Article

Precision-Based Weighted Blending Distributed Ensemble Model for Emotion Classification

Gayathri Soman [1], M. V. Vivek [1], M. V. Judy [1], Elpiniki Papageorgiou [2,*] and Vassilis C. Gerogiannis [3,*]

[1] Department of Computer Applications, Cochin University of Science and Technology, Kochi 682022, Kerala, India; gayathrisoman0250@cusat.ac.in (G.S.); vivekmv@cusat.ac.in (M.V.V.); judynair@cusat.ac.in (M.V.J.)

[2] Department of Energy Systems, University of Thessaly, Gaiopolis, 41500 Larissa, Greece

[3] Department of Digital Systems, University of Thessaly, Gaiopolis, 41500 Larissa, Greece

* Correspondence: elpinikipapageorgiou@uth.gr (E.P.); vgerogian@uth.gr (V.C.G.)

Abstract: Focusing on emotion recognition, this paper addresses the task of emotion classification and its performance with respect to accuracy, by investigating the capabilities of a distributed ensemble model using precision-based weighted blending. Research on emotion recognition and classification refers to the detection of an individual's emotional state by considering various types of data as input features, such as textual data, facial expressions, vocal, gesture and physiological signal recognition, electrocardiogram (ECG) and electrodermography (EDG)/galvanic skin response (GSR). The extraction of effective emotional features from different types of input data, as well as the analysis of large volume of real-time data, have become increasingly important tasks in order to perform accurate classification. Taking into consideration the volume and variety of the examined problem, a machine learning model that works in a distributed manner is essential. In this direction, we propose a precision-based weighted blending distributed ensemble model for emotion classification. The suggested ensemble model can work well in a distributed manner using the concepts of Spark's resilient distributed datasets, which provide quick in-memory processing capabilities and also perform iterative computations effectively. Regarding model validation set, weights are assigned to different classifiers in the ensemble model, based on their precision value. Each weight determines the importance of the respective classifier in terms of its performing prediction, while a new model is built upon the derived weights. The produced model performs the task of final prediction on the test dataset. The results disclose that the proposed ensemble model is sufficiently accurate in differentiating between primary emotions (such as sadness, fear, and anger) and secondary emotions. The suggested ensemble model achieved accuracy of 76.2%, 99.4%, and 99.6% on the FER-2013, CK+, and FERG-DB datasets, respectively.

Keywords: emotion classification; distributed machine learning; blending ensemble model; Spark's resilient distributed datasets

1. Introduction

Emotions have a vital role in the development of human–computer interaction and humanoid robots. This is due to the fact that, by incorporating the idea of emotion understanding, intelligent software systems become more efficient and intuitive, as well as more akin to human–human communication [1]. Human beings convey their emotions in everyday encounters through vocal, hand, face, and body gestures. Because most human-to-computer interfaces (HCI) rely on simple interactions via traditional devices, such as keyboards, mice, touch screens, and so on, emotion identification and classification can help HCI devices function more efficiently. Textual data, facial expressions, verbal, gesture, and physiological signal recognition, electrocardiogram (ECG) and electrodermography (EDG)/galvanic skin response (GSR), and so on, can all be used to recognize and classify an individual's emotional state. Emotional state recognition can help increase access to

identification. A recent study has revealed that facial expressions is the most expressive method in which humans display emotions [2]. Facial expression detection has numerous uses, including non-intrusive sensor creation, medical teaching, telecommunications, police enforcement, lifelike software agents, and so on [2].

Many scholars have investigated so far various methods of emotion classification, with the most basic ones being happiness, sadness, anger, fear, disgust and surprise, all of which are based on a two-dimensional plane known as the valence-arousal plane. The discrete fundamental emotion description approach is used for this classification [3]. Primary and secondary emotions are two further types of emotions that are frequently employed [4]. Fear, joy, contempt, sadness, and surprise are examples of primary emotions, while feelings that create a mental image that links to memory or another primary emotion are examples of secondary emotions. Emotions are classified into two (valence and arousal) or three (valence, arousal, and dominance) dimensions using dimension-based techniques. Valence describes a person's level of positivity or negativity, whereas arousal describes the level of enthusiasm or indifference of emotion [5]. The dominance scale runs from submissive (no control) to dominance (empowered).

Various methods are utilized to recognize and classify human face expressions, with ensemble machine learning techniques outperforming other methods in terms of performance. Ensemble methods are machine learning algorithms that combine many base models to improve each base model's stability and predictive capability [6]. They are attempting to improve forecast performance by merging predictions from multiple models. Rather than relying on a single model and hoping for the best, a decision maker can utilize an ensemble method, which involves taking a sample of models into account, calculating which features to employ, and determining the final prediction based on the sampled models' aggregated results. There is no limit to the amount of ensembles a decision maker can create for a particular predictive modelling task; nonetheless, two types of approaches dominate the field of ensemble learning: bagging [7] and boosting [8]. The majority of machine learning models work differently, and each model may perform well on some data while performing poorly on others; however, when all of them are combined, the ensemble model may cancel out each other's flaws. As a result, ensemble learning is frequently effective and can be used for both prediction and classification tasks. For example, several machine learning approaches can be coupled to recognize and classify human facial emotions. The performance of the chosen machine learning technique, however, degrades as the amount of the used datasets grows. To tackle this difficulty, we might adopt a distributed environment in which a deep machine learning algorithm can perform well. A distributed computing system consists of a set of connected computers that perform tasks in parallel and communicate with one another as needed. Distributed machines are intended to improve performance and accuracy while also scaling to larger data sets. By expanding the input data amount, we may dramatically minimize the learning error for many algorithms [9].

In this paper, we present a precision-based weighted blending distributed ensemble model for emotion categorization. The proposed ensemble model uses a distributed approach by inheriting the main concepts of Spark's resilient distributed datasets, which allow fast in-memory processing and can conduct effectively iterative computations. Weights are assigned to the different classifiers in the ensemble model depending on their precision value using the model validation set. Each weight specifies the importance of each classifier in making a prediction, and a new model is built using the resultant weights. The final prediction on the test dataset is made using the model that was created. The prediction results show that the proposed ensemble model is capable of distinguishing between primary (such as sadness, fear, and rage) and secondary emotions.

The main contributions of the paper are the following:

- A precision-based weighted blending distributed ensemble model for emotion classification is proposed and tested on three datasets, as well as on a combination of them.

- The suggested ensemble model can work in a distributed manner using the concepts of Spark's resilient distributed datasets, which provide quick in-memory processing capabilities and also perform iterative computations in an effective way.
- The proposed ensemble model outperforms other approaches because not only does it consider the probabilities of each class, but also the precision value of each classifier, when generating the final prediction, thus giving greater weight to the classifier that performs well throughout each run.

The paper is structured as follows. In Section 2, we present related research on the use of machine learning and ensemble models for emotion detection and classification. In Section 3, we describe our proposed precision-based weighted blending ensemble model for emotion classification. The description emphasizes particularly on the explainability issues of the proposed ensemble model. In Section 4, we present the results produced after testing the model on some available datasets. In the last section, conclusions and future work are included.

