Adaptive Visual Sentiment Prediction Model Based on Event Concepts and Object Detection Techniques in Social Media

Yasser Fouad[1], Ahmed M. Osman[2], Samah A. Z. Hassan[3], Hazem M. El-Bakry[4], Ahmed M. Elshewey[5]
Department of Computer Science-Faculty of Computers and Information, Suez University, Suez, Egypt[1, 5]
Department of Information Systems-Faculty of Computers and Information, Suez University, Suez, Egypt[2, 3]
Department of Information Systems-Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt[4]

***Abstract*—Now-a-days, the increasing number of smartphones has caused the immediate sharing of photographs capturing current events on social media. The sentimental content of pictures from social events starts to be obtained from visual material, so visual sentiment analysis is a vital research topic. The research aims to reach valuable criteria to modify the visual sentiment prediction model based on event concepts and object detection techniques. In addition to adapting the approach for designing the method for predicting visual sentiments in a social network according to concept scores and measuring the performance of the model for predicting visual sentiments as accurately as possible, approach obtains a visual summary of social event images based on the visual elements that appear in the pictures which exceed sentiment-specific features. By this method, attributes (color, texture) are assigned to sentiments with discovering affective objects that are used to obtain emotions related to a picture of a social event by mapping the top predicted qualities to feelings and extracting the prevailing emotion connected with a photograph of a social event. This method is valid for a wide range of social events. This strategy also demonstrates the social event's effectiveness for a difficult social event image collection by using techniques for classifying complicated event images into sentiments, whether positive or negative.**

***Keywords*—Sentiment Analysis (SA); visual sentiment analysis; image analysis; object recognition; event concepts; events concepts with object detection**

I. Introduction

Online social networks have integrated significantly into our daily lives. It is being designed to become a crucial resource for gathering and disseminating information in a variety of industries, including politics, business, entertainment, and crisis management. Social media Big Data is the result of the enormous increase in social media usage [1], which has resulted in a growing accumulation of data from which we are unable to profit. Numerous options for data formats are available on social networking sites like Instagram, Flickr, Twitter, and Facebook, including text, photographs, videos, sounds, and geospatial data.

Social media network users share a vast amount of written and visual content to communicate their emotions and thoughts, allowing us to construct a large collection of feelings and opinions. Analyzing user-generated material can aid in understanding and forecasting user behavior and emotions. Examining this information is crucial in the behavioral sciences, including areas like opinion mining, affective computing, and sentiment analysis. These disciplines strive to comprehend and anticipate human decision-making, enabling various practical applications such as monitoring brands, predicting stock market trends, and forecasting political voting patterns. As a result, scholars have recently become interested in these topics, and much research has been conducted in the "sentiment analysis" age of web mining. in light of the explosive spread of camera-enabled smartphones and the growth of social media and online visual content. It became easier than ever to create and share images. This led to an increase in the volume of images on the web, which continues to grow exponentially. Images have great power; they are more impactful than text, memorable, more engaging, and more likely to be shared and re-shared. This is due to the fact that the human brain is built for visual communication. Humans handle visual elements faster and remember them longer, and they elicit a stronger emotional response. In fact, visuals are processed 60,000 times faster than text. It's also more efficient for the person who is communicating.

Opinion mining, also referred to as "sentiment analysis," is an automated approach [2] to identifying opinions expressed in text. The increasing prevalence of mobile devices with cameras and social media platforms like Facebook, Twitter, and Weibo emphasizes the significant role played by multimedia content such as images and videos in conveying people's sentiments and opinions within social networks.

In recent years, numerous innovative concepts have emerged in the promising field of visual sentiment analysis. One notable advancement in artificial intelligence is deep learning, which has made significant strides [3, 4, 5, 6]. Researchers have begun to utilize deep learning techniques for sentiment analysis across various forms of social media data.

Traditionally, sentiment analysis has primarily concentrated on analyzing textual content. However, images, as a vital element of multimedia web data, play a significant role in conveying, expressing, communicating, comprehending, and illustrating people's opinions or sentiments to viewers. The increasing significance of image sentiment analysis, or predicting sentiments from images, is becoming increasingly apparent.

