1$. Histogram probabilities include the mean, standard deviation, skew, Kurtosis, energy, and entropy. Mean (μ) is the average value, which informs us more about the ROI's overall brightness. The mean of a light ROI will be high, while the mean of a dark ROI will be low. Standard deviation (σ) shows contrast and represents the distribution of data; a high-contrast image has a high variance, while a low-contrast image has a low variance. Skewness (P_3) is a metric for symmetry or, more specifically, its absence. A distribution, or data collection, is said to be symmetrical if it appears identical on both sides of the middle point. Kurtosis (P_4) is a statistic that indicates when data are heavily or lightly skewed in comparison to a standard distribution. Energy (ζ) reveals knowledge about how the gray levels are distributed. For an image with a constant value, the measure of energy has a maximal value of 1 and gets increasingly smaller as the pixel values are distributed across grayer level values. The greater this value, the more easily the ROI data can be compressed. If the energy is high, it indicates that the ROI has a limited number of gray levels, implying that the distribution is concentrated in just a few different gray levels. Entropy (η) is a measure that informs the number of bits required to code the ROI data. The entropy of the ROI grows as the pixel values are spread over more gray depths. This value is usually inversely proportional to the energy [43].

$$\mu = \sum_{g=0}^{L-1} g\, P(g) \tag{2}$$

$$\sigma = \sum_{g=0}^{L-1} (g - P_1)^2\, P(g) \tag{3}$$

$$P_z = \sum_{g=0}^{L-1} \frac{(g-\mu)^z\, P(g)}{(\mu_1)^{\frac{z}{2}}}\ z \in \{3,4\} \tag{4}$$

$$\zeta = \sum_{g=0}^{L-1} [P(g)]^2 \tag{5}$$

$$\eta = -\sum_{g=0}^{L-1} P(g) \log_2[P(g)] \tag{6}$$

3.2.2. Alpha and Beta Features

Alpha and beta are the comparisons between the area of teeth and lips. In order to reduce the amount of redundant information, the oral region needs to be extracted, and the lip area, teeth area, and eye area are taken as regions of interest. A method based on a localized active contour model can segment the mouth area by general structure and face proportion (see [44] for all method details). Figure 5 illustrates the steps to pick the lips area.