2. Related Work

Emotion is among the most difficult things to define in psychology. There are several different definitions of emotions in the scientific literature. Mood, temperament, disposition and motivation are frequently linked to emotions. Emotion is defined in psychology as a complicated state of feeling that causes physical and psychological changes that influence behavior and thought. Many studies have been conducted to determine which emotions are fundamental. Ekman [10] proposed a list of six primary emotions. He noted that each emotion functions as a distinct category rather than an individual emotional state. Some theories view emotions as a synthesis of multiple psychological characteristics, with three axes: (1) tension versus relaxation, (2) enjoyable versus unpleasant, and (3) arousing versus subduing [11]. Russell [12] suggested a two-dimensional emotion model incorporating arousal and valence.

To categorize emotions, Wiem et al. [13] employed the support vector machine (SVM) technique with multiple kernels. ECG and the respiration volume were employed in this method. To elicit emotions, Liu and Sourina [14] employed visual and audio stimuli. Higher order crossings (HOC) derived from EEG signals offered the best accuracy for identifying emotions.

Bahari and Janghorbani [15] extracted 13 non-linear characteristics from EEG using recurrence plot analysis. The collected features were then categorized into emotional states using the k-nearest neighbors algorithm based on the arousal-valence plane. To extract a set of geometric features from a face, Murugappan and Mutawa [16] devised a triangulation method. Using a random forest (RF) classifier, they used the inscribed circle area of the triangle (ICAT) feature, which resulted in a higher classification rate. Using the instability property of EMG data, Cheng and Liu [17] devised a wavelet transform approach for recognizing emotions. The extracted maximum and minimum values of the wavelet coefficients were given as input to a back propagation (BP) neural network.

In [18], a facial image threshing (FIT) machine was presented, which employs sophisticated features of pre-trained facial recognition, using also the Xception method for training. In addition to the data-augmentation technique, the FIT machine involves deleting extraneous facial photographs, gathering facial images, correcting misplaced face data, and integrating original information on a vast scale. An unsupervised learning dataset is transformed into a supervised learning dataset by the FIT machine. Using the FIT machine to create facial pictures for the FER dataset or a facial-related dataset could be less expensive for many FER developers. They employed the multi-task cascade neural network (MTCNN) [19], a contemporary face detection system that outperforms the Haar cascade classifier approach.

Face expressions were discovered in [20,21] by collecting features from small size grids generated on the face. However, a minor misalignment of the face lowers recognition accuracy because features are retrieved from the wrong places. In [22], a local binary

pattern histogram of various block sizes was employed as feature vectors from the global face region. The facial expressions were classified using principal component analysis (PCA). Local variations of the facial components could not be reflected in the feature vector using this method. Ghimire et al. [23] split the entire face region into domain-specific local regions and retrieved appearance features, particular to each region. By utilizing an incremental search strategy that minimizes the feature dimensions, recognition accuracy was improved. Hammal introduced a facial expression classification model based on the transferable belief model (TBM) architecture in [24]. Classification was done by performing analysis on facial deformations. Devi and Prabhu [25] introduced a new technique for extracting facial features called advanced maximally stable extremal regions (AMSER). This classification method proved successful in extracting more accurate facial expressions.

Corchs et al. [26] look at the role of ensemble learning in emotion classification from both visual and linguistic perspectives. They employed five state-of-the-art classifiers as independent models: naive Bayes (NB), Bayesian network (BN), closest nearest-neighbor (NN), decision tree (DT) and linear support vector machine (SVM). To assess the predictions provided by all these different models, the authors considered two distinct ensemble methodologies based on the Bayesian model averaging method (BMA). A study of low-level and mid-level features in convolutional neural networks (CNNs) for facial expression identification was undertaken by Nguyen et al. [27]. They suggested a model that included three different forms of mid-level links as part of an ensemble. The results revealed that the ensemble model can achieve a final classification accuracy of 74.09% in facial representation. They also used a method for face emotion identification that concatenated feature vectors from three multi-level networks, followed by fully connected layers. For face expression recognition, Fan et al. [28] also presented an ensemble approach. Multi-region ensemble CNN is the approach employed in this study (MRE-CNN). The final recognition rate can be improved over the initial single network utilizing a weighted sum operation of the prediction scores from each subnetwork. The suggested work is assessed using the databases AFEW 7.0 and RAF-DB. The authors in [29] suggested a foreground extraction-based FER system based on an Xception model. The foreground extraction-based FER technique extracts the FER features reliably, whereas the system's deep learning model successfully uses these features for model training. The Xception model, utilized in this study, makes the most of the foreground extracted face images from the FER dataset and improves FER performance. In [30], a feature vector extraction technique is introduced, which integrates facial landmarks into facial image pixel values, while the deep learning model employs these combined features as input. The FER deep learning model efficiently uses facial landmark characteristics to classify user emotions with the least amount of classification error.

Compared to the previous approaches, when generating the final prediction for emotion classification, not only does the proposed ensemble network consider the probabilities of each class, but also the precision value of each classifier, giving more weight to the classifier that performs well throughout each run and thus improving performance. As the amount of data grows, processing such a vast amount of data becomes harder, which may have an impact on the model's performance. So, in the proposed approach, we used the SparkML [31] pipeline, which enhances the model's performance even more. It is common in machine learning to run a series of algorithms to process and learn from data. A pipeline in Spark ML is a process that consists of a series of pipeline stages (transformers and estimators) that must run in a specified order. The proposed model is described in the next section.

3. The Proposed Method

The proposed system is a precision-based weighted blending ensemble model, whose overall framework is shown in Figure 1.
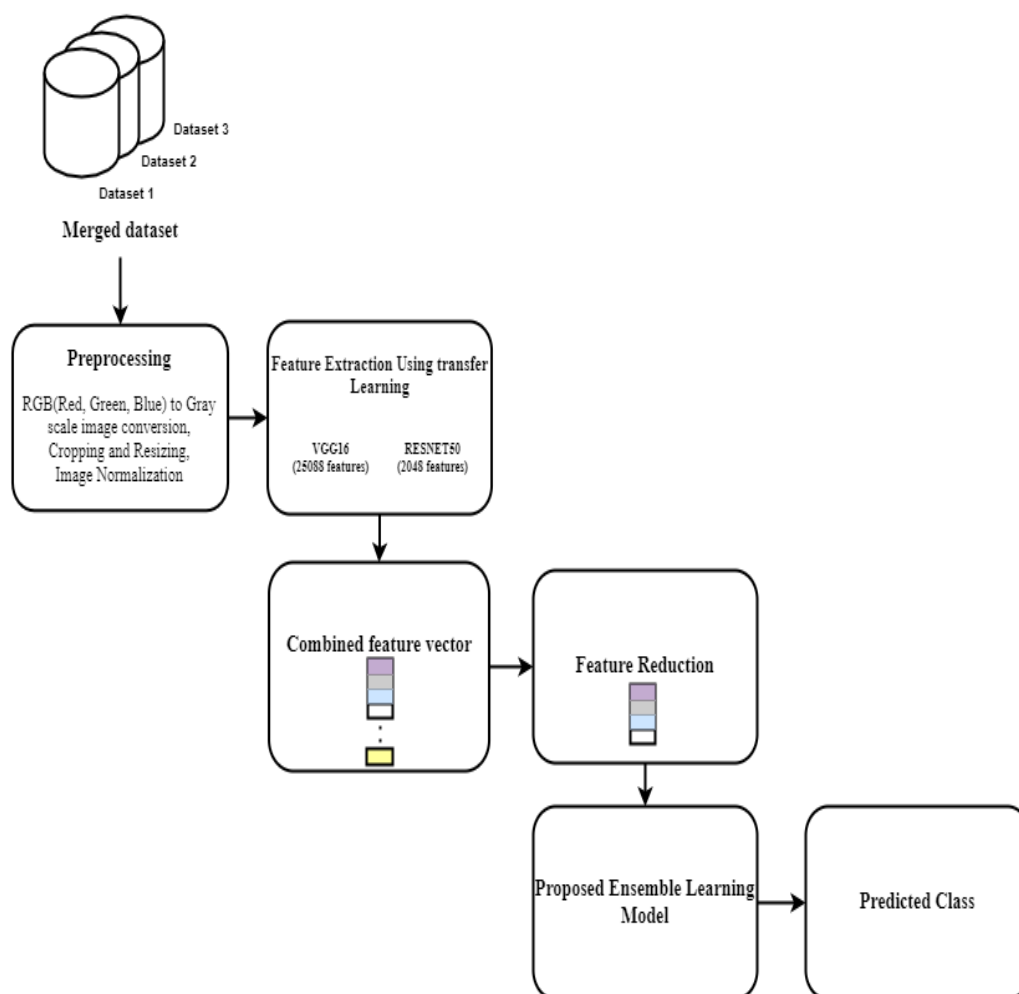
Figure 1. Overall framework of the proposed model.