Sentiment analysis refers to the computational process of classifying and categorizing sentiments expressed in "multimedia web data," including both textual and non-textual elements. Its objective is to determine the attitude or opinion of a speaker or writer regarding a specific topic [7], as well as the overall contextual polarity or emotional response towards a document, image, interaction, or event.

Deep learning and machine learning are two interconnected fields that have become increasingly important in various domains [8, 9, 10]. Here are some key reasons for their significance as data-driven decision-making tools where deep learning and machine learning algorithms enable organizations to analyze and make sense of large amounts of data. They can uncover patterns, extract valuable insights, and make data-driven decisions. This is particularly crucial in today's era of big data, where traditional methods of analysis may not be sufficient for automation and efficiency. Machine learning algorithms can automate repetitive tasks and streamline processes, leading to increased efficiency and productivity. For example, in industries like manufacturing and logistics, machine learning can optimize supply chains, predict maintenance needs, and improve overall operational efficiency. In personalization and recommendations, deep learning and machine learning algorithms power personalized recommendations in various applications, such as e-commerce, streaming services, and social media platforms. These algorithms analyze user preferences, behaviors, and historical data to deliver tailored suggestions, enhancing the user experience and increasing customer engagement.

Recognizing attitudes elicited by photos from social media is more challenging than many other visual identification tasks, such as object categorization, scene recognition, and so on. For visual sentiment prediction, a diverse range of cues must be considered. Visual sentiment analysis is the process of identifying an object, scene, or activity and their emotional context.

Visual sentiment analysis has recently become one of the areas of computer vision, and it is a clue to solving the picture sentiment prediction problem. The most powerful computer vision approaches improve the process of recognizing human sentiment from low-level features to high-level features. The state-of-the-art in traditional computer vision tasks has lately experienced fast transformations as a result of deep learning techniques such as convolutional neural networks (CNN), which are utilized for image recognition activities. The network employs a multi-layered architecture that, through layer-wise processing, can represent features from raw pixels. This led to the application of similar techniques to forecast visual sentiment, in which we strive to recognize the emotion that an image would elicit in a human observer. In visual recognition, CNN models are reaching human-level performance. Several researchers have also used CNN to classify image sentiment and determined the difference between the amazing performance of deep features and hand-tuned features for sentiment classification. As a result, image sentiment analysis is regarded as an essential topic of research in online multimedia big data. However, visual sentiment analysis research is still in its infancy.

The paper's contributions are as follows: an approach to predicting the sentiment of complex event photographs using visual content and event concept detector scores with object detection, with no text analysis on test images required; Without extracting sentiment-specific information from the photos, the method outperforms state-of-the-art sentiment prediction algorithms. We conducted extensive tests on a difficult social event image dataset that has been tagged with sentiment labels (positive and negative) from different social media engines.

II. Related Work

Yang, J., et al. [11] addressed the difficulty of automatically recognizing sentiments in images. A framework is suggested to find out affective regions and gather information through CNN, inspired by the observation that the entire image as well as local sections have the ability to convey that sentimental information, which is most important. Considering both the objectless score and the sentiment score, the level of sentiment content in some region can be measured, noticing that the objectless score typically contains rich texture information, and that the sentiment score analyses the sentiment of that region at the affective level. The experimental results demonstrated that the proposed strategy outperformed state-of-the-art methods on common emotional datasets. Campos, V. et al. [12] performed extensive tests to compare different fine-tuned convolutional neural networks (CNNs) for visual sentiment prediction. The results demonstrated that deep architectures have the ability to learn new features and effectively understand the visual sentiment conveyed in social photos. The researchers developed multiple models that surpassed the current state-of-the-art performance on a dataset consisting of Twitter photos. They also highlighted the significance of pre-training in model initialization, particularly when dealing with small datasets. Furthermore, the researchers provided visualizations of the network's learned local patterns, which helped in understanding how these models perceive visual positivity or negativity and provided insights into their recognition capabilities. Ahsan, U., et al. [13] introduced a framework based on visual content alone to predict complicated image sentiment. They presented a dataset of an annotated social event and demonstrated that the features of the suggested event can be effectively assigned to sentiments. Accordingly, there was a proposed method to predict complicated image sentiment using visual content and event detector scores without having to analyze text on tested images. Not only did this proposed approach classify complicated event images into sentiments in a way surpassing state-of-the-art approaches, but it also demonstrated the effectiveness of a dataset of challenging social event images. Islam, J., and Zhang, Y. [14] introduced a novel framework for visual sentiment analysis using a transfer learning approach. They employed hyperparameters obtained from a highly deep convolutional neural network as the initialization for their network model. This choice aimed to mitigate overfitting issues. The researchers conducted a comprehensive set of experiments on a dataset of Twitter images, showcasing the superior performance of their proposed model compared to the current state-of-the-art approaches in sentiment analysis. Wang, Y., et