$$Alpha\ (\alpha) = teeth\_Area / Lips\_Area \tag{7}$$

$$Beta(\beta) = Lip\_Width / Eye\_Length \tag{8}$$

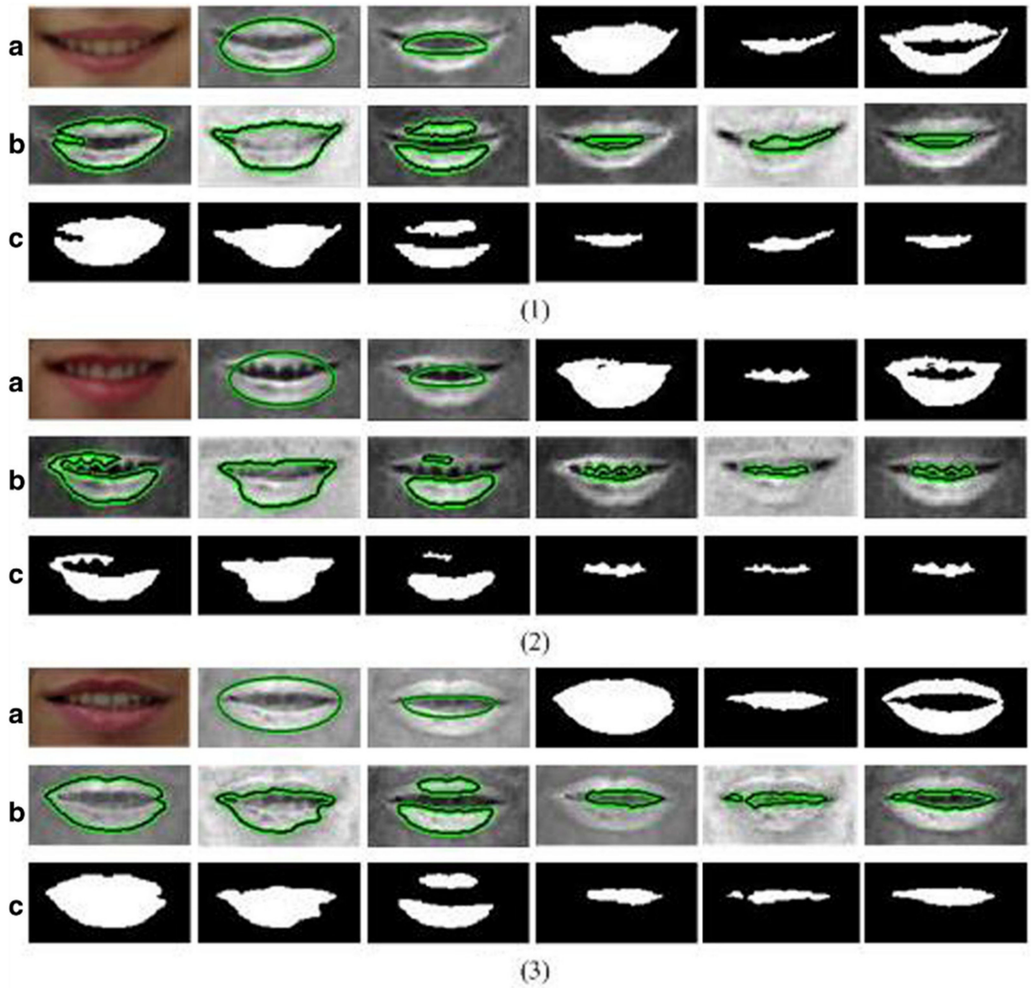

Figure 5. For each smile sample in (1), (2), and (3) (**a**) From left to right: RGB images after brightness equalization, outside contours, internal contours, outcomes of the outside boundary, outcomes of the internal boundary, and finishing segmentation outcomes; (**b**) convergence outcomes of the outside and internal contours; (**c**) segmentation outcomes of images in (**b**).

3.3. Cascade Classifier

The cascade classifier consists of several strong classifiers. The classifiers in the earlier stages are simple and can speedily filter out the background regions. They are more complex in the classifiers of the later stages to spend more computational time on the promising face-like regions. In our case, the features are combined to be used in the MATLAB toolbox. The proposed system is employed with different classifiers such as AdaBoostM2, RUSBoost, and Bagging. AdaBoost is flexible, so the learners may be fine-tuned after being misclassified in the prior classifications of the initial learner. Such issue types might be more prone to overfitting than others. As long as each individual's output has a higher than random chance of creating improved results, the final model may be considered a convergent learner. Although improving accuracy can improve the learner's score, it sacrifices understanding and simplification. Additionally, it can be difficult to execute because of the increased computing requirements. Herein, these classifiers were used as a black box with their default parameters (see [11] for more details). Figure 6 illustrates some different smile categories.

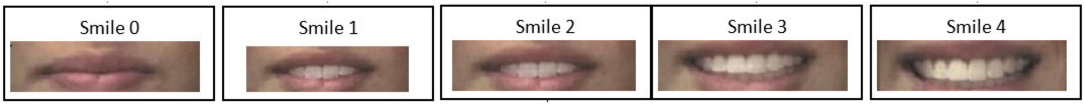

Figure 6. Different smiles for a person.