3.1. Dataset Description

The FER-2013 dataset was obtained from the kaggle data repository [32]. This dataset was created by performing a Google image search of each emotion and synonyms of the emotions. The dataset consists of 48 × 48 pixel gray scale images of 35,685 example faces. The training set consists of 28,709 examples and the test set consists of 3589 examples. Each face is more or less centered, since all faces were automatically registered, occupying the same amount of space in each image. Each image is labeled as: happy, sad, angry, afraid, surprise, disgust and neutral, representing each of the seven emotions, while happy seems to be the most prevalent emotion.

The extended Cohn–Kanade (known as CK+ [33]) facial expression dataset consists of seven expressions. It considers both posed and non-posed expressions. The CK+ comprises a total of 593 sequences from a total of 123 subjects. They are recorded at 30 frames per second (FPS) with a resolution of either 640 × 490 or 640 × 480 pixels. Out of these sequences, 327 are labelled with seven expression classes: anger, contempt, disgust, fear, happiness, sadness, and surprise.

FERG-DB dataset [34] contains 55,767 annotated face images of six stylized characters. The characters were modeled using MAYA. The images for each character are grouped into seven types of expressions: anger, contempt, disgust, fear, neutral, sadness and surprise.

3.2. Preprocessing

In the preprocessing stage, images from the three datasets, FER-2013 [32], CK+ [33] and FERG-DB [34], are combined into a single dataset following the data preparation methods.

FER-2013 and CK+ consist of gray scale images, whereas the FERG-DB dataset consists of RGB images. Thus, the FERG-DB dataset was converted into gray scale images. The images in the datasets have mismatching image sizes which possibly creates problems. So, in the next stage, all images from these datasets were cropped correctly and resized. The last stage involved image normalization. Moreover, the provided datasets contained different classes of emotions. In particular, the FER-2013 dataset consists of seven expressions including anger, disgust, fear, happiness, sadness, surprise and neutral. The CK+ dataset consists of seven expressions including anger, disgust, fear, happiness, sadness, surprise and contempt. FERG-DB is a database of six stylized characters, namely: Ray, Malcolm, Jules, Bonnie, Mery and Aia, in which seven types of expressions are used to group each character's images. These are: anger, disgust, fear, joy, neutral, sadness and surprise. Out of these seven types, we have considered six emotions from all these three datasets: anger, disgust, fear, happiness, sadness and surprise. Each emotion class was mapped into integer labels in the range 0–5, in the following manner: 0—anger, 1—disgust, 2—fear, 3—happy, 4—sad and 5—surprise.

3.3. Extracting Features Using Transfer Learning

Transfer Learning (TL) is a technique that builds a new model using the knowledge learned from previously trained models. To address complex issues, most deep learning and machine learning methods necessitate a large amount of data. Transfer learning techniques handle issues like having insufficient data for the new task. In this technique, we first train a base dataset, and then apply the learnt features to a target network. This strategy will work only if the characteristics utilized are generic, meaning they are appropriate for both the base and destination datasets. Machine learning and deep learning algorithms are typically trained to solve specific problems. As a result, if the feature space distribution changes, the models must be rebuilt from the ground up, or they may suffer a considerable performance loss, if not outright failure. This solitary learning paradigm is overcome through transfer learning.

Transfer learning can be used as a feature extractor in deep learning. Deep learning models have a layered architecture and learn different features at different layers. These layers are then finally connected to a last layer (a fully connected layer) to get the final output. In transfer learning, a pre-trained network without its final layer is utilized as a fixed feature extractor for other tasks. The key idea is that in the case of training the model with new data for a new task, the pre-trained model's weighted layers are used for feature extraction rather than updating the weights of the model's layers. That is, it allows the extraction of features from a new domain task and the utilization of the knowledge from a source-domain task.

Spark ML provides fast, distributed implementations of common learning algorithms. In our case, we need to perform transfer learning on a huge set of images. So, to facilitate quick transfer learning, we used the parallel processing power, the Apache Spark. The pre-trained CNN, VGG16 and ResNet50 models, which are two pre-trained architectures on ImageNet, were used in this study. The transformer peels off the final layer of a pre-trained neural network and uses the output from all previous layers as features in the classification method. The number of features given as output of VGG16 and RESNET 50 constitute the input to classifiers after applying a feature reduction approach (please see Section 3.4). For classification, a precision-based ensemble model is used.

3.4. Feature Reduction

The amount of features that transfer learning generates is enormous. The "curse of dimensionality" may affect the performance of machine learning algorithms. So, using any dimensionality reduction method, we should lower the amount of input characteristics. For dimensionality reduction, we employed principal component analysis (PCA) in this study. PCA is a projection-based method for reducing dimensionality by transforming a big collection of features into a smaller one that contains the majority of the information

from the larger set. The number of features given as output of VGG16 and RESNET 50 are 25,088 and 2048, respectively. The combined feature vector of size 27,136 is given as input to PCA to get a reduced feature vector.

3.5. Precision Based Weighed Blending Ensemble Learning

Ensemble approaches generate several models, which are then combined to get better results. Ensemble approaches are capable of producing more precise results than a single model. The production of base learners is the initial stage in the ensemble model. The production of base learners can be done in two ways: sequential or parallel. The base learners are generated sequentially in the sequential technique, while they are generated in parallel in the parallel method. The selection of component classifiers based on their efficiency is a crucial stage in the ensemble technique. The second key stage in ensemble learning is to combine the predictions of individual classifiers, which is done using a variety of strategies.

