symmetry

MDPI

Article

TGSL-Dependent Feature Selection for Boosting the Visual Sentiment Classification

Usha Kingsly Devi Karuthakannan [1] and Gomathi Velusamy [2,*]

[1] Department of Electronics and Communication Engineering, Anna University Regional Campus, Tirunelveli 627007, Tamil Nadu, India; ushakingslydevi.k@auttvl.ac.in
[2] Department of Computer Science and Engineering, National Engineering College, Kovilpatti 628503, Tamil Nadu, India
* Correspondence: vgcse@nec.edu.in

Abstract: The automatic recognition of the emotions in still images is inherently more challenging than other visual recognition tasks, such as scene recognition, object classification and semantic image classification, as it involves a higher level of abstraction in the human cognition perspective. Symmetry can be found in many objects in the nature and can be used for many purposes such as object detection and recognition. Furthermore, rotating and flipping of the image is employed based on symmetry for training the classifier for the most accurate classification. Hence, there is a need to handle effectively large intra-class variance, scalability and subjectivity during recognition, and it is inherently ambiguous as an image can evoke multiple emotions. To address these issues, many of the existing works focus on improving the image representations. It is motivated by the observation that both global distributions and local image regions carry massive sentiments. In this research, three different pre-trained architectural models are implemented, and the classification performance of binary sentiment classification is examined on five widely-used effective datasets. Moreover, the features from the pre-trained models are selected optimally using the proposed Teaching Gaining Sharing Learning (TGSL) algorithm, which is the major contribution of the research. Extensive experiment results on the five datasets demonstrate that the proposed Visual sentiment analysis based on the TGSL algorithm with data augmentation achieved an improved performance compared to all other conventional techniques. The proposed framework uses the pre-trained model and never utilized any hand-crafted features, boosting the mean accuracy, sensitivity, and specificity to 99.11%, 99.31%, and 99.22%, respectively, for abstract dataset.

Keywords: visual sentiment analysis; pre-trained models; feature selection; affective computing; data augmentation

check for updates

Citation: Karuthakannan, U.K.D.; Velusamy, G. TGSL-Dependent Feature Selection for Boosting the Visual Sentiment Classification. *Symmetry* **2021**, *13*, 1464. https://doi.org/10.3390/sym13081464

Academic Editor: Jan Awrejcewicz

Received: 23 June 2021
Accepted: 26 July 2021
Published: 10 August 2021

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

Copyright: © 2021 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

Nowadays Affective computing draws significant attention from various researchers as it renders great service to various research communities ranging from computer vision to psychology. Advanced computer vision and machine learning techniques automatically classify the object, scenes in images and videos but often lack an affective interpretation of such content. The automatic symmetric detection of the objects is employed for the emotion recognition [1], which provides the most accurate and robust performance for the recognition system. Furthermore, symmetry is widely used for the real emotional recognition combined with convolutional neural network [2]. Initially, affective computing research was towards detecting and recognizing emotions to learn facial expression and gesture recognition. Sentiment analysis belongs to the active field of affective computing research domain [3]. The objective of automatic visual sentiment analysis is to recognize the sentiments that are evoked in images. Sentiment and emotion are closely related entities, but they do not express the same thing. Sentiment refers to an opinion or mental attitude produced by the reflection of a negative or positive feeling. On the other hand, emotion is

usually defined as a complicated, multidimensional integrated feeling state that reflects humans' psychological and emotional changes.

Online visual contents in social networks and websites play a prominent role in the users' day-to-day life. More significantly, the users keep a record of their daily activities through different types of media. Moreover, due to the widespread popularity of social networks, internet users intend to express their views through various media types, such asvideos, images, etc.For instance, one can reveal their travel experience or any view on the habitual events through these visual media. Hence, visual sentiment analysis comes into existence, enabling the automatic analysis of the sentiments from animage posted via social networks and other online media. Images and videos display strong sentiments, which strengthens the opinion portrayed in the content, and the sentiments underlying in the visual contents. Therefore, it benefits society in many means, such as targeted marketing, opinion mining [4], biomedical studies (stress/pain monitoring), autism-related assistive technology, e-business, image captioning [5], image retrieval [3] and so on. In computer vision, studies insist on the modelingof generic visual ideas. In contrast, the study on the adjectives concerning the visual sentiments is a question and seemed impossible due to the huge affective gap existing between the low and high-level features [6]. However, visual sentiment analysis is quite in the development stage, and it seems to be a tedious analysis compared with text-based sentiment analysis [7].

In visual sentiment analysis, the semantics are usually hidden in the images, in contrast withthe text semantic, and this seems to be a high-level visual semantic. However, it is worth noting that there is no high-level visual semantic dictionary similar to Wordnet, a semantic dictionary for text analysis. Therefore, there is a need for high-level visual semantic ontology features instead of low-level visual features for acquiring the visual sentiments from images. In the present scenario, the prediction of the visual sentiments is made based on the features acquired from the entire image as per the inspiration from the psychology theory and principles of arts. Moreover, it is peculiar to note that the emotions evoked by an image are from its global appearance and its local regions. Figure 1 shows the example images for the positive and negative sentiment.

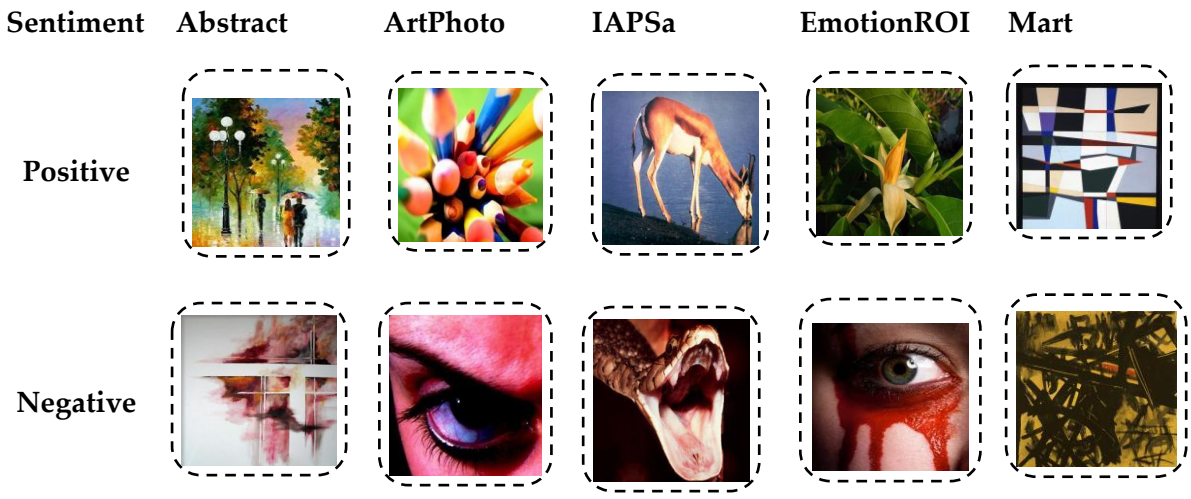


Figure 1. Example images of five different dataset Top Row: positive sentimentand Bottom Row: negative sentiment.

Recently, with the tremendous growth of high-performance computing (HPC) resources, various deep approaches have been developed to analyze sentiment [8–13]. The efficiency of machine learning-oriented in-depth features has been viewed over handcrafted features, such as color, texture, content and composition [6] on visual sentiment prediction. In comparison with the conventional recognition tasks, visual sentiment analysis is highly challenging due to a higher level of subjectivity in the human recognition process. It is also quite challenging as images with different looks can have the same emo-

tions, and some images with similar looks can have different emotions. Thus, interpreting visual information at the affective level is uncertain, and it is also hard to predict emotions directly from visual information.

In this research, visual sentiment analysis is concentrated. The feature selection is carried outusing the proposed TGSL algorithm, which inherits the Instructor characteristics and the mutual gaining characteristics of the Trainee to explore the highly relevant features from the extracted features using the pre-trained model. Generally, Visual sentiment analysis comprises two major groups; the first is feature extraction, and the second is sentiment classification. In this research, feature selection is placed as an intermediate step that boosts classification performance. Some of the contributions of the research are enlisted in the following statements.

- To analyze the effectiveness of the three dynamic frameworks such as ResNet50 [14], VGG [15] and Alexnet [16]. In addition, the ImageNet [17] dataset is utilized for training the models above and the classification of the visual sentiments.
- The three data augmentation, such as Scaling, rotation and Intensityvariation, are explored and analyzed for data augmentation in the visual sentiment classification.
- Proposed a novel feature selection algorithm TGSL, which is developed by integrating Instructor and Trainee's characteristic features from the Teaching learning-based algorithm [18] and Gaining-sharing based algorithm [10]. It is used to selectively select the huge feature vector produced from the pre-trained CNN models, which helps improve the classification of sentiments from images.

Following this introduction section, the rest of the paper is organized as Section 2 provides the literature review. Section 3 elaborates the proposed method of visual sentiment prediction. Experimental results and discussions are analyzed in Section 4. Finally, Section 5 concludes the work with future research scope.

2. Related Works

The progress and impressive performance of CNN is drawing much research attention in computer vision-related research. It has recently been employed to understand the sentiment from visual content and has attracted significant attention. There are two main ways to handle visual sentiment prediction: dimensional models [19] and categorical models [20]. Categorical approaches are easier for a human to understand, and hence we aim to make categorical sentiment predictions. Categorical models are associated with one or more labels. In contrast, in dimension models, the affective states are represented by the 3-D valence-arousal-control emotion space [21], or 2-D valence-arousal space [19,20], or activity-weight-heat space [22]. However, other emotions evoked in an image can also be described using emotional adjectives. Here, adjectives are a common way to express human emotional feelings and impressions or the quality of any facts or events.