4. Experimental Results

The proposed facial expression recognition system is tested with a data set of benchmark data sets that includes the Japanese Female Facial Expression (JAFFE), Extended Cohn–Kanade (CK+), and CK+48 data sets [43–45]. JAFFE is a Japanese database containing 7 facial expressions with a 256 × 256-pixel resolution of 213 images. With 10,414 images with a resolution of 640 × 490 pixels, the CK+ database has 13 expressions. The CK+48 data set has 7 facial expressions with a resolution of 48 × 48 pixels with 981 images. Features are extracted from ROIs using histograms and lip, teeth, and eye areas, which produce a 21-dimensional feature vector. Herein, 80 percent is selected for training, and 20 percent is for the testing of each data set considered. The prototype classification methodology was developed in a modular manner and implemented and evaluated on a Dell^{TM} Inspiron^{TM} N5110 laptop device, manufactured by Dell Computer Corporation in Round Rock, Texas, U.S. with specifications Intel(R) Core(TM) i5-2410M processor running at 2.30 GHz, 4.00 GB of RAM, Windows 7 64-bit. Herein, recognition rate, accuracy, sensitivity, recall, specificity, precision, $F_{measure}$, and sensitivity are used to evaluate the efficiency of the suggested model. See [39] for more details.

$$Rate = \frac{No.\,of\ expressions\ \ classified\ correctly}{Total\ no.\,of\ images} \times 100 \tag{9}$$

$$Precision = Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP} \times 100 \tag{10}$$

$$Sensitivity = Recall = Hit\ rate = True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} \times 100 \tag{11}$$

$$F_{measure} = 2 \times \frac{PPV \times TPR}{PPV + TPR} \times 100 \tag{12}$$

$$Specificity = Selectivity = True\ Negative\ Rate\ (TNR) = \frac{TN}{TN+FP} \tag{13}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \times 100 \tag{14}$$

where *TP*, *TN*, *FP*, and *FN* are the true positive, true negative, false positive, and false negative, respectively. Herein, 80% of samples for each class were used for training, and the remaining 20% of samples were used for testing.

Performance Analysis

The first set of experiments was conducted to verify the efficiency of the suggested model under different cascade classifiers for different data sets that represent different conditions such as occlusions, pose changes, and extreme lightings. Tables 1–3 illustrate the statistical results of applying the suggested smile classification model using Adaboost, Bagging, and RUSBoost classifiers, respectively. In our evaluation using standard benchmarks, the suggested model achieved accuracy over 98%, 91%, and 65% for CK+ and CK+48 and JAFFE. respectively. On average, the suggested model requires about 1 s to execute loading, preprocessing, and feature building for each image of size 256 × 256. The cascade classification process requires about 13 s for training and to perform the classification of 2000 images. This runtime depends on the capabilities of the device used, and this runtime can be reduced by using a device that has higher specifications.

It is noted that the proposed system does not achieve good results in the case of the JAFFE data set, as this benchmark database includes an insufficient number of images for each class. In general, utilizing cascading classifiers need more data for correct training. Results in the tables reveal that there are no clear differences between the uses of the different cascade classifiers for the proposed model of the same database in terms of different objective measurements. The results confirm the research hypothesis that using cascade classifiers based on discriminative features will enhance the classification accuracy. The previous results were confirmed through confusion matrix analysis for the Bagging classifiers using the three benchmark data sets, as illustrated in Tables 4–6. During the training stage, Bagging and Boosting get N learners by generating additional data. N new training data sets are produced by random sampling with substitution from the original set. Some observations may be replicated in each new training data set by sampling with replacement. In Bagging, any variable has the same potential of appearing in a new data set. To expand on this idea, though, some of the observations are updated or supplemented as often as Boosting requires. This learning algorithm will be trained using multiple sets of multiple samples to avoid issues that could arise due to multiple classifiers training on the same data [45].

Table 1. Statistical analysis of various data sets using the AdaboostM1 classifier.

Data Set	TP	TN	FP	FN	Accuracy	TPR	TNR	PPV	$F_{measure}$
CK+	2048.9	1916.59	33.1	33.1	98.35	0.98	0.98	0.98	0.98
CK+48	180.2	165.74	15.8	15.8	91.64	0.92	0.92	0.92	0.92
JAFFE	26.5	34.22	16.5	16.5	65.02	0.62	0.68	0.62	0.62

Table 2. Statistical analysis of various data sets using the Bagging classifier.

