

Article

Improving Facial Emotion Recognition Using Residual Autoencoder Coupled Affinity Based Overlapping Reduction

Sankhadeep Chatterjee ¹, Asit Kumar Das ¹, Janmenjoy Nayak ² and Danilo Pelusi ^{3,*}

¹ Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711101, India; chatterjeesankhadeep.cu@gmail.com (S.C.); akdas@cs.iests.ac.in (A.K.D.)

² Department of Computer Science, Maharaja Sriram Chandra Bhanja Deo (MSCBD) University, Baripada 757003, India; mailforjnayak@gmail.com

³ Department of Communications Sciences, University of Teramo, 64100 Teramo, Italy

* Correspondence: dpelusi@unite.it

Abstract: Emotion recognition using facial images has been a challenging task in computer vision. Recent advancements in deep learning has helped in achieving better results. Studies have pointed out that multiple facial expressions may present in facial images of a particular type of emotion. Thus, facial images of a category of emotion may have similarity to other categories of facial images, leading towards overlapping of classes in feature space. The problem of class overlapping has been studied primarily in the context of imbalanced classes. Few studies have considered imbalanced facial emotion recognition. However, to the authors' best knowledge, no study has been found on the effects of overlapped classes on emotion recognition. Motivated by this, in the current study, an affinity-based overlap reduction technique (AFORET) has been proposed to deal with the overlapped class problem in facial emotion recognition. Firstly, a residual variational autoencoder (RVA) model has been used to transform the facial images to a latent vector form. Next, the proposed AFORET method has been applied on these overlapped latent vectors to reduce the overlapping between classes. The proposed method has been validated by training and testing various well known classifiers and comparing their performance in terms of a well known set of performance indicators. In addition, the proposed AFORET method is compared with already existing overlap reduction techniques, such as the OSM, ν -SVM, and NBU methods. Experimental results have shown that the proposed AFORET algorithm, when used with the RVA model, boosts classifier performance to a greater extent in predicting human emotion using facial images.

Keywords: facial emotion recognition; class overlapping; residual networks; variational autoencoders



Citation: Chatterjee, S.; Das, A.K.; Nayak, J.; Pelusi, D. Improving Facial Emotion Recognition Using Residual Autoencoder Coupled Affinity Based Overlapping Reduction. *Mathematics* **2022**, *10*, 406. <https://doi.org/10.3390/math10030406>

Academic Editors: Christophe Guyeux and Daniel-Ioan Curiac

Received: 1 December 2021

Accepted: 24 January 2022

Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human emotion identification is a growing area in the field of Cognitive Computing that incorporates facial expression [1], speech [2], and texts [3]. Understanding human feelings is the key to the next era of digital evolution. Recent developments in the field have realized its potential in fields such as mental health [4], intelligent vehicles [5], and music [6]. Recognizing emotions from facial expressions is a trivial task for the human brain, but it associates a higher level of complexity when carried out using machines. The reason for this intricacy is the non-verbal nature of the communication that is enacted through facial cues. Emotion prediction through other forms of data sources such as texts are comparatively easier tasks because of the word-level expressions that can be easily annotated through hashtags or word dictionaries [7–9].

Emotion recognition through facial images has been comprehensively studied in the last decade. The studies conducted in the recent years are mostly focused on the application of Deep Neural models. This is mostly because of the variance in the real-world sets. In [10], the use of two residual layers (each composed of four convolutional layers, two short-connection, and one skip-connection) with traditional Convolutional

Neural Networks (CNNs) resulted in an average enhancement in performance of 94.23% accuracy. Lin et al. [11] proposed a model utilizing multiple CNNs and utilized an improved Fuzzy integral to find out the optimal solution among the ensemble of CNNs. Facial Emotion Recognition has also been utilized in medical applications. Specifically, Facial Emotion analysis has been mostly utilized in psychiatric domains such as Autism and Schizophrenia. Sivasangari et al. [12] illustrated an IoT-based approach to understand patients suffering from Autism Spectrum Disorder (ASD) by integrating facial emotions. Their framework is built to monitor the patients and is equipped to propagate information to the patient's well-wisher. The emotion identification module developed using a Support Vector Machine is designed to help the caretaker to understand the emotional status of the subject. Jiang et al. [13] proposed an approach to identify subjects with ASD by utilizing facial emotions detected using an ensemble model of decision trees. Their approach was found to be 86% accurate in the appropriate classification of subjects. One study by Lee et al. [4] performed emotional recognition on 452 subjects (with 351 patients with schizophrenia and 101 healthy adults). Facial Emotion Recognition Deficit (FERD) is a common deficit found in patients with Schizophrenia. In [14], the authors highlighted the drawbacks of FERD screeners and proposed an ML-FERD screener to undertake a concrete discrimination between Schizophrenia patients and healthy adults. The ML-FERD framework was built using an Artificial Neural Network (ANN) and trained using 168 images. Their approach demonstrated a high True Positive Rate (TPR) and True Negative Rate (TNR). Recent studies have also focused on the emotion inspection from videos. Hu et al. [15] concentrated their study on extracting facial components from a video sequence. The authors developed a model that modifies Motion History Image (MHI) by understanding the local facial aspects from a facial sequence. One interesting approach proposed by Gautam and Thangavel [16] trains the CNN with 3000 facial images using an iterative optimization and tested the model on a video of American Prison. The primary interest of the authors was to develop an automated prison surveillance system, and the proposed approach recorded an average accuracy of 93.5% over the video tests. Haddad et al. [17] tried to preserve the temporal aspect of video sequences by using a 3D-CNN architecture and optimized it using a Tree-structured Parzen Estimator. Another approach called Contrastive Adversarial learning [18] was recently proposed by Kim and Song to perform a person-independent learning by capturing the emotional change through adversarial learning. Their approach resulted in reliable results on video sequence data. Auto-encoder networks in emotion recognition has also been accentuated in recent years [19]. In 2018, two studies [20,21] addressed the problem of computational complexity in Deep Networks and proposed a Deep Sparse Autoencoder Network (DSAN) to re-construct the images and integrated it with a softmax classifier capable of sorting out seven emotional categories that can be determined from the faces. Convolutional Autoencoders were found to be useful in continuous emotion recognition from images [22]. One approach using Generative Adversarial Stacked Convolutional Autoencoders was illustrated by Ruiz-Gracia et al. [23] in the context of Emotion Recognition. The pose and illumination invariant model was found to achieve 99.6% accuracy on a bigger image dataset. Sparse autoencoders were also explored with Fuzzy Deep Neural Architectures by Chen et al. [24]. The authors obtained reliable results on three popular datasets by applying a 3-D face model using Candide3. In another recent work by Lakshmi and Ponnusamy [25], the authors used Support Vector Machine (SVM) with Deep Stacked Autoencoder (DSAE) to predict the emotions from facial expressions. The pre-processing approach proposed by the authors is developed on a spatial and texture information extraction using a Histogram of Oriented Gradients (HOG) and a Local Binary Pattern (LBP) feature descriptor. Multimodal applications in emotion recognition have also been explored with autoencoders. In [26], the authors developed a novel autoencoder-based framework to integrate visual and audio signals and classified emotions using a two-layered Long Short-Term Memory network. Label distribution learning has been explored in [27,28] for chronological age estimation from human facial images.