Visual sentiment analysis can also be viewed as a high-level visual content classification problem. Several researchers have employed traditional generic image processing features color [23], texture [24,25] and shape [26]. Yanulevskaya et al. [27] employed Wiccest and Gabor feature, global/local RGB histogram color histogram, color moment and color correlogram. Furthermore, the SIFT-based bag of features [28] and Gist features [28,29], and the model was developed with machine learning to achieve sentiment categorization. Machajdik et al. [6] tried to extract eight different pixel-level hand-crafted features based on art and psychology theory to represent the affective content of images. Inspired by the art theories, Wang et al. [30] extracted the aesthetic features such as color patterns, shapes and compositions that are more suitable for effective prediction. Sentiment analyses were also carried out using midlevel features by researchers [29]. In [29] scene, low level features were suggested to bridge the affective gap between the low-level features and high-level sentiments. These features were trained using an SVM classifier to find 102 mid-level attributes of an image to predict its sentiment. Zhao et al. [31] designed a multi-task hypergraph learning framework to predict emotion which considers different factors that influence emotion such as social context, temporal evaluation and location influence.

In recent years, the increase in the computation ability of GPU and inspired by the great success of various CNN models, such as AlexNet, VGGNet and GoogLeNet, in many computer vision tasks attempted to develop deep architectures for the task of visual sentiment analysis. AshimaYadav and DineshKumarVishwakarma [32] utilize the residual attention-based deep learning network (RA-DLA network) to evaluate the emotion recognition process in the still network. The powerful emotions are conveyed by the local regions of the specific image. Hence, the residual attention strategy is accomplished in the local region, the emotion-rich region in the image. The high detection accuracy is the prime advantage of the RA-DLA system. However, the RLDA-system is not adaptable for the multi-modal data to recognize multi-modal emotions. Papiya Das et al. [33] utilize the SVM classifier to recognize emotions in still images. The SVM classifier in this method replaces the SoftMax layer of the DNN.

The hashtags or the descriptive textual methods are utilized in the system to map the local image. The SVM classifier is not suitable for non-linear issues and fails to handle high-dimensional datasets. Jie Chen et al. [34] presented the active learning curve to attain effective recognition of emotion in the image. The Deep-CNN is incorporated with active learning for the sentiment analysis process. The active learning methods utilize few standard samples to train the system. The active learning system is needed to be integrated with the uncertainty sampling method to obtain better detection accuracy in the complex sentiment analysis process. Xiong et al. [35] modeled a region-based convolutional neural network (R-CNNGSR) using sparse group regularization for image sentiment classification. In addition, some methods incorporate the model weights learned from a large-scale general dataset and further fine-tuned the CNNs for visual sentiment prediction. Shaojing et al. [36] designed a DNN based model to learn how the image sentiment is related to human attention using the EMOd dataset. Kaikai Song et al. [37] presented a novel Sentiment Networks with visual Attention (SentiNet-A) and a multi-layer saliency detection neural network framework to estimate the attention distribution in the task of visual sentiment analysis. These methodologies have attained adequate performance on visual sentiment analysis. However, the complexity present in visual content extraction hinders the performance of visual sentiment analysis and still lags behind textual sentiment analysis.

Problem Statement

- The prime challenge encountered in the active learning framework is in analyzing the sentiments in the complex environment. Hence, uncertainty sampling techniques need to be employed in this active learning technique to analyze the sentiments in the complex environment.
- The RA-DL Net architecture faces a hectic challenge in integrating the multi-modal data in the form of speech and analyzing the multi-modal for sentiment classification.
- One of the prime challenges experienced in the Visual analysis using the SVM classifier is to restrain the issues related to the multi-modal sentiment analysis. Therefore, the Deep learning model is developed to restrain the issues related to the multi-modal analysis [7].

3. Proposed Framework for Sentiment Analysis

The prime intention of the Visual sentimental analysis is to analyze the expressions of a person concerning the visualized image, which invokes the emotions such as disgust, excitation and fear. The diagrammatic representation of the proposed framework is shown in Figure 2. The entire process of visual sentiment analysis is implemented in the four-step process, such as Data augmentation, feature extraction, feature selection and image classification. First, the clipped images are generated in the data augmentation module using the processes, such as intensity variations, rotation and Scaling of the input images, subject to the feature extraction using the pre-trained models, such asVGG16, Alexnet and ResNet50. The extracted features using the pre-trained model are subjected to the feature

selection using the proposed TGSL algorithm that selects the most relevant features to reduce computational complexity. Finally, the visual sentiment classification is carried outusing the SVM classifier.

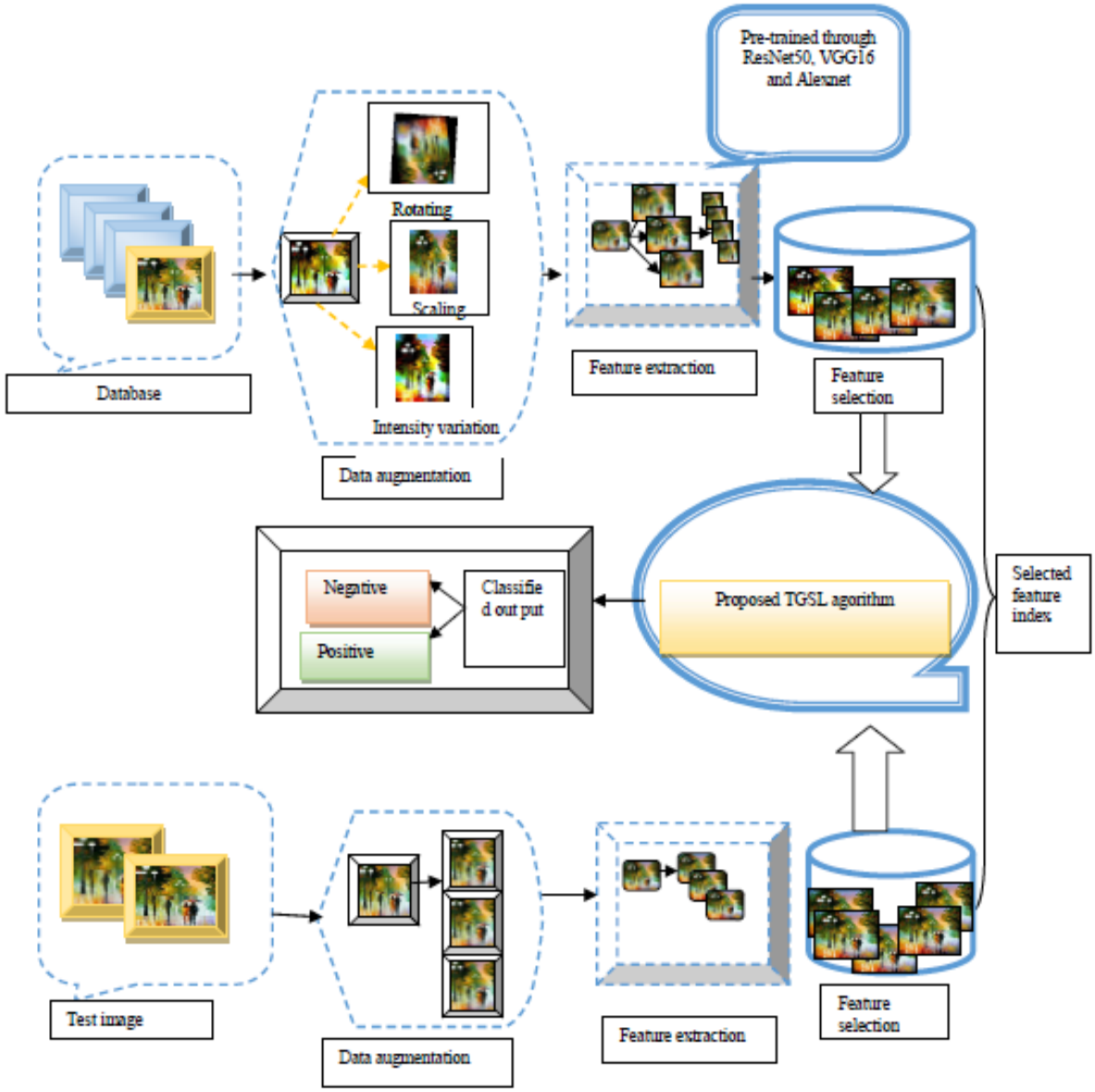


Figure 2. Framework of the proposed methodology.

3.1. Data Augmentation

The effect of image augmentation has been learned in deep neural networks [38]. Deep learning models require many training samples to build a robust model, which is hard to collect training samples with a large variety. Therefore, deep learning models developed for applications having small datasets have the risk of overfitting. Data augmentation is a strategy that facilitates the addition of replicas of the original samples using techniques such ascropping, flipping androtating, thus, increasing the variation of the training samples. Applying image augmentation to the dataset helps develop a better generalized model for large and small datasets. Furthermore, the model performance can be improved when the pre-trained models are fine-tuned and trained on the augmented data. The commonly used data augmentation techniques are scaling, rotating and Intensity variation. In this research, Scaling, Intensity variation and rotation are applied to analyze visual

sentiment. After preparing the data augmentation database, pre-processing is employed. The individual images are resized to a suitable size and converted to the RGB image for the feature extraction.

3.2. Feature Extraction through Pre-Trained Models

Feature extraction is one of the significant tasks to be implemented in visual sentiment analysis to reduce redundant data. Most relevant information from the massive datasets is extracted through the feature extraction process, which restrains the system's computational complexities. In this research, for effective feature extraction, the images are pre-trained through standard training models such as ResNet50, Alexnet and Visual Geometry Group (VGG). A brief description of the training model is given below.

3.2.1. Alexnet

The Alexnet framework restrains the working of the deep neural network, comprised of eight layers involving the five convolutional layers and three fully connected layers. Alexnet is excellent over the traditional CNNs as the possibility of extracting the features is higher than any other deep network. Moreover, it is interesting to note that the features at different layers carry various abstract features. Therefore, the feature extraction is better at the top-level layers and at the output layer, where the dimension is $[1 \times 1000]$.

Figure 3 depicts the architectural model of the Alexnet. The Alexnet utilizes the Rectifier linear units (RLUs) as an alternative to the tanh function, reducing computational time.