Data Set	TP	TN	FP	FN	Accuracy	TPR	TNR	PPV	$F_{measure}$
CK+	2049.2	1919.32	32.8	32.8	98.38	0.98	0.98	0.98	0.98
CK+48	179.8	165.66	16.2	16.2	91.44	0.92	0.91	0.92	0.98
JAFFE	27.3	34.47	15.7	15.7	66.62	0.65	0.69	0.63	0.64

Table 3. Statistical analysis of various data sets using the RUSBoost classifier.

Data Set	TP	TN	FP	FN	Accuracy	TPR	TNR	PPV	$F_{measure}$
CK+	2047.5	1919.19	34.5	34.5	98.29	0.98	0.98	0.98	0.98
CK+48	179.7	165.67	16.3	16.3	91.39	0.92	0.91	0.92	0.92
JAFFE	26.2	34.46	16.8	16.8	64.64	0.61	0.68	0.61	0.61

Table 4. Confusion matrix using the Bagging classifier for the CK+48 data set (20% of samples for testing).

Emotions	E1	E2	E3	E4	E5	E6	E7
E1	31	0	0	0	0	0	0
E2	2	5	0	0	0	0	0
E3	0	5	24	0	0	0	0
E4	0	0	5	16	0	0	0
E5	0	0	0	1	35	0	0
E6	0	0	0	0	3	13	0
E7	0	0	0	0	0	4	52

Table 5. Confusion matrix using the Bagging classifier for the JAFFE data set (20% of samples for testing).

Emotions	E1	E2	E3	E4	E5	E6	E7
E1	15	0	0	0	0	0	0
E2	3	3	0	0	0	0	0
E3	0	3	3	0	0	0	0
E4	0	0	2	2	0	0	0
E5	0	0	0	1	1	0	0
E6	0	0	0	0	1	6	0
E7	0	0	0	0	0	1	2

Table 6. Confusion matrix using the Bagging classifier for the CK+ data set (20% of samples for testing).

Emotions	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
E1	418	0	0	0	0	0	0	0	0	0	0	0	0
E2	4	371	0	0	0	0	0	0	0	0	0	0	0
E3	0	3	341	0	0	0	0	0	0	0	0	0	0
E4	0	0	3	329	0	0	0	0	0	0	0	0	0
E5	0	0	0	4	231	0	0	0	0	0	0	0	0
E6	0	0	0	0	3	209	0	0	0	0	0	0	0
E7	0	0	0	0	0	5	85	0	0	0	0	0	0
E8	0	0	0	0	0	0	5	25	0	0	0	0	0
E9	0	0	0	0	0	0	0	5	3	0	0	0	0
E10	0	0	0	0	0	0	0	0	2	8	0	0	0
E11	0	0	0	0	0	0	0	0	0	2	14	0	0
E12	0	0	0	0	0	0	0	0	0	0	2	5	0
E13	0	0	0	0	0	0	0	0	0	0	0	2	3

The second set of experiments demonstrated how the classification rate of the proposed model is dependent on the number of facial image samples enrolled in the smile's class, so the chance of a correct hit rises as the smile's class enrolls more samples. In Tables 7–9 for each cascade classifier, as expected, the classification rate goes up in line with the interclass term variation. The accuracy rate increases by about 4% on average with every 100 samples added to the data set. By combining both samples to develop the proposed model, the precision is increased (on average) by 98.29%, 91.64%, and 65.02% for CK+, CK+48, and JAFFE data sets. respectively.

Table 7. Relationship between accuracy rate and the number of samples using the AdaboostM1 classifier.