The proposed method selects a group of classifiers and applies them to a dataset, then combines their separate predictions using a new weighted methodology. Logistic regression (LR), decision trees, and naïve Bayesian classifiers were employed in this study. LR will not get caught in a local minimum because it uses a convex loss function. The logistic regression model is a simple, rapid, and straightforward classification method. The parameters are used to explain the degree and direction of the independent variable's influence on the dependent variable. It can also be used to classify multiple classes. A decision tree is a tree-based method for dealing with problems such as regression and classification. An inverted tree with branches spreading off from a homogeneous probability distributed root node to severely heterogeneous leaf nodes is used to produce the result. Regression trees are used for dependent variables with continuous values, while classification trees are used for dependent variables with discrete values. The decision tree enables automatic feature interaction, while also being faster. Decision trees are more successful when there are a large number of categorical characteristics. A naïve Bayes classifier, in basic words, posits that the existence of one feature in a class is independent to the presence of any other feature. The text categorization industry is the primary focus of naïve Bayes. It is mostly used for grouping and classification purposes, and it is based on the conditional chance of something happening.

As shown in Figure 2, we employed a weighted blending-based ensemble technique to classify emotions. As a result, the emotion dataset was split into two parts: training and testing. Out of the whole dataset, 80% was taken for training and 20% for testing. The validation set is made up of a small piece of the training set. The features obtained from transfer learning and the labels on emotion face images were fed into the three classifiers, resulting in three trained models. The validation set is given as input to these models in the second stage. The precision value of each model is calculated based on the predictions provided by the three models on the validation set. The weighted precision value of a classifier in a multiclass classification problem can be calculated by taking the precision for each class label and finding their average weighted by a support factor, which can be the number of true instances for each class label. The weighted precision value of a classifier in a multiclass classification problem is shown below in Equation (1) [35]:

$$\frac{1}{\sum s\varepsilon S|\hat{y}_l|} \sum s\varepsilon S|\hat{y}_l| P(y_j, \hat{y}_j) \tag{1}$$

where,

Y is the set of predicted (sample, label) pairs,

$\hat{y}$ is the set of true (sample, label) pairs,

$\hat{y}_j$ is the subset of $\hat{y}$,

y_j the subset of y with label j,

S is the set of labels, and

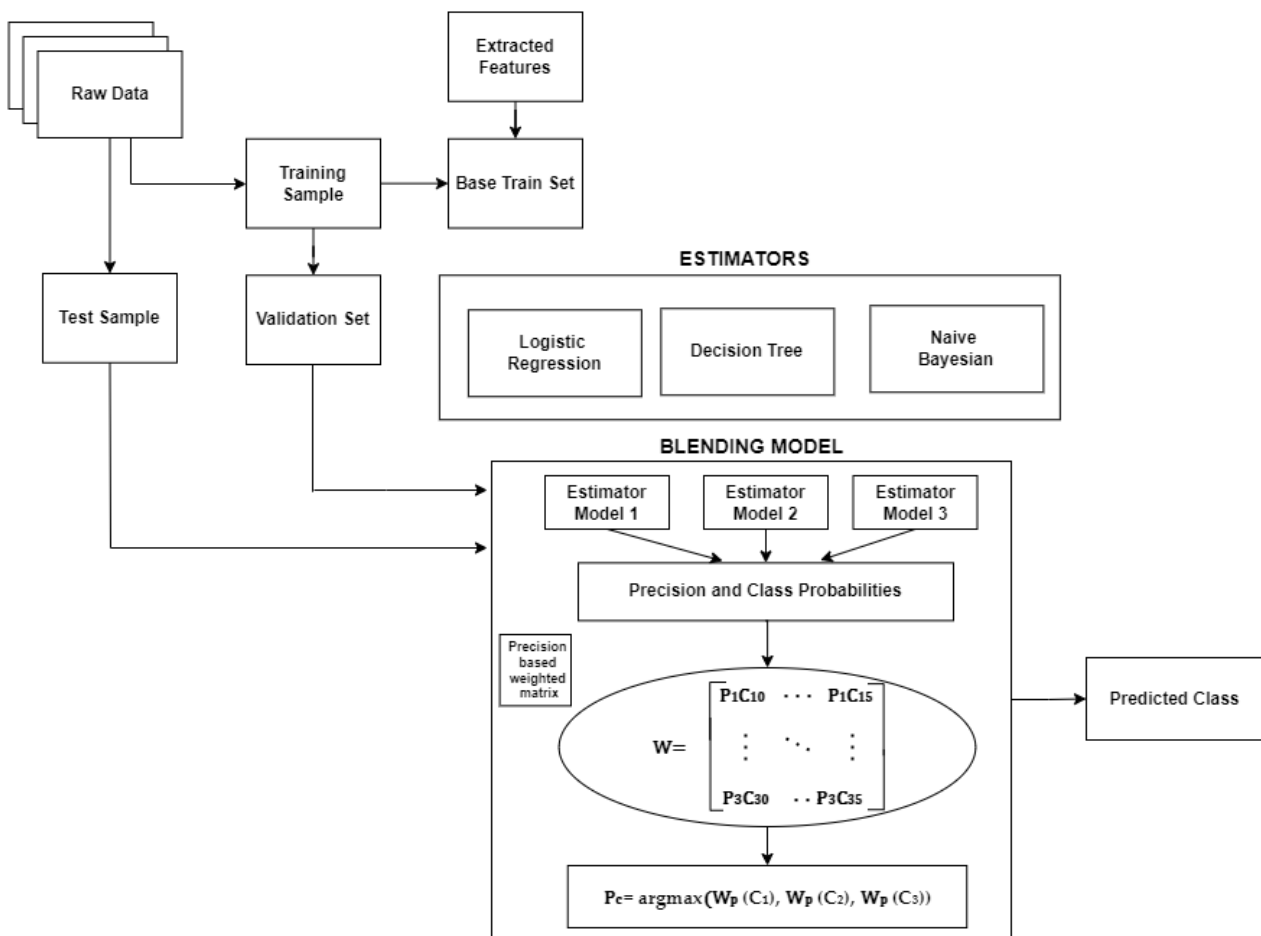$P(Q, W) = \frac{|Q \cap W|}{|W|}$ for some sets Q and W.

Figure 2. Precision-based blending ensemble model.

In the next stage, the weighted precision value that we found earlier is taken as input. This step uses a weight matrix, as shown below in Equation (2):

$$W = \begin{bmatrix} P_1C_{10} & \cdots & P_1C_{15} \\ \vdots & \ddots & \vdots \\ P_3C_{30} & \cdots & P_3C_{35} \end{bmatrix} \tag{2}$$

where, P_i value represents the precision value of the classifier i, and C_{ij} represents the predicted class probability of class j by classifier i. Then, the average of each column is taken, which gives the weighted average probability of each class. The class of a sample is identified, based on this weighted average probability. If for a sample, class K has the highest weighted average probability, then we classify the sample as class K. The predicted class P_c is given in Equation (3).

$$P_c = \text{argmax}(W_p\,(C_1),\ W_p\,(C_2),\ W_p\,(C_3)) \tag{3}$$

where, W_p is the weighted average probability values of each column c_i of the precision weight matrix W.

During the last stage, the blending model is fit on the validation set and predictions are made on the test set. The pseudo code for the proposed ensemble model is shown in Figure 3.

Pseudo code of blending ensemble model:

Input: D_t -> **Training Dataset**

$\quad\quad S_t$ -> **Test Dataset**

Step 1: Split D_t into base train set *B* and validation set *V*.