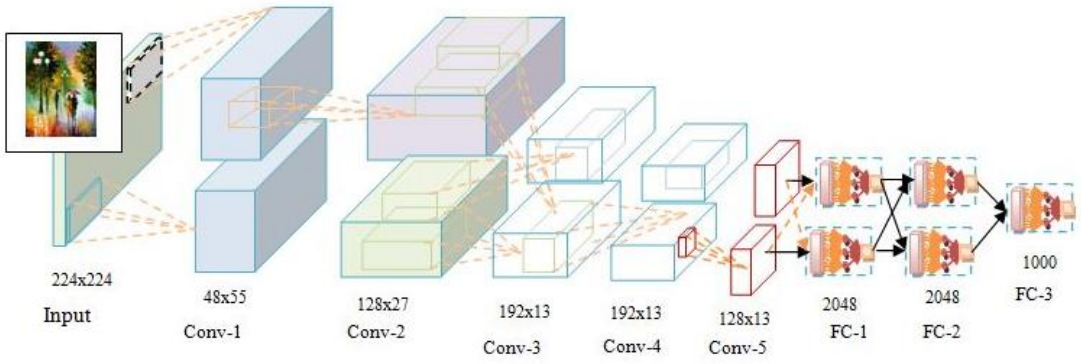


Figure 3. Architecture of Alexnet.

The input visual image is first fed to the convolutional layers in which the filter is applied to the input image. The filters utilized in the convolutional layers summarize the existence of the recognized features and thus, generate the feature maps. The first convolutional layer extracts the low-level features, such asedges, lines and corners, and the higher layer extracts the higher-level features. The feature map obtained from the convolutional layers is then subjected to the pooling layer, which is utilized to sub-sample the feature maps by summing up features in patches of the element map. On the other hand, the output from the pooling layers is fed to the fully connected layers in which the weights are applied to each input to classify the image. The final features are extracted from the eighth fully connected layer, considered the features for sentimental visual classification.

3.2.2. ResNet50

ResNet is the widely utilized training model among the deep network, which acquires the first position in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and Common Objects in Context (COCO). The ResNet is considered one of the innovative solutions for the vanishing gradient issues that result in performance saturation. ResNet has several stacked residual units and developed many different networks with deep layers: 18, 34, 50, 101, 152 and 1202. A deeper CNN stacked with more layers has the vanishing gradient problem, and this architecture overcomes it by the residual learning

module. The single residual block of the ResNet50 is demonstrated in Figure 4. The overall architecture of ResNet50 is illustrated in Figure 5. ResNet-50 uses the bottleneck architecture; it has six units, the convolutional layer followed by four stacked residual blocks of 3×3 convolutional layers and a fully connected layer; this can quickly train the network. Finally, the 100th fully connected layer features are gathered for establishing the feature vector, where the dimension is $[1 \times 1000]$ for an input image.

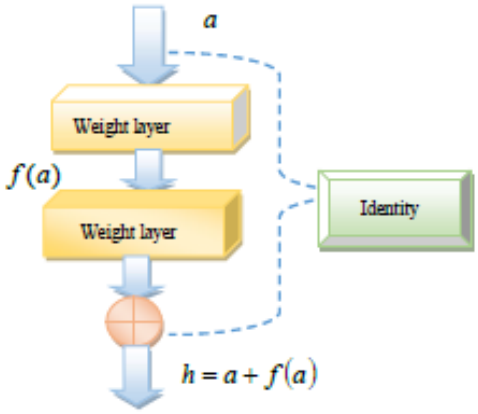


Figure 4. Single Residual block.

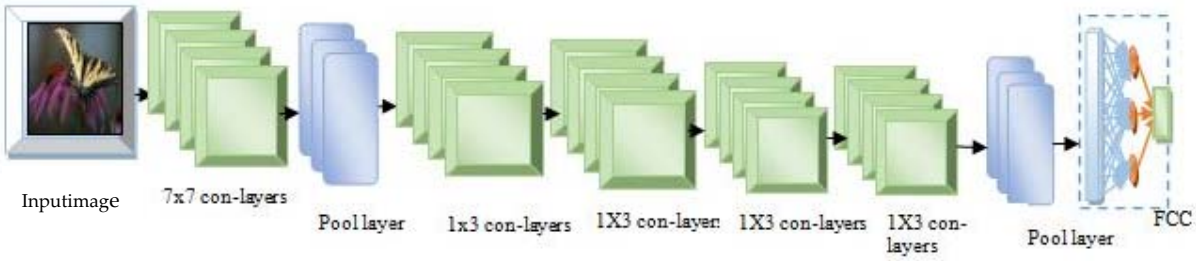


Figure 5. ResNet50 architecture.

3.2.3. VGG

VGG is one of the advanced frameworks of the Convolutional Neural network, which focuses on enhancing the precision of the network. VGG 16 architecture comprises 13 convolutional layers and three fully connected layers. The input image is fixed at $[224 \times 224]$, which is fed to the stack of the convolutional layers, and a small filter of size $[3 \times 3]$ is utilized in the VGG16 architecture as an alternative to large filters. The final fully connected layer features are derived as the feature whose dimension is $[1 \times 1000]$, and the very last layer is the SoftMax layer. The architecture of VGG16 is depicted in Figure 6.

3.3. Proposed TGSL Algorithm for Feature Selection

The feature selection plays a significant role in the classification problems as it tackles the dimensional reduction issues thereby, boosting the classification performance. The algorithm called TGSL optimization is proposed in this research to restrain the highly significant features for classification. The major need for feature selection is to reduce the features that minimize the computational complexity associated with the classification. In the proposed TGSL algorithm, the characteristic features of the instructor, teaching ability and upgrading Trainee's tendency [10,18] are combined to hold advanced features in the proposed algorithm such that the proposed algorithm holds the higher tendency to render a globally optimal solution.

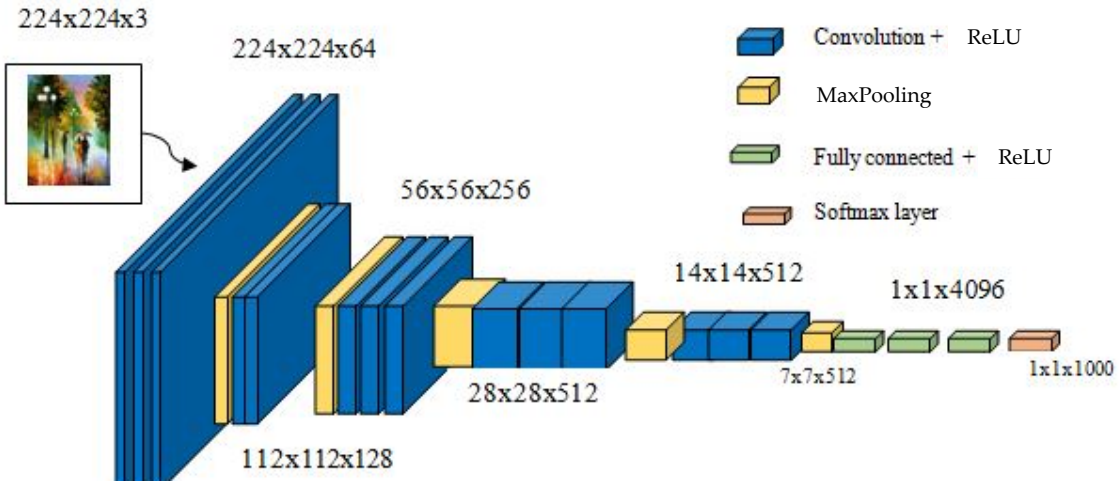


Figure 6. VGG-16 architecture.

Motivation

The TGSL algorithm is inspired by the learning and sharing characteristics of human beings. A trainee holding better outcomes in terms of grades is considered the best observer and upgraded to the instructor level. Now, the upgraded Trainee shares the knowledge gained during the learning phase with the other trainees. Thus, the proposed algorithm is comprised of both the instructor phase and the trainee phase. In the trainee phase, the Trainee obtains deep knowledge from their instructor, and they make an effort to enhance their grades. Once the Trainee is the best observer, the Trainee is upgraded to instructor level. Once the Trainee enters the instructor level, knowledge sharing occurs between the instructor and other trainees. In the proposed approach, two ways are adopted by the Trainee to enhance the knowledge. One way is that the Trainee gains knowledge from his/her instructor, and the other way is enhancing the trainees' knowledge through the mutual interaction between the other trainers. The interaction may be in the form of formal communication, group discussion and presentation. The stepwise procedure of the proposed TGSL algorithm is demonstrated in the following section below.

Step 1: Initialization of the Optimization Parameters: Initializing the parameters is the major step involved in the proposed TGSL algorithm. Let us consider the population size as S, the number of generation as N_G and the Design variables as Q_v within limits l_u and l_w. The optimization problem is characterized as min $Z_0(P)$ where, $Z_o(P)$ is the objective function and P represents the vector for design variables such that $l_u \leq x \leq l_w$.

Step 2: Population Initialization: Generating the random population is the next step to the initialization process of parameters. The generated population is based on the population size and the design variables. For example, in the proposed TGSL, the population size represents the number of Trainees under the instructor. The Generated population is expressed in the following equations.

$$Population = \begin{bmatrix} P_{1,1} & P_{1,2} & \ldots\ldots & P_{1,Q} \\ P_{2,1} & P_{2,2} & \ldots\ldots & P_{2,Q} \\ . & . & & . \\ . & . & \ldots\ldots & . \\ . & . & & . \\ P_{S,1} & P_{S,2} & \ldots\ldots & P_{S,Q} \end{bmatrix} \quad (1)$$

Step 3: Fitness Evaluation: The fitness evaluation of the solutions to be fittest depends on the maximal value of the fitness. In this phase, the instructor's solutions are decided for

which the Trainee gains knowledge to becomepromoted as the instructor. The fitness is based on accuracy and is formulated as,

$$Acc = \frac{Tru_p + Tru_N}{Pos_r + Neg_r} \tag{2}$$

where, Acc represents the accuracy, Tru_p refers to the true positive value, Tru_N refers to the true negative value, Neg_r refers to the real negative cases, Pos_r refers to the real positive cases.

