

A Semantic Classification of Images by Predicting Emotional Concepts from Visual Features

Tamil Priya.D¹, Divya Udayan.J^{2*}

^{1,2} School of Information Technology and Engineering,
Vellore Institute of Technology, Vellore, Tamil Nadu, India.

*Corresponding author- divya.udayan@vit.ac.in

Article Info

Volume 81

Page Number: 42 - 70

Publication Issue:

November-December 2019

Abstract: Classification of emotion based on the image is a trivial task and hence a challenging issue in Content-Based Image Retrieval (CBIR) technique in the framework called Emotion-Based Image Retrieval (EBIR) system. This paper discusses emotion predication system that automatically predicts the semantic meaning of human emotion and uses visual features in order to ease the process of feature extraction. The emotion-based on the image is the subjectivity which plays a major role in the image retrieval process. The main motivation of this study is to predict the human emotion from an image. In the proposed work, the color, texture and shape are used as feature concepts to predict the semantics of emotion which are related with an image and features extracted by dominant color structure descriptors for color feature extraction, homogeneous texture descriptor(edge histogram) for texture feature extraction and region-based or contour-based descriptor(shape spectrum) for the shape feature extraction. Deep convolution Neural Network (Deep CNN) is used for classification into 6 basic emotion classes like anger, disgust, fear, happiness, sadness, and surprise by using color emotion factors. For training and testing the images, real-time data such as wallpaper, textile, and painting database is used. Experimental results of approach are evaluated and compared with classifier algorithms by using the measures of precision, recall, and accuracy. The prediction system also measures the arousal, valence and dominance value in order to measure the performance of EBIR approach/system.

Keywords: Content-Based Image Retrieval, color, texture, shape, descriptor, deep convolution neural network, visual features.

Article History

Article Received: 3 January 2019

Revised: 25 March 2019

Accepted: 28 July 2019

Publication: 22 November 2019

I. INTRODUCTION

Basically each human have their own habit and taste of color, pattern, the way of dressing, etc depending on the location, age, the environment they are born in. People feel cheap and manageable that is found in the design store such

as textile design, wallpaper design and painting more accurately. Textile design is the process of creating and producing structured fabric appearances which dream up the designers to suit their emotions in order to design to knitted or woven into cloths or printed patterns, curtains, and containers such as bags, table covers, beds

(coverings) and also other surface materials, art galleries and so on. Likewise, wallpaper is fabric, foil and vinyl materials that have their dominant role in the field of engineering and used as a ceiling covering which covers the ceilings of rooms, hallways, and walls. Digital wallpaper is a design or picture in the background of the display screen of a graphical user interface (GUI). Similarly, painting is an art of applying color, pigment, and paint on the surface of the wall, glass, wood, pottery, canvas, lacquer and concrete which incorporates other multiples materials including clay, paper, plaster, gold leaf, sand as well as objects.

Consequently, images play a major role in the field of image processing, multimedia and computer vision, which is recognized to be important in conveying the human emotional messages and opinion in the area of image retrieval, image analysis, emotion classification and sentiment analysis and so on. Also, the image is a powerful tool for conveying the moods and emotion of the people. Due to rapid accumulation of images on the web and in the social media, there may be an explosive growth of the multimedia data which has brought great challenges on how effectively and efficiently user indexes, retrieves and manages these rich set of resources based on their requirements. Multimedia data are heterogeneous in nature and there might be a lack of understanding and analysis of the behavior of data based on content-based search. Therefore, CBIR plays the main role in searching, analyzing and retrieving the images from a large scale database [1, 2]. CBIR is the technique which uses visual features such as color, texture, shape, and spatial layout. In this study, CBIR extracts the visual features like color, shape, and texture based on semantic concepts from the textile, wallpaper and painting images. CBIR is implemented to retrieve a similar kind of images using those

features. However, it influences the semantic gap between the low-level features such as the information requested by the client and the features extracted by machine on images. Therefore, in order to reduce the semantic gap, certain factors are taken into account such as color, pattern, objects or emotion that can be associated with a given image in a semantic manner. Emotion depends on subjectivity which deals with the emotional content of the image itself or perception it produces on the human eye. This kind of system is called Emotion-Based Image Retrieval (EBIR), which is one of the subcategories of CBIR [2].

This main focus of this work is to train the system to predict emotion based color, texture, and shape by using textile, wallpaper, and painting (artwork) images. This work of art comes with an array of colors, patterns, shapes, width or sometimes with heights due to its orientation and used to provide a finished look to materials, objects, things, and so on, in the form of designing, painting, coloring, decorating. This is the great challenge to predict the human emotion based on features concept in the field of art and design which uses the semantic concepts of emotion, which have been applied in many disciplines to study the method of image analysis, retrieval, classification and also annotation. The elements of art which are used can be listed as form, shape, line, color, value, space, and texture. Shape refers to the two dimensional, enclosed areas in the form geometric primitives. Color is the visual perception of an object which produces different sensation on the human eyes as results of the way it reflects or emits light. According to art, color is defined as an art of an element that which is produced, when the light striking on the objects and it is reflected back to the eyes. Color has three properties like hue, intensity, and value. The hue refers to the color like red, yellow, blue, green, etc. The intensity

refers to the vividness of the color and also called as colourfulness such as saturation, purity, strength, etc. The third property of the color is value is referred to as how light or dark it is and sometimes it is termed as shade and tint which refers to the value changes in color. The middle value between shades and tints is known as half-tone which is found on the value scale. In art painting, shade and tint are created by adding black and white to color. The texture is referred to as feel, visual, appearance or consistency, and characteristics structure of a surface or a substance like furry, bumpy, smooth, rough, soft, and hard. The main two forms of texture are actual and visual. The actual texture can be the feeling and cannot be visible and increases transitioning from two- dimensional to three- dimensional above the surface. Visual texture can be visible and as it is strict to two-dimensional and seems like the texture. Value in the art is defined as the degree of lightness and darkness in a color, otherwise called as contrast based on the difference in value. The sample images used in the study for emotion classification are shown in Fig.1.



Fig.1. Randomly selected sample images.
(a) Painting images (b) Textile images (c)
Wallpaper images.

The MPEG-7 visual descriptor [3, 4, 5], is applied in this study in order to extract the feature content to classify the emotional based on the visual content of an image. Hence, visual descriptors are commonly grouped as follows such as color descriptors, texture descriptors, shape descriptors, motion descriptors, and location descriptors and the properties of visual descriptors are shown in Fig.2.

In psychologically, emotion is well-defined as a multifaceted state of feeling, which has significant changes in psychologically and physically, which affects the thoughts and also the behavior of human. According to physiologist David G. Meyers, human emotion is expressed as physiological arousal, sensitive conducts, and as a conscious experience. According to the James-Lange theory of emotion [6], emotions take place due to physiological reactions towards certain events, a theory which was proposed by the psychologist William James and Carl Lange. According to the theory of emotion suggested by Cannon-Bard, Walter Cannon distressed the theory of emotion by James-Lange in numerous ways, not the same grounds. Moreover, he suggested that people can know-how physiological reactions accompanying emotions of the human without feeling those emotions really and also put forward that emotional responses can occur more quickly for those people, which can easily simply the products of physical states. According to Schechter-Singer theory suggestion [6], first physiological arousal is taking place, and then and there individual must experience feelings and identify the reason for that arousal and label those feelings as an emotion.

Accordingly, appraisal theories of emotion before live through the emotion, thinking must be occurring first. The facial feedback theory of emotions says that emotional experiences are

connected to facial expressions. According to the physiologist, Charles Darwin and William James [6], consistently physiological responds had taken

a direct effect on emotion, instead of emotional consequences. Many attempts had been

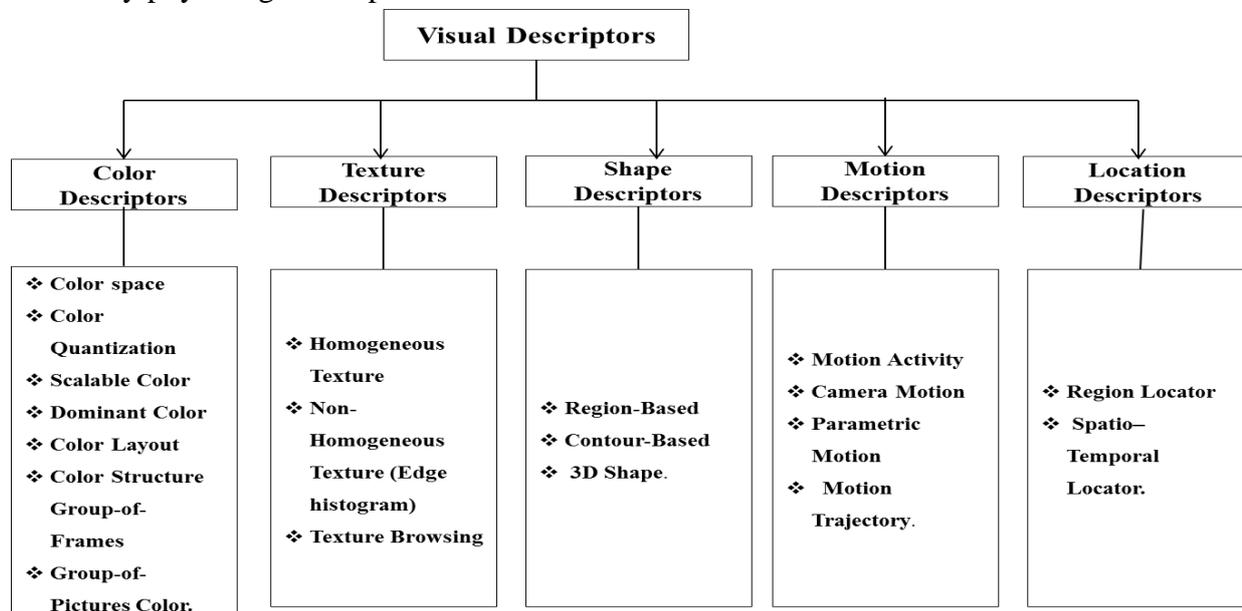


Fig.2. Summary of MPEG-7 Visual descriptors

made to analyze and to predict the reaction of human and their emotion towards images. According to survey of 20th century and by the theory of American psychologists Paul Ekman and Wallace V. Friesen the six basic emotions such as anger, fear, disgust, surprise, joy, sadness has been studied [6], which is used to study and to predict the human emotion and arousal given by particular portion of visual content.