Step 2: Learn base models BM_t, where *t*=3 indicates number of base models.

$\quad\quad$ **for *t*=1 to 3 do**

$\quad\quad\quad\quad$ **Learn BM_t based on *B***

$\quad\quad$ **end for**

Step 3: Make prediction on *V* using all the BM_t models

$\quad\quad$ **for *i*=1 to 3 do**

$\quad\quad\quad\quad$ P_i = **Tranform *V* based on BM_t**

$\quad\quad$ **end for**

Step 4: Calculate the precision value for each model BM_t using (1).

Step 5: Calculate the weight matrix *W* using (2).

Step 6: Calculate weighted average probability values of of each column c_i and save in vector *R*.

$\quad\quad$ **R<-[ ]**

$\quad\quad$ **for *i*=1 to 3 do**

$\quad\quad\quad\quad$ **C<-0**

$\quad\quad\quad\quad$ **for *j*=1 to 3 do**

$\quad\quad\quad\quad\quad\quad$ **C=C+W[*j*][*i*]**

$\quad\quad\quad\quad$ **end for**

$\quad\quad\quad\quad$ **R<- R+[C]**

$\quad\quad$ **end for**

Step 7: Take argmax(*R*), it returns the prediction as in (3).

Step 8: Ensemble model *M* is build using steps 2–7.

Step 9: Fit M on validation set *V*.

Step 10: Make predictions on test set S_t using the learned model *M*.

Output: Predicted class label *y* of a test sample.

Figure 3. Pseudo code for the proposed ensemble model.

3.6. Considerations to Explainability Issues of the Proposed Ensemble Model

The use of different models to combine judgments has helped to improve the overall performance of the model by lowering noise, bias, and variance. To increase accuracy, the classifiers of logistic regression, which has a high bias and low variance, decision tree, which has a low bias and high variance, especially, in large datasets, and naïve Bayesian, which has a high bias and low variance, are utilized. In particular, the naïve Bayes classifier learns its parameters by explicitly calculating them, rather than iteratively modifying them to minimize a loss function via gradient descent, as the vast majority of machine learning models do. Due to this, the naïve Bayes classifier trains faster than logistic regression; the basic counting and calculation of its parameters takes far less time than gradient descent. The ensemble's ability to explain the final model decision is experimentally reduced, but it has better generalizability.

4. Results

4.1. Classification Accuracy and Confusion Matrix on the FER-2013 Dataset

In the FER-2013 dataset, we considered images of six expressions, namely, anger, disgust, fear, happiness, sadness and surprise. The comparison between the proposed ensemble method and some of the previous works on the FER-2013 dataset is provided in Table 1. As shown in Table 1, the proposed approach reached 76.2% classification accuracy and has a 0.4% accuracy improvement, compared to the previous works on the FER-2013 dataset. The proposed classification method outperforms, in terms of accuracy, the previously proposed classification methods on the same dataset. Figure 4 shows the proposed ensemble model confusion matrix on the FER-2013 dataset. The model shows best classification on the "happiness" and "sad" emotion. On the other hand, it makes the most mistakes when classifying on the "disgust" and "surprise" emotion. In the models, the classification error is higher for classes such as disgust and surprise, because they have fewer samples.

Table 1. Classification accuracy on the FER-2013 dataset.

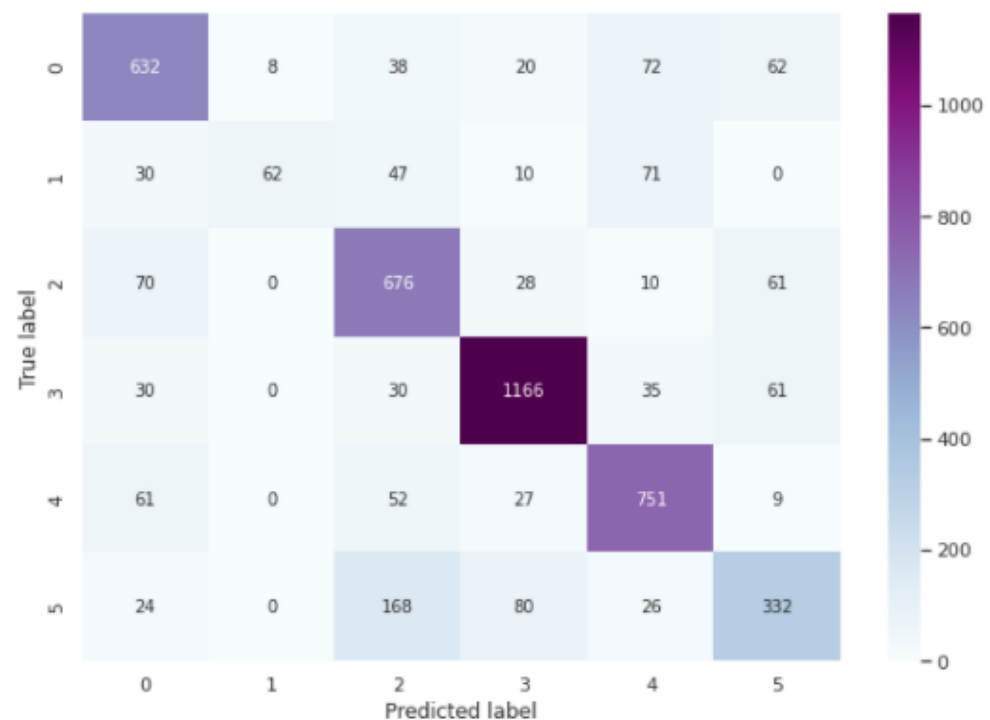| SNO	Classification Model	Accuracy %
1	Proposed method (weighted ensemble model)	76.2
2	Ensemble CNN [36]	75.8
3	Ensemble CNN [37]	75.2
4	LHC-Net [38]	74.42
5	VGG [37]	72.70
6	Resnet [37]	72.40
7	Inception [37]	71.60
8	ARM [39]	71.38
9	CNN + SVM [40]	71.20
10	Attentional ConvNet [41]	70.02
11	GoogLeNet [42]	65.20

Figure 4. Confusion matrix on the FER-2013 dataset.

4.2. Classification Accuracy and Confusion Matrix on the CK+ Dataset

In the CK+ dataset, we considered images of six expressions (anger, disgust, fear, happiness, sadness and surprise). The comparison between the proposed ensemble method and some of the previous works on the CK+ dataset is provided in Table 2. In Table 2, it can be seen that the proposed approach reached 99.4% classification accuracy and has a 1.17% accuracy improvement, compared to the previous works on the CK+ dataset. Figure 5 shows the confusion matrix of the proposed ensemble model on the CK+ dataset. In the models, the classification error is higher for classes such as "fear", because it has fewer samples.

Table 2. Classification accuracy on the CK+ dataset.