1.1. Motivation

The class overlapping problem is well-known in the research community, however, very few research works have addressed it. The majority of research work focuses on the effects of class overlapping in the presence of imbalanced classes. Apart from these, few domain-specific works have been reported. The class overlapping problem in the context of face recognition has been studied in [29]. The proposed method used Fisher's Linear Discriminant combat majority biased face recognition; however, in the presence of overlapping classes, a new distance-based technique has been proposed. The study also pointed out the challenges in learning overlapped classes by various classifiers such as ANNs. Fuzzy rules have been used to address the same [30], where both imbalanced and overlapped classes are learned. The fuzzy membership values of data points have been used to partition the data points into several fuzzy sets. Batista et al. [31] found that classifiers may find difficulty in learning imbalanced classes in presence of overlapped classes, especially the minority classes. Similar studies [32,33] have also pointed out this issue where the performance of classifiers have been tested by varying the degree of overlapping. Another study [34] reported the effect of overlapped classes, where the overlapping region has majorly occupied minority samples. It has been found that the presence of overlap makes class-biased learning difficult. Later, Garcia et al. [35] studied the problem in detail and recorded the effects of overlapping classes in the presence of overlapping. It has been reported that the imbalance ratio might not be the primary cause behind the dramatic degradation of the classifier, whereas overlapped classes play a vital role. It established the fact that class overlapping is more important to classifier performance than class imbalance. Lee et al. [36] proposed an overlap sensitive margin classifier by taking the leverage of fuzzy support vector machines and k-nearest neighbor classifiers. The degree of overlap for individual data points are then calculated using the KNN classifier and used in a modified objective function to train the fuzzy SVM in order to split the data space into two regions, known as the Soft overlap and Hard overlap regions. Devi et al. [37] adopted a similar approach, where a ν -SVM was used as one class classifier to identify novel data instances from a dataset. However, the explicit detection of data points in an overlapping region is not reported. Neighborhood-based strategies have also been employed to undersample data points in the overlapping region and subsequently removing those data points to improve classifier performance [38].

1.2. Contribution

In the context of emotion recognition, the effect of class overlapping has not been precisely addressed. The challenge of overlapped classes appear as studies have revealed [39] that the presence of multiple facial expression is common in humans. Hence, facial images categorized in a particular class may have close similarity with other categories, which leads to the severe overlapping of classes. In order to address this problem, in the current study, a residual variational autoencoder (RVA) has been used to represent a facial image in latent space. After training the RVA model, only the encoder part transforms the images of all classes to a latent vector form. Now, to overcome the overlapped classes, an affinity-based overlap reduction technique (AFORET) has been proposed in the current article. The proposed method reduces the overlapping of classes in latent space. After modifying the dataset, it has been used to train a wide range of well-known classifiers. The performances of the classifiers have been tested by using well-known performance indicators. A thorough comparative analysis has been conducted to understand how the degree of overlap affects the classifiers' performance. The ingenuity of the proposed algorithm has been compared with the OSM [36], ν -SVM [37], and Neighborhood Undersampling (NBU) techniques, which have also attempted to address the overlapping problem in general. Overall the contributions of the current study are as follows:

1. To address the overlapped classes in emotion recognition, an affinity-based class overlapping reduction technique has been proposed.

2. An affinity-based metric is used to identify the data points in overlapping regions. Unlike previous methods [37,38], affinity values of data points provide a better understanding of whether a data point belongs to an overlapping region or not.
3. As it is evident from the work described in [36] that the removal of data points from the initial dataset is essential to improve classifier performance, hence, a similar approach is also adopted in the current study. However, it may be noted that the removal of too many data points from the original dataset may cause the classifier to improperly learn the underlying decision boundary. Thus, extensive analyses have been carried out in order to clearly understand how much data removal is optimal in the case of facial emotion recognition.

The rest of the article is arranged as follows: Section 2 introduces the residual variational autoencoder model, which is followed by the affinity-based overlap reduction technique. Next, in Section 4, these two methods are combined together to address the class overlapping problem in facial emotion recognition. Section 5 begins with a discussion on experimental setup, and the classifier and overlapping techniques are compared in terms of experimental performances. Finally, the conclusions are made in Section 6.

2. Residual Variational Autoencoder

Among various generative models, autoencoders are designed to transform inputs into a low-dimensional latent vector representation and transform them back to their original form. Such networks are trained in unsupervised mode in order to extract the most useful features of the input using unlabeled data [40]. A typical autoencoder consists of two components, viz., an encoder and a decoder. The encoder usually takes an input and eventually reduces its shape through a series of convolutional layers. The output of the encoder is a latent vector which can be passed to the decoder to reconstruct the original input. For every instance y^i in the training dataset $D = \{y^1, y^2, \dots, y^N\}$, where y^i and N represent the input vector of the i th sample and the number of instances, respectively. The encoding layer can be represented as:

$$l_i = f(y^i) = s_e(W_e y^i + b_e) \quad (1)$$

where $s_e(\cdot)$, W_e , and b_e represent the activation function, the weight matrix, and the bias vector of the encoding layer, respectively. In the same manner, the decoding layer can be defined as:

$$g(l^i) = s_d(W_d l^i + b_d) \quad (2)$$

where $s_d(\cdot)$, W_d , and b_d denote the activation function, the weight matrix, and the bias vector of the decoding layer, respectively. Hence, the output of the autoencoder for the instances can be defined as:

$$z = g(f(y^i)) \quad (3)$$

The Variational Autoencoders (VAE) have proved to be a major improvement while dealing with the feature representation capability [41]. The VAEs are generative models that are based on the Variational Bayes Inference [42] and combine deep neural networks which aim to regulate the encoding pattern during training so that the latent space has good properties to enable the process of the instance generation using a probabilistic distribution. The VAE has had many applications in the domain of image synthesis [43], video synthesis [44], and unsupervised [45], respectively. As described in [46], numerous data points with similar characteristics to the input can be created by sampling different points from the latent space and decoding them for use in downstream tasks. However, a constraint is imposed on learning the latent space to store the latent attribute as a probability distribution in order to generate new high-quality data points.

In VAE model, the input is as follows:

$$p_\theta(x|z) = f(x; z, \theta) \quad p(z) = N(z|0, I) \quad (4)$$

where f is a posterior probability function that uses a deep neural network to perform a non-linear transformation. The exact computation of the posterior $p_{\theta}(z|x)$ in this model is not mathematically feasible. Instead, a distribution $q_{\phi}(z|x)$ [41] is used to approximate the true posterior probability. This inference network $q_{\phi}(z|x)$ is parameterized as a multivariate normal distribution as shown below:

$$q_{\phi}(z|x) = N(z|\mu_{\phi}(x), \text{diag}(\sigma_{\phi}^2(x))) \tag{5}$$

where both $\sigma_{\phi}^2(x)$ and $\mu_{\phi}(x)$ represent the vector’s variance and means, respectively.

In case of deep networks, convergence of may lead to degradation problem [47]. With the increase in the depth of the networks, performance saturates to an unsatisfactory level. Furthermore, in case of autoencoders, proper reconstruction of the input may not be achieved; thereby, the essential features cannot be captured in the latent vectors. This problem is solved by introducing skip connections (Figure 1). Such residual blocks enable the autoencoder to learn a layer-wise identity relation which does not incur the cost of learning any extra parameters. Moreover, the applications of autoencoders have been successfully studied in facial image restoration and emotion recognition. Motivated by this, in the current study, we employ a residual variational autoencoder model to extract the most important features in a latent space. The proposed RVA model architecture has been depicted in Figure 2.

