International Conference on Advanced Engineering, Science, Management and Technology – 2019 (ICAESMT19)

Sarc-M: Sarcasm Detection in Typo-graphic Memes

Akshi Kumar[a*], Geetanjali Garg[b]

[a,b]Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India-110042
**akshikumar@dce.ac.in*

ARTICLEINFO

Article history:

Keywords:
Sarcasm
Memes
OCR
Sentiment

ABSTRACT

Detecting sarcastic tone, which conveys a sharp, bitter, or cutting expression, remark or taunt in natural language is tricky even for humans, making its automated detection more arduous. The growing use of typo-graphic images, that is text represented as an image further characterizes the power of expressiveness in online social data. This research proffers a model Sarc-M, a sarcastic meme predictor, for sarcasm detection in typo-graphic memes using supervised learning based on lexical, pragmatic and semantic features. The learning model is evaluated using five different classifiers and the results are evaluated using a balanced dataset of typo-graphic images, called MemeBank, scrapped from Instagram. The contribution of the research is two-fold, firstly, typo-graphic text is extracted using optical character recognizer and then analyzed for sarcasm and secondly for detecting sarcasm the need of contextual information is explored, that is, contextual cues such as frequency of punctuations and sentiment words are considered as features. The best sarcasm prediction model for typo-graphic memes is built using Multi Layer perceptron which achieves an accuracy of approximately 88%.

Peer review under responsibility of Uttaranchal University, Dehradun.

1. Introduction

More recently, as memes and GIFs dominate the social feeds; typo-graphic visual content has become a considerable element of social media. 'Meme' is a viral image or video often altered by internet users for humorous effect. These convey human expressions but with a wide range of emotions, and often require context to fully understand humor and sarcasm. "Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text" [Bharti et al., 2016]. It is an expression representative of conflict between the apparent and the applied. Memes are topic-dependent and highly contextual, therefore, polarity shift and other contextual clues can help detect sarcasm from text and improve the generic sentiment classification of typo-graphic social data. For example, as in the post shown in Fig. 1, "Being nice" demonstrates a conflict between the obvious state of "hardest part of the job". This inconsistency, contrast and shifts within the polarities of sentiments validate sarcasm as a distinctive case of sentiment analysis.

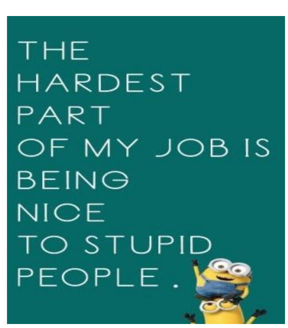

Fig. 1- Example of sarcastic meme

Detecting sarcastic tone is very difficult to accomplish without having a sufficient knowledge of the 'context' of the situation, the specific topic, and the environment [Kumar & Garg, 2019]. Textual sentiment analysis has been widely studied [Pang & Lee, 2008; Kumar & Jaiswal, 2019; Kumar & Sharma, 2017; Kumar & Jaiswal, 2017a]; few related studies have also reported visual analysis of images to predict sentiment [Kumar & Jaiswal, 2017b; Gajarla

&Gupta, 2015; Katsurai & Satoh, 2016] but the domain of visual text which combines both text and image has been least explored in literature. This combination can be observed in two ways as follows:

- *Typo-graphic:* Artistic way of text representation (Fig. 1 & 2). It is primarily textual content but in an image form using variety of fonts and styling options to change the apperance of text.

Fig. 2- Example of sarcastic typo-graphic meme

- *Info-graphic:* Text embedded along with an image (Fig.3). The image may either intensify the textual expresssion or contradict the written content.

Fig. 3- Example of sarcastic info-graphic meme

The research presented in this work attempts to detect sarcasm in typo-graphic memes. We propose a model, Sarc-M, sarcastic meme predictor, to analyze sarcasm from visual language of Instagram memes. The model has five components:

- Data Acquisition Module: Input typo-graphic meme to the model. MemeBank is built with the collected memes.
- Text Extraction Module: Text is retrieved using Computer Vision API for Optical Character Recognition (CV/OCR)
- Text analysis Module: Text is preprocessed and lexical, pragmatic and semantic features are extracted as features
- Learning Module: The features are trained over five classifiers (K-Nearest Neighbor, Support Vector Machine, Random Forest, Decision Tree and Multi Layer Perceptron).
- Prediction Module: Test and performance analysis of classifier prediction accuracy

The rest of the paper is organized as follows: The next section, section 2 describes the related work followed by a detailed illustration of the proposed Sarc-M model for sarcasm detection in memes in section 3. Section 4 gives the results and finally section 5 concludes the research conducted.