As an analogy, in order to predict emotion based on visual features of an image to bridging the gap in perceptive content analysis in extracting the content information emotionally from images requires bridging of the affective gap measured as an emotion that can be induced in human by an image is highly subjective and different [7].

Consequently, this paper deliberates the complex challenges involved in the prediction of emotion which establishes human emotional concepts and a novel predictive framework for classification of emotion based on the feature vector of image

content by deep convolution neural network is proposed. Emotion prediction system which automatically predicts human emotion in particular image database, by which the performance of the prediction system can be improved by integrating the features vectors of an image such as color, texture, and shape, etc., instead of using a single feature, Hence, color, texture, and shape are used as feature vectors which are extracted by dominant color descriptors, homogeneous texture edge histogram descriptors, and region/contour based descriptors (shape spectrum), with the categories of three color emotion factors like activity, weight, and heat.

The remainder paper is discussed as follows: The section I, deals with introduction part which discussed about the basic introduction about textile, wallpaper, and painting, and content-based image retrieval techniques, EBIR system and description about color, texture, shape and MPEG-7 visual descriptors and final one about

description of emotion and its categories. Section II deals with the related works, section III, deals with methodology, and section IV, which deals

II. RELATED WORKS

Emotion is subjectivity according to physiological concern. This section which discusses the various studies carried out so far to predict the human emotion by prediction system framework and the semantic meaning of emotion associated with an image in the various image dataset. It also studies various classification algorithms and feature extraction methods used for image classification, image retrieval, image analysis, the image recognizing and etc., in the field of image processing and computer vision. It also analyzes various performance measures of each algorithm based on feature vectors of image concepts. Various deep learning framework, multitask-learning framework, transfer learning framework, and content-based image retrieval techniques, etc., for classification of emotion, music, voice, text, face, etc., video retrieval, speech/handwriting recognition, indexing, visual concept detection, and etc. are discussed further.

Shinet al. [1] developed a system that automatically predicts human emotion concepts from textile images. In this study, color and texture are used as feature vectors to predict emotion concepts associated with an image semantically. Consequently, color quantization and multi-level wavelet transform techniques [2] are used to extract those features and then these extracted features are applied to three typical classifier algorithms namely, K-means clustering algorithm, Naive Bayesian, and a multi-layered perceptron (MLP). However, this emotion prediction system shows better performance of accuracy of 92% for MLP in textile images for color and pattern features. Machajdik et al. [7]

with experimental results and discussion, and finally section V, briefly discuss the conclusion of the work.

developed an approach which is applied to extract the low-level features of an image collectively to represent the emotional content of an image for emotion classification. By using psychology and art theory, the author experimented theoretical and empirical concepts so as to extract the image feature in the domain of artworks with emotional expressions. Three different types of the dataset are used as the International Affective Picture System (IAPS), set of photographs from photo sharing site and painting datasets for testing and training the datasets. The experimental result shows that the IAPS dataset which shows better classification results when compared with other datasets. Szirtes et al. [8], technically predicts emotion feelings by linear ensemble classifier with 10-fold cross validations to predict the sales rate of the products via moments to moments responses across commercial media.

Su et al. [10] studied and developed a framework for indexing and classification for wallpaper and textiles images for predicting human emotions and visual impressions. In this method, two main features are used such as the core colors is used which is known as representative colors and texture (pattern) of an image is used as foreground complexity termed as pattern complexity which are implemented by using Touch Four Sides (TFS) and Touch Four Sides (TUS) algorithm used in this study to extract the exact foreground and background images. The semantic features of images are shown in Fig.3.

In this series of papers, J. Kim, Babu Kaji Baniya, Y. A Mokhsin, Y.H. Yang et al. [11] [12] [13] [14] [15] presents the classification of emotion based on music. Both tempo feature and timbre

features are combined with AdaBoost algorithm [11] to detect the highlighted segment of music and to extract music emotion classification. Rough set (RS) approach [12] is used to extract four different sets of music features such as dynamic, rhythm, spectral and harmony in which extra features are removed in order to achieve the exact music emotions. Naive Bayes classifier [13] is applied to extract musical features of emotion based on lyrics. The lyrics present in the music which represent the mood or emotion, which classify and analyze the emotion more effectively.

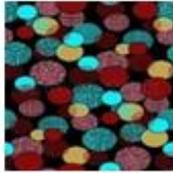
High-level semantics (the terms of abstracting emotions)	Calm	Love	Classic
High-level semantics (the terms of an object)	Flower	Tree	Bubble
Visual keywords (the terms representing features)	Blue, purple, green, violet and fan-shape	Pink, grey, brown, green and bell-shaped, curve line	Brown, orange, blue, red, pink and circle
Images			

Fig.3. Semantic feature level of images.

In Music Emotion Classification (MEC) system [14], in a song detection of the emotion features remains a great challenge in the field of research. Machine learning algorithm plays an intellectual role to learn the features of data based on definite mood or emotion, in order to classify the emotion in the selected song with the features of emotion more accurately and to extract music emotion classification [15].

Lindborg et al. [16] presented a content analysis on well-organized participants in spoken interviews, which provide the suggestion for an important emotion negotiation mechanism. Therefore, the user can tend to match the color relationship through the seeming emotion in the music. Hence, emotion can facilitate the cross-modal relationship between music and visual color based on this mixed method approaches. Deep learning framework proposed by Kim et al. [17] designed a system for emotion recognition. The framework which is designed is used to predict the emotion that extracts the objects and background features semantically by means of high-level features, so-called the affective-gap.

Emotion recognition system performance is improved better once the high level and low-level features are applied together. Emotion recognition system performance is improved better once the high level and low-level features are applied together.

Liu et al. [18] anticipated a color transfer method among images that can be fine-tuned by emotion words or by reference image. The system can adjust an image to the target emotion (for instance lovely) automatically based on the emotion word given by the user as an input. The different kinds of emotion such as fresh, antique, and happy, etc., are represented by each emotion word. Different color can arouse different emotion for each image to represent various combinations of color and to precise the different emotions. In this study, the author recommended emotion calculation method in order to calculate the target emotion from the reference image by using three color-emotion model datasets so as, it can find the appropriate color combinations. This leads to proposing a novel color transfer algorithm [19] [20] [21] [22] by using color adjustment and color combination

techniques for undertaking the color gradient and spontaneity.

Social Network takes as a suitable platform for expressing the feelings, opinion of the human through posting messages, tweets, and image with a short text or by uploading of media files. Pang et al. [23] discover the learning method for emotion prediction between low-level features through various modalities with an extremely non-linear relationship. Therefore, a joint density model is developed by using Deep Boltzmann Machine (DBM) through the space of multimodal inputs such as visual, audio and text modalities. Without any efforts of labelling, the model is trained by using user-generated contents (UGC) data [24] [25] [26]. However, the lack of certain modalities can be influenced, when the model has learned a joint representation through multimodal inputs. Moreover, this joint model which assists emotion-oriented cross-model retrieval such as, for e.g. by using text query "crazy cat" in order to retrieve videos. Hence, perhaps it retrieves any combination of media based on non-restricted types of input and output. Baohan et al. [25] studied that understanding of emotion is complex based on video due to amorphous nature of the user-generated content (UGC) and sparsity of video frames which expresses the emotion in each frame of video. Emotion is the crucial components of user-generated videos. In this paper, the author conveys the knowledge as heterogeneous sources as an external one, such as image and text data is difficult, in order to assist three associated task such as emotion recognition, emotion attribution and

emotion-oriented summarization in understanding video emotion. Specifically, this anticipated context can be used in following ways such as to learn video encoding from an auxiliary emotional image dataset that increases supervised video emotion recognition. And it also used to transfer the knowledge from auxiliary textual corpora for zero-shot recognition during the training stage of invisible emotion classes.

According to Wollmer et al. [26] automatically scrutinizing the sentiment of speaker's based on online videos which comprising of movie reviews. Additionally, it uses the audio features naturally cast-off in speech for speech-based emotion recognition along with the video which encodes valued valence information carried out by the speaker, in addition to text data. The database called Multi-Modal Movie Opinion (ICT-MMMO) corpus is used in examining the cross-validation and multi-modal experimental setup in analyzing different approaches for linguistic sentiment analysis in domain analysis. For training, the large database which contains written movie reviews for cross-domain analysis is used and then, it finally searches the online knowledge sources for gathering the speaker's sentiment in order to contest with linguistic analysis for audio-visual analysis. Campos et al. [27] experimented on fine-tuned convolution neural network [28] [29] for predicting visual sentiment analysis on social multimedia images. In this paper, author aid in investigation the visual sentiment analysis in subsequent manner, such as, first author experimented on large image dataset through more

ambivalent techniques of annotations and next it investigated by influence of initializing the weight of network by changing the source domain from fine-tuned CNN architecture based on observed perceptions and then, finally visualization process takes place on the local image regions, which contributes the overall sentiment prediction of visual images. Feng et al. [30] studied the relationship between color-emotion, to predict the emotion based on social tags. In which users can easily select the image from their upload along with the annotation tags. To encode the chromatic contents of images, two color representations have been deliberated, to choose the most dominant one used for determining color –emotions associations, based on their act of classification algorithm. Then, the image is extracted based on certain encoding and associations rules which describe the relationships between color and emotions. Classification algorithm called Apriori algorithm is used to emphasize the certain consequences of presence/absence of color on emotion and comments tagged by the users. According to Lucassen et al. [31] by addition of texture to the color samples, color emotions are changed by using three types of visual complexity such as UC, gray-scale textures, and color textures (CTs).