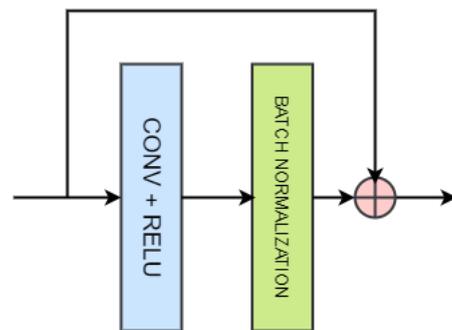


Figure 1. Proposed Residual Learning Block.

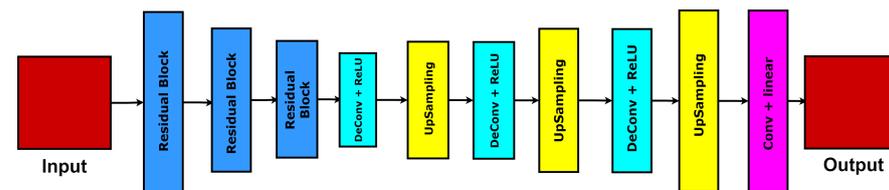


Figure 2. Proposed Residual Variational Autoencoder Architecture.

3. Affinity-Based Overlapping Detection

In the current article, the detection of an overlapping region between different classes has been achieved by using the notion of affinity. Let us assume a labeled dataset $D = \{(p_1, y_1), (p_2, y_2), \dots, (p_m, y_m)\}$, where the i th data point p_i denotes a point in R^n , and y_i is the label associated with it. We assume that data points belong to $k (k \geq 2)$ classes. Hence, for any i th data point, the class label $y_i \in [1, k]$. Data points belonging to a particular class are considered as a labeled cluster. The entire cluster is represented by a cluster representative. The cluster membership is calculated by taking the mean of data points in a cluster and calculated using Equation (6):

$$C_j = \frac{1}{n_j} \sum_{l=1}^{n_j} p_l \tag{6}$$

where n_j denotes the number of data points in j th cluster. In the initial dataset, the membership of the data points are crisp. However, such crisp label information does not reveal how close a data point is to its cluster representative. Therefore, we define an affinity score associated with every data point for all class representatives. The affinity score is designed to reflect the confidence of membership of a data point. The affinity score of a data point is calculated using Equation (7):

$$a_{ij} = e^{\frac{-d_{ij}^2}{2\sigma^2}} \tag{7}$$

where a_{ij} represents the affinity between the i th data point and the j th cluster, and d_{ij} denotes the distance between the same. The scaling parameter σ decides the the closeness between one data point and class representative within σ units. The affinity score between any data point and class representative becomes high when they are close and becomes eventually small as the becomes far apart. Now, we define a metric β , which is used to decide whether a data point is in an overlapping region or not. It is defined in Equation (8):

$$\beta_i = \frac{a_{ic}}{\sum_{j=1}^k a_{ij}} \tag{8}$$

To elaborate further, a binary classification has been considered in Figure 3, where C_1 and C_2 represents class representatives of two classes, viz., ‘1’ and ‘2’, respectively. The elliptical boundary denotes the class data distributions. Data points p_1 and p_2 both belong to class ‘1’. However, p_1 is outside the overlapping region. and p_2 is inside the overlapping region. The affinity of these data points with respect to both class representatives can be calculated by using Equation (7), and subsequently, the β value is calculated using Equation (8). Affinity between p_1 and C_1 is denoted by a_{11} and the other affinity values are mentioned on the line joining them in Figure 3.

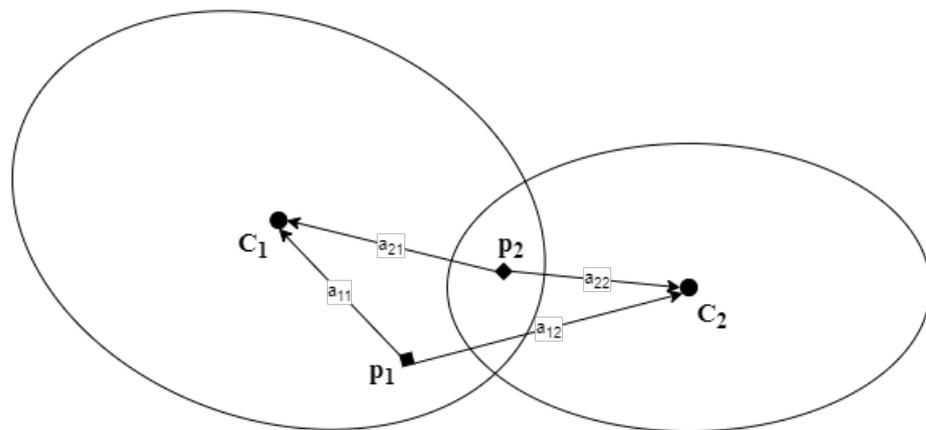


Figure 3. Affinity between data points and class representatives, where $k = 2$ and $n = 2$.

4. Proposed Method

In Section 3, the preliminary concept of affinity-based overlapping region detection has been discussed. In this section, we introduce the overlap reduction method and the overall proposed scheme for emotion recognition. The proposed method has been explained in Figure 4. The initial facial emotion dataset is first used to train the Residual Variational Autoencoder (RVA). After training the RVA, only the encoder has been used to convert all images to a latent vector form. These latent vectors corresponding to various emotion categories are overlapped. Hence, the affinity score is calculated for all latent vectors using Equation (7), and the corresponding β values are calculated using Equation (8). Now, the β value increases as the data point becomes closer to the overlapping region. Therefore, the data point having β value greater than a predefined threshold β_t has been removed from the dataset. After that, a set of well-known classifiers have been trained with both

overlapped and overlapped reduced modified datasets. The performances in both cases are calculated based on test phase confusion matrix.

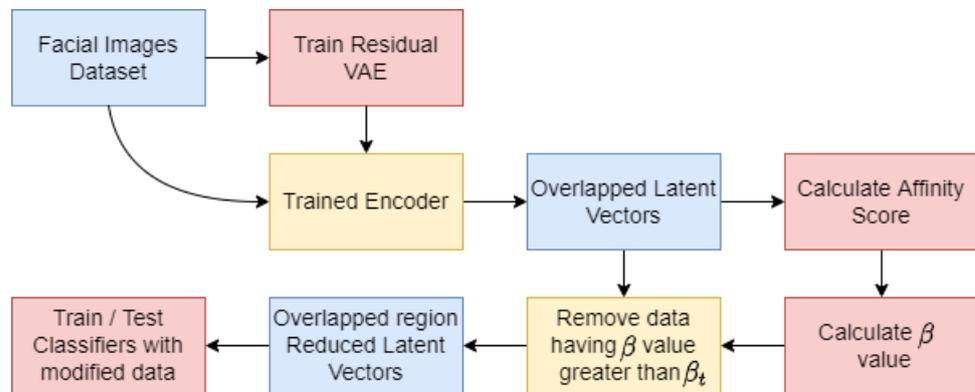


Figure 4. Proposed Residual Variational Autoencoder based Affinity score coupled overlapping region reduction method.