2. Related work

Textual sarcasm detection has been studied by researchers as a specialized case of sentiment classification. Both machine learning [Joshi, Bhattacharyya & Carman, 2017; Hercig & Lenc, 2017; Camp, 2012; Riloff et al., 2013] and deep learning [Girshick et al., 2014; Cai & Xia, 2015] models have been explored to predict sarcastic tone in online user-generated content, especially on twitter. The growth in use of images to express opinion online makes text-based sentiment analysis and subsequently sarcasm detection restricted in terms of capturing the sentiment associated. Image-based sentiment analysis has emerged as a significant research domain with the research in this field falling under three areas which are: aesthetics [Datta et al., 2006; Jia et al., 2012, Marchesotti et al., 2011], emotion detection [Zhao et al., 2014a; Zhao et al., 2014b; Zhao et al. 2014c; Li et al., 2012a; Li et al., 2012b; Machajdik & Hanbury, 2010; Lu et al., 2012; Yanulevskaya et al., 2008; Yanulevskaya et al., 2012; Vonikakis & Winkler, 2012] and sentiment ontology [Borth et al., 2013]. To the best of our knowledge no work on visual language based typo-graphic images has been done. This work is an attempt to capture the sarcastic tone in typo-graphic memes.

3. The proposed Sarcastic Meme Predictor Model: Sarc-M

Automatic sarcasm detection as a typical classification task has the primary goal to distinguish between sarcastic or non-sarcastic content. This research probes the problem of detecting sarcasm in memes from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. Context is defined using various features such as n-grams with TF-IDF encoding, frequency encoding for punctuations, sentiment based hand-crafted features. These linguistic context markers are used to train and test classifiers which predict sarcasm effectively and efficiently in the MemeBank dataset built.The following figure 4 depicts the architecture of the proposed Sarc-M model for saracsm detection in tyopographic memes.