Jutila et al. [32] investigated the color factor that influences in creating the color preference and product preference in the field of advertisements. In this study, the HSL color system is used to carry out the saturation degree for typical blue and yellow color by medium lightness. In field advertisement, black and white color is

more predominant with the six hue-saturation combinations. The analysis is intended to accomplish in dimensions of preference, arousal, sadness, and their purchase committed by the contestants on a 7-point scale and the result which indicates that blue color is more important than yellow color to specify the sadness in this field. Elliot et al. [33] studied the prediction of emotion based on color is a relationship between color and psychological functioning of human beyond the color aesthetics. According to Wilms et al. [34] in previous studies consequences of emotion based on color is failed to control the three color dimensions such as hue, saturation, and brightness. In this study, three chromatic colors are used discretely such as hue has blue, green, red, saturation has low, medium, high, and brightness has dark, medium, bright as a color factor for designing the emotion prediction system. To analyze and predict the color emotion which is indicated by saturation and brightness by greater arousal to evaluate emotion. However, the hue also has some substantial effect on arousal which improves to red from blue and green. The valence ratings are at peak for saturated and bright colors, and moreover, it depends on hue. For both arousal and valence, the number of interaction effects has been observed for three color dimensions. For example, ratings of valence are greater for blue than the remaining hue. Gajarla et al. [35] classification of emotion generated by images has an application of automatic tagging through emotion categories of images and, automatic tagging of video

sequences into genres such as thriller, comedy, romance, etc.

Yang et al. [36] developed a multi-task framework for emotion classification, considering deep Convolutional Neural Networks (DCNNs) [37], which categories the emotion and assign a label to each image for visual sentiment prediction. Simonyan et al. [37] studied the DCNNs in large scale image recognition which evaluate the network by increasing the depth by using architecture to the scale of (3*3) convolution filters, which illustrates the major enhancement takes place in prior-art configurations that can be accomplished through pushing the depth to 16–19 weight layers. Rao et al. [38] anticipated a novel network that studies the multi-level deep representations for classification of emotion on images. In this study, the emotion is not only pretentious by high-level image semantics but also associated midlevel and a low-level, visual feature, such as color, texture and image aesthetics, is established. The deep convolutional network combines the deep representations extracted from different layer for image emotion classification and achieves constant progress in image emotion classification accuracy using fewer convolutional layers related to popular CNN models for various kind of image emotion datasets. Kim et al. [39] present a method which changes the overall mood of image in a deterministic way called an affective image recoloring method which semantically segments the source image and a target emotion based on reference image found by system commencing on the collection of images which have been tagged through

crowdsourcing using mathematically measured emotion labels. At that time, recolor the source segments by using the colors selected from target segments; however, it can conserve the gradient of the source image to produce the expected result based on image mood. The author [40] studied the various semantic based video and image retrieval techniques based on images.

Kossaifi et al. [41] address the problem of estimating valence and arousal dimensional factor in AFEW-VA dataset which provides highly precise annotations for valence and arousal values together with facial benchmarks videos intended for 600 video clips per frame that are extracted from feature films. The result shows that it is convenient for static models in estimating valence and arousal in a wide range; nevertheless, it is not as much for dynamic models, because some video clips have short duration with the low inter-variability of expressions which displayed between the frames. Moreover, it extends RCICA method which is used towards handling temporal absurdities increasing in the data. In order to solve the optimization problem in four applications such as face recognition in the heterogeneous image, multi-modal feature fusion of human behaviour analysis which is called the prediction of audio-visual content in conflict interest, face clustering, and temporal alignment of facial expressions. Panagakis et al. [42] studied a method called Robust Correlated and Individual Component Analysis (RCICA) with the presence of gross non-Gaussian noise and sparse errors called temporally misaligned data which are used to extract

features of the real world datasets in an efficient manner.

Y. Fu et al. [43], focus the issue of estimating subjective visual properties of video and image in order to recognize the visual features. For learning a prediction model and annotating visual property of image/video, uses crowdsourcing tool to collect pairwise comparison labels by unified robust learning method and to identify outliers by detection method, in order to make the annotation as more reliable method and as the rank problem. Altogether, it minimizes the cost of global inconsistency and ranking order. Robust regularization path algorithm is used to avoid exceptions completely, to handle the singularities of the key matrix, and to fit the entire solution path in a finite number of steps, B.Gu et al. [44]. The result of the algorithm shows that it fits the entire solution path in fewer steps and it reduces running time complexity than the original one does.

B.Gu et al. [45] studied the method to represent prior knowledge of the real-world problem and the structural information of data is more important. To solve this, discriminative classifier called the mini-max probability machine (MPM) is used which calculate probabilistic accuracy by minimizing the maximum probability of misclassification. Structural MPM (SMPM) which solves the sequence of the second-order programming problems more effectively. It also inferred as a large margin classifier and it can be transmuted to support vector machine (SVM) and maxi-min margin machine in definite distinct circumstances. Z. Zhou et

al [46] developed a method which is used to detect an illegal copy of images through detection method where local features are quantized into visual words intended for image matching. SIFT collects matching image between images by the Bag-of-Word (BOW) quantization process and it verifies these matching images in order to filter copy matches. The overlapping region-based method on global context descriptor (OR-GCD) is used, which increase the accuracy and efficiency inefficient manner in the filtering of copy images.

The overall review study shows that, the feature extraction without using CNN by various methods, discussed in this series of paper [4], [5], [8] [27], and [40] shows less significant result on emotion classification in the range of 57% to 72%, but CNN with feature extraction [17] [18] has accuracy result of 90% to 92%, then without using feature extraction with CNN [28] [32] [34] has accuracy range of 80% to 82% and training the CNN without feature extraction methods [23] [36] [37] and [38] shows the accuracy results in range of 77% to 79%. The overall analysis shows that, CNN with feature extraction methods shows significance performance results on emotion classification, when compared with other methods. Hence, in our work we adopt CNN with feature extraction methodology. Also, we explore deep learning techniques together with CNN to achieve better accuracy.

III. METHODOLOY

Realistically, there exist different types of colors such as red, orange, yellow, green, blue, indigo, and purple color. In view of

psychology, four primary psychological colors are present such as red, blue, yellow, and green which relates to the human body, mind, and the emotions respectively. Likewise, other psychological colors are present which emanates the properties of psychological colors which basically comprises of eleven types such as red, blue, yellow, green, violet, orange, pink, grey, black, white, and brown and all are not have same properties of emotion like positive and negative emotions. Hence, emotion based on the colors is not unique and it varies depends on the culture, age, gender, and behavior, etc.

In the field of marketing, branding, and painting, etc., color psychology is used widely. Many marketing people perceive the color as an important factor for their marketing. Since color plays a major role which impacts the customer's emotion and perceptions of their goods and services. Similarly, most of the companies prefer color when deciding their brand logos. Hence, the logos give the impression to the people and attract more customers, if the color present in the brand logos ties the personalities of the goods and services.

Preprocessing:

Each image is represented by certain emotional concepts by $E = \{1, 2, 3 \dots N\}$ that can include various emotions. Assume that each image is labeled with six emotional concepts which belong to the image i such that emotion vector is represented as $\mathbf{e}_i = (\mathbf{e}_{i, 1}, \mathbf{e}_{i, 2}, \dots, \mathbf{e}_{i, N})$. Given input image i is normalized, histogram equalization is performed and then visual feature extraction takes place.

The emotion prediction method is composed of the feature extraction process, emotion color information model and emotion classification.

Feature Extraction process:

To extract the feature vector for an input image i , which belongs to, $(i = 1, 2, 3, \dots, m)$ set of features and feature vector is denoted as f_i , which is represented as $f_i \in \Omega$. The extraction of the feature vector consists of three parts: the color feature, the texture feature and the shape feature.

Image color feature extraction is based on MPEG-7 descriptors, in which feature is extracted based on color, texture, and shape which means the color feature is extracted by dominant color descriptors, the texture feature is extracted by homogeneous texture descriptor (Edge Histogram), whereas shape feature is extracted by shape spectrum known as contour and region based descriptor. The function of each descriptor is discussed as follows:

1. Dominant Color Descriptor

Both dominant color descriptor (DCD) and color structure descriptor (CSD) are combined together as Dominant color structure descriptor (DSCS), to sense the set of scales among computational complexity and spatial redundancy of accuracy. DCD stores only the dominant colors and it is more compact while CSD is prominently accurate due to structure block scanning. To generate a dominant color structure histogram two main processes are needed, one is color quantization is performed by DCD and

second is structure block scanning is achieved using structuring window by scanning the image region. The function of the dominant color structure descriptor is discussed below:

Color Quantization:

A set of quantizing colors is represented as, $C_q = \{c_1, c_2, \dots, c_n\}$ achieved by the k-means clustering algorithm, and by closest quantizing color, is shown in equation (1) and (2)

$$c_{\min} = c_i \mid \min(d(c_{xy} \mid c_i)), c_i \in C_q \quad (1)$$

where C_{xy} , the color value found for an image pixel at the location x and y.

$$c_{xy} = \begin{cases} c_{\min}, & \text{if } d(c_{xy}, c_{\min}) < T_d \\ \text{null}, & \text{otherwise} \end{cases} \quad (2)$$

Where $d(c_{xy}, c_{\min})$ the distance among the original color and closest representative color in the dominant set and where T_d the threshold value range from 15 to 20.

c_{xy} Quantized with c_{\min} , is shown in equation (3)

$$\text{if } d(c_{xy}, c_{\min}) < T_d \quad (3)$$

Structure blocks scanning:

Each image region is scanned using a structuring window with structure block of the 8*8 structuring element. A histogram will be plotted with the count of bin value, and then it can be normalized using the number of pixels with structuring element scanned throughout the region.