The analogy behind using the β value to determine the overlapped region can be conceptualized by using Figure 5, which plots the posterior densities of two classes for a binary classification problem. Data instances of this binary classification problem have a single feature only, which is plotted along the horizontal axis. The density of class ‘1’ is plotted in blue, and the same for class ‘2’ has been plotted in red. The posterior densities reveal that for all patterns within range $[-1, 3.5]$ will incur some error in the decision-making process. Furthermore, at the point at which both densities intersect with each other, the data points having a feature value of 1.8 will have equal probability of being in class ‘1’ and ‘2’. In addition, a region around that point in feature space will have data points for which membership to a particular class is uncertain as posterior densities indicate that they have almost equal chance of being a member of both classes. Along with the densities, a black dashed plot has been depicted in Figure 5. This line plots the β values corresponding to every data point. This plot reveals that the β value increases as the uncertainty about the membership of data points increase. At the intersection of densities, the corresponding data point achieves the highest β value. Hence, by using a threshold on β , data points having less confidence about their membership can be discarded from the dataset, thereby reducing the overlapping region of the dataset.

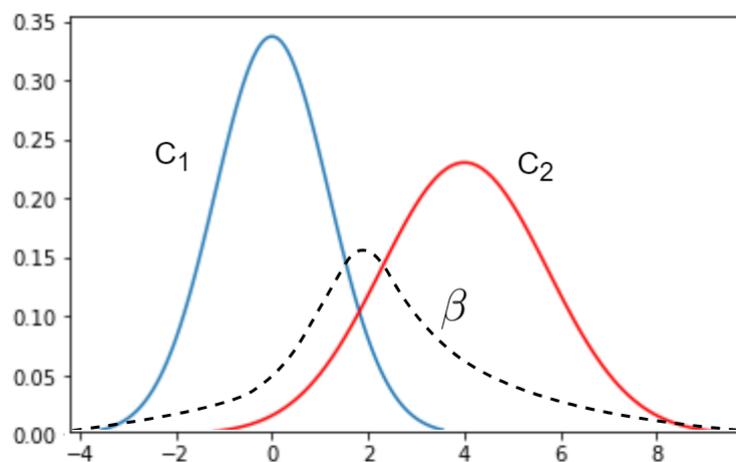


Figure 5. Posterior density plot for a binary classification with β value.

Figure 6a depicts a similar dataset which has two categories of data instances. Both categories of data instances are plotted using different colored markers. It can be observed

that there is a substantial amount of overlapping between the classes. After applying the proposed affinity-based overlap reduction method, the modified dataset is shown in Figure 6b. The class representatives of both the classes are marked using red marker. The β_t was set to 0 to obtain this dataset which has almost no overlap between the classes. The contribution of the affinity score in this process can be further elaborated by the affinity plots depicted in Figure 7a,b. These figures depict the affinity values of individual data points of the dataset (Figure 6a) with respect to class representatives ‘1’ and ‘2’, respectively. Figure 7a reveals that the affinity of data points with respect to class ‘1’ increases as it becomes closer to the class representative of class ‘1’. A similar trend can be observed in the case of class ‘2’, as well (Figure 7b).

Algorithm 1 explains the proposed RVA model supported AFORET method. From steps 1 to 9, RVA model training has been explained. In line 10, the trained encoder has been used to obtain the overlapped latent vectors. Lines 11 to 14 calculate the affinity of the data points for all classes. From lines 15 to 21, the β values for all data points have been calculated. Finally, from lines 22 to 26, data points having β values greater than the threshold β_t have been selected in the final latent vector set.

Algorithm 1 Residual Variational Autoencoder (RAV) based Affinity-based Overlap Reduction Technique (AFORET)

Input: Dataset $I = \{I^1, \dots, I^n\}$

Output: Overlap reduced latent vectors $P = \{p_1, \dots, p_n\}$

```

1:  $\theta, \phi \leftarrow \text{Initialize}$ 
2: repeat
3:   for  $k \leftarrow 1$  to  $N$  do
4:     Draw Samples  $S$  from  $\epsilon \sim \mathcal{N}(0, 1)$ 
5:      $z^{(k,s)} = h_\phi(\epsilon^{(k)}, x^{(k)})$ 
6:   end for
7:    $E = \sum_{k=1}^N -D_{KL}(q_\phi(z|x^{(k)})||p_\theta(z)) + \frac{1}{S} \sum_{s=1}^S (\log p_\theta(x^{(k)}|z^{(k,s)}))$ 
8:    $\phi, \theta$   $\triangleright$  Update the parameters using Stochastic Gradient Descent
9: until The parameters  $\phi, \theta$  converges
10:  $L = P_\phi(I)$   $\triangleright P_\phi$  is trained encoder. Set of latent vectors  $L = \{l_1, \dots, l_n\}$ 
11: for  $i \leftarrow 1$  to  $n$  do
12:    $a_{ij} = e^{\frac{-d_{ij}^2}{2\sigma^2}}$   $\triangleright d_{ij}$  is the euclidian distance between  $i$ th data instance
13:    $\triangleright$  and  $j$ th class representative
14: end for
15: for  $i \leftarrow 1$  to  $n$  do
16:    $s_i \leftarrow 0$ 
17:   for  $j \leftarrow 1$  to  $k$  do
18:      $s_i = s_i + a_{ij}$ 
19:   end for
20:    $\beta_i = \frac{a_{ic}}{s_i}$   $\triangleright a_{ic}$  is the affinity of  $i$ th data point where it belongs to  $c$ th class
21: end for
22: for  $i \leftarrow 1$  to  $n$  do
23:   if  $\beta_i \geq \beta_t$  then
24:      $P \leftarrow P \cup l_i$ 
25:   end if
26: end for

```

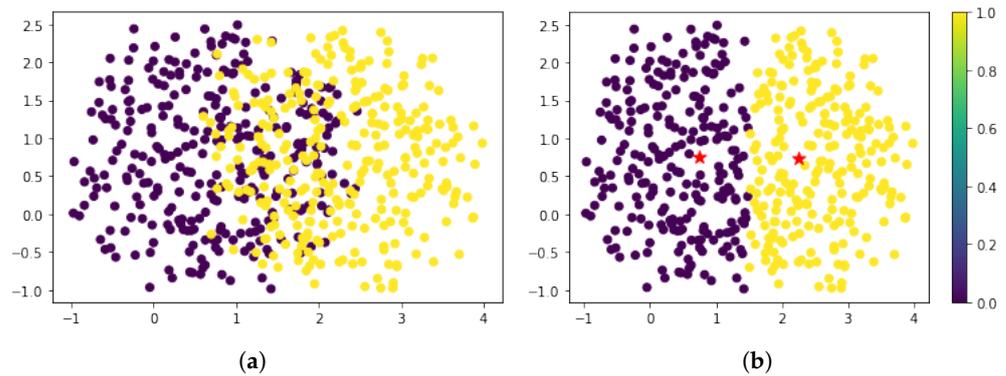


Figure 6. (a) Sample overlapped binary class dataset. (b) Dataset after removing overlapping region when $\beta_t = 0$.