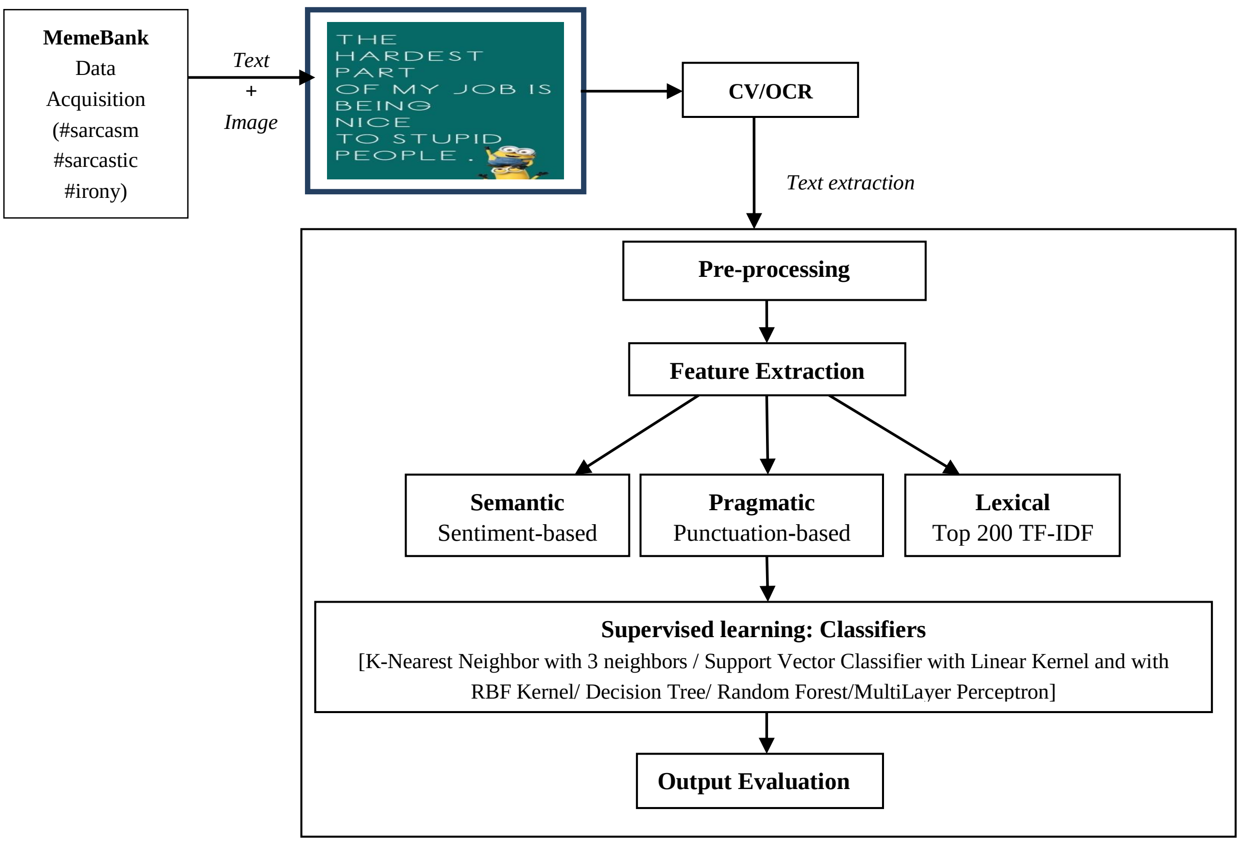

Fig. 4- Architecture of the proposed Sarc-M model

The following sub-sections present the details:

3.1. Data Acquisition

The visual language corpus, MemeBank, is constructed using images from Instagram. This MemeBank is a balanced dataset consisting of 1200 images, which includes600 typo-graphic images collectedwith hashtags #sarcastic, #sarcasm and #irony and negative class data of 600 images collected using hashtags #motivational and #inspirational.

3.2. Text Extraction

For typo-graphic text, we have used the Computer Vision API to extract text using optical character recognition (OCR) from the image. We used Open CV version 3.4.2, EAST text detector proposed by Zhou et.al, 2017, which is a deep learning model, based on a novel architecture and training pattern. It is capable of running at near real-time at 13 FPS on 720p images and obtains state-of-the-art text detection accuracy. The extracted text is passed through an OCR for recognition. The following figure 5 shows the sample text extraction using the API.

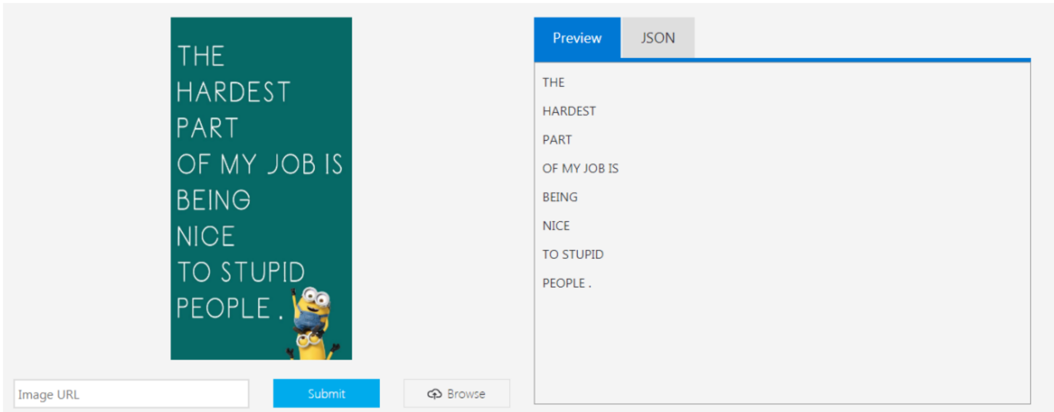

Fig.5- Sample text extraction using the Computer Vision API

3.3. Text Analysis

The text analysis is done to convert the unstructured extracted text into a set of representative features which are used to train and test the classifiers. It includes pre-processing which cleans and transforms the text for feature extraction.