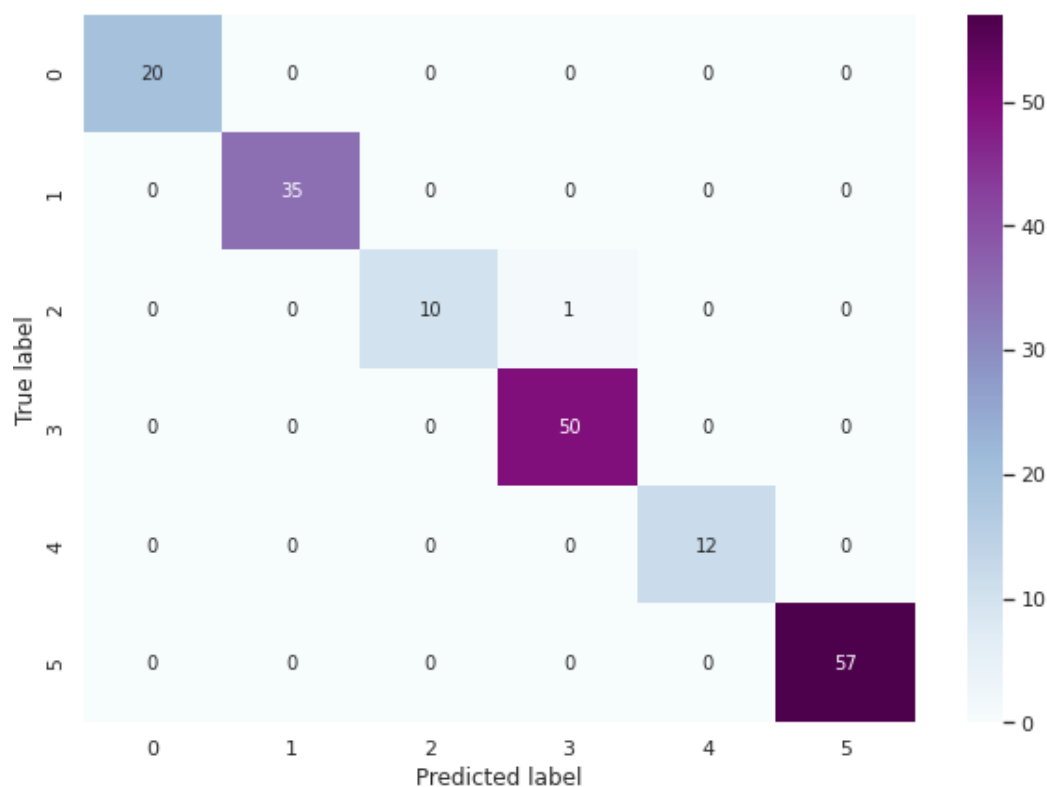| SNO	Classification Model	Accuracy %
1	Proposed method (weighted ensemble model)	99.4
2	Using Self-Supervised Auxiliary Tasks to Improve Fine-Grained Facial Representation [43]	98.23
3	FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition [44]	98.6
4	DTAGN [45]	97.2
5	IACNN [46]	95.37
6	IB-CNN [47]	95.1

Figure 5. Confusion matrix on the CK+ dataset.

4.3. Classification Accuracy and Confusion Matrix on the FERG-DB Dataset

The comparison between the proposed ensemble method and some of the previous works on the FERG-DB dataset is provided in Table 3. The proposed approach reached 99.6% classification accuracy and has a 1.4% accuracy improvement compared to the previous works on the FERG-DB dataset. In Table 3, it is obvious that the proposed method outperforms, in terms of accuracy, the other artificial intelligent methods applied in the same dataset. Figure 6 shows the confusion matrix of the proposed ensemble model on the

FERG-DB dataset. In the models, the classification error is higher for classes such as "fear" and "happiness", since they have fewer samples compared to other classes.

Table 3. Classification accuracy on the FERG-DB dataset.

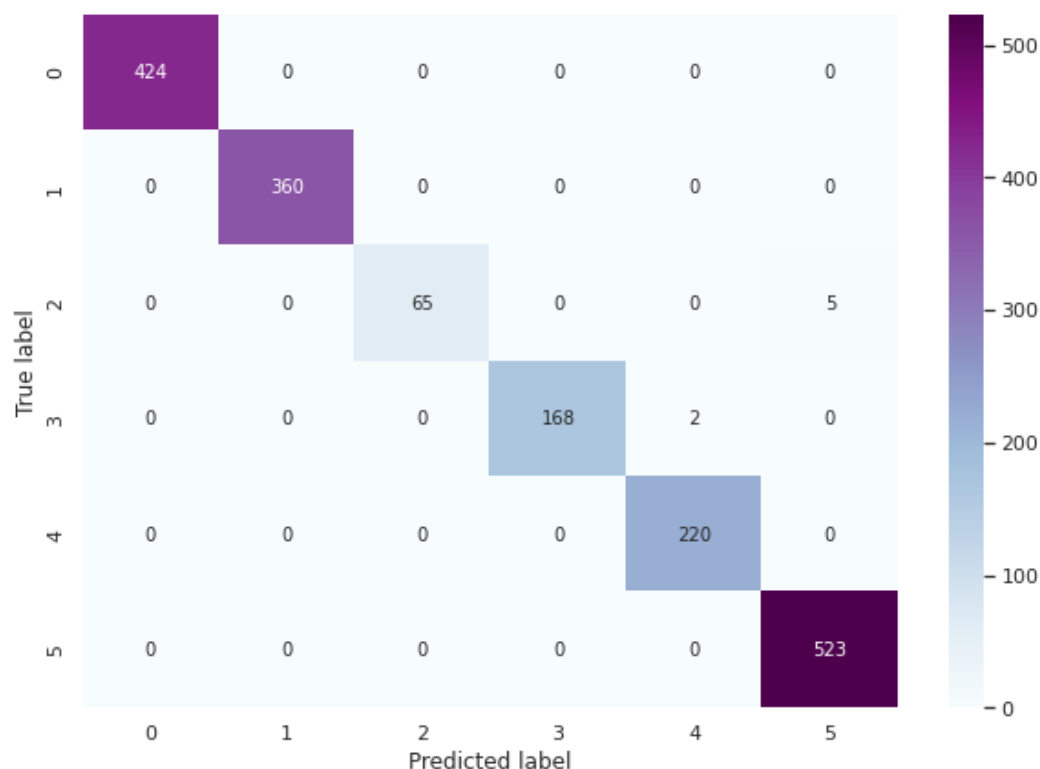| SNO	Classification Model	Accuracy %
1	Proposed method (weighted ensemble model)	99.6
2	Adversarial NN [48]	98.2
3	Ensemble Multi-feature [49]	97
4	DeepExpr [34]	89.02

Figure 6. Confusion matrix on the FERG-DB dataset.

Finally, it could be noticed that the classification accuracy values for the FER-2013 (Table 1) is far less than those values for the other datasets (Tables 2 and 3). The main reason for that is because the FER-2013 dataset suffers from extreme crowdsourcing as it comprises non-face photos, text images, drowsy faces, profile photographs, etc., that are discernible by humans, as well as a substantial number of incorrectly labeled images.