2. Homogeneous Texture Descriptor (HTD)

For the similarity-based matching process of each image, HTD provides a computable categorization to texture feature. In order to compute the descriptor,

the image is first filtered by a bank of orientation and scale-sensitive filters, and then it computes mean and standard deviation for filtered outputs. To filter an image, a bank of Gabor filter is used by HTD. During the process of computing, the frequency space is divided into 30 channels at 30° interval into equivalent partitions in angular direction and octave partition in the radial direction as equally as 5 octaves. In the angular direction, the center frequencies of normalized frequency space $0 \leq w \leq 1$ of the feature channels is spaced equally in 30° interval. While in the radial direction, the center frequencies of the neighboring feature channels are spaced one octave apart and it finally calculates the mean and standard deviation of the texture feature of an image which leads to a feature vector of 62 values. Furthermore, the function of HTD is discussed as below:

In $N_{fs}(0 \leq W \leq N) = C_f$, where feature space of 30° angular direction is given in equation (4),

$$q_r = 30^\circ * r, \quad (4) \text{ In where } r \text{ is angular index with } r \in \{0,1,2,3,4,5\}$$

the radial direction, the center frequencies spaced one octave apart such that, given in equation (5),

$$\omega_s = \omega_0 * 2^{-s}, s \in \{0,1,2,3,4\} \quad (5)$$

Where s is radial index and $\omega_0 = 3/4$ is the highest center frequency. Then, image is filtered by a bank of Gabor filter is shown in equation (6) and (7),

$$G_c[i, j] = B e^{\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i \cos\theta + j \sin\theta)) \quad (6)$$

$$G_s[i, j] = C e^{\frac{(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i \cos\theta + j \sin\theta)) \quad (7)$$

HTD is then performed and shown in the equation (8),

$$T_d = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}] \quad (8)$$

f_{DC} and f_{SD} are the mean and the standard deviation of the texture feature on images. Compute the overall mean and standard deviation of the filtered images. The image texture energy (e) and then the deviation of the energy (d) are computed for the filtered channels.

3. Contour and region based descriptor (Shape spectrum)

The shape is denoted as a 2-dimensional bounded area which is used to quantitatively describe an object's shape. The shape could be geometric such as square, shape, circle, triangle, and so on or it might be organic and curvilinear. An object of the shape spectrum can be achieved by histogram creation of the shape index values. The shape index of a surface is not only independent of its location and alignment in space but also independent of its scale.

To extract the shape feature, the edge detection method is applied, and then determines the edge of the object in the scene by the following equation (9) and (10),

$$G_x = [-1 \ 0 \ 1; -2 \ 0 \ 2; -1 \ 0 \ 1] \quad (9)$$

$$G_y = [-1 \ -2 \ 1; 0 \ 0 \ 0; 1 \ 2 \ -1] \quad (10)$$

Then, sobel edge detector is applied and edge detection is performed is given in equation (11), (12) and

$$(13), X1 = \sum(\sum(G_x * A(m:m+2, n:n+2))) \quad (11)$$

$$(12) \quad X2 = \sum(\sum(G_y * A(m:m+2, n:n+2))) \quad (12)$$

$$mag(m+1, n+1) = \sqrt{X1^2 + X2^2} \quad (13)$$

So, RGB image is converted into to gray image and then edge detection is done. Hence, the shape index is extracted and a histogram of the shape index is prepared.

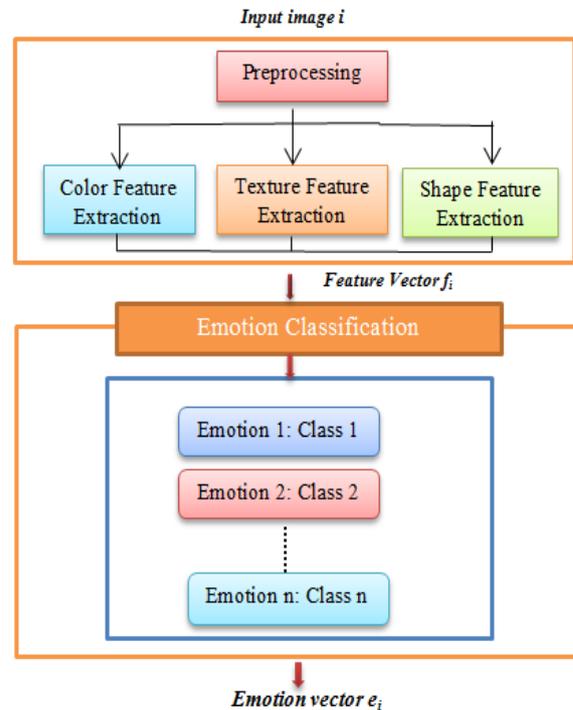


Fig.4. The pipeline of an Emotion classification method

The process of shape spectrum works as follows: (i) Edge detection is the first step to extract shape feature, if shape is used as the feature vector, (ii) to determine the edge of the object in the scene, the Sobel edge detector is used, (iii) Then, to convert RGB image to gray image, the RGB image is first pre-processed, (iv) Finally edge detection is finished. As a result, the shape index is extracted and a histogram of the shape index is prepared and the process of feature extraction is discussed. The process of shape spectrum results in a feature vector of 128 values. The pipeline of emotion classification process during training process is illustrated in the Fig.4.

Emotion information color model:

Color emotional model for single-color or multi-color combination is derived from the physiological experiment. Viewers are requested to measure ten color emotion scales based on ten colors, which are normally articulated as the semantic words, namely "warm", "soft", "active", and so on. Emotions founded color, which composed of color memory, color harmony, color meaning, and so on, which belong to the cognitive color aspects. Emotional prediction cannot be conveyed correctly for every feasible image based on color content. However, its result shows that the resultant correlates with user study on the image descriptor. The emotional color model points out that factor analysis could reduce the color emotions scales into three categories, or three color emotion factors such as *activity*, *weight*, and *heat* [9]. The experiment results of color emotion factors or color emotion scales such as activity, weight, and heat are defined in the equation (14), (15) and (16),

$$Activity = -2.1 + 0.06 * \left[(L' - 50)^2 + (x^* - 3)^2 + \left(\frac{y^* - 17}{14} \right)^2 \right]^{\frac{1}{2}} \quad (14)$$

$$Weight = -1.8 + 0.04(100 - L') + 0.45 \cos(h - 100^\circ) \quad (15)$$

$$Heat = -0.5 + 0.02(C^*)^{1.07} \cos(h - 50^\circ) \quad (16)$$

Such that L^* , x^* and y^* are coordinates, h is hue angle and C^* is Chroma of CIELAB, is shown in the equation (17) and (18). The proposed architecture for emotion classification framework is shown in Fig.5.

$$h = \arctan\left(\frac{y^*}{x^*}\right) \quad (17)$$

$$C^* = \sqrt{x^{*2} + y^{*2}} \quad (18)$$

Classification of emotion:

The feature vectors f_i of given images belongs to each emotional concept such as $n \in \{1, 2, \dots, N\}$: $\{f_i | e_{i,n} = 1, 1 \leq i \leq M\}$, to construct the classifier K is $(K: \Omega \rightarrow E)$. Here classifier is constructed using K-means clustering algorithm to derive different set of clusters.

Classification algorithm

Procedure

Input: Input image i extracts feature vector f_i

Output: Classified emotion output e_i .

Step 1: The CNN convolution layer and pooling layer is used to reduce the size of the features.

- Input layer size=number of input images i ;
- Scaling size=3;
- Output map size and kernel size=1;

Step 2: The two auto-encoder layers is used to reduce the given features further.

- The hidden layer of size is 100 which representation the auto-encoder size that specify as positive integer value. This number denotes the number of neurons in the hidden layer.
- Encoder and decoder transfer function is 'purelin', represented as below,

Linear transfer function $f(z) = z$.

- The coefficient for the L_2 weight regularizer in the cost function (Loss Function) is specified as the comma-separated by pair which consisting of 'L2WeightRegularization' and 'a positive scalar value'.
- Loss function is used for training, definite as the comma-separated by