3.3.1. Pre-processing

Post text extraction, pre-processing is done to clean and transform the text for extraction of features. This pre-processing includes:

- Removal of stop words, placeholders, mentions etc.
- Replacement of URLs, special characters such as @, #.
- Use of Natural Language tool-Kit (NLTK)* for tokenization.
- Use of Porter's stemmer for stemming to the root word.
- Removal of non-ASCII English character.
- Part-of- Speech tagging is also done to extract common structural patterns such as verb, adverb, adjective and noun.
- Although all the punctuation marks in a tweet are removed as the part of the cleaning process; however, the count of each punctuation mark is kept as we use them as pragmatic features to train the model. In this work, five punctuation-based features that represent figurative text and provide symbolic clues within the tweet are used. These include exclamation marks (!), question marks (?), periods (.), capital letters and use of "or".

3.3.2. Feature Extraction

As discussed, three categories of features, namely, lexical, pragmatic and semantic features are used in this study.

- Firstly, the conventional lexicon based statistical weighing term frequency-inverse document frequency (TF-IDF) [Bhatia & Kumar, 2008; Jain et al., 2019] measure is used and the top 200 entries are filtered.
- Secondly, frequency of punctuation-based pragmatic features is used. Punctuations, wordplay or uppercase alphabets characterize symbolic clues which help comprehend the context within the text. Frequency encoding has thus been done for these pragmatic features as shown in the following table 1 [Bouazizi & Ohtsuki, 2016]

Table 1- Pragmatic features

Feature	Decription
rep	frequency of repetitive alphabets, that is, if alphabet repetition > 2, then set feature to true else false
excl	frequency of exclamation marks
ques	frequency of question marks
dots	frequency of dots
caps	frequency of capital letters
quotes	frequency of '' or ""

*https://www.nltk.org/

- Thirdly, while pragmatic features are concerned with the language use, sentiment-based semantic features typically convey the conceptual relationship between words. The sarcastic sentences are skewed in the sense that the sentiment polarity of the literal meaning may differ radically from the intended figurative meaning. The shift in sentiment (positive to negative or negative to positive) within these sentences is a strong indicator of sarcasm. For example, in the sentence "I love everybody. Some I love to be around, some I love to avoid and others I would love to punch on their face" clearly expresses sentiment polarity shifts. Therefore, any type of inconsistency or contradiction between sentiments within the text can be considered hinting sarcasm. The sentiment-based features defined in a previous reported work [Bouazizi & Ohtsuki, 2016] have been used in this study. The authors characterize two lists of words classified as "positive words" and "negative words" which contain words with the positive emotional content i.e. emotional positive terms (e.g., "love", "enjoy", "happy", etc.) and that with negative emotional content i.e. emotional negative terms (e.g., "hate", "worry", "sad", etc.). These lists of words are created using the pos_tag library under NLTK [Loper & Bird, 2002]. Next, using these lists, for each textual content the number of positive words (pw) and negative words (nw) are counted. Also, the adjectives, adverbs and verbs have higher emotional content as compared to nouns, therefore, all the words, either positive or negative that have the associated POS tag, are counted another time to create two additional features that are represent the number of highly emotional positive terms (PW) and highly emotional negative terms(NW) respectively. Sometimes, emotional content is conveyed through Hashtags too. For example, in the textual post, "Thanks a lot for always helping me #ihateyou", the hashtag"#ihateyou" tells that the user is not actually thanking the addressed user, but was rather extremely disliking him for not helping him. In addition to the already mentioned features, some features that are related to the contrast between the sentimental components are also extracted. Contrast means the co-existence of both a negative and a positive component in the same tweet. The ratio of emotional words is calculated denoted as ρ(t) expressed in the following equation 1:

$$\rho(t) = ((\delta \cdot PW + pw) - (\delta \cdot NW - nw)) / ((\delta \cdot PW + pw) + (\delta \cdot NW - nw)) \quad (1)$$

Where,

pw is the count of words with positive sentiment
nw is the count of words with negative sentiment
PW is the count of words with highly positive emotional content
NW is the count of words with highly negative emotional content
ρ is the score determined to find contrast between the above sentimental components
t is the tweet
δ is a weight bigger than 1 given to the highly emotional words and is set to 3.
The sentiment related features considered for this study are summarized in Table 2.