Data Set	Samples Numbers	Accuracy Rate
CK+	1000	80.01
	2000	82.10
	3000	85.00
	4000	87.00
	5000	88.02
	6000	89.02
	7000	91.01
	8000	95.00
	9000	97.05
	10,414	98.35

Table 7. *Cont.*

Data Set	Samples Numbers	Accuracy Rate
CK+48	100	70.01
	200	75.06
	300	76.08
	400	79.80
	500	81.02
	600	81.06
	700	85.00
	800	89.90
	981	91.64
JAFFE	40	48.00
	80	49.00
	120	60.03
	160	62.50
	213	65.02

Table 8. Relationship between accuracy rate and the number of samples using the Bagging classifier.

Data Set	Samples Numbers	Accuracy Rate
CK+	1000	81.81
	2000	83.15
	3000	84.50
	4000	86.00
	5000	85.02
	6000	88.02
	7000	90.01
	8000	92.00
	9000	96.05
	10,414	98.38
CK+48	100	71.01
	200	73.05
	300	75.28
	400	77.85
	500	80.52
	600	82.46
	700	85.50
	800	88.95
	981	91.44
JAFFE	40	42.00
	80	44.00
	120	55.03
	160	60.50
	213	66.62

Table 9. Relationship between accuracy rate and the number of samples using the RUSBoost classifier.

Data Set	Samples Numbers	Accuracy Rate
CK+	1000	82.82
	2000	84.35
	3000	84.58
	4000	85.55
	5000	86.08
	6000	88.72
	7000	91.51
	8000	92.07
	9000	97.07
	10,414	98.29
CK+48	100	71.61
	200	72.65
	300	75.26
	400	76.85
	500	82.72
	600	84.47
	700	85.57
	800	89.75
	981	91.39
JAFFE	40	47
	80	55.05
	120	60.03
	160	62.5
	213	64.64

The third set of experiments was conducted to verify how the proposed model's verification rate varies with the amount of noise in the picture. In this scenario, the image is reinforced with Gaussian noise (noise amount between 1 and 10). As seen in Table 10, the probability of a correct hit decreases with increasing noise. Up to a degree of noise, however, output gains decline with the noise levels in the class's image. When noise alters the gray level of the images, a variation in the derived features occurs. In these three data sets, we observe the same disparity in precision.

The final series of experiments validated the proposed model's efficiency in comparison to the state-of-the-art models mentioned in Table 11 using the CK+ data set. The findings corroborate the proposed model's dominance. Despite the proposed model's convergence with the 3D Shape-based recognition model's performance, the suggested model is descriptor-independent (geometric descriptor), and it employs a number of translation- and scale-invariant functions. By and large, the 3D Shape descriptor performs poorly, while the data collection contains more noise, i.e., target groups overlap. Furthermore, employing a Deep neural network needs adjusting network configuration parameters that, in turn, need more effort.

Table 10. Relationship between accuracy rate and noise amount.

Data Set	Noise Amount	Accuracy Rate
CK+	1	98.01
	2	96.5
	5	95.02
	7	93.54
	10	92.68

Table 10. *Cont.*

Data Set	Noise Amount	Accuracy Rate
CK+48	1	90.72
	2	89.23
	5	87.25
	7	85.03
	10	84.38
JAFFE	1	65.73
	2	62.32
	5	61.96
	7	61.61
	10	60.63

Table 11. Comparisons with different methods on CK+.

Method	Accuracy Rate
Neural Network [36]	94.4
Deep Neural Network [37]	97.8
3D Shape [46]	86.8
Gabor [47]	91.81
LBP [47]	82.38
MSDF [47]	94.34
Simple BoW [47]	92.67
SS-SIFT+BoW [47]	93.28
MSDF+BoW [47]	95.85
3D Shape + GA [48]	97.6
3D Shape + GA + KSS [48]	97.9
Gabor [49]	93.8
2D Shape [50]	92.4
The proposed work	98.01