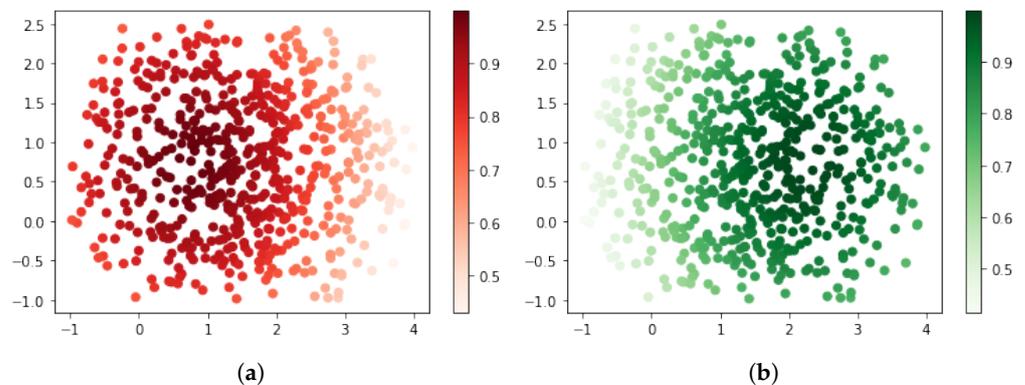


Figure 7. (a) Affinity plot of class 1 of dataset depicted in Figure 6a. (b) Affinity plot of class 2 of dataset depicted in Figure 6a.

5. Results and Discussion

5.1. Experimental Setup

The proposed affinity-based overlap reduction technique (AFORET) coupled with the initial stage of the RVA model has been tested by using the popular Affectnet Facial Expression Dataset [48]. Out of the original 11 categories of facial emotion images of Affectnet, 7 categories of emotions, viz., ‘Neutral’, ‘Happy’, ‘Sad’, ‘Surprise’, ‘Fear’, ‘Disgust’, and ‘Anger’, have been considered in the current study. As evident from previous studies [49,50], the presence of overlapped classes in the dataset have significantly reduced the classifier performance in predicting facial emotions. Thus, in the current study, the modified dataset is first used to train the proposed RVA model. Later, the encoder of the trained RVA model is used to convert the input images to a latent form. The shape of the latent vectors are decided by a separate experiment.

To reduce the overlapped region of the latent vectors, the affinity-based overlap region reduction technique has been applied. The β_t threshold for the study has been decided by conducting an extensive analysis. The performances of the classifiers have been checked for β_t values such that the total amounts of data loss are 5%, 10%, and 15%. The performances of the classifiers have been checked in terms of performance indicators such as ‘Accuracy’, ‘Sensitivity’, ‘Specificity’, ‘Balanced Accuracy’, ‘G-mean’, ‘Area Under Curve’ (AUC), and ‘Matthews Correlation Coefficient’ (MCC). Firstly, the original latent vectors with overlapping have been used to train and test the classifier. After that, the modified dataset obtained after applying overlap reduction technique has been used to test the classifiers for data losses of 5%, 10%, and 15%. For all experiments, 10-fold cross validation has been used.

The classifiers used in the current study are ‘Logistic Regression’ (LR), ‘Naive Bayes’ (NB), ‘Random Forest’ (RF), ‘K-Nearest Neighbor’ (KNN), ‘Multilayer Perceptron’ (MLP), ‘Support Vector Machine’ (SVM), and XGBoost. All these classifiers have been compared with modified datasets after various degrees of data reduction by AFORET. The parametric setup of the aforementioned classifiers are as in [8].

In addition, AFORET has been compared with three well-known overlap region reduction techniques, viz., OSM [36], ν -SVM [37], and Neighborhood Undersampling (NBU) [38]. After converting the images to latent vector form using RVA, the aforementioned algorithms have been applied to reduce the overlapping region present in the latent dataset. The modified datasets corresponding to individual algorithms have been employed to train and test the best-performing classifier to compare their performance in terms of all the performance metrics.

5.2. Analysis Using Classifiers

The classifiers used in the current study have been trained using the latent vectors obtained from the RVA model. In this section, the performances of the classifiers have been compared by training them using modified datasets with varying degrees of data loss. Instead of comparing in terms of varying β_t value, a more logical alternative of comparing in terms of the amount of data loss has been considered to understand in a better way as to how the performance change. For this purpose, AFORET algorithm has been applied on the initial latent vectors and datasets with 5%, 10%, and 15% data loss have been obtained. For each modified dataset, all the classifiers have been trained.

Table 1 depicts the performance of classifiers in terms of accuracy. For the original dataset with the overlapped region being untouched, the performance of the classifiers has been found to be poor. The best performance is achieved by XGBOOST, which has achieved an accuracy of 0.61. After modifying the dataset by applying AFORET and reducing 5% of the data from the original dataset, it is used to train the classifiers. The performances of all classifiers have been found to have improved with an average of 0.94. On the other hand, the dataset with 10% data loss does not improve the performance beyond 0.94. However, KNN has reported slight improvement with an accuracy of 0.98. next, with the 15% data loss, the performance is improved further to 0.95 with an overall improvement in almost all classifiers.

Table 1. Performance comparison of classifiers in terms of accuracy.

	Overlapped	5%	10%	15%
LR	0.42	0.92	0.92	0.94
NB	0.51	0.96	0.96	0.96
RF	0.34	0.93	0.93	0.93
KNN	0.4	0.97	0.98	0.99
MLP	0.31	0.91	0.91	0.93
SVM	0.5	0.94	0.94	0.95
XGBOOST	0.61	0.98	0.98	0.95
Average	0.44	0.94	0.94	0.95

Table 2 reports the performance of all classifiers in terms of sensitivity. It has been observed that the performances of almost all classifiers are not satisfactory for the original overlapped dataset (column “Overlapped”). The best performance is achieved by XGBOOST with a sensitivity score of 0.57. On average, the classifiers have been able to achieve a sensitivity of 0.45 for the overlapped dataset. Next, the modified dataset with 5% data loss is used to train the classifiers. Significant improvement can be observed in the performance in terms of sensitivity. The best performance is achieved by the NB classifier with a sensitivity of 0.95, whereas the average performance is improved to 0.92. For 10% data loss, the performance improves further with an average sensitivity of 0.94. Finally,

after reducing 15% of data, the average performance improves, whereas a few classifier’s performance decreases. The average performance becomes 0.95, which is slightly better than a 10% data loss.

Table 2. Performance comparison of classifiers in terms of sensitivity.

	Overlapped	5%	10%	15%
LR	0.43	0.89	0.88	0.97
NB	0.46	0.95	0.93	0.98
RF	0.39	0.9	0.96	0.91
KNN	0.41	0.93	0.97	0.99
MLP	0.34	0.89	0.96	0.93
SVM	0.52	0.94	0.96	0.94
XGBOOST	0.61	0.94	0.93	0.94
Average	0.45	0.92	0.94	0.95

Table 3 reports the performance of classifiers in terms of specificity. The classifiers’ performance have been unsatisfactory for original latent vectors with overlapped regions, where the average performance in terms of specificity has been recorded to be 0.45. However, after applying AFORET, the performance of the classifiers gradually improves from an average of 0.94 for 5% data loss to 0.96 at 15% data loss. The performance of all individual classifiers have also reflected a similar trend of improvement. Among all, XGBOOST has been found to perform the best with an accuracy of 0.99.

Table 3. Performance comparison of classifiers in terms of specificity.