Step 4: Instructor Phase: The main intention of the instructor is to enhance the Mean score of the trainers to upgrade the trainer's position. The best instructor always provides significant consideration to their Trainee, and they make an effort to shift the grade of their Trainee. Hence, the mean of the Trainee for each subject is estimated in this phase, and the mathematical model for the estimated mean is expressed as,

$$\mu_Q = \left[\mu_1, \mu_2 \ldots\ldots \mu_Q\right] \tag{3}$$

The best solution act as the instructor in this level and the upgraded Trainee is expressed as,

$$P_I = P_{Z_0(p)} = \min \tag{4}$$

As mentioned above, the instructor makes an effort to shift the mean of his Trainee to the instructor I, and the new mean obtained is mathematically expressed as,

$$\mu_Q^{t+1} = P_{I,Q} \tag{5}$$

The variations between the new and previous means are estimated in this phase to determine the Trainee's progress. The difference or variation of the two means is mathematically expressed as,

$$\partial_Q = V\left(\mu_Q^{t+1} - I'.\mu_Q\right) \tag{6}$$

where, ∂_Q represents the variation of mean, V is the random variable, I' is the instructing factor. The obtained difference is added with the existing solution to obtain the update its value.

$$P_Q^{t+1} = P_Q^t + \partial_Q \tag{7}$$

If the fitness of the newly generated solution is better, the updated solution P_Q^{t+1} is restored as the best solution of the current instance, while if the P_Q^{t+1} holds lower fitness than the fitness of the best solution corresponding to the previous iteration, the best solution of the previous iteration is maintained as the best solution and discards P_Q^{t+1} during this time.

Step 5: Trainee Phase: The observer can increase his/her knowledge through two different strategies. The first strategy of increasing the knowledge happens through their instructor and the second strategy of knowledge gaining occurs through the interaction between the observers. Let as consider the two trainees T_i and T_j. The updated equation is given as,

$$P_{i,Teach}^{t+1} = \begin{cases} P_i^t + V_i\left(T_i - T_j\right); if\ Z(T_i) < Z\left(T_j\right) \\ P_i^t + V_i\left(T_j - T_i\right); else \end{cases} \tag{8}$$

where, $Z(T_i)$ and $Z\left(T_j\right)$ denote the fitness measure of the trainees. In the mutual interaction, the senior trainees provide more effective solutions as they gain more knowledge than the junior trainees. Hence, the solutions shared by the senior level are utilized in this TGSL algorithm, and the updated gained function value is expressed in the following equation.

$$P_{i,gain}^{t+1} = \begin{cases} P_i^t + G_d * (P_{best} - P_{worst}) + \left(P_j - P_i^t\right); if\ Z\left(P_i^t\right) \le Z\left(P_j\right) \\ P_i^t + G_d * (P_{best} - P_{worst}) + \left(P_i^t - P_j\right); else \end{cases} \tag{9}$$

where, the G_d represents the knowledge factor in the above equation.

The above equation is the equation that highlights the mutual sharing features, and the mutual interaction of the trainees is considered through which the trainees gain knowledge sooner to become the instructor and the tendency to avoid the global optimal convergence of the solution is boosted. The integration of the mutual interaction with the trainee phase is discussed below.

Case 1. $Z(T_i) < Z(T_j)$

In this phase, the fitness of the trainees is verified and the trainee with the best fitness becomes the Instructor of the iteration, which is the best solution.

$$P_i^{t+1} = 0.5\left[P_i^t + V_i(T_i - T_j)\right] + 0.5\left[P_i^t + G_d * (P_{best} - P_{worst}) + \left(P_j - P_i^t\right)\right] \quad (10)$$

$$P_i^{t+1} = P_i^t + 0.5\left[V_i(T_i - T_j) + G_d * (P_{best} - P_{worst}) + \left(P_j - P_i^t\right)\right] \quad (11)$$

Case 2. $Z(T_i) > Z(T_j)$

$$P_i^{t+1} = P_i^t + 0.5\left[V_i(T_j - T_i) + G_d * (P_{best} - P_{worst}) + \left(P_i^t - P_j\right)\right] \quad (12)$$

Hence, the solution update for the proposed TGSL is given by,

$$P_i^{t+1} = \begin{cases} P_i^t + 0.5\left[V_i(T_i - T_j) + G_d * (P_{best} - P_{worst}) + \left(P_j - P_i^t\right)\right]; Z(T_i) < Z(T_j) \\ P_i^t + 0.5\left[V_i(T_j - T_i) + G_d * (P_{best} - P_{worst}) + \left(P_i^t - P_j\right)\right]; Z(T_i) > Z(T_j) \end{cases} \quad (13)$$

Step 6: Termination: The steps are repeated for the maximal iterations to reveal the best solution, which are the highly relevant features for classification.

The proposed TGSL Algorithm shown in Algorithm 1.

Algorithm 1. Pseudocode of the proposed TGSL algorithm.

Proposed TGSL Algorithm

Input: P_i^t; v **solutions**
Output: P_i^{t+1}
Initialize optimization parameter
Initialize the population of the Trainee
Evaluate the fitness of the solutions
//Instructor phase
Estimate the mean μ_Q for each design variables
Update the mean with the shifted mean μ_Q^{t+1}
Estimate the difference in the mean ∂_Q
Find the update solution P_Q^{t+1}
Update the trainee phase P_i^{t+1}
If $\left(Z(T_i) < Z\left(T_j\right)\right)$
{
$P_i^{t+1} = P_i^t + 0.5\left[V_i\left(T_i - T_j\right) + G_d * (P_{best} - P_{worst}) + \left(P_j - P_i^t\right)\right]$
}
Else
{
$P_i^{t+1} = P_i^t + 0.5\left[V_i\left(T_i - T_j\right) + G_d * (P_{best} - P_{worst}) + \left(P_i^t - P_j\right)\right]$
}
Terminate

The solution dimension obtained by the comparative method such as AlexNet, ResNet 50, VGG16, VGG with SA feature selection, VGG with GOA feature selection, ResNet with TLBO feature selection and the proposed TGSL algorithm is 1000, 1000, 1000, 500, 500, 500 and 1000, respectively. The maximum iteration taken by the comparative methods to attain the solution dimension is 50.

3.4. Visual Sentiment Classification Using the Conventional SVM Classifier

The selected features using the proposed optimization are fed to the sentiment classification model for classification, for which the SVM classifier [33] is utilized in this research. The SVM classifier is well suited for unstructured and semi structured datasets. It is also helpful to solve the complex problems using the kernel function. Furthermore, the overfitting is reduced and is useful for high dimensional data classification. The SVM classifier is one of the supervised artificial intelligence, which is utilized to restrain the rectification and the classification issues. The SVM-classifier comprised of the set of the categorized data, which is utilized to classify the undefined data. The SVM classifier estimates the closeness of the data; it transfers the features into a feature space where it classifies the similar data into the same class with kernel functions.

4. Experimental Results and Discussion

The results and discussion of the proposed TGSL method are elucidated in this section. Furthermore, the performance evaluation and the comparative analysis are implemented in this research to manifest the supremacy of the proposed TGSL method.

4.1. Datasets Description

This section portrays the information about the dataset considered for the experimental setup and evaluates the proposed work's performance by comparing the state of art approaches. The visual sentiment analysis system is implemented in MATLAB over the benchmark datasets Abstract [7], ArtPhoto [7], IAPSa [39] and MART [40], and EmotionROI [41].

Abstract:includes 279 peer-rated abstract paintings consisting of color and texture without contextual content.

ArtPhoto: contains Machajdik and Hanbury built 806 artistic photographs of professional artists from a photo-sharing site.

IAPSa: a subset of the IAPSdataset comprises 395 images chosen from the standard emotion-evoking image set categorized into eight emotions.

MART: dataset contains paintings by professional artists. It contains 500 abstract painting collections obtained from the Museum of Modern and Contemporary Art of Trento and Rovereto (MART).

EmotionROI (EmoROI): has a total of 1980 images collected from Flickr in seven discrete emotion categories (happy, sad, anger, disgust, fear, joy and neutral), 330image samples for each category, along with Valance-Arousal scores. Emotion6 dataset has an emotion distribution for each image, which gives a realistic appearance. The emotions anger, disgust, fear, sadness based on valence score are considered negative emotions; remaining emotions are treated as positive. Note that we do not include images with neutral sentiment in the experiment.

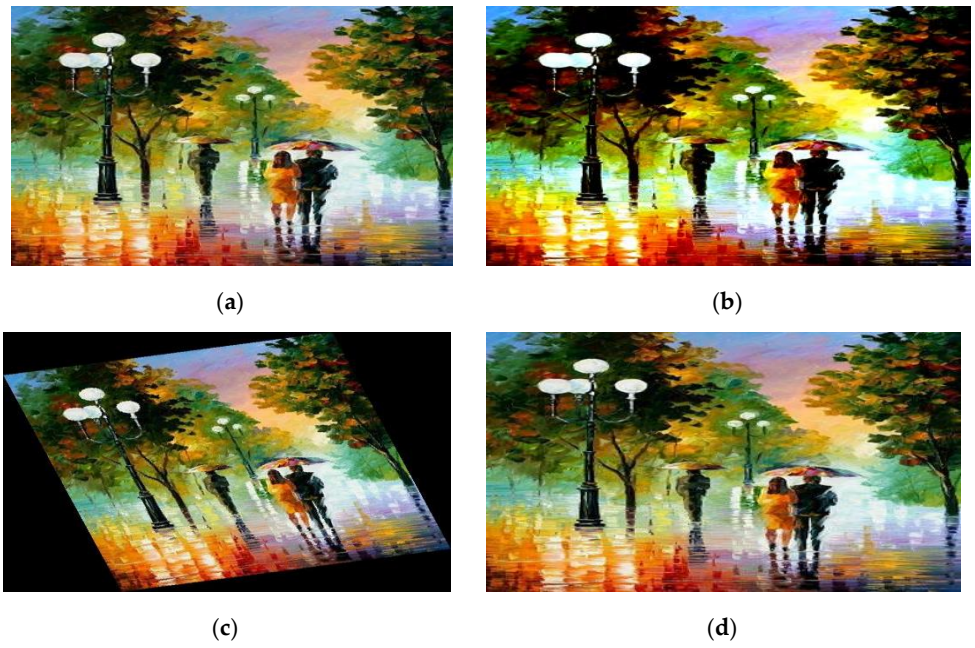
All these datasets are highly imbalanced; hence classifying into positive and negative emotion category is challenging. The dataset (IAPSa, Abstract and ArtPhoto) has eight emotions: amusement, awe, contentment, excitement, anger, disgust, fear and sadness. The first four emotions are categorized as positive, and meanwhile, the remaining four emotions as negative. The original images in each dataset had different resolutions; we resized the images to fit the standard size of the AlexNet, VGG and ResNet50 models.