5. Conclusions

Facial expression classification is a very challenging and open area of research. This paper developed a simple yet effective smile classification approach based on a combination of a row-transform-based feature extraction algorithm and a cascade classifier. Utilizing row transformation helps to remove some unnecessary coefficients from the extracted features' vector to reduce computational complexity. By taking a weighted average of the decisions made by poor learners, cascading assists in training a highly reliable classifier. The model's objective is to achieve the lowest possible recognition error, the shortest possible run time, and the simplest layout. For various samples, the model achieves a strong identification accuracy of 98.69%. The proposed model is characterized by simplicity in implementation, in contrast to deep-learning-based classification methods that depend on adjusting multiple variables to achieve reliable accuracy. On the other hand, the limitation of this work appeared in the JAFFE data set because of the insufficient number of samples. In the future, a mobile application shall be created to find expressions in each video frame automatically. Furthermore, speech detection includes both audios from a speaker tone, and video responses can further improve detection accuracy.

Author Contributions: Conceptualization, O.A.H., N.A.A., and Z.Z.A.; methodology, O.A.H., N.A.A., and Z.Z.A.; software, O.A.H. and S.M.D.; validation, O.A.H., N.A.A., and S.M.D.; formal analysis, O.A.H., N.A.A., and Z.Z.A.; investigation, O.A.H. and S.M.D.; resources, O.A.H., N.A.A., and Z.Z.A.; data curation, O.A.H., N.A.A., and Z.Z.A.; writing—Original draft preparation, O.A.H. and S.M.D.; writing—Review and editing, S.M.D. visualization, O.A.H., N.A.A., and Z.Z.A.; supervision, N.A.A. and Z.Z.A.; project administration, O.A.H., N.A.A., and Z.Z.A.; funding acquisition, O.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The study did not report any data.