	Overlapped	5%	10%	15%
LR	0.42	0.96	0.9	0.93
NB	0.48	0.92	0.98	0.93
RF	0.38	0.97	0.97	0.98
KNN	0.45	0.95	0.98	0.97
MLP	0.34	0.89	0.95	0.97
SVM	0.55	0.98	0.99	0.99
XGBOOST	0.58	0.92	0.93	0.99
Average	0.45	0.94	0.95	0.96

Next, in Table 4, the performances of the classifiers have been compared in terms of balanced accuracy. As observed earlier, the performance of classifiers for the overlapped dataset is unsatisfactory. On average, the classifiers achieved a score of 0.45 in terms of balanced accuracy. However, as AFORET is applied and the initial dataset is modified, the performance improves. At 5% data loss, the performance, on average, is 0.93. Further loss of data to 10% and 15% improves it even further to 0.95 and 0.96, respectively.

Table 5 tabulates the performance in terms of the G-mean. This performance metric reflects the combined effect of sensitivity and specificity. Hence, a similar trend of performance has been recorded. XGBOOST remains the best performer for the overlapped dataset. However, KNN has been found to be the best after applying AFORET. This indicates that the latent space embedding produced by the proposed RVA model is efficient enough that local information could be sufficient to distinguish between different emotions.

Tables 6 and 7 report the performance of classifiers in terms of the AUC and MCC scores. These two performance metrics reveals that the performances of the classifiers at 10% data loss are slightly improved with 15% data loss. Thus, in order to minimize the amount of data loss and at the same time to achieve the best classification performance, 10% data loss is sufficient and optimal. It can be further noted from Tables 1–7 that the performances of the classifiers for the original overlapped dataset are significantly lower

compared to their performance when the dataset is processed with AFORET. This reveals that the original latent vector form of dataset has all classes highly overlapped with each other. After the reduction of the overlapping region even by 5%, the performance of the classifiers improves significantly. Table 8 reports the accuracy scores of the individual classes for all classifiers. The classifiers are trained with the overlapped dataset and test phase performance in terms of accuracy have been measured. Next, the same experiment has been repeated with the overlapped reduced dataset. Previous experiments have already revealed that a 10% data reduction would be sufficient to alleviate the overlapped class problem. Hence, AFORET with 10% data loss has been considered for this class-wise comparison. Table 8 reveals that for all seven categories of emotions which are considered in the current study, it is observed that AFORET significantly improves classifier performance in detecting individual categories.

Table 4. Performance comparison of classifiers in terms of balanced accuracy.

	Overlapped	5%	10%	15%
LR	0.43	0.93	0.89	0.95
NB	0.47	0.94	0.96	0.96
RF	0.39	0.94	0.97	0.95
KNN	0.43	0.94	0.98	0.98
MLP	0.34	0.89	0.96	0.95
SVM	0.54	0.96	0.98	0.97
XGBOOST	0.6	0.93	0.93	0.97
Average	0.45	0.93	0.95	0.96

Table 5. Performance comparison of classifiers in terms of G-Mean.

	Overlapped	5%	10%	15%
LR	0.42	0.92	0.89	0.95
NB	0.47	0.93	0.95	0.95
RF	0.38	0.93	0.96	0.94
KNN	0.43	0.94	0.97	0.98
MLP	0.34	0.89	0.95	0.95
SVM	0.53	0.96	0.97	0.96
XGBOOST	0.59	0.93	0.93	0.96
Average	0.45	0.93	0.94	0.95

Table 6. Performance comparison of classifiers in terms of AUC.

	Overlapped	5%	10%	15%
LR	0.39	0.98	0.92	0.95
NB	0.47	0.9	0.9	0.98
RF	0.43	0.84	0.99	0.92
KNN	0.41	0.91	0.97	0.98
MLP	0.37	0.89	0.95	0.97
SVM	0.51	0.89	0.98	0.97
XGBOOST	0.56	0.86	0.9	0.94
Average	0.45	0.9	0.94	0.96

Table 7. Performance comparison of classifiers in terms of Matthews correlation coefficient.

	Overlapped	5%	10%	15%
LR	0.39	0.88	0.9	0.9
NB	0.56	0.95	0.98	0.99
RF	0.35	0.96	0.98	0.99
KNN	0.44	0.97	0.96	0.95
MLP	0.29	0.9	0.98	0.98
SVM	0.59	0.98	0.98	0.99
XGBOOST	0.64	0.83	0.85	0.89
Average	0.47	0.92	0.95	0.95

Table 8. Accuracy achieved by classifiers for all emotion classes before and after applying AFORET.

		LR	NB	RF	KNN	MLP	SVM	XGBOOST
Neutral	Overlapped	0.72	0.81	0.24	0.5	0.31	0.6	0.71
	AFORET(10%)	0.99	0.76	0.83	0.99	0.81	0.94	0.99
Happy	Overlapped	0.62	0.51	0.24	0.6	0.41	0.8	0.31
	AFORET(10%)	0.92	0.76	0.93	0.88	0.99	0.99	0.98
Sad	Overlapped	0.32	0.31	0.04	0.5	0.01	0.4	0.81
	AFORET(10%)	0.99	0.76	0.73	0.78	0.99	0.99	0.99
Surprise	Overlapped	0.72	0.21	0.44	0.6	0.41	0.6	0.81
	AFORET(10%)	0.99	0.99	0.73	0.99	0.81	0.99	0.99
Fear	Overlapped	0.52	0.31	0.34	0.1	0.51	0.5	0.41
	AFORET(10%)	0.92	0.86	0.99	0.88	0.99	0.99	0.99
Disgust	Overlapped	0.62	0.81	0.34	0.7	0.21	0.8	0.51
	AFORET(10%)	0.72	0.99	0.93	0.88	0.99	0.84	0.99
Anger	Overlapped	0.52	0.21	0.54	0.1	0.41	0.8	0.81
	AFORET(10%)	0.99	0.99	0.99	0.78	0.99	0.94	0.99

5.3. Comparative Study of Overlap Reduction Methods

In Section 5.2, various classifiers have been compared in terms of several performance indicators to understand the ingenuity of the proposed AFORET method used to mitigate the overlapped classes. In the current section, the proposed AFORET is compared with three well known overlapping removal techniques, viz., OSM [36], ν -SVM [37], and Neighborhood Undersampling (NBU) [38]. It has been observed in Section 5.2 that the performance of a majority of the classifiers are close to each other. Hence, in this section, the overlapped latent vectors are processed using the overlap reduction/removal techniques, and the modified dataset is then used to train and test all the previously used classifiers. The performances of the classifiers in terms of all performance indicators have been recorded. In order to compare the algorithms, the data losses in all methods have been restricted to 10% of the original set only.

Table 9 reports the performance of the overlap removal algorithms in terms of various performance metrics for all classifiers. In terms of accuracy, it has been observed that the performance of almost all classifiers have achieved the best results for the proposed AFORET method. However, the LR, KNN, and MLP classifiers have performed equally well for NBU as well. Next, a sensitivity-based comparison reveals that the performance of AFORET remains the best for all classifiers except KNN. In addition, the ν -SVM performed equally well for NB, RF, and KNN. However, the average performance of classifiers remains

best for AFORET only. After that, the performance analysis for specificity reveals a similar trend. In case of balanced accuracy, OSM and NBU performed equally well on all classifiers, whereas the performance of ν -SVM and the proposed method is close for a few classifiers. However, the average performance of AFORET with 0.95 is significantly better than ν -SVM.