Table 2-Sentiment-based features

Feature	Description
pw	Count of words with positive sentiment
nw	Count of words with negative sentiment
PW	Count of words with highly emotional positive content
NW	Count of words with highly emotional negative content
ρ score	Finds contrast between the above sentimental components, here delta is equal to 3.

3.4. Supervised Learning

Classifiers refer to the type of task-specific algorithms where the feature extraction is problem specific and the model is trained using the handcrafted features. The model extracts 11 features (sentiment +punctuation) which are concatenated with the top TF-IDF features generated. This feature matrix is used to train the five baseline classifier namely, K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) and Multi Layer Perceptron (MLP). We considered K-NN with 3 neighbors and the support vector classifier with both Linear and RBF Kernel. The description of these techniques is available in relevant literature studies [Kumar & Jaiswal, 2017a; Kumar & Sangwan, 2019a;Kumar & Sangwan, 2019b].

4. Results

This section illustrates the performance achieved by the sarcasm classifier using based on accuracy and F1 score. The following table 3 illustrates the performance results achieved. It can be observed from the above results that the MLP gives the highest accuracy for the dataset, that is, 87.95 % followed by the RF classifier. The least accuracy of 73.25% is observed using SVC with RBF kernel.

Table 3-Performance Results of the proposed model

Performance measure	K-NN with 3 neighbors	SVC with Linear Kernel	SVC with RBF Kernel	Decision Tree	Random Forest	Multi Layer Perceptron
Accuracy	76.12	73.25	78.34	79.37	80.35	87.95
F1 Score	71.58	80.19	72.67	70.70	80.56	80.09

The figures 6 and 7 depict the results graphically.