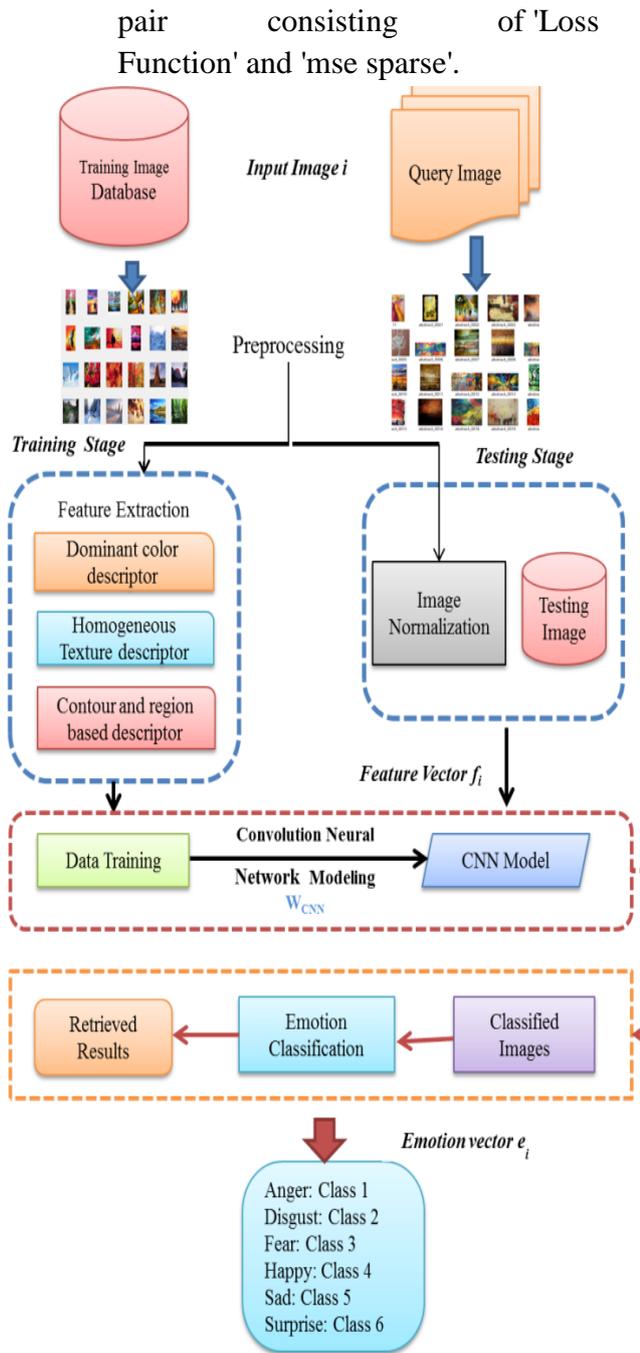


Fig.5. The proposed architecture of Emotion classification Framework.

It corresponds to the mean squared error function adjusted for training a sparse auto encoder.

$$E_i = \frac{1}{M} \sum_{m=1}^M \sum_{l=1}^L \left(X_{lm} - \hat{X}_{lm} \right)^2 + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity}, \quad (19)$$

where λ is the coefficient for the L2 regularization term and β is the coefficient for the sparsity regularization term. The values of λ and β can be specified by using the L2WeightRegularization and SparsityRegularization name-value pair arguments, respectively, while training an auto-encoder, is shown in equation (19).

Step 3: Finally DNN is trained with 2 auto-encoder layer and softmax layer to acquire good and appropriate results.

- Loss function of the softmax layer is specified as the comma-separated by pair consisting of 'Loss Function' and either 'cross-entropy' or 'mse'. mse stands for mean squared error function, which is given by the equation (20),

$$E_i = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^l \left(t_{ji} - y_{ji} \right)^2, \quad (20)$$

Where, m is the number of training examples, and l is the number of classes. t_{ji} is the ji th entry of the target matrix, T , and y_{ji} is the i th output from the auto-encoder when the input vector is x_j . The cross entropy function is given in the equation (21),

$$E_i = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^l t_{ji} \ln y_{ji} + (1 - t_{ji}) \ln(1 - y_{ji}) \quad (21)$$

For instance 'loss function' and 'mse'.

IV. EXPERIMENTAL RESULTS

This paper discusses the challenges of the novel emotional predictive framework of

emotional concepts for the classification of emotion based on the features vector of an image by using deep convolution neural network. For automatic emotion prediction system from given image database in which emotion is subjective, in which performance of the prediction system can be improved by integrating features vector of an image rather than a single feature such as color, texture, shape, and etc. Therefore, dominant color descriptors, homogeneous edge histogram descriptors, and region/contour based descriptors are used to extract the features like color, texture, and shape by three kinds of color factors like activity, weight, and heat.

Feature extraction methods:

For an image dataset and query image, preprocessing is performed in which histogram equalization and normalization is taking place. Then feature extraction process is implemented by dominant color descriptors, homogeneous edge histogram descriptors, and region/contour based descriptors in order to extract the features such as color, texture, and shape. Finally, Emotion classification is achieved by convolution neural network, in which emotion is subjective one by which the performance of the prediction system can be improved. Hence, color, texture and shape features are used to extract the features vector of emotion concepts from an image.

1. Color feature

Color is a powerful and rudimentary element to express emotions. It is effectively used by artists to persuade

emotional effects. Several studies have been conveyed to change the image color to emotion. Color is observed as a low-level dimensional feature which can directly resolve an affective-gap. However, color is still considered as a crucial factor for emotion recognition. RGB mean value and HSV color space are extracted from the color and it also calculates the HSV histogram of color. The effect of brightness is more or less similar for chromatic and achromatic colors such as (blue, blue-green, green, red-purple, purple and purple-blue) are most pleasant hues, but (yellow and green-yellow) are slightest pleasant. Green-yellow, blue-green, and green are peak arousing, while purple-blue and yellow-red are slightest arousing and then green-yellow is definite better dominance than red-purple.

Subsequently, the effect of hue is obviously transverse the sample group of the image such as Blue and green color is good whereas yellow is weak and bad, but red is strong and active and black is bad and strong and it also inactive. Grey is bad, weak, and inactive whereas white is good, weak and active. Moreover, the evaluations are interconnected intensely and positively with brightness, potency interconnected positively with darkness, and activity was associated strongly with color vs. no color.

2. Texture feature

The texture is another type of feature which can segment the image into the region of interest and to classify those regions. Texture provides information

about the spatial arrangement of colors or the intensities in an image. The texture is commonly found in natural scenes, particularly in outdoor scenes which contain both natural and man-made objects. Sand, stones, grass, leaves, bricks, and many more objects create a textured appearance in images like masculine–feminine, hard–soft, and heavy–light. Three types of sample features are used such as uniform color (UC), gray-scale textures, and color textures (CTs) which increases the visual complexity of features. Texture scales of emotions features are warm–cool, masculine–feminine, hard–soft, like –dislike, and heavy–light.

3. Shape feature

The shape is denoted as 2-dimensional, bounded area and which used to quantitatively describe an object's shape. The shape could be geometric such as square, shape, circle, triangle, and so on or it might be organic and curvilinear. The information present in an image with respect to shape on the basis of significant concepts of an edge. A histogram of the edge directions is used to represent the shape attribute. The information present in an image dataset about edge attributes is generated in the preprocessing stage using the canny edge operator. A histogram intersection technique is used to retrieve shape-based features. Matching the histograms of the edge directions is not characteristics of scale invariant or rotation. Normalizing this histogram with respect to the number of edge points in an image solves the scale invariance problem. Rotation of an image only shifts the

histogram bins, so a match across all possible shifts (a function of bin quantization level) solves the rotation invariance problem. A histogram intersection technique also allows for matching only parts of the two images.

Emotion classification method:

Our task of emotion prediction involves the use of neural network with many layer like input layer, encoder, decoder, softmax layer and output layer. Hence, we prefer deep CNN to classify the emotion based on feature vectors such as color, texture, and shape in textile, wallpaper, and painting images have been discussed. As well as, the step by step process of the neural network is discussed briefly in following subsections.

CNN model for Emotion classification:

CNN is usually composed of two layers. Layer 1, is the encoder layer which used to train the different set auto-encoder to extract the different set of features based on training data set. Layer 2, called softmax layer which stacks the encoders from the auto-encoders together with the softmax layer to form a stacked network for classification of emotion. CNN convolutional layer and pooling layer is used to reduce the size of the features. Two layers of auto-encoder are used to further reduce the given features. Finally, DNN is trained with 2 auto-encoder layer and softmax layer to get good and appropriate results.

(i) Input Layer. Training data is loaded in the input layer of the network and it is forwarded to the next hidden layer called Encoder layer.

(ii) Encoder Layer. Training an auto-encoder with a hidden layer of size 10, then perform a linear transfer function for the decoder and it sets L2 weight regularizer as 0.001, and sparsity regularizer as 4 and also sparsity proportion as 0.05. This auto-encoder uses regularizers to learn a sparse representation of the first layer. The influence of these regularizers is controlled by setting up the various parameters such as L2 weight regularization which controls the impact of L2 regularizer for the network weights. The sparsity regularization controls the impact of a sparsity regularizer. Moreover, both Sparsity Proportion and sparsity regularization attempt to controls the sparsity of the output from the hidden layer. A low value meant for sparsity proportion which typically leads to each and every neuron of the hidden layer which provides high-value output for the small number of training data and then it extracts features present in the hidden layer.

The training data contains original vectors of 579 dimensions and it is passed into the first encoder, which reduces feature vectors to 100 dimensions. Then, when it is passed to the second encoder again it reduces to 100-dimensions and then, now it trains the final layer of the network to classify these 100-dimensions of vectors into various classes. Then, train a second auto-encoder by using the features extracted from the first auto-encoder and there is no need to scale the data, and then extract the features of the hidden layer. The part of the encoder from an auto-encoder is used to learn mapping function

and to extract the features of data. Every neuron present in the encoder layer has a vector of weights; associated with each other has been tuned to respond to a particular visual feature. Then, next auto-encoder is trained by these set of extracted feature vectors by training data. And it might use the encoder by the trained auto-encoder, in order to generate the features. Similarly, train the second auto-encoder as like first auto-encoder. The main difference is that, the features that are extracted from the first auto-encoder which can use the same training data in the second auto-encoder. Moreover, it can decrease the size of the hidden layer representation so that, the encoder in the second auto-encoder can learn the even smaller representation of the input data. Next set of features can be extracted by passing the previous set of features into the encoder of second auto-encoder.

(iii) Softmax Layer. Softmax layer is trained for image classification by using features such as features1, features2, and feature3 from the second autoencoder2 and then train the network. Unlike the auto-encoders, the softmax layer is trained in a supervised fashion using labels of training data. As an individual, it can also train three dissimilar components of a stacked neural network and it can be viewed as three different types of neural networks such as encoder1, encoder2, and softmax and which might be trained and used for further process.