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

References

1. Andrian, R.; Supangkat, S.H. Comparative Analysis of Deep Convolutional Neural Networks Architecture in Facial Expression Recognition: A Survey. In Proceedings of the International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 19–20 November 2020; pp. 1–6.
2. Nestor, M.S.; Fischer, D.; Arnold, D. Masking our emotions: Botulinum toxin, facial expression, and well-being in the age of COVID-19. *J. Cosmet. Dermatol.* **2020**, *19*, 2154–2160. [CrossRef] [PubMed]
3. Geng, Z.; Cao, C.; Tulyakov, S. Towards Photo-Realistic Facial Expression Manipulation. *Int. J. Comput. Vis.* **2020**, *128*, 2744–2761. [CrossRef]
4. Zhang, F.; Zhang, T.; Mao, Q.; Xu, C. Geometry guided pose-invariant facial expression recognition. *IEEE Trans. Image Process.* **2020**, *29*, 4445–4460. [CrossRef] [PubMed]
5. Gogić, I.; Manhart, M.; Pandžić, I.S.; Ahlberg, J. Fast facial expression recognition using local binary features and shallow neural networks. *Vis. Comput.* **2020**, *36*, 97–112. [CrossRef]
6. Escalera, S.; Puertas, E.; Radeva, P.; Pujol, O. Multi-modal laughter recognition in video conversations. In Proceedings of the IEEE Conference on Computer Vision, Miami, FL, USA, 20–25 June 2009; pp. 110–115.
7. Gong, B.; Wang, Y.; Liu, J.; Tang, X. Automatic facial expression recognition on a single 3D face by exploring shape deformation. In Proceedings of the 17th ACM international conference on Multimedia, Beijing, China, 23 October 2009; pp. 569–572.
8. Cotter, S.F. MobiExpressNet: A deep learning network for face expression recognition on smart phones. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 4–6 January 2020; pp. 1–4.
9. Law, S.; Seresinhe, C.I.; Shen, Y.; Gutierrez-Roig, M. Street-Frontage-Net: Urban image classification using deep convolutional neural networks. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 681–707. [CrossRef]
10. Alarifi, A.; Tolba, A.; Al-Makhadmeh, Z.; Said, W. A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. *J. Supercomput.* **2020**, *76*, 4414–4429. [CrossRef]
11. Ashir, A.M.; Eleyan, A.; Akdemir, B. Facial expression recognition with dynamic cascaded classifier. *Neural Comput. Appl.* **2020**, *32*, 6295–6309. [CrossRef]
12. Cossetin, M.J.; Nievola, J.C.; Koerich, A.L. Facial expression recognition using a pairwise feature selection and classification approach. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 5149–5155.
13. Happy, S.L.; Routray, A. Automatic facial expression recognition using features of salient facial patches. *IEEE Trans. Affect. Comput.* **2015**, *6*, 1–12. [CrossRef]
14. Mistry, K.; Zhang, L.; Neoh, S.C.; Lim, C.P.; Fielding, B. A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition. *IEEE Trans. Cybern.* **2017**, *47*, 1496–1509. [CrossRef] [PubMed]
15. Samara, A.; Galway, L.; Bond, R.; Wang, H. Affective state detection via facial expression analysis within a human–computer interaction context. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 2175–2184. [CrossRef]
16. Mitra, S.; Acharya, T. Gesture Recognition: A Survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 311–324. [CrossRef]
17. Hassen, O.; Abu, N.; Abidin, Z. Human identification system: A review. *Int. J. Comput. Bus. Res. IJCBR* **2019**, *9*, 1–26.
18. Buciu, I.; Kotropoulos, C.; Pitas, I. ICA and Gabor representation for facial expression recognition. In Proceedings of the Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 14–17 September 2003; p. II-855.
19. Minsky, M.; Papert, S.A. *Perceptrons: An Introduction to Computational Geometry*; MIT press: Cambridge, MA, USA, 2017.
20. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
21. Sikkandar, H.; Thiyagarajan, R. Soft biometrics-based face image retrieval using improved grey wolf optimization. *IET Image Process.* **2020**, *14*, 451–461. [CrossRef]
22. Woźniak, M.; Połap, D. Bio-inspired methods modeled for respiratory disease detection from medical images. *Swarm Evol. Comput.* **2018**, *41*, 69–96. [CrossRef]
23. Woźniak, M.; Połap, D. Adaptive neuro-heuristic hybrid model for fruit peel defects detection. *Neural Netw.* **2018**, *98*, 16–33. [CrossRef] [PubMed]
24. Kanan, H.R.; Faez, K.; Hosseinzadeh, M. Face recognition system using ant colony optimization-based selected features. In Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications, Honolulu, HI, USA, 1–5 April 2007; pp. 57–62.
25. Karaboga, N. A new design method based on artificial bee colony algorithm for digital IIR filters. *J. Frankl. Inst.* **2009**, *346*, 328–348. [CrossRef]
26. Ababneh, J.I.; Bataineh, M.H. Linear phase FIR filter design using particle swarm optimization and genetic algorithms. *Digit. Signal Process.* **2008**, *18*, 657–668. [CrossRef]