G-mean, AUC, and MCC reveal a similar trend of performance. It has been observed that in terms of all performance metrics, the average performance of OSM is almost same as NBU, whereas few classifiers have reported equal performance of ν -SVM with the proposed AFORET. However, upon taking the average performance obtained by all classifiers, it has been found that the performance of proposed AFORET remains better than all other remaining methods. This extensive comparative analysis with existing overlap removal technique establishes the fact that the proposed AFORET-based method to reduce overlap between classes has significantly improved the performance of the classifiers to detect human emotions based on an RVA model.

Table 9. Comparison of AFORET with OSM, NBU, and ν -SVM algorithms.

Performance Metric	Classifiers	OSM	ν -SVM	NBU	AFORET
Accuracy	LR	0.92	0.91	0.92	0.92
	NB	0.93	0.93	0.92	0.96
	RF	0.92	0.88	0.92	0.93
	KNN	0.97	0.97	0.98	0.98
	MLP	0.87	0.9	0.91	0.91
	SVM	0.94	0.89	0.91	0.94
	XGBOOST	0.95	0.96	0.93	0.98
Sensitivity	LR	0.87	0.85	0.86	0.88
	NB	0.88	0.93	0.9	0.93
	RF	0.94	0.96	0.91	0.96
	KNN	0.95	0.97	0.97	0.97
	MLP	0.96	0.95	0.92	0.96
	SVM	0.94	0.96	0.95	0.96
	XGBOOST	0.91	0.93	0.9	0.93
Specificity	LR	0.9	0.94	0.9	0.94
	NB	0.96	0.95	0.98	0.98
	RF	0.93	0.95	0.97	0.97
	KNN	0.98	0.98	0.98	0.98
	MLP	0.91	0.94	0.93	0.95
	SVM	0.96	0.98	0.99	0.99
	XGBOOST	0.89	0.92	0.92	0.93
Balanced Accuracy	LR	0.86	0.87	0.86	0.89
	NB	0.93	0.95	0.95	0.96
	RF	0.94	0.96	0.95	0.97
	KNN	0.97	0.98	0.98	0.98
	MLP	0.94	0.95	0.93	0.96
	SVM	0.95	0.97	0.97	0.98
	XGBOOST	0.9	0.92	0.91	0.93

Table 9. Cont.

Performance Metric	Classifiers	OSM	ν -SVM	NBU	AFORET
G-Mean	LR	0.87	0.88	0.86	0.89
	NB	0.91	0.93	0.93	0.95
	RF	0.93	0.95	0.93	0.96
	KNN	0.96	0.97	0.97	0.97
	MLP	0.93	0.94	0.92	0.95
	SVM	0.95	0.97	0.97	0.97
	XGBOOST	0.9	0.93	0.91	0.93
AUC	LR	0.92	0.91	0.91	0.92
	NB	0.86	0.9	0.88	0.9
	RF	0.97	0.99	0.97	0.99
	KNN	0.97	0.97	0.93	0.97
	MLP	0.93	0.94	0.91	0.95
	SVM	0.94	0.98	0.95	0.98
	XGBOOST	0.86	0.85	0.9	0.9
MCC	LR	0.85	0.88	0.9	0.9
	NB	0.98	0.93	0.96	0.98
	RF	0.96	0.97	0.93	0.98
	KNN	0.96	0.92	0.94	0.96
	MLP	0.94	0.95	0.93	0.98
	SVM	0.97	0.97	0.93	0.98
	XGBOOST	0.8	0.84	0.81	0.85

6. Conclusions

The current article has proposed a novel overlapping reduction technique to improve classification performance in emotion recognition using facial images. The class overlapping problem in facial emotion detection has been solved by using an affinity-based overlap reduction technique. The proposed AFORET method has been used to reduce the overlapping region so that performance of classifiers in emotion recognition can be improved. AFORET has been tested for various degrees of data loss starting from 5% up to 15%. The original facial image dataset is transformed to a latent vector form to capture the most important features for the classification task. These latent vectors are then modified using AFORET to reduce the overlapping region. After reducing the overlapping region, a set of well-known classifiers have been trained and tested to establish the ingenuity of the proposed model. Experimental results have revealed that 10% data loss using AFORET sufficiently reduces the overlap regions and improves classifier performance. Any extra data loss beyond 10% does not improve classifier performance further. In addition, a comparative analysis with existing overlap removal techniques, viz., OSM, ν -SVM, and NBU, has been conducted. The comparative study revealed that the proposed AFORET is better than all other methods in addressing the class overlapping problem in facial emotion recognition. Overall, the proposed RVA model combined with AFORET has been able to significantly improve classification performance to a greater extent.

Author Contributions: Data curation, S.C.; Formal analysis, A.K.D.; Investigation, D.P.; Methodology, S.C.; Project administration, D.P.; Resources, J.N.; Supervision, A.K.D.; Validation, J.N.; Writing (original draft), S.C.; Writing (review and editing), A.K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Zhang, H. Expression-EEG based collaborative multimodal emotion recognition using deep autoencoder. *IEEE Access* **2020**, *8*, 164130–164143. [[CrossRef](#)]
2. Sajjad, M.; Kwon, S. Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. *IEEE Access* **2020**, *8*, 79861–79875.
3. Wu, J.L.; He, Y.; Yu, L.C.; Lai, K.R. Identifying emotion labels from psychiatric social texts using a bi-directional LSTM-CNN model. *IEEE Access* **2020**, *8*, 66638–66646. [[CrossRef](#)]
4. Lee, S.C.; Liu, C.C.; Kuo, C.J.; Hsueh, I.P.; Hsieh, C.L. Sensitivity and specificity of a facial emotion recognition test in classifying patients with schizophrenia. *J. Affect. Disord.* **2020**, *275*, 224–229. [[CrossRef](#)]
5. Zepf, S.; Hernandez, J.; Schmitt, A.; Minker, W.; Picard, R.W. Driver emotion recognition for intelligent vehicles: A survey. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–30. [[CrossRef](#)]
6. Panda, R.; Malheiro, R.M.; Paiva, R.P. Audio features for music emotion recognition: A survey. *IEEE Trans. Affect. Comput.* **2020**, *1*, 1. [[CrossRef](#)]
7. Huang, C.; Trabelsi, A.; Qin, X.; Farruque, N.; Mou, L.; Zaiane, O.R. Seq2Emo: A Sequence to Multi-Label Emotion Classification Model. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 4717–4724.
8. Banerjee, A.; Bhattacharjee, M.; Ghosh, K.; Chatterjee, S. Synthetic minority oversampling in addressing imbalanced sarcasm detection in social media. *Multimed. Tools Appl.* **2020**, *79*, 35995–36031. [[CrossRef](#)]
9. Ghosh, K.; Banerjee, A.; Chatterjee, S.; Sen, S. Imbalanced twitter sentiment analysis using minority oversampling. In Proceedings of the 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, 23–25 October 2019; pp. 1–5.
10. Jain, D.K.; Shamsolmoali, P.; Sehdev, P. Extended deep neural network for facial emotion recognition. *Pattern Recognit. Lett.* **2019**, *120*, 69–74. [[CrossRef](#)]
11. Lin, C.J.; Lin, C.H.; Wang, S.H.; Wu, C.H. Multiple convolutional neural networks fusion using improved fuzzy integral for facial emotion recognition. *Appl. Sci.* **2019**, *9*, 2593. [[CrossRef](#)]
12. Sivasangari, A.; Ajitha, P.; Rajkumar, I.; Poonguzhali, S. Emotion recognition system for autism disordered people. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *1*, 7. [[CrossRef](#)]
13. Jiang, M.; Francis, S.M.; Srishyla, D.; Conelea, C.; Zhao, Q.; Jacob, S. Classifying individuals with ASD through facial emotion recognition and eye-tracking. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6063–6068.
14. Lee, S.C.; Chen, K.W.; Liu, C.C.; Kuo, C.J.; Hsueh, I.P.; Hsieh, C.L. Using machine learning to improve the discriminative power of the FERD screener in classifying patients with schizophrenia and healthy adults. *J. Affect. Disord.* **2021**, *292*, 102–107. [[CrossRef](#)]
15. Hu, M.; Wang, H.; Wang, X.; Yang, J.; Wang, R. Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks. *J. Vis. Commun. Image Represent.* **2019**, *59*, 176–185. [[CrossRef](#)]
16. Gautam, K.; Thangavel, S.K. Video analytics-based facial emotion recognition system for smart buildings. *Int. J. Comput. Appl.* **2019**, *43*, 858–867. [[CrossRef](#)]
17. Haddad, J.; Lézoray, O.; Hamel, P. 3d-cnn for facial emotion recognition in videos. In *International Symposium on Visual Computing*; Springer: Cham, Switzerland, 2020; pp. 298–309.
18. Kim, D.H.; Song, B.C. Contrastive Adversarial Learning for Person Independent Facial Emotion Recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 5948–5956.
19. Chen, L.; Wu, M.; Pedrycz, W.; Hirota, K. Deep Sparse Autoencoder Network for Facial Emotion Recognition. In *Emotion Recognition and Understanding for Emotional Human-Robot Interaction Systems*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 25–39.
20. Zeng, N.; Zhang, H.; Song, B.; Liu, W.; Li, Y.; Dobaie, A.M. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* **2018**, *273*, 643–649. [[CrossRef](#)]