(iv) Output Layer. As explained above, the encoders of the auto-encoders have been used to extract feature vectors. A stacked network can be formed by the

encoders of auto-encoders along with the softmax layer for emotion classification based image concepts. Finally, it stacks

the encoders and the softmax layer to form a deep network, once the minimum

Table 1. Color Palette

Color map regions	Proportional Palette	Hex color	Area	Closest color name
Color map regions 		 #a04dda	26.6%	Medium Orchid (Violet)
Source image 		 #6f5bad	23.7%	Blue Marguerite (Violet)
		 #4b60e2	16.4%	Royal Blue (Blue)
		 #ac4751	9.2%	Hippie Pink (Brown)
		 #d5d9d5	8.8%	Aqua Haze (Grey)
		 #a1427f	4.9%	Royal Health (Red)
		 #c1252a	3.7%	Fire Brick (Red)
		 #609ab7	3.4%	Shakespeare (Blue)
		 #836675	2.2%	Old Lavender (Violet)
		 #42d6bb	1.2%	Medium Turquoise (blue)

gradient reached by the network by training samples. Images with the stacked network can be used and it can reshape the test images into a matrix and it could be ensured by stacking the columns of an image, to form a vector, and then it forms a matrix from these vectors.

Color palette:

The way of extracting a color palette from an image is a technique which would be averaged over the color values inside specific areas. Averaging of color values is practically the same as that of as averaging numbers, excluding further initial step which resulting in components such as red, green and blue of the color. The color palette representation is shown in the Table. 1., which consists of color map regions, proportional palette, hex color, coverage area and closest color name for an image.

Results and Discussion:

In 1947, Albert Mehrabian and James A. Russell developed and defines the psychological model called PAD emotional state model which measures emotional state. PAD model is defined by three numerical dimensions such as Pleasure, Arousal and Dominance are used to represent emotions class. Table. 2 discuss the two classes of PDA model, Table. 3 discuss three defined classes of PDA model and, Table. 4 discusses the defined classes of PDA model using an emotional keyword, which is briefly discussed in the below section. Pleasure or Displeasure scale is used to measures approximately pleasant or unpleasant feelings to something, for example, anger and fear together are unpleasant emotions and score on the displeasure side. On the other hand, joy is a pleasant emotion.

Likewise, Arousal or Non-arousal emotion scale measures how invigorated or tranquilizing of individual feeling. For example, anger and rage are unpleasant emotions, while rage has a greater intensity or else a greater arousal state. However, boredom is also an unpleasant state but then it has a low arousal value. Dominance or submissiveness scale represents the controlling and dominant versus controlled or submissive of individual feeling. For example, fear and anger are equally unpleasant emotions, whereas anger is a dominant emotion and fear is a submissive emotion. Block diagram of the Proposed method is shown in Fig.6., in which (a) represents the original image (b) represents preprocessing image (c) represents feature extracted image for color, shape, and texture by using color quantization, Sobel edge detector and Gabor filter feature extraction (d) represents the emotion classification.

Table 2. Two classes of PDA model.

Categorization			Rating Values “r”
Pleasure	Valence	Dominance	
High	Negative	Dominant	$r \leq 4.5$
Low	Positive	Submissive	$4.5 \leq 2$

An emotional response towards color is based on hue, saturation, and brightness color value of Pleasure-Arousal-Dominance (PDA) emotion model, but Saturation (S) and Brightness (B) supports vigorously with unswerving effects on emotions. Regression equations are expressed by the standard variables such as Pleasure, Arousal, and Dominance and the formula to measure the rating value of pleasure(P), arousal(A), and dominance(D) and it to compute the value of ‘r’ is shown in the equation (22), (23) and (24) as follows:

$$P = 0:69Y + 0:22S \quad (22)$$

$$A = 0:31Y + 0:60S \quad (23)$$

$$D = 0:76Y + 0:32S \quad (24)$$

To compare the performance of various features, all the images are sampled frequently by each image of feature vector f_i of the whole vector from the given dataset by various feature extraction techniques as discussed in section 3.

Table 3. Three defined classes of PDA model

Categorization			Rating Values “r”
Pleasure	Valence	Dominance	
Pleasant	Calm	Dominant	$1 \leq r \leq 3$
Neutral	Medium	Unpleasant	$4 \leq r \leq 6$
Unpleasant	Excited	Submissive	$7 \leq r \leq 9$

Table 4. Defined classes of PDA model using emotional keyword

Emotional classes	Affective Classes	Discrete Emotion Tagging
Pleasure	Pleasant	Joy, Happy
	Neutral	Neutral, Surprise
	Unpleasant	Anger, Disgust, Sad, Fear
Valence	Calm	Sad, Disgust, Neutral
	Medium	Happy, Joy
	Excited	Surprise, Fear, Anger
Dominance	Dominant	Anger
	Unpleasant	fear and anger
	Submissive	Fear

Then it performs a classification process to predict the emotion which iteratively uses the set of training data in order to predict the valence and arousal values for each image set using a deep convolution neural network process. The prediction values for color, texture and shape feature vector of emotion concepts for images based arousal, pleasure and dominance values, and also emotion classification respectively, by deep convolution neural network, is shown in Figure.7. The two-dimensional models for valence and arousal are shown in Fig.8.



Fig.7. Values displayed below each image show the prediction results of DCNN based on PDA model and prediction of emotion respectively, (a) the color, (b) the texture and (c) the shape feature vectors.

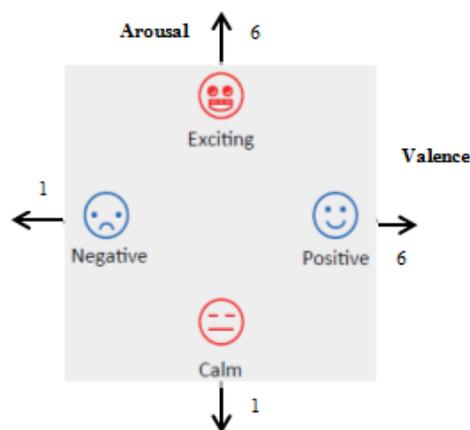


Fig.8. Valence and Arousal (VA) emotion model

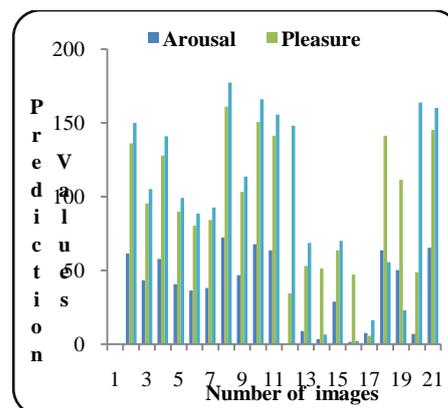


Fig.9. Distribution of the arousal, pleasure and dominance results by color feature extraction method.

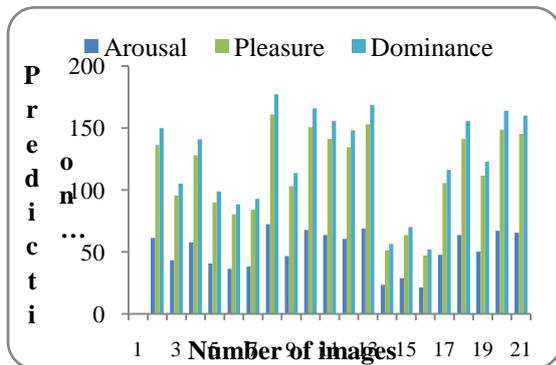


Fig.10. Distribution of the arousal, pleasure and dominance results of texture feature extraction method

It was observed that the overall color feature vector performs well and gives the preminent results for dominance emotion state when compared with another feature vector of emotional factor. This is not a surprising one such as dominance is typically characterized by dominant emotional state, directly redirected in the position of the various kinds of databases. And it also shows that the exact classification of emotion is vital. On the other hand, pleasure can be much more subtle is typically characterized the pleasant feeling of emotion, in which color feature vectors show better feeling when compared with texture and shape and arousal on low subtle does not necessarily translate in variations of emotions for colors. While texture and shape feature vectors have the same intensity level, hence it prerequisite the appearance of features and emotion classification. Among all these, unsurprisingly dominant color descriptors give the best results, what it encodes and it not only provides appearance information but it also intrinsically provides some geometric information about data when compared with other descriptors.

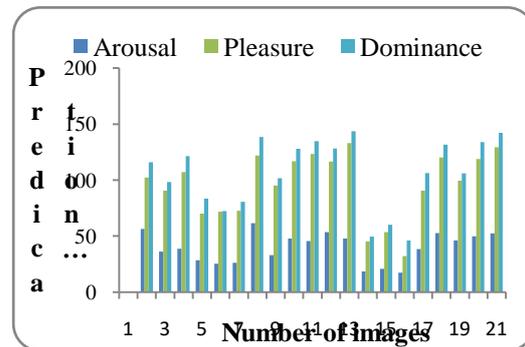


Fig.11. Distribution of the arousal, pleasure and dominance results of shape feature extraction method.

Overall, homogenous texture and region-based or contour-based descriptor (shape spectrum) have similar kind of emotion felt by the values of PDA model for both texture and shape feature vectors, whereas dominant color descriptor shows best for the values PDA method. However, three different features vectors are used in the study; color feature vectors will predict the emotions very accurately than the texture and shape and in the distribution of values.

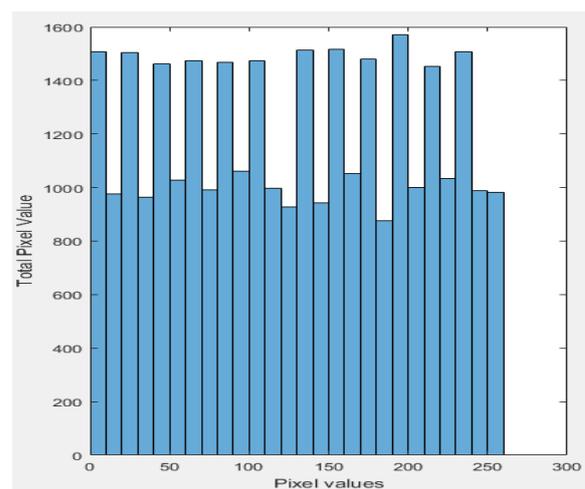


Fig.12. Histogram diagram for K-means clusters.