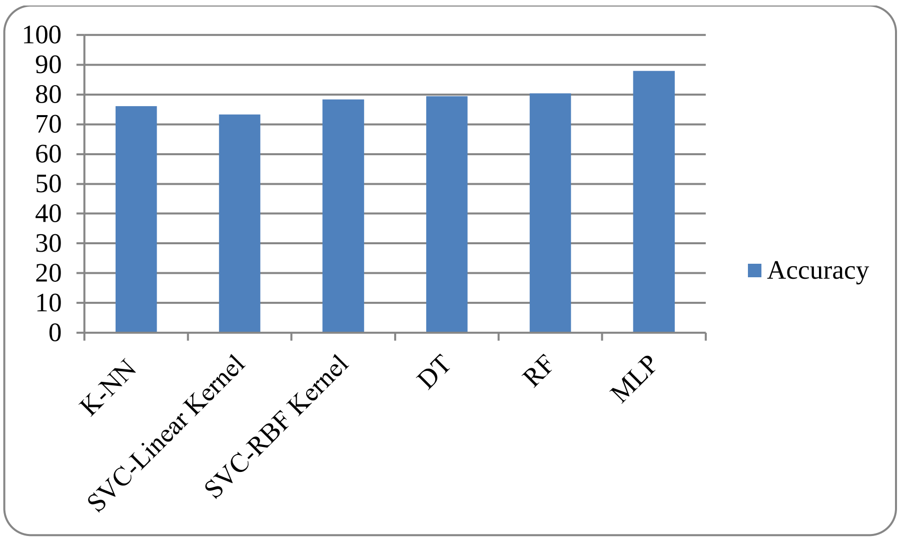

Fig.6- Accuracy achieved by various classifiers

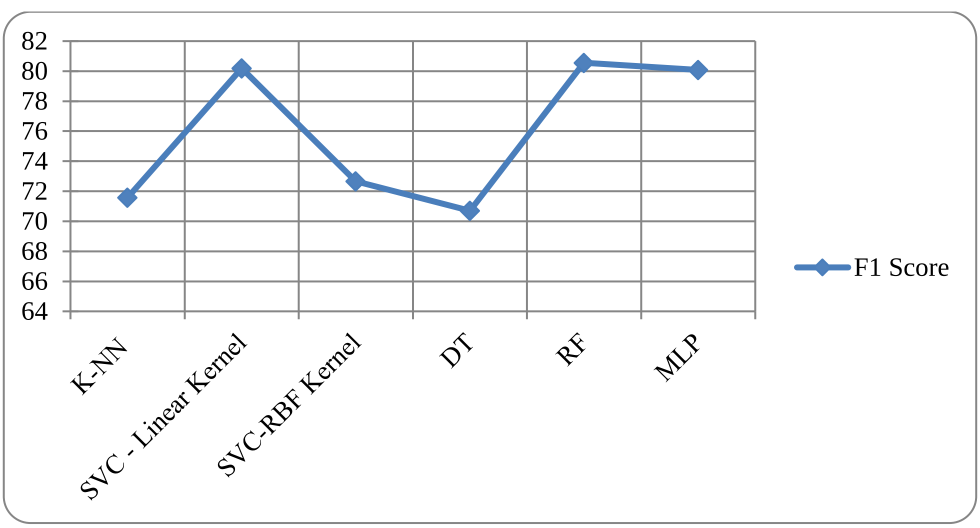

Fig.7- F1 Score of various classifiers

As observed from fig.7, the F1 score of 80.56 has been reported for random forest classifier.

5. Conclusion

Visual language defines the current trend of expression on social platforms and memes further underlines this growing internet culture. Automatic saracsm detection in text has been a dynamic area of research. This work presented automatic detection of sarcastic tone in typo-graphic text, that is memes with asthetically typed text. A MemeBank containing 1200 typo-graphic images was built from Instagram. Computer vision API was used to extract text for analytics from the typo-graphic memes. A combination of hand-crafted features was used to train and test classifiers and it was observed that the best sarcasm classification was achieved by MLP whereas the least was observed for SVC with linear kernel. The main limitation of the model is that the text recognition is restricted by the capability of the Computer Vision API. As a potential direction of research, info-graphic memes, that is memes with both image and text can be evaluated for sentiment and subsequently sarcasm.

REFERENCES

Bharti S.K., Vachha B., Pradhan R.K., Babu K.S., Jena S.K. (2016). Sarcastic sentiment detection in tweets streamed in real time: a big data approach. *Digital Communications and Networks.*2(3), 108-21.

Kumar A., Garg G. (2019) Systematic Literature Review on Context-Based Sentiment Analysis in Social Multimedia, *Multimedia tools and Applications*, https://doi.org/ 10.1007/s11042-019-7346-5

Pang B., Lee L. (2008). Opinion mining and sentiment analysis, *Foundations and trends in information retrieval*, vol. 2, no. 1–2, 1-135.

Kumar A., Jaiswal A. (2019). Systematic literature review of sentiment analysis on Twitter using soft computingtechniques, *ConcurrencyComputatPractExper*. https://doi.org/10.1002/cpe.5107

Kumar A., Sharma A., (2017). Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence. *Webology*, Vol.14, No. 2.

Kumar A, Jaiswal A. (2017a). Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques. *Proceedings of the World Congress on Engineering and Computer Science*, 1:1-5

Kumar A, Jaiswal A. (2017b). Image sentiment analysis using convolutional neural network. *International Conference on Intelligent Systems Design and Applications*, 464-473.

Gajarla V, Gupta A. (2015). Emotion Detection and Sentiment Analysis of Images. *Georgia Institute of Technology*.

Katsurai M, Satoh SI. (2016). Image sentiment analysis using latent correlations among visual, textual, and sentiment views. *IEEE International Conference onAcoustics, Speech and Signal Processing (ICASSP)*, 2837-2841).

Joshi A, Bhattacharyya P, Carman M. (2017). Automatic Sarcasm Detection: A Survey. *ACM Computing Survey (ACM-CSUR*), Article No. 73, Volume 50 Issue 5.

Hercig T, Lenc L. (2017). The Impact of Figurative Language on Sentiment Analysis. *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP*, 301–308

Camp E. (2012). Sarcasm, pretense, and the semantics/pragmatics distinction. *Noûs2012;* 46(4):587-634.

Riloff E, Qadir A, Surve P, Silva LD, Gilbert N, Huang R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. *EMNLP*; 13:704–714.

Girshick R, Donahue J, Darrell T, Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580-587.