4.2. Experimental Settings and Simulation Results

This section enumerates the simulation result of the proposed visual sentiment analysis. For effective sentiment analysis, the images from the five different datasets that induce the positive sentiments and the negative sentiments are utilized in this research. Due to inadequate training images per class, the proposed model uses data augmentation techniques to expand the training datasets. To validate the effectiveness of the data augmentation, the experiments were carried out in two approaches. The first approach of experiments uses the original datasets to train the CNN. The second approach first performs the data

augmentation to the original datasets and then uses augmented datasets to train the same CNN. The first and second approaches are termed as without augmentation and with augmentation, respectively. Applying image augmentation to the dataset helps develop a better-generalized model for both large and small datasets. To further investigate the data augmentation approach, classifiers are built using 80% of the augmented images for training and the remaining 20% for testing.

Figures 7 and 8 illustrates the data augmentation process of the positive sentiments and negative sentiments of the proposed Visual sentiment analysis. In this research, the data augmentation process, such as Scaling, rotation and Intensity variation, is solicited to restrain the over-fitting issues. Figure 7b depicts the Intensity variation of the input image, in which the intensity of the input image is assorted to obtain precision. Figure 7c demonstrates the rotation of the input image in which the source image is rotated either in clockwise or the anti-clockwise direction to change the position of the source image in the frame. Finally, Figure 7d demonstrates the Scaling of the input image in which the image is resized to the desired size. For the effective analysis of the negative images, the augmentation process, such as Intensity variation, rotating and Scaling is applied to the input sample, which boosts classification performance.



(**a**) (**b**)

(**c**) (**d**)

Figure 7. Data augmentation of the positive sentiment images (**a**) input image, (**b**) Intensity variation, (**c**) Rotating and (**d**) Scaling.

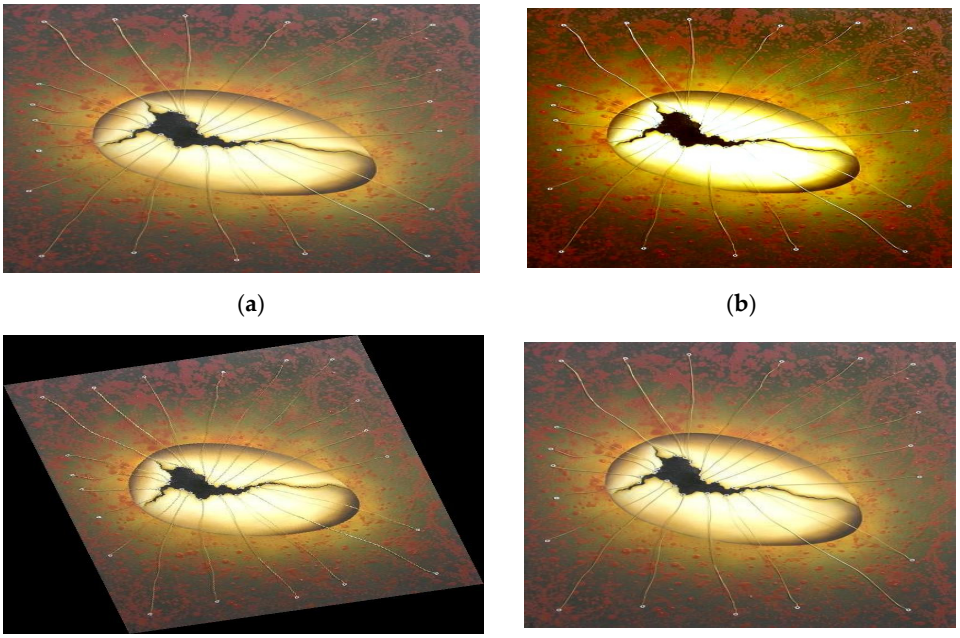
4.3. Performance Metrics

The performance metrics, such as accuracy, sensitivity and specificity, are evaluated in the Visual sentiment analysis method. A brief description of the performance metrics is given below.

Specificity: The specificity is characterized as the ratio of True Negatives to the number of real negative cases in the data as formulated as,

$$Spe = \frac{Tru_N}{Neg_r} \tag{14}$$

where, Spe in the equation demonstrates the specificity, Tru_N represents the true negative value and Neg_r represents the real negative cases.



(**a**) (**b**)

(**c**) (**d**)

Figure 8. Data augmentation of the negative sentiment images, (**a**) input image, (**b**) Intensity variation, (**c**) Rotating and (**d**) Scaling.

Sensitivity: The sensitivity is characterized as the ratio of True Positive values to the number of real positive cases, and it is mathematically expressed as,

$$Sen = \frac{Tru_p}{Pos_r} \tag{15}$$

where, Sen in the equation demonstrates the sensitivity, Tru_p represents the true positive value and Pos_r represents real positive cases.

Accuracy: The system's accuracy is characterized as the degree of closeness of obtained quantity to the real quantity. It is expressed in Equation (2).

4.4. Concurrent Methods

In order to show the effectiveness of feature fusion and the feasibility of the proposed method, two experiments were conducted on the five datasets before and after concatenating the features extracted from three pre-trained CNN models. Even though CNN is a good feature extractor, there is still a need to explore how to find the relevant features from the multi-CNN dimensional features. Therefore, the TGSL method was proposed to optimally separate the features that do not contribute much to classification. Finally, the proposed method is compared to three other methods for optimized feature selection: SA, GOA and TLBO.

The competent methods utilized in this research in order to reveal the supremacy of the proposed TGSL feature selection methods are [], Alexnet [], ResNet50 VGG16 [], Resnet50 [] +SA [42], Alexnet [] + SA [42], VGG16 [] + SA [42], Resnet50 [] + GOA [42], Alexnet + GOA [42], VGG16 + GOA [43], Resnet50 + TLBO [18], Alexnet + TLBO [10], VGG16 + TLBO [10], Hybrid features + SA [42], Hybrid features + GOA [43], Hybrid features +TLBO [18], Hybrid features + PSO, Hybrid features + GA and Hybrid features + TGSL.

4.5. Performance Evaluation

The performance evaluation of the proposed hybrid features +TGSL technique is elucidated in this section. Three metrics, namely, accuracy, sensitivity and specificity, were utilized to determine the proposed techniques' effectiveness. The performance evaluation

is carried out in two circumstances: augmentation and without augmentation for the three different CNN models considered in work. For any model to be effective, more images are desirable to train the model. Therefore, data augmentation is applied to expand the dataset by adding virtual images from each original, which eventually upgrades the proposed model performance to attain better accuracy.

Table 1 shows the performance evolution of the hybrid features + TGSL method in terms of training percentage and the 5-fold value for with and without augmentation approaches. For the 80% of training, the maximum values of the accuracy achieved by the proposed hybrid features + TGSL method with augmentation in terms of accuracy while using the datasets such as Abstract, Art photo, IAPSa, MART and Emotion ROI are 99.11%, 85.5%, 97.25%, 87.75% and 80.75%, respectively, while 99.49%, 84.50%, 98.25%, 90.15% and 79.35%, respectively, for without augmentation. The sensitivity values obtained by the proposed method with augmentation are 99.31%, 86.5%, 100%, 87.5% and 79.50%, respectively, by the datasets mentioned above, which is better compared to without data augmentation. In specificity, the maximum value obtained by the proposed method with augmentation is 99.22% for the abstract datasets, whereas the maximum specificity values without augmentation are 69.57%. Hence, from the table, the proposed hybrid features + TGSL method in case of training percentage with augmentation performs better than the hybrid features + TGSL method without augmentation. From Table 1, it can be perceived that for 5-fold cross-validation, the proposed hybrid features + TGSL method with augmentation has achieved the maximum accuracy of 99.49%, 84.50%, 98.25%, 90.15% and 79.35% for the datasets as mentioned above in comparison without data augmentation approach. Moreover, it can be observed from the results and above discussion that the hybrid features lead to richer representation of images than processing features from an individual CNN net, and besides with data augmentation effectively improves the accuracy of sentiment classification.

Table 1. Performance evaluation of the proposed hybrid features + TGSL method.

Database	Training Percentage (80%)						K-Fold (5)					
	With Augmentation			Without Augmentation			With Augmentation			Without Augmentation		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
Abstract	99.11	99.31	99.22	73.58	86.67	69.57	99.49	98.95	99.89	77.44	92.80	66.60
Artphoto	85.5	86.5	85.5	80.39	76.06	84.15	84.50	84.90	84.40	77.73	71.22	83.37
IAPSa	97.25	100	96	94.38	93.80	95.7	98.25	100	98.20	86.06	93.79	90.59
MART	87.75	87.5	90	82.11	91.66	91.70	90.15	89.00	91.70	83.57	92.64	70.48
EmoROI	80.75	79.50	82.00	80.25	79.50	96.01	79.35	79.50	96.02	78.30	63.95	96.09

4.6. Comparative Analysis

As mentioned above, the comparative analysis is performed in this research to manifest the effectiveness of the proposed TGSL feature selection method. The comparative analysis is carried out in two different scenarios, illustrated in the following sub-section.

4.6.1. Performance of Sentiment Classification with Augmentation

Comparative Analysis concerning Training Percentage

The comparative analysis is executed concerning the training percentage in the first scenario. The comparative analysis is carried out at 80% of training to obtain the peak metric values.The comparative analysis for five datasets, namely, Abstract, Artphoto, IAPSa, MART and EmoROI datasets with the comparative methods such as Alexnet (M1), Resnet 50 (M2), VGG16 (M3), Resnet50 +SA (M4), Alexnet + SA (M5), VGG16 + SA (M6), Resnet50 + GOA (M7), Alexnet + GOA (M8), VGG16 + GOA (M9), Resnet50 + TLBO (M10), Alexnet + TLBO (M11), VGG16 + TLBO (M12), Hybrid features + SA (M13), Hybrid features + GOA(M14), Hybrid features +TLBO (M15) feature selection, Hybrid features + PSO (M16), Hybrid features + GA (M17) and proposed Hybrid features +TGSL(M18) methods is illustrated in Table 2.

Table 2. Comparative analysis.