Distribution of the arousal, pleasure and dominance values by color feature extraction method is shown in Fig.9.

Fig.10. depicts the distribution of the arousal, pleasure and dominance values by texture feature extraction method and distribution of the arousal, pleasure, and dominance by shape feature extraction method is shown in Fig.11. In this paper, the K-means algorithm is used to form a cluster set based on the features vector of images concepts. K-means algorithm started to process the feature vectors of an image and to classify the images into the number of index and vectors based on pixel values of data in order to measure the performance of the feature vector. Fig.12. represents the histogram diagram of each pixel values of feature vectors and the performance output for each cluster of the K-means algorithm is depicted in Fig.13. Histogram diagram based on the number of pixels with certain values versus pixel values for each image of various cluster set of feature vectors is illustrated in Fig.14.

Comparison with various classification algorithms:

Various studies have been learned for emotion classification models by pre-trained weights of vectors designed for image classification.

Here, our proposed emotion prediction model is compared using SVM, GOSVM and LSSVM emotion prediction model and also with deep learning [2], [7], [9], [17], [23], [29] and by transfer learning [18]-[22]. The weight of the neural network is initialized first with the weights learned by deep learning network for the image classification. The convolution layers and the fully connected layers use the existing model structure by excluding the final output layer. Subsequently, the number of the output layer of the CNN model is

based on the number of images, which can change the number of output layers to 1. It can also calculate valence and arousal value and it measures the performance of the classification algorithm to evaluate the recall, precision, and F-measure too.

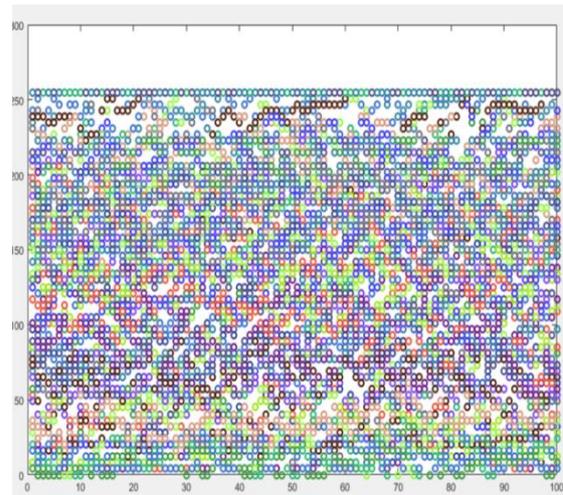


Fig.13. Performance of clustered output for each cluster by K-means clustering algorithm.

Besides, various results can be obtained in deep learning through DCNN which has encoders and softmax layers to classify emotion based on the training image given to CNN. We experiment with two conditions. Foremost, in first learning, training data is loaded in the input layer and it is forwarded to next hidden layer called Encoder layer, When it passed to the first encoder, it reduces dimensions vector and it passes the second encoder, this again reduces dimensions vectors then it can train a final layer of network to classify these dimensional vectors into different classes. The second auto-encoder layer is

trained using the features extracted from the first auto-encoder and extract the features and not to scale the data. In the second is learning, the softmax layer is trained for classification of an image using the features1, features2, and features 3 from the second autoencoder2 and train the network. It can train the three separate components of a stacked neural network separately such as encoder1, encoder2, and Softmax. Finally, stack the encoders and the softmax layer to form a deep network retrieves the image based on emotion. While train the CNN models, we used the same training and test dataset, including our proposed framework. It can train the three separate components of a stacked neural network separately such as encoder1, encoder2, and Softmax. Finally, stack the encoders and the softmax layer to form a deep network retrieves the image based on emotion. While train the CNN models, we used the same training and test dataset, including our proposed framework.

When compared with emotion prediction model of SVM, GOSVM, and LSSVM, our proposed learning model achieves best performance measure of 95%, which is shown in Fig.15. The novel proposed framework which improves the performance by 4% of accuracy when

compared with other methods by using color, texture and shape features using deep CNN by efficient feature extraction method and classification algorithm. The emotion rating also measured accurately and effectively by color factor mechanism. The emotion consists of many components like customer preference, action, subjectivity feeling that tend to increase the sales, purchases and customer needs. Emotions that bridge the gap between perceptive content analysis and human emotion induced by images called "Affective gap".

V.CONCLUSION

In this work, we presented a new emotion prediction system that automatically predicts human emotion from textile, wallpaper and painting images database with deep learning framework. The emotion-based on the image is the subjectivity which plays a major role in image classification/retrieval process. In this work the color, texture and shape feature are used as a combined feature to predicts the human emotion associated with an images whereas these features are extracted by descriptors such as dominant color descriptors for color feature extraction, homogeneous texture descriptor (edge histogram) for texture feature extraction and region-based or contour-based descriptor (shape spectrum) for the shape feature extraction.

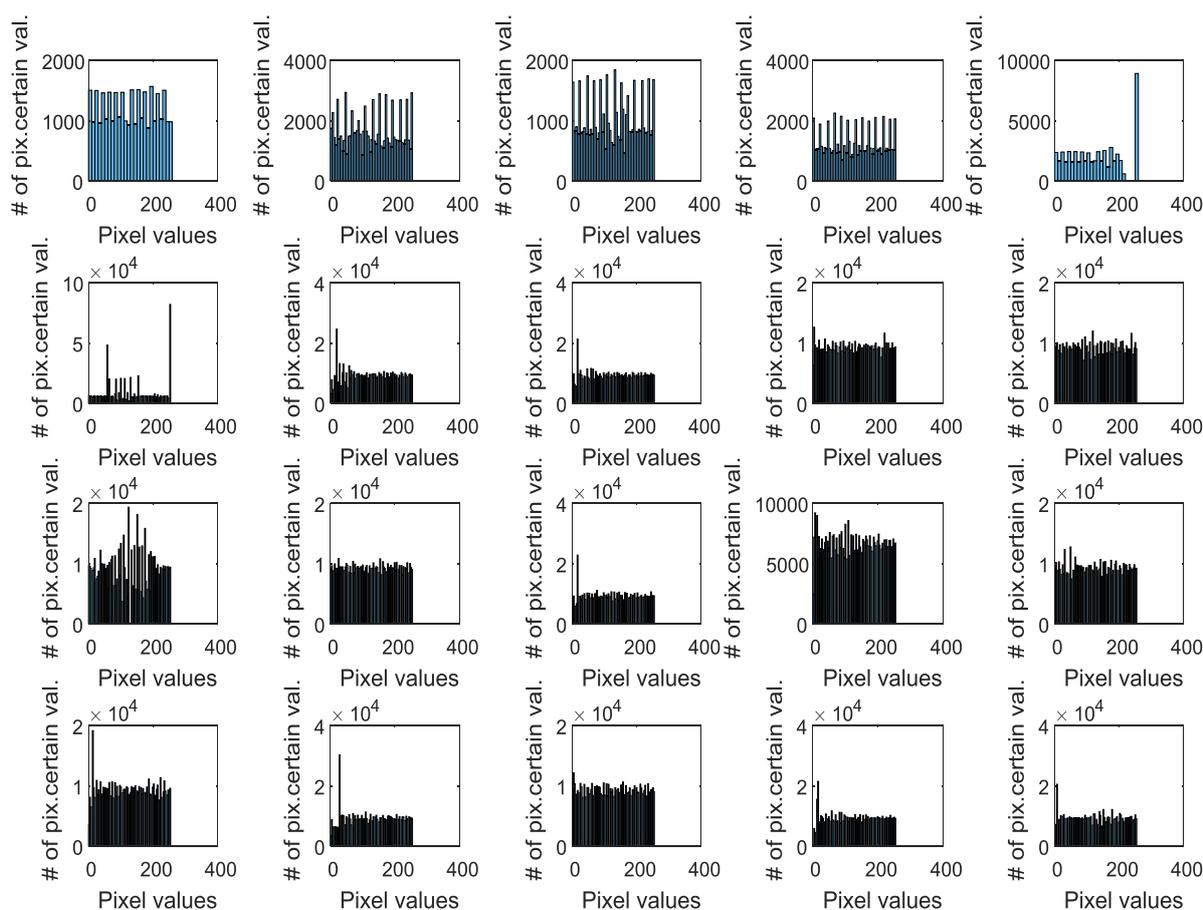


Fig.14. Illustration of a histogram diagram based on the number of pixels with certain values vs. pixel values of each image for a various cluster set of different feature vectors

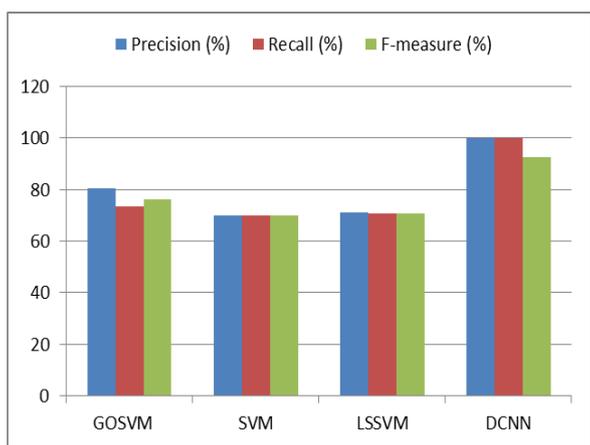


Fig.15. Performance evaluation of different classification algorithm.