4.4. Accuracy Comparison Based on the Model Used for Transfer Learning on Each Dataset

Features were extracted by performing transfer learning using RESNET50, VGG16 and by combining features extracted from both the models. Experiments were done on all three datasets and the accuracy for each case was then calculated. Figure 7 illustrates the accuracy comparison for the three sets of features extracted (RESNET50, VGG16, combination of both) on the datasets. It can be seen that the accuracy is higher when a combined feature vector is used for transfer learning in the case of all three datasets.
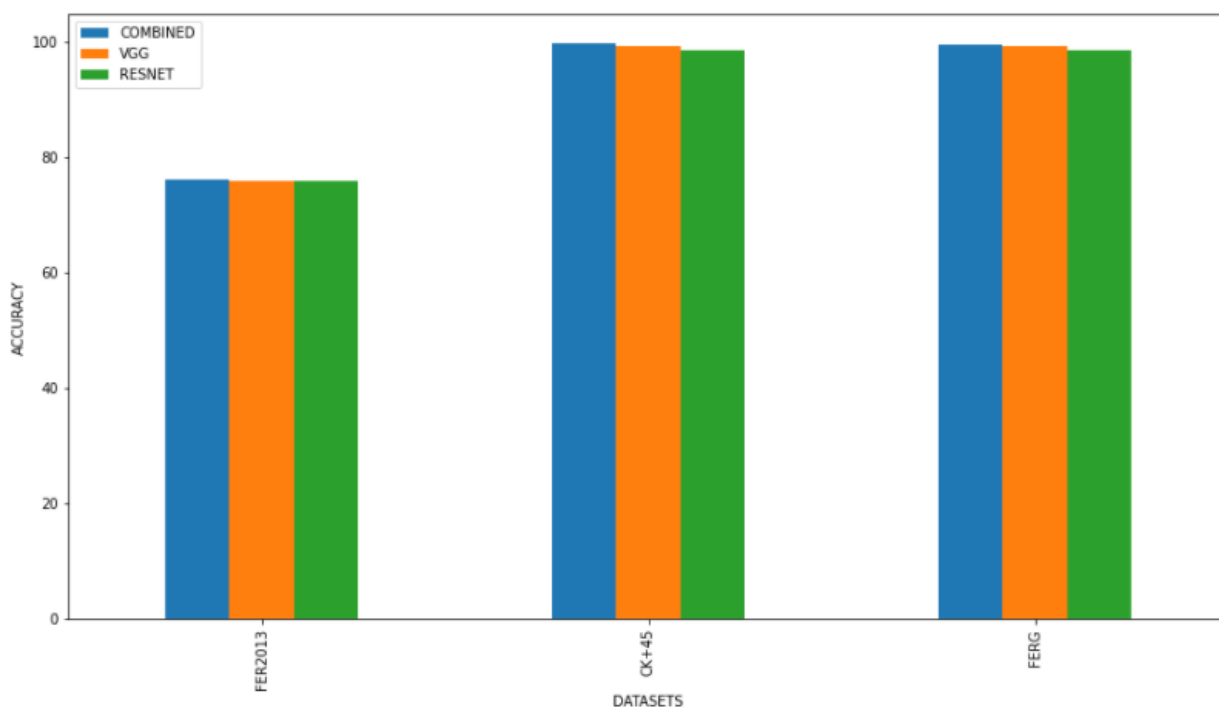
Figure 7. Accuracy comparison based on the model used for transfer learning.

4.5. Classification Accuracy of Individual Classifiers and the Proposed Approach on the Three Datasets and on the Combined Dataset

The three datasets, FER2013, CK+ and FERG-DB, were combined so that our proposed approach was also applied on the combined dataset. The three classifies used to form the ensemble classifier were logistic regression, naïve Bayesian and decision tree. Table 4 depicts the classification accuracy of the individual classifiers and the proposed ensemble model on the three datasets, as well as on the combined dataset.

Table 4. Classification accuracy of the individual classifiers on the three datasets.

SNO	Dataset	Logistic Regression Model	Naïve Bayesian	Decision Tree	Proposed Ensemble Model
1	FER 2013	74.2	75.46	76.1	76.2
2	CK+	98.9	99.12	99.38	99.4
3	FERG-DB	99.02	99.4	99.52	99.6
4	Combined	87.5	87.98	88.2	88.68

Figure 8 shows the confusion matrix of the proposed ensemble model on the combined dataset. The model runs the best classification on the "happiness" and "sadness" emotions. On the other hand, it has the poorest performance when classifying on the "disgust" and "surprise" emotions. In these models, the classification error is higher for classes such as "disgust" and "surprise" due to fewer samples.

Furthermore, the precision, recall and F1-score values of the proposed ensemble classifier on the provided three datasets are shown in Table 5.

Figure 8. Confusion matrix on the combined dataset.

Table 5. Precision, recall and F1-score values on the three datasets.

SNO	Dataset	Precision %	Recall %	F1-Score %
1	FER-2013	77.01	76.95	75.9
2	CK+	98.96	98.47	98.68
3	FERG-DB	99.5	99.57	99.5

5. Conclusions

For emotion categorization, a weighted blended distributed ensemble model is presented in this study. The proposed strategy is tested on three datasets (FER-2013, CK+, and FERG-DB), as well as on their combination. For emotion classification, the proposed network employs the SparkML pipeline approach. Transfer learning is utilized to extract features from emotion facial images. The proposed ensemble model network outperforms the competition as it considers not only the probabilities of each class, but also the precision value of each classifier, when generating the final prediction, thus giving greater weight to the classifier that performs well throughout each run. SPARK distributed computing is used in the proposed technique, so that it can work efficiently and effectively on any large dataset. In the intermediate step of the ensemble, the proposed approach can be extended by using any number of classifiers. Our future directions will focus on applying the model for real-time face emotion classification in video sequences.

Author Contributions: Conceptualization, M.V.J.; methodology, G.S. and M.V.J.; formal analysis and investigation, G.S., M.V.V. and M.V.J.; writing—original draft preparation, G.S. and M.V.V.; writing—review and editing, G.S., M.V.V., M.V.J., E.P. and V.C.G.; supervision: M.V.J., E.P. and V.C.G.; project administration, V.C.G. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Three datasets were analyzed in this study which are available upon request for non-commercial use. In particular, the FER-2013 dataset is available at: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data (accessed on 31 December 2021). The FERG-DB dataset is available at: http://grail.cs.washington.edu/projects/deepexpr/ferg-2d-db.html (accessed on 31 December 2021). The CK+ dataset is available at: http://www.jeffcohn.net/Resources/ (accessed on 31 December 2021).