Methods/Metrics			M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
With data agumentation	Training Percentage	Accuracy																		
		Abstract	59.52	99.11	60.12	74.26	79.61	78.72	84.97	85.12	84.52	97.47	95.98	95.98	74.26	84.97	97.47	84.45	92.18	99.11
		Artphoto	65.75	80.25	53.50	75.25	75.00	74.50	77.75	76.75	79.25	83.25	80.25	80.25	75.25	77.75	83.25	76.47	80.79	85.50
		IAPSa	67.75	75.00	72.00	76.25	75.00	75.00	75.75	75.00	75.00	75.75	75.00	75.00	76.25	75.75	75.75	76.09	78.68	97.25
		MART	60.25	87.75	51.50	77.50	77.25	74.50	82.25	78.75	78.50	85.00	85.50	85.50	77.50	82.25	85.00	78.55	84.07	87.75
		EmoROI	56.00	76.00	53.50	75.25	73.00	72.50	75.00	74.25	73.50	75.00	74.00	74.00	75.25	75.00	75.00	72.38	75.57	80.75
		Sensitivity																		
		Abstract	56.25	99.31	62.85	68.06	74.31	73.96	78.13	79.51	81.25	96.88	94.79	94.79	68.06	78.13	96.88	81.40	89.83	99.31
		Artphoto	67.50	80.50	69.50	74.50	75.00	77.00	78.50	73.50	80.50	81.00	80.00	80.00	74.50	78.50	81.00	77.38	80.21	86.50
		IAPSa	54.50	100.00	85.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	96.22	100.00	100.00
		MART	68.00	87.50	60.50	78.50	74.00	70.50	84.00	77.00	79.00	86.50	84.00	84.00	78.50	84.00	86.50	79.38	84.43	87.50
		EmoROI	62.50	78.00	65.00	71.50	73.00	71.50	76.50	74.00	74.00	76.50	79.00	79.00	71.50	76.50	76.50	74.03	76.93	79.50
		Specificity																		
		Abstract	61.98	98.96	58.07	78.91	83.59	82.29	90.10	89.32	86.98	97.92	96.88	96.88	78.91	90.10	97.92	86.75	93.97	99.22
		Artphoto	64.00	80.00	37.50	76.00	75.00	72.00	77.00	80.00	78.00	85.50	80.50	80.50	76.00	77.00	85.50	75.63	81.50	85.50
		IAPSa	81.00	50.00	59.00	52.50	50.00	50.00	51.50	50.00	50.00	51.50	50.00	50.00	52.50	51.50	51.50	56.06	57.57	96.00
		MART	52.50	88.00	42.50	76.50	80.50	78.50	80.50	80.50	78.00	83.50	87.00	87.00	76.50	80.50	83.50	77.84	84.00	90.00
		EmoROI	49.50	74.00	42.00	79.00	73.00	73.50	73.50	74.50	73.00	73.50	69.00	69.00	79.00	73.50	73.50	70.72	74.21	82.00
	K-fold	Accuracy																		
		Abstract	71.70	64.15	67.92	71.70	67.92	43.40	69.81	60.38	54.72	69.81	67.92	64.15	71.70	69.81	69.81	66.16	69.54	73.58
		Artphoto	73.86	64.71	55.56	67.97	71.24	75.16	67.32	71.24	73.86	63.40	60.78	66.01	67.97	67.32	63.40	68.14	67.04	80.39
		IAPSa	71.05	73.68	72.37	76.32	73.68	72.37	76.32	73.68	72.37	76.32	73.68	72.37	76.32	76.32	76.32	74.84	76.50	84.21
		MART	69.47	70.53	52.63	71.58	68.42	72.63	69.47	61.05	68.42	64.21	67.37	65.26	71.58	69.47	64.21	68.03	69.17	82.11
		EmoROI	56.38	66.49	57.98	77.93	72.07	81.12	75.80	72.61	78.99	77.13	67.82	74.47	77.93	75.80	77.13	73.17	75.91	81.12

Table 2. *Cont.*

Methods/Metrics			M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
											Sensitivity									
		Abstract	86.67	80.00	86.67	80.00	80.00	43.33	76.67	66.67	43.33	76.67	83.33	73.33	80.00	76.67	76.67	74.79	79.05	86.67
		Artphoto	71.83	69.01	43.66	66.20	74.65	67.61	63.38	73.24	69.01	57.75	61.97	59.15	66.20	63.38	57.75	65.05	63.18	76.06
		IAPSa	80.95	100.00	71.43	57.14	100.00	50.00	57.14	100.00	50.00	57.14	100.00	50.00	57.14	57.14	57.14	71.58	68.37	100.00
		MART	81.67	75.00	48.33	81.67	76.67	83.33	75.00	66.67	71.67	71.67	71.67	68.33	81.67	75.00	71.67	74.48	75.95	91.67
		EmoROI	41.60	52.80	40.80	59.20	46.40	58.40	56.80	53.60	55.20	66.40	56.00	62.40	59.20	56.80	66.40	56.15	61.94	66.40
											Specificity									
		Abstract	52.17	43.48	43.48	60.87	52.17	43.48	60.87	52.17	69.57	60.87	47.83	52.17	60.87	60.87	60.87	55.71	59.01	69.57
		Artphoto	75.61	60.98	65.85	69.51	68.29	81.71	70.73	69.51	78.05	68.29	59.76	71.95	69.51	70.73	68.29	70.81	70.38	84.15
		IAPSa	58.82	41.18	73.53	100.00	41.18	100.00	100.00	41.18	100.00	100.00	41.18	100.00	100.00	100.00	100.00	81.07	91.60	100.00
		MART	48.57	62.86	60.00	54.29	54.29	54.29	60.00	51.43	62.86	51.43	60.00	60.00	54.29	60.00	51.43	56.96	57.55	65.71
		EmoROI	63.75	73.31	66.53	87.25	84.86	92.43	85.26	82.07	90.84	82.47	73.71	80.48	87.25	85.26	82.47	82.12	83.95	96.02
Without data agumentation	**Training Percentage**										Accuracy									
		Abstract	59.52	98.45	60.68	78.01	77.05	75.77	84.85	84.79	83.13	91.40	90.57	90.57	78.01	84.85	91.40	83.03	89.47	99.49
		Artphoto	63.30	83.45	55.70	72.20	74.15	74.20	75.70	75.35	75.95	79.45	77.05	77.05	72.20	75.70	79.45	74.71	77.91	84.50
		IAPSa	68.05	76.25	67.90	76.00	76.45	76.45	75.95	76.30	76.25	75.95	76.25	76.25	76.00	75.95	75.95	76.51	79.23	98.25
		MART	61.70	87.75	51.65	77.50	78.35	77.50	81.55	80.55	80.20	85.05	80.60	80.60	77.50	81.55	85.05	78.58	82.93	90.15
		EmoROI	57.75	77.20	56.45	74.30	71.40	71.40	76.95	74.10	73.10	78.00	73.50	73.50	74.30	76.95	78.00	72.89	76.23	79.35
											Sensitivity									
		Abstract	56.53	98.40	53.54	70.69	70.07	68.26	82.64	79.31	78.75	88.96	88.82	88.82	70.69	82.64	88.96	79.13	86.84	98.96
		Artphoto	66.50	82.50	70.70	73.00	73.60	75.20	75.80	75.60	77.50	78.60	75.30	75.30	73.00	75.80	78.60	75.74	77.36	84.90
		IAPSa	51.50	100.00	75.80	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	95.46	100.00	100.00
		MART	71.10	89.00	66.60	74.70	74.00	75.60	81.50	78.90	78.20	84.80	80.40	80.40	74.70	81.50	84.80	79.08	82.23	89.00
		EmoROI	72.20	75.30	73.50	74.50	73.00	72.60	77.10	73.50	74.40	77.40	74.50	74.50	74.50	77.10	77.40	75.06	76.41	79.50

Table 2. *Cont.*

Methods/Metrics		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18
										Specificity									
	Abstract	61.77	98.49	66.04	83.49	82.29	81.41	86.51	88.91	86.41	93.23	91.88	91.88	83.49	86.51	93.23	85.96	91.44	99.90
	Artphoto	60.10	84.40	40.70	71.40	74.70	73.20	75.60	75.10	74.40	80.30	78.80	78.80	71.40	75.60	80.30	73.70	78.51	84.40
	IAPSa	84.60	52.50	60.00	52.00	52.90	52.90	51.90	52.60	52.50	51.90	52.50	52.50	52.00	51.90	51.90	57.68	58.70	98.20
	MART	52.30	86.50	36.70	80.30	82.70	79.40	81.60	82.20	82.20	85.30	80.80	80.80	80.30	81.60	85.30	78.11	83.69	91.70
	EmoROI	43.30	79.10	39.40	74.10	69.80	70.20	76.80	74.70	71.80	78.60	72.50	72.50	74.10	76.80	78.60	70.72	76.04	79.20
										Accuracy								
	Abstract	63.91	65.05	63.91	60.91	64.63	60.91	64.64	62.03	58.29	61.26	64.68	65.80	60.91	64.64	61.26	63.77	65.14	77.44
	Artphoto	68.54	64.75	62.39	69.20	71.04	70.12	70.77	67.63	69.73	65.66	66.06	64.74	69.20	70.77	65.66	68.37	68.55	77.73
	IAPSa	72.13	85.01	68.69	77.38	81.06	76.05	80.80	81.85	76.05	78.96	83.43	76.05	77.38	80.80	78.96	78.79	80.23	86.06
	MART	69.90	67.80	49.27	75.80	69.46	72.21	74.74	68.22	68.82	70.72	68.63	69.70	75.80	74.74	70.72	70.63	73.41	83.57
	EmoROI	64.70	72.30	61.99	78.10	76.71	77.56	79.27	76.50	77.83	77.62	75.06	75.71	78.10	79.27	77.62	75.53	77.66	80.22
										Sensitivity									
	Abstract	72.95	70.47	90.13	66.49	75.72	66.52	65.76	65.23	58.58	66.47	63.27	66.49	66.49	65.76	66.47	69.98	69.68	92.80
	Artphoto	70.66	61.33	56.24	64.44	66.69	66.11	70.04	63.85	67.22	64.68	64.71	61.00	64.44	70.04	64.68	65.46	65.82	71.22
K-fold	IAPSa	82.82	80.34	73.70	79.83	82.72	82.38	76.49	84.15	82.38	82.68	86.53	82.38	79.83	76.49	82.68	81.82	83.48	93.79
	MART	87.30	72.89	39.11	83.30	79.59	82.31	77.27	75.92	75.27	75.25	73.90	73.93	83.30	77.27	75.25	76.53	78.79	92.64
	EmoROI	59.18	61.24	56.28	58.05	54.23	54.54	63.48	56.30	59.49	63.95	57.57	60.13	58.05	63.48	63.95	59.62	61.58	63.95
										Specificity									
	Abstract	51.58	57.94	29.13	53.36	50.04	53.56	63.08	57.83	57.87	54.51	66.60	64.98	53.36	63.08	54.51	56.13	60.52	66.60
	Artphoto	66.75	67.72	67.74	73.36	74.81	73.58	71.41	70.91	71.88	66.51	67.22	67.96	73.36	71.41	66.51	70.91	70.90	83.37
	IAPSa	59.06	90.59	62.66	74.12	78.82	68.54	85.88	78.82	68.54	74.12	79.41	68.54	74.12	85.88	74.12	75.86	78.11	90.59
	MART	40.43	59.21	66.52	63.05	52.29	55.08	70.48	55.14	57.92	63.06	59.81	62.40	63.05	70.48	63.06	60.78	64.62	70.48
	EmoROI	67.46	77.83	64.83	88.12	87.96	89.07	87.16	86.60	87.00	84.45	83.81	83.50	88.12	87.16	84.45	83.98	86.80	96.09