21. Chen, L.; Zhou, M.; Su, W.; Wu, M.; She, J.; Hirota, K. Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. *Inf. Sci.* **2018**, *428*, 49–61. [[CrossRef](#)]
22. Allognon, S.O.C.; Britto, A.D.S.; Koerich, A.L. Continuous Emotion Recognition via Deep Convolutional Autoencoder and Support Vector Regressor. In Proceedings of the IEEE 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
23. Ruiz-Garcia, A.; Palade, V.; Elshaw, M.; Awad, M. Generative adversarial stacked autoencoders for facial pose normalization and emotion recognition. In Proceedings of the IEEE 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
24. Chen, L.; Su, W.; Wu, M.; Pedrycz, W.; Hirota, K. A fuzzy deep neural network with sparse autoencoder for emotional intention understanding in human-robot interaction. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 1252–1264. [[CrossRef](#)]
25. Lakshmi, D.; Ponnusamy, R. Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders. *Microprocess. Microsyst.* **2021**, *82*, 103834. [[CrossRef](#)]
26. Nguyen, D.; Nguyen, D.T.; Zeng, R.; Nguyen, T.T.; Tran, S.; Nguyen, T.K.; Sridharan, S.; Fookes, C. Deep Auto-Encoders with Sequential Learning for Multimodal Dimensional Emotion Recognition. *IEEE Trans. Multimed.* **2021**, *1*, 1. [[CrossRef](#)]
27. Zhao, P.; Zhou, Z.H. Label distribution learning by optimal transport. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
28. Akbari, A.; Awais, M.; Fatemifar, S.; Khalid, S.S.; Kittler, J. A Novel Ground Metric for Optimal Transport-Based Chronological Age Estimation. *IEEE Trans. Cybern.* **2021**, *1*, 14. [[CrossRef](#)]
29. Er, M.J.; Wu, S.; Lu, J.; Toh, H.L. Face recognition with radial basis function (RBF) neural networks. *IEEE Trans. Neural Netw.* **2002**, *13*, 697–710.
30. Visa, S.; Ralescu, A. Learning imbalanced and overlapping classes using fuzzy sets. In Proceedings of the ICML, Washington, DC, USA, 21–24 August 2003; Volume 3, pp. 97–104.
31. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
32. Prati, R.C.; Batista, G.E.; Monard, M.C. Class imbalances versus class overlapping: An analysis of a learning system behavior. In *Mexican International Conference on Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 312–321.
33. Batista, G.E.; Prati, R.C.; Monard, M.C. Balancing strategies and class overlapping. In *International Symposium on Intelligent Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 24–35.
34. García, V.; Mollineda, R.A.; Sánchez, J.S.; Alejo, R.; Sotoca, J.M. When overlapping unexpectedly alters the class imbalance effects. In *Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 499–506.
35. García, V.; Sánchez, J.; Mollineda, R. An empirical study of the behavior of classifiers on imbalanced and overlapped data sets. In *Iberoamerican Congress on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 397–406.
36. Lee, H.K.; Kim, S.B. An overlap-sensitive margin classifier for imbalanced and overlapping data. *Expert Syst. Appl.* **2018**, *98*, 72–83. [[CrossRef](#)]
37. Devi, D.; Biswas, S.K.; Purkayastha, B. Learning in presence of class imbalance and class overlapping by using one-class SVM and undersampling technique. *Connect. Sci.* **2019**, *31*, 105–142. [[CrossRef](#)]
38. Vuttipittayamongkol, P.; Elyan, E. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. *Inf. Sci.* **2020**, *509*, 47–70. [[CrossRef](#)]
39. Du, S.; Tao, Y.; Martinez, A.M. Compound facial expressions of emotion. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1454–E1462. [[CrossRef](#)] [[PubMed](#)]
40. Turchenko, V.; Chalmers, E.; Luczak, A. A deep convolutional auto-encoder with pooling-unpooling layers in caffe. *arXiv* **2017**, arXiv:1701.04949.
41. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
42. Svensén, M.; Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Amsterdam, Netherlands, 2007.
43. Gregor, K.; Danihelka, I.; Graves, A.; Rezende, D.; Wierstra, D. Draw: A recurrent neural network for image generation. In Proceedings of the International Conference on Machine Learning (PMLR), Lille, France, 7–9 July 2015; pp. 1462–1471.
44. Babaeizadeh, M.; Finn, C.; Erhan, D.; Campbell, R.H.; Levine, S. Stochastic variational video prediction. *arXiv* **2017**, arXiv:1710.11252.
45. Sønderby, C.K.; Raiko, T.; Maaløe, L.; Sønderby, S.K.; Winther, O. Ladder variational autoencoders. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3738–3746.
46. Nguyen, T.T.D.; Nguyen, D.K.; Ou, Y.Y. Addressing data imbalance problems in ligand-binding site prediction using a variational autoencoder and a convolutional neural network. *Brief. Bioinform.* **2021**, *22*, bbab277. [[CrossRef](#)] [[PubMed](#)]
47. Mao, X.; Shen, C.; Yang, Y.B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2802–2810.
48. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.* **2017**, *10*, 18–31. [[CrossRef](#)]

-
49. Pons, G.; Masip, D. Multitask, Multilabel, and Multidomain Learning With Convolutional Networks for Emotion Recognition. *IEEE Trans. Cybern.* **2020**, *1*, 8. [[CrossRef](#)]
 50. Bendjoudi, I.; Vanderhaegen, F.; Hamad, D.; Dornaika, F. Multi-label, multi-task CNN approach for context-based emotion recognition. *Inf. Fusion* **2021**, *76*, 422–428. [[CrossRef](#)]