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

References

1. SimilarNet: The European Taskforce for Creating Human–Machine Interfaces Similar to Human–Human Communication. Available online: http://www.similar.cc/ (accessed on 31 December 2021).
2. Kayalvizhi, S.; Kumar, S.S. A neural networks approach for emotion detection in humans. *IOSR J. Electr. Comm. Engin.* **2017**, 38–45.
3. Van den Broek, E.L. Ubiquitous emotion-aware computing. *Pers. Ubiquit. Comput.* **2013**, *17*, 53–67. [CrossRef]
4. Ménard, M.; Richard, P.; Hamdi, H.; Daucé, B.; Yamaguchi, T. Emotion Recognition based on Heart Rate and Skin Conductance. In Proceedings of the 2nd International Conference on Physiological Computing Systems, Angers, France, 13 February 2015; pp. 26–32.
5. Zheng, W.-L.; Lu, B.-L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev.* **2017**, *7*, 162–175. [CrossRef]
6. Ganaie, M.A.; Hu, M.; Tanveer, M.; Suganthan, P.N. Ensemble deep learning: A review. *arXiv* **2021**, arXiv:2104.02395.
7. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
8. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
9. Halevy, A.; Norvig, P.; Pereira, F. The unreasonable effectiveness of data. *IEEE Intell. Syst.* **2009**, *24*, 8–12. [CrossRef]
10. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [CrossRef]
11. Matilda, S. Emotion recognition: A survey. *Int. J. Adv. Comp. Res.* **2015**, *3*, 14–19.
12. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
13. Wiem, M.; Lachiri, Z. Emotion classification in arousal valence model using MAHNOB-HCI database. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 318–323.
14. Liu, Y.; Sourina, O. EEG Databases for Emotion Recognition. In Proceedings of the 2013 International Conference on Cyberworlds, Yokohama, Japan, 21–23 October 2013; pp. 302–309.
15. Bahari, F.; Janghorbani, A. EEG-based Emotion Recognition Using Recurrence Plot Analysis and K Nearest Neighbor Classifier. In Proceedings of the 2013 20th Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 18–20 December 2013; pp. 228–233.
16. Murugappan, M.; Mutawa, A. Facial geometric feature extraction based emotional expression classification using machine learning algorithms. *PLoS ONE* **2021**, *16*, e0247131.
17. Cheng, B.; Liu, G. Emotion Recognition from Surface EMG Signal Using Wavelet Transform and Neural Network. In Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE), Shanghai, China, 16–18 May 2008; pp. 1363–1366.
18. Kim, J.H.; Poulose, A.; Han, D.S. The extensive usage of the facial Image threshing machine for facial emotion recognition performance. *Sensors* **2021**, *21*, 2026. [CrossRef] [PubMed]
19. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
20. Rivera, A.R.; Castillo, J.R.; Chae, O.O. Local directional number pattern for face analysis: Face and expression recognition. *IEEE Trans. Image Process.* **2013**, *22*, 1740–1752. [CrossRef] [PubMed]
21. Moore, S.; Bowden, R. Local binary patterns for multi-view facial expression recognition. *Comput. Vis. Image Underst.* **2011**, *115*, 541–558. [CrossRef]
22. Happy, S.L.; George, A.; Routray, A. A Real Time Facial Expression Classification System Using Local Binary Patterns. In Proceedings of the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), Kharagpur, India, 27–29 December 2012; pp. 1–5.
23. Ghimire, D.; Jeong, S.; Lee, J.; Park, S.H. Facial expression recognition based on local region specific features and support vector machines. *Multimed. Tools Appl.* **2017**, *76*, 7803–7821. [CrossRef]
24. Hammal, Z.; Couvreur, L.; Caplier, A.; Rombaut, M. Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. *Int. J. Approx. Reason.* **2007**, *46*, 542–567. [CrossRef]
25. Devi, M.K.; Prabhu, K. Face Emotion Classification Using AMSER with Artificial Neural Networks. In Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 148–154.
26. Corchs, S.; Fersini, E.; Gasparini, F. Ensemble learning on visual and textual data for social image emotion classification. *Int. J. Mach. Learn. Cyber.* **2019**, *10*, 2057–2070. [CrossRef]

27. Nguyen, H.D.; Yeom, S.; Lee, G.S.; Yang, H.J.; Na, I.S.; Kim, S.H. Facial emotion recognition using an ensemble of multi-level convolutional neural networks. *Int. J. Pattern Recognit. Artif. Intell.* **2019**, *33*, 1940015. [CrossRef]
28. Fan, Y.; Lam, J.C.K.; Li, V.O.K. Multi-region Ensemble Convolutional Neural Network for Facial Expression Recognition. In Proceedings of the 27th International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp. 84–94.
29. Poulose, A.; Reddy, C.S.; Kim, J.H.; Han, D.S. Foreground Extraction Based Facial Emotion Recognition Using Deep Learning Xception Model. In Proceedings of the 12th International Conference on Ubiquitous and Future Networks (ICUFN), Jeju Island, Korea, 17–20 August 2021; pp. 356–360.
30. Poulose, A.; Kim, J.H.; Han, D.S. Feature Vector Extraction Technique for Facial Emotion Recognition Using Facial Landmarks. In Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 20–22 October 2021; pp. 1072–1076.
31. Meng, X.; Bradley, J.; Yavuz, B.; Sparks, E.; Venkataraman, S.; Liu, D.; Freeman, J.; Tsai, D.B.; Amde, M.; Owen, S.; et al. MLlib: Machine learning in Apache Spark. *J. Mach. Learn. Res.* **2016**, *17*, 1235–1241.
32. Challenges in Representation Learning: Facial Expression Recognition Challenge. Available online: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data (accessed on 31 December 2021).
33. 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 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
34. Aneja, D.; Colburn, A.; Faigin, G.; Shapiro, L.; Mones, B. Modeling Stylized Character Expressions via Deep Learning. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 136–153.
35. Metrics and Scoring: Quantifying the Quality of Predictions. Available online: https://scikit-learn.org/stable/modules/model_evaluation.html#precision-recall-f-measure-metrics (accessed on 31 December 2021).
36. Khanzada, A.; Bai, B.; Celepcikay, F.T. Facial expression recognition with deep learning. *arXiv* **2020**, arXiv:2004.11823.
37. Pramerdorfer, C.; Kampel, M. Facial expression recognition using convolutional neural networks: State of the art. *arXiv* **2016**, arXiv:1612.02903.
38. Pecoraro, R.; Basile, V.; Bono, V.; Gallo, S. Local multi-head channel self-attention for facial expression recognition. *arXiv* **2021**, arXiv:2111.07224.
39. Shi, J.; Zhu, S.; Liang, Z. Learning to amend facial expression representation via de-albino and affinity. *arXiv* **2021**, arXiv:2103.10189.
40. Tang, Y. Deep learning using linear support vector machines. *arXiv* **2013**, arXiv:1306.0239.
41. Minaee, S.; Minaei, M.; Abdolrashidi, A. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors* **2021**, *21*, 3046. [CrossRef]
42. Giannopoulos, P.; Perikos, I.; Hatzilygeroudis, I. Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013. In *Advances in Hybridization of Intelligent Methods*; Springer: Cham, Switzerland, 2018; pp. 1–16.
43. Pourmirzaei, M.; Montazer, G.A.; Esmaili, F. Using self-supervised auxiliary tasks to improve fine-grained facial representation. *arXiv* **2021**, arXiv:2105.06421.
44. Ding, H.; Zhou, S.K.; Chellappa, R. FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 118–126.
45. Han, Z.; Meng, Z.; Khan, A.-S.; Tong, Y. Incremental boosting convolutional neural network for facial action unit recognition. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 109–117.
46. Meng, Z.; Liu, P.; Cai, J.; Han, S.; Tong, Y. Identity-aware Convolutional Neural Network for Facial Expression Recognition. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 558–565.
47. Jung, H.; Lee, S.; Yim, J.; Park, S.; Kim, J. Joint Fine-tuning in Deep Neural Networks for Facial Expression Recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2983–2991.
48. Clément, F.; Piantanida, P.; Bengio, Y.; Duhamel, P. Learning anonymized representations with adversarial neural networks. *arXiv* **2018**, arXiv:1802.09386.
49. Hang, Z.; Liu, Q.; Yang, Y. Transfer Learning with Ensemble of Multiple Feature Representations. In Proceedings of the IEEE 2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA), Kunming, China, 13–15 June 2018; pp. 54–61.