From the comparative analysis based on training and the k-fold with and without data augmentation, the proposed method outperformed all the other existing methods for all the datasets.

4.6.2. Comparative Analysis in Terms of Area under ROC Curve (AUC)

Table 3 demonstrates the comparative evaluation of the Visual sentiment analysis method concerning the five different datasets. The AUC obtained by the proposed TGSL methods with datasets, such as Abstract, Art photo, IAPSa, MART, EmotionROIis 0.9998, 0.9843, 0.9986, 0.987594 and 0.9753, respectively, whereas the conventional methods, such as Alexnet and VGG, obtain the AUC of0.9991, 0.9731, 0.9968, 0.987594 and 0.9540, respectively. Furthermore, the proposed method obtains the maximum value of AUC of 0.9998in the Abstract dataset. Hence, it is clear from the table that the proposed TGSL feature selection technique exceeds all other conventional techniques in terms of AUC.

Table 3. Comparative analysis in terms of AUC with five different datasets.

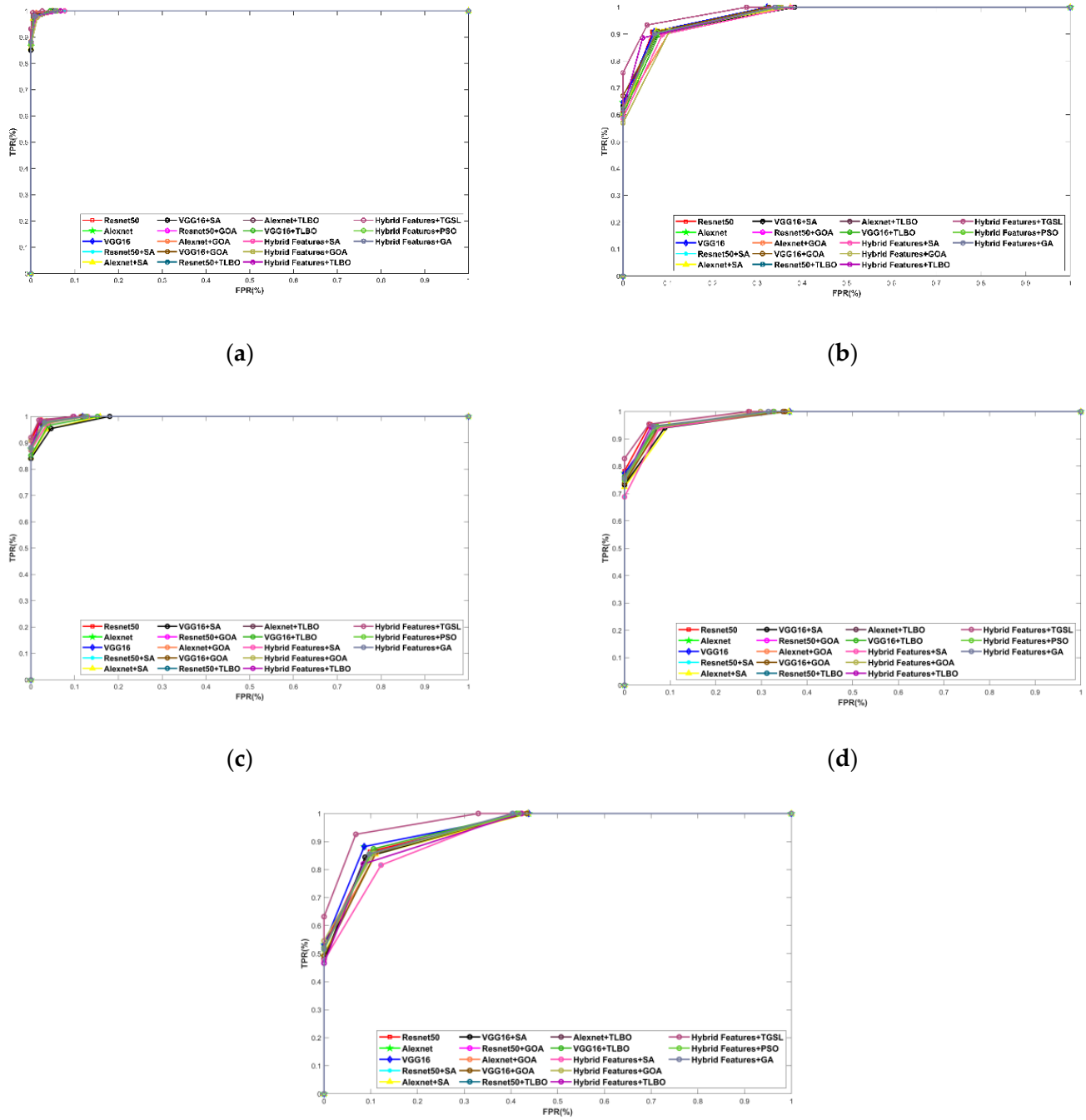
Methods	Abstract	Artphoto	IAPSa	Mart	EmoROI
Alexnet	0.9987	0.9712	0.9983	0.987594	0.9478
Resnet 50	0.9991	0.9731	0.9968	0.987594	0.9540
VGG16	0.9991	0.9731	0.9967	0.987594	0.9540
Resnet50 + SA	0.9984	0.9626	0.9959	0.987594	0.9309
Alexnet + SA	0.9989	0.9687	0.9932	0.987594	0.9406
VGG16 + SA	0.9985	0.9671	0.9922	0.987594	0.9444
Resnet50 + GOA	0.9983	0.9621	0.9974	0.987594	0.9461
Alexnet + GOA	0.9990	0.965	0.9959	0.987594	0.9436
VGG16 + GOA	0.9991	0.9725	0.9972	0.987594	0.9420
Resnet50 + TLBO	0.9987	0.9723	0.9979	0.987594	0.9396
Alexnet + TLBO	0.9990	0.9663	0.9945	0.987594	0.948
VGG16 + TLBO	0.9990	0.9663	0.9945	0.987594	0.948
Hybrid features + SA	0.9984	0.9627	0.9959	.987594	0.9309
Hybrid features + GOA	0.9983	0.9620	0.9974	0.987594	0.9461
Hybrid features + TLBO	0.9987	0.9723	0.9979	0.987594	0.9396
Hybrid features + PSO	0.8675	0.7738	0.9622	0.7938	0.8398
Hybrid features + GA	0.9397	0.8150	100	0.8443	0.8680
Hybrid features + TGSL	0.9998	0.9843	0.9986	0.987594	0.9753

4.6.3. ROC Analysis of the Comparative Methods

The receiver operating characteristic (ROC) curve is useful to validate the performance of the binary class image classification task achieved by figuring the TPR and FPR by setting different thresholds. Figure 9 demonstrates the comparative analysis of the proposed TGSL techniques in ROC to different algorithms on the five datasets. It can be discerned from the Figure 9a on the Abstract dataset TPR value of 0, 0.9319, 0.9931,1 and 1 are obtained by the proposed TGSL feature selection method for the FPR value of 0, 0, 0.0042, 0.0261 and 1, respectively. The maximum TRP value of 0.9931 is achieved through the proposed TGSL method, while the maximum TRP obtained by the comparative method such as Alexnet and VGG16 is 0.975.

4.7. Computational Complexity

The computational complexity of the proposed method and the existing methods is depicted in Table 4, shown below. The proposed method has a low computational complexity of 150 sec compared to the existing technique.



(**a**) (**b**)

(**c**) (**d**)

(**e**)

Figure 9. ROC analysis with (**a**) Abstract, (**b**) Artphoto, (**c**) IAPSa, (**d**) MART and (**e**) EmoROI.

Table 4. Computational complexity.

Methods	Computational Complexity (s)
Alexnet	268
Resnet 50	253
VGG16	186
Resnet50 + SA	212
Alexnet + SA	196
VGG16 + SA	230
Resnet50 + GOA	174
Alexnet + GOA	165
VGG16 + GOA	185
Resnet50 + TLBO	199
Alexnet + TLBO	175
VGG16 + TLBO	245
Hybrid features + SA	223
Hybrid features + GOA	189
Hybrid features + TLBO	195
Hybrid features + PSO	163
Hybrid features + GA	178
Hybrid features + TGSL	150

5. Conclusions

A visual sentiment analysis method based on the TGSL selection algorithm is proposed in this research for effective sentiment classification. In this research, three different pre-trained CNN architectural models are implemented on five widely-used affective datasets, IAPSa, Abstract, ArtPhoto, MART and EmoROI dataset, to validate the performance of the proposed method. To gain enough training images, three different data augmentation techniques were performed. Proper feature selection and classification are important in visual sentiment classification. This is accomplished by a well-organized algorithm known as TGSL is proposed to select the most relevant features from the input data. Extensive experiment results on the five datasets demonstrate that the proposed Visual sentiment analysis based on the TGSL algorithm with data augmentation hasimproved performance compared to all other conventional techniques. Furthermore, the proposed framework is effective and general because it does not require any hand-crafted feature extraction and achieves a maximum accuracy value of 99.49%, 85.5%, 98.25%, 90.15%, and 80.75% for five datasets, respectively. In the future, the robustness of the model could be improved by fusing the features with different weights and devising an effectively supervised classifier for the visual sentiment classification.