27. Chu, S.C.; Tsai, P.W. Computational intelligence based on the behavior of cats. *Int. J. Innov. Comput. Inf. Control* **2007**, *3*, 163–173.
28. Aziz, M.A.E.; Ewees, A.A.; Hassanien, A.E. Multi-objective whale optimization algorithm for content-based image retrieval. *Multimed. Tools Appl.* **2018**, *77*, 26135–26172. [CrossRef]
29. Ibrahim, R.A.; Elaziz, M.A.; Lu, S. Chaotic opposition-based grey-wolf optimization algorithm based on differential evolution and disruption operator for global optimization. *Expert Syst. Appl.* **2018**, *108*, 1–27. [CrossRef]
30. Aziz, M.A.E.; Hassanien, A.E. Modified cuckoo search algorithm with rough sets for feature selection. *Neural Comput. Appl.* **2018**, *29*, 925–934. [CrossRef]
31. Mostafa, A.; Hassanien, A.E.; Houseni, M.; Hefny, H. Liver segmentation in MRI images based on whale optimization algorithm. *Multimed. Tools Appl.* **2018**, *76*, 24931–24954. [CrossRef]
32. Krithika, L.B.; Priya, G.L. Graph based feature extraction and hybrid classification approach for facial expression recognition. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 2131–2147. [CrossRef] [PubMed]
33. Fan, X.; Tjahjadi, T. A dynamic framework based on local Zernike moment and motion history image for facial expression recognition. *Pattern Recognit.* **2017**, *64*, 399–406. [CrossRef]
34. Roomi, M.; Naghasundharam, S.A.; Kumar, S.; Sugavanam, R. Emotion recognition from facial expression-A target oriented approach using neural network. In Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing, Kolkata, India, 16–18 December 2004; pp. 1–4.
35. Fuentes, C.; Herskovic, V.; Rodríguez, I.; Gerea, C.; Marques, M.; Rossel, P.O. A systematic literature review about technologies for self-reporting emotional information. *J. Ambient Intell. Humaniz. Comput.* **2017**, *8*, 593–606. [CrossRef]
36. Langeroodi, B.Y.A.; Kojouri, K.K. Automatic Facial Expression Recognition Using Neural Network. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), Las Vegas, NV, USA, 18–21 July 2011; pp. 1–5.
37. Mollahosseini, A.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.
38. Walecki, R.; Rudovic, O.; Pavlovic, V.; Schuller, B.; Pantic, M. Deep structured learning for facial expression intensity estimation. *Image Vis. Comput.* **2017**, *259*, 143–154.
39. Yolcu, G.; Oztel, I.; Kazan, S.; Oz, C.; Bunyak, F. Deep learning-based face analysis system for monitoring customer interest. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 237–248. [CrossRef]
40. Zhao, K.; Chu, W.-S.; Zhang, H. Deep region and multi-label learning for facial action unit detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3391–3399.
41. Viola, P.; Jones, M. Robust real-time object detection. *Int. J. Comput. Vis.* **2001**, *4*, 34–47.
42. Silva, C.; Schnitman, L.; Oliveira, L. Detection of facial landmarks using local-based information. In Proceedings of the 19th Brazilian Conference on Automation, Campina Grande, Brazil, 2–6 September 2012; pp. 1–5.
43. Garali, I.; Adel, M.; Bourennane, S.; Guedj, E. Histogram-based features selection and volume of interest ranking for brain PET image classification. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–12. [CrossRef]
44. Lu, Y.; Liu, Q. Lip segmentation using automatic selected initial contours based on localized active contour model. *EURASIP J. Image Video Process.* **2018**, *7*, 1–12. [CrossRef]
45. Khoshgoftaar, T.M.; Hulse, J.V.; Napolitano, A. Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2011**, *41*, 552–568. [CrossRef]
46. Jeni, L.A.; Lőrincz, A.; Nagy, T.; Palotai, Z.; Sebők, J.; Szabó, Z.; Takács, D. 3Dshape estimation in video sequences provides high precision evaluation of facial expressions. *Image Vis. Comput.* **2012**, *30*, 785–795. [CrossRef]
47. Sikka, K.; Wu, T.; Susskind, J.; Bartlett, M. Exploring bag of words architectures in the facial expression domain. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 250–259.
48. Jeni, L.A.; Lőrincz, A.; Szabó, Z.; Cohn, J.F.; Kanade, T. Spatio-temporal event classification using time-series kernel based structured sparsity. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 135–150.
49. Mahoor, M.; Zhou, M.; Veon, K.L.; Mavadati, S.; Cohn, J. Facial action unit recognition with sparse representation. In Proceedings of the Automatic Face Gesture Recognition and Workshops, Santa Barbara, CA, USA, 21–25 March 2011; pp. 336–342.
50. Zafeiriou, S.; Petrou, M. Sparse representations for facial expressions recognition via l1 optimization. In Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 32–39.