Cai G, Xia B. (2015). Convolutional neural networks for multimedia sentiment analysis. *In Natural Language Processing and Chinese Computing*. 159-167.

Datta R, Joshi D, Li J, Wang JZ. (2006). Studying aesthetics in photographic images using a computational approach. *European Conference on Computer Vision*, 288-301.

Jia J, Wu S, Wang X, Hu P, Cai L, Tang J. (2012). Can we understand van gogh'smood?: learning to infer affects from images in social networks. *Proceedings of the 20th ACM international conference on Multimedia* 857-860.

Marchesotti L, Perronnin F, Larlus D, Csurka G. (2011). Assessing the aesthetic quality of photographs using generic image descriptors. *IEEE International Conference on Computer Vision*, 1784-1791.

Zhao S, Yao H, Wang F, Jiang X, Zhang W. (2014a). Emotion based image musicalization. *InMultimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference*, 1-6.

Zhao S, Yao H, Yang Y, Zhang Y. (2014b). Affective image retrieval via multi-graph learning. *Proceedings of the 22nd ACM International Conference on Multimedia*, 1025-1028.

Zhao S, Gao Y, Jiang X, Yao H, Chua TS, Sun X. (2014c). Exploring principles-of-art features for image emotion recognition. *Proceedings of the 22nd ACM International Conference on Multimedia*, 47-56.

Li B, Feng S, Xiong W, Hu W. (2012a). Scaring or pleasing: exploit emotional impact of an image. *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 1365-1366.

Li B, Xiong W, Hu W, Ding X. (2012b). Context-aware affective images classification based on bilayer sparse representation. *Proceedings of the 20th ACM International Conference on Multimedia*, 721-724.

Machajdik J, Hanbury A. (2010). Affective image classification using features inspired by psychology and art theory. *Proceedings of the 18th ACM international conference on Multimedia*, 83-92)

Lu X, Suryanarayan P, Adams Jr RB, Li J, Newman MG, Wang JZ. (2012). On shape and the computability of emotions. *Proceedings of the 20th ACM International Conference on Multimedia*, 229-238.

Yanulevskaya V, van Gemert JC, Roth K, Herbold AK, Sebe N, Geusebroek JM. (2008). Emotional valence categorization using holistic image features. *In ICIP*, 101-104.

Yanulevskaya V, Uijlings J, Bruni E, Sartori A, Zamboni E, Bacci F, Melcher D, Sebe N. (2012). In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. *Proceedings of the 20th ACM International Conference on Multimedia*, 349-358.

Vonikakis V, Winkler S.(2012). Emotion-based sequence of family photos. *Proceedings of the 20th ACM International Conference on Multimedia*, 1371-1372.

Borth, Damian, RongrongJi, Tao Chen, Thomas Breuel, and Shih-Fu Chang. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. *Proceedings of the 21st ACM international conference on Multimedia*, 223-232.

Zhou X, Yao C, Wen H, Wang Y, Zhou S, He W, Liang J. (2017). EAST: an efficient and accurate scene text detector. *In Proc. CVPR* 2017, 2642-2651.

Bhatia M.P.S, Kumar, A. (2008). A Primer on the Web Information Retrieval Paradigm. *Journal of Theoretical & Applied Information Technology* 1;4(7).

Jain D. K., Kumar, A., Sangwan, S. R., Nguyen G. N. and Tiwari, P.(2019). A Particle Swarm Optimized Learning Model of Fault Classification in Web-Apps. *IEEE Access*,7: 18480-18489.

Bouazizi M, OhtsukiT.(2016). A pattern-based approach for sarcasm detection on Twitter. *IEEE Access*; 4:5477-88.

Loper E, Bird S. (2002). NLTK: The natural language toolkit. *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics*, 1:63-70.

Kumar A, Sangwan S R. (2019a). Rumor Detection Using Machine Learning Techniques on Social Media. *International Conference on Innovative Computing and Communications*, 213-221.

Kumar A, Sangwan S R., Nayyar, A. (2019b). Rumor Veracity Detection on Twitter Using Particle Swarm Optimized Shallow Classifiers. *Multimedia tools and Applications*. https://doi.org/ 10.1007/s11042-019-7398-6.