Author Contributions: Conceptualization U.K.D.K.; methodology, U.K.D.K.; software, U.K.D.K.; validation, G.V.; formal analysis, G.V.; investigation, G.V.; resources, U.K.D.K.; data curation, U.K.D.K.; writing—original draft preparation, U.K.D.K.; writing—review and editing, G.V.; visualization U.K.D.K.; supervision, G.V. 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: The data presented in this study are openly available in Abstract, ArtPhoto, IAPSa, MART, and EmotionROI datasets at https://doi.org/10.1109/ICIP.2016.7532430, https://doi.org/10.1145/2393347.2393399, https://doi.org/10.3758/BF03192732, https://doi.org/10.1007/978-3-642-34475-6_29, respectively, accessed on 23 April 2021, Reference number [7,39–41].

Acknowledgments: The authors would like to thank the National Engineering College, Department of Computer Science Engineering, Kovilpatti and Anna University Regional Campus-Tirunelveli, for providing the resources for carrying out the research work.

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

References

1. Lu, X.; Suryanarayan, P.; Adams, R.B.; Li, J.; Newman, M.; Wang, J.Z. On shape and the computability of emotions. In Proceedings of the 20th ACM international conference on Information and Knowledge Management—CIKM'11, Nara, Japan, 29 October–2 November 2012. [CrossRef]
2. Ghayoumi, M.; Bansal, A.K. Improved Human Emotion Recognition Using Symmetry of Facial Key Points with Dihedral Group. *arXiv* **2017**, arXiv:1706.07757.
3. Pang, L.; Zhu, S.; Ngo, C.W. Deep Multimodal Learning for Affective Analysis and Retrieval. *IEEE Trans. Multimedia* **2015**, *17*, 2008–2020. [CrossRef]
4. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends®Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
5. Mathews, A.P. Captioning Images Using Different Styles. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 665–668.
6. Machajdik, J.; Hanbury, A. Affective image classification using features inspired by psychology and art theory. In Proceedings of the International Conference on Big Data and Internet of Thing—BDIOT2017, London, UK, 20–22 December 2017; Association for Computing Machinery (ACM): New York, NY, USA, 2010; pp. 83–92.
7. Peng, K.-C.; Sadovnik, A.; Gallagher, A.; Chen, T. Where do emotions come from? Predicting the Emotion Stimuli Map. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 614–618. [CrossRef]
8. Campos, V.; Salvador, A.; Giró-I.-Nieto, X.; Jou, B. Diving Deep into Sentiment. In Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia, Brisbane, Australia, 30 October 2015; pp. 57–62. [CrossRef]
9. Quanzeng, Y.; Luo, J.; Jin, H.; Yang, J. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
10. Baykasoğlu, A.; Gindy, N. A simulated annealing algorithm for dynamic layout problem. *Comput. Oper. Res.* **2001**, *28*, 1403–1426. [CrossRef]
11. Yang, J.; She, D.; Sun, M. Joint image emotion classification and distribution learning via deep convolutional neural network. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3266–3272. [CrossRef]
12. She, D.; Yang, J.; Cheng, M.-M.; Lai, Y.-K.; Rosin, P.L.; Wang, L. WSCNet: Weakly Supervised Coupled Networks for Visual Sentiment Classification and Detection. *IEEE Trans. Multimedia* **2019**, *22*, 1358–1371. [CrossRef]
13. Rao, T.; Li, X.; Xu, M. Learning Multi-level Deep Representations for Image Emotion Classification. *Neural Process. Lett.* **2020**, *51*, 2043–2061. [CrossRef]
14. He, L.; Zhang, X.; Ren, S.; Sun, J. Deep Residual learning for image recognition. *arXiv* **2020**, arXiv:1512.03385v1. [CrossRef]
15. Liu, X.; Wang, C.; Wang, H.; Fu, M.; Feng, Y.; Bourennane, S.; Sun, Q.; Ma, L. Target Detection of Hyperspectral Image Based on Faster R-CNN with Data Set Adjustment and Parameter Turning. In Proceedings of the OCEANS 2019, Marseille, France, 17–20 June 2019; pp. 1–8.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
17. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
18. Saremi, S.; Mirjalili, S.; Lewis, A. Grasshopper Optimisation Algorithm: Theory and application. *Adv. Eng. Softw.* **2017**, *105*, 30–47. [CrossRef]
19. Nicolaou, M.A.; Güneş, H.; Pantic, M. A multi-layer hybrid framework for dimensional emotion classification. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Scottsdale, AZ, USA, 28 November–1 December 2011; p. 933.
20. He, X.; Zhang, W. Emotion recognition by assisted learning with convolutional neural networks. *Neurocomputing* **2018**, *291*, 187–194. [CrossRef]
21. Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.* **1954**, *61*, 81–88. [CrossRef] [PubMed]
22. Martin, S.; Lenz, R. Color based bags-of-emotions. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Munster, Germany, 2–4 September 2009; pp. 573–580. [CrossRef]
23. Zhao, S.; Gao, Y.; Jiang, X.; Yao, H.; Chua, T.-S.; Sun, X. Exploring principles-of-art features for image emotion recognition. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 47–56. [CrossRef]
24. Zhang, H.; Augilius, E.; Honkela, T.; Laaksonen, J.; Gamper, H.; Alene, H. Analyzing emotional semantics of abstract art using low-level image features. In Proceedings of the International Symposium on Intelligent Data Analysis X, Porto, Portugal, 29–31 October 2011; pp. 413–423. [CrossRef]
25. Li, B.; Feng, S.; Xiong, W.; Hu, W. Scaring or pleasing. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management—CIKM'11, Glasgow, UK, 24–28 October 2012; pp. 1365–1366. [CrossRef]
26. Liu, N.; Dellandrea, E.; Tellez, B.; Chen, L. Associating textual features with visual ones to improve affective image classification. *Computer Vision* **2011**, *6974*, 195–204. [CrossRef]

27. Yanulevskaya, V.; van Gemert, J.; Roth, K.; Herbold, A.; Sebe, N.; Geusebroek, J. Emotional valence categorization using holistic image features. In Proceedings of the 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 101–104. [CrossRef]
28. Chen, T.; Yu, F.X.; Chen, J.; Cui, Y.; Chen, Y.-Y.; Chang, S.-F. Object-Based Visual Sentiment Concept Analysis and Application. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 367–376. [CrossRef]
29. Yuan, J.; McDonough, S.; You, Q.; Luo, J. Sentribute. In Proceedings of the Second International Workshop on Exploratory Search in Databases and the Web, Chicago, IL, USA, 11 August 2013; p. 10.
30. Wang, X.; Jia, J.; Yin, J.; Cai, L. Interpretable aesthetic features for affective image classification. In Proceedings of the IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 3230–3234. [CrossRef]
31. Sicheng, Z.; Yao, H.; Yang, Y.; Zhang, Y. Affective image retrieval via multi-graph learning. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 1025–1028. [CrossRef]
32. Yadav, A.; Vishwakarma, D.K. A deep learning architecture of RA-DLNet for visual sentiment analysis. *Multimedia Syst.* **2020**, *26*, 431–451. [CrossRef]
33. Das, P.; Ghosh, A.; Majumdar, R. Determining Attention Mechanism for Visual Sentiment Analysis of an Image using SVM Classifier in Deep learning based Architecture. In Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 4–5 June 2020; pp. 339–343. [CrossRef]
34. Chen, J.; Mao, Q.; Xue, L. Visual Sentiment Analysis with Active Learning. *IEEE Access* **2020**, *8*, 185899–185908. [CrossRef]
35. Xiong, H.; Liu, Q.; Song, S.; Cai, Y. Region-based convolutional neural network using group sparse regularization for image sentiment classification. *EURASIP J. Image Video Process.* **2019**, *2019*, 30. [CrossRef]
36. Shaojing, F.; Shen, Z.; Jiang, M.; Koenig, B.L.; Xu, J.; Kankanhalli, M.S.; Zhao, Q. Emotional attention: A study of image sentiment and visual attention. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; p. 75217531.
37. Song, K.; Yao, T.; Ling, Q.; Mei, T. Boosting image sentiment analysis with visual attention. *Neurocomputing* **2018**, *312*, 218–228. [CrossRef]
38. Aquino, N.M.R.; Gutoski, M.; Hattori, L.T.; Lopes, H.S. The effect of data augmentation on the performance of convolutional neural network. In Proceedings of the Brazilian Society of Computational Intelligence, Niterói, Rio de Janeiro, 30 October–1 November 2017.
39. Yanulevskaya, V.; Uijlings, J.; Bruni, E.; Sartori, A.; Zamboni, E.; Bacci, F.; Melcher, D.P.; Sebe, N. In the eye of the beholder: Employing statistical analysis and eye tracking for analyzing abstract paintings. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management—CIKM'11, Glasgow, UK, 24–28 October 2011. [CrossRef]
40. Mikels, J.A.; Fredrickson, B.L.; Samanez-Larkin, G.; Lindberg, C.M.; Maglio, S.J.; Reuter-Lorenz, P.A. Emotional category data on images from the international affective picture system. *J. Behav. Res. Methods* **2005**, *37*, 626–630. [CrossRef]
41. Meftah, I.T.; Le Thanh, N.; Ben Amar, C. Emotion recognition using KNN classification for user modeling and sharing of affect states. In Proceedings of the International Conference on Neural Information, Doha, Qatar, 12–15 November 2012; pp. 234–242. [CrossRef]
42. Mohamed, A.W.; Hadi, A.A.; Mohamed, A.K. Gaining-sharing knowledge based algorithm for solving optimization problems: A novel nature-inspired algorithm. *Int. J. Mach. Learn. Cyber.* **2020**, *11*, 1501–1529. [CrossRef]
43. Rao, R.V.; Savsani, V.J.; Vakharia, D.P. Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput. Des.* **2011**, *43*, 303–315. [CrossRef]