Deep convolution Neural Network (Deep CNN) is used for classification of emotion into 6 basics emotion like anger, disgust, fear, happiness,

sadness, and surprise by using color emotion factors and color harmony factors. The real-time wallpaper, textile, and painting database is used for training and testing the images. Experimental results of each are evaluated and compared with different classifier algorithms, the result of the performance evaluation shows that measures of precision is 98%, recall is 98%, F-measure is 92.31% and accuracy is 95%.

The prediction system also measures the arousal value of 0.62, valence value of 0.69, to analyze the performance of EBIR approach/system. The main application of the system is to predict the emotion based on the image, to infer emotion in designing interior design, decorative wallpaper,

advertisement, artwork design, fabric design, logo and so on.

In future, the proposed work can be extended by applying multicolour combination, color transfer, and transfer learning method. Further, the discussed methodology can be extended for audio data and more complex computer vision problem like image captioning and NLP.

Acknowledgement:

The authors thank VIT University for providing “VIT SEED GRANT” for carrying out this research work.

Conflict of Interest:

The authors declare that they have no conflict of interest.

REFERENCES:

- [1]. Shin, Yunhee, Youngrae Kim, and Eun Yi Kim. "Automatic textile image annotation by predicting emotional concepts from visual features." *Image and Vision Computing* 28.3 (2010): 526-537.
- [2]. Kamatchi.T, Geethu.G.S et al. " Color, Texture, Shape Features Based Retrieval of Emotional Scenes." *International Journal of Innovative Research in Science, Engineering, and Technology*. Vol. 5, Issue 8, August 2016:14544-14552.
- [3]. Martinez, Jose M. "Standards-MPEG-7 overview of MPEG-7 description tools, part 2." *IEEE MultiMedia* 9.3 (2002): 83-93.
- [4]. Manjunath, Bangalore S., Philippe Salembier, and Thomas Sikora, eds. *Introduction to MPEG-7: multimedia content description interface*. John Wiley & Sons, 2002.
- [5]. Chang, Shih-Fu, Thomas Sikora, and Atul Purl. "Overview of the MPEG-7 standard." *IEEE Transactions on circuits and systems for video technology* 11.6 (2001): 688-695.
- [6]. <https://www.verywellmind.com/theories-of-emotion-2795717>.
- [7]. Machajdik, Jana, and Allan Hanbury. "Affective image classification using features inspired by psychology and art theory." *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010.
- [8]. Szirtes, Gabor, et al. "Behavioral cues help predict the impact of advertising on future sales." *Image and Vision Computing* 65 (2017):49-57.
- [9]. Solli, Martin, and Reiner Lenz. "Color emotions for image classification and retrieval." *Conference on Colour in Graphics, Imaging, and Vision*. Vol. 2008. No. 1. Society for Imaging Science and Technology, 2008.
- [10]. Su, Yuan-Yuan, and Hung-Min Sun. "Emotion-Based Classification and Indexing for Wallpaper and Textile." *Applied Sciences* 7.7 (2017): 691.
- [11]. J. Lee, J. Kim and H. Kim, "Music Emotion Classification Based on Music Highlight Detection," 2014 International Conference on Information Science & Applications (ICISA), Seoul, 2014, pp. 1-2.
- [12]. Babu Kaji Baniya and Joonwhoan Lee, "Rough Set-Based Approach for Automatic Emotion Classification of Music," *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 400~416, 2017.
- [13]. Y. An, S. Sun and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," *IEEE Transaction on Computer and Information Science*, (2017), pp. 635-638.
- [14]. Mokhsin, M., Rosli, N. B., Zambri, S., Ahmad, N. D., & Hamidi, S. R. (2014). Automatic music emotion classification using artificial neural network based on vocal and

- instrumental sound timbres. *Journal of Computer Science*, 10(12), 2584-2592.
- [15]. Y.H. Yang and H. Chen, "Ranking-based emotion recognition for music organization and retrieval," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 4, pp. 762–774, May 2011.
- [16]. Lindborg, PerMagnus, and Anders K. Friberg. "Colour association with music is mediated by emotion: evidence from an experiment using a CIE lab interface and interviews." *PloS one* 10.12 (2015): e0144013.
- [17]. Kim, Hye-Rin, et al. "Building Emotional Machines: Recognizing Image Emotions through Deep Neural Networks." *IEEE Transactions on Multimedia* (2018).
- [18]. Liu, Shiguang, and Min Pei. "Texture-aware emotional color transfer between images." *IEEE Access* 6 (2018): 31375-31386.
- [19]. M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326-337, 2016.
- [20]. E. Reinhard, M. Adhikari, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34-41, Sep./Oct. 2001.
- [21]. V. H. Jimenez-Arredondo, J. Cepeda-Negrete, and R. E. Sanchez-Yanez, "Multilevel color transfer on images for providing an artistic sight of the world," *IEEE Access*, vol. 5, p. 15390-15399, 2017.
- [22]. B. Xie, C. Xu, Y. Han, and R. K. F. Teng, "Color transfer using adaptive second-order total generalized variation regularizer," *IEEE Access*, vol. 6, pp. 6829-6839, 2018.
- [23]. Pang, Lei, Shiai Zhu, and Chong-Wah Ngo. "Deep multimodal learning for affective analysis and retrieval." *IEEE Transactions on Multimedia* 17.11 (2015): 2008-2020.
- [24]. X. Wang, J. Jia, J. Tang, B. Wu, L. Cai, and L. Xie, "Modeling emotion influence in image social networks," *IEEE Trans. Affect. Comput.* vol. 6, no. 3, pp. 286-297, Jul. 2015.
- [25]. Xu, Baohan, et al. "Heterogeneous knowledge transfer in video emotion recognition, attribution, and summarization." *IEEE Transactions on Affective Computing* 9.2 (2018): 255-270.
- [26]. Wollmer, Martin, et al. "Youtube movie reviews: Sentiment analysis in an audio-visual context." *IEEE Intelligent Systems* 28.3 (2013): 46-53.
- [27]. Campos, Victor, Brendan Jou, and Xavier Giro-i-Nieto. "From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction." *Image and Vision Computing* 65 (2017): 15-22.
- [28]. T. Chen, D. Borth, T. Darrell, S.-F. Chang, *DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks*, 2014.
- [29]. K.He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [30]. Feng, Haifeng, Marie-Jeanne Lesot, and Marcin Detyniecki. "Using association rules to discover color-emotion relationships based on social tagging." *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, Berlin, Heidelberg, 2010.
- [31]. Lucassen, Marcel P., Theo Gevers, and Arjan Gijzenij. "Texture affects color emotion." *Color Research & Application* 36.6 (2011): 426-436.
- [32]. Jutila, Valtteri. "Colour emotion effects in creating product preference in advertisements: Predicting purchase intent of blue and yellow colors, in advertisements of low-involvement products." (2018).
- [33]. Elliot, Andrew J., and Markus A. Maier. "Color psychology: Effects of perceiving

- color on psychological functioning in humans." Annual review of psychology 65 (2014): 95-120.
- [34]. Wilms, Lisa, and Daniel Oberfeld. "Color and emotion: effects of hue, saturation, and brightness." Psychological research 82.5 (2018): 896-914.
- [35]. Gajarla, Vasavi, and Aditi Gupta. "Emotion Detection and Sentiment Analysis of Images." Georgia Institute of Technology (2015).
- [36]. Yang, Jufeng, Dongyu She, and Ming Sun. "Joint image emotion classification and distribution learning via a deep convolutional neural network." Int. J. Artif. Intell. 2017.
- [37]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv: 1409.1556 (2014).
- [38]. Rao, Tianrong, Min Xu, and Dong Xu. "Learning multi-level deep representations for image emotion classification." arXiv preprint arXiv: 1611.07145 (2016).
- [39]. Kim, Hye-Rin, Henry Kang, and In-Kwon Lee. "Image Recoloring with Valence-Arousal Emotion Model." Computer Graphics Forum. Vol. 35. No. 7. 2016.
- [40]. Tamil Priya, D., Divya Udayan J (2019). A comprehensive survey on various semantic based video/image retrieval techniques". International Journal of Innovative Technology and Exploring Engineering 8(6), pp- 185-196.
- [41]. Kossaiji, Jean, et al. "AFEW-VA database for valence and arousal estimation in-the-wild." Image and Vision Computing 65 (2017): 23-36.
- [42]. Y. Panagakis, M.A. Nicolaou, S. Zafeiriou, M. Pantic, Robust correlated and individual component analysis, IEEE Trans. Pattern Anal. Mach. Intell. 38 (8) (2016) 1665–1678.
- [43]. Y. Fu, Yanwei, et al. "Robust subjective visual property prediction from crowdsourced pairwise labels." IEEE transactions on pattern analysis and machine intelligence 38.3 (2016): 563-577.
- [44]. B. Gu and V. S. Sheng. A robust regularization path algorithm for v-support vector classification. IEEE Transactions on Neural Networks & Learning Systems, 2016.
- [45]. B. Gu, X. Sun, and V. S. Sheng. Structural minimax probability machine. IEEE Transactions on Neural Networks & Learning Systems, 2016.
- [46]. Z. Zhou, Y. Wang, Q. M. J. Wu, C. N. Yang, and X. Sun. Effective and efficient global context verification for image copy detection. IEEE Transactions on information forensics and security, 2016.

Author Biography:

Tamil Priya D received the M.E degree in Computer Science and Engineering from Anna University, India. She is an Assistant Professor in the Department of Information Technology, VIT University, India. Her research interests include image processing, data mining, computer vision and multimedia, Semantic Indexing, and Machine learning.

Divya Udayan J received a Ph.D. degree in Internet and Multimedia Engineering from Konkuk University, South Korea. She is an Associate Professor in the Department of Information Technology, VIT University, India. Her research interests include semantic modeling, image processing, recognition, and classification, augmented reality and HCI.