al. [15] focused on the task of recognizing human sentiments using a combination of image features and contextual social network information, such as friend comments and user descriptions, within a large collection of Internet images. They developed a novel technique for visual sentiment analysis that leveraged various forms of prior knowledge, including sentiment lexicons, sentiment labels, and visual sentiment strength. The researchers devised a two-stage method for universal affective norms for pictures (ANPs), which involved detecting mid-level qualities to bridge the "affective gap" between low-level image features and high-level image emotions. They introduced a multiplicative updating technique to identify optimal solutions for model inference and demonstrated its convergence. Through experiments conducted on two large-scale datasets, they demonstrated that their proposed model surpassed previous state-of-the-art approaches in both sentiment inference and fine-grained sentiment prediction. You, Q., et al. [16] used the recently developed convolutional neural networks to find a solution to the problem of visual sentiment analysis. A new architecture wasn't only designed but also new training strategies, and that was a way to overcome the noisy nature of the large-scale training samples. The results emphasized that the proposed CNN surpassed not only classifiers that use predefined low-level features but also those with mid-level visual attributes. And its performance in image sentiment analysis was superior to that of other competing algorithms. Chen, T., et al. [17] introduced a hierarchical system that focuses on modelling object-based visual sentiment concepts, such as "crazy car" and "shy dog," to extract emotion-related information from social multimedia content. This system operates in an object-specific manner, enabling sentiment concept classification and addressing the challenge of concept localization. By leveraging an online commonsense knowledge base and introducing novel classification techniques to model concept similarity, the proposed framework significantly improved classification performance compared to previous approaches, achieving up to a 50% improvement. Moreover, the system identifies discriminative features, enabling the interpretation of the classifiers.

III. Approach

An overview of the model is described in the following sections: The proposed method contains four phases, as shown in Fig. 1. The model consists of four main steps: A) elicit social networking images; B) discover affective objects; C) perform feature extraction; and D) make sentiment predictions.

A. Elicit Social Networking Images

Collect images from popular social networking sites to assess the proposed method. The dataset includes pictures from user content, which is collected using eight sentiment categories as keywords on social websites.

B. Discovering Affective Objects

Localize the object regions in the input image. Then, for these regions, extract features. using object recognition techniques. Finally, the region that has a high score of object recognition indicates an affective object. The phase consists of three steps:

1) Object detection: This step used techniques to generate a set of candidate windows that detect visual objects with a rigid structure, such as a car or bike, and non-rigid objects, such as pedestrians and dogs. During the past decades, the object detection problem has been handled by many object proposal methods. These methods are Deformable Part Model (DPM), EdgeBoxes, and BING. The model used EdgeBoxes to detect objects.

2) Object feature: During this stage, object recognition techniques are employed to extract features from the objects detected within the bounding box regions. Deep learning algorithms, such as convolutional neural networks (CNNs), have become widely utilized for object recognition. These models have the ability to automatically learn the intrinsic properties of objects and can distinguish between different categories, such as cats and dogs, by analyzing large volumes of images and identifying distinguishing features. In addition to deep learning, machine learning techniques offer alternative approaches to object detection. Examples of traditional machine learning methods include using histogram of oriented gradients (HOG) features in combination with support vector machine (SVM) models, as well as employing bag-of-words models such as SURF and MSER. Another well-known algorithm, the Viola-Jones algorithm, can recognize various objects, including faces and upper bodies. The model used the (HOG) features in combination with support vector machine (SVM) models, as well as employing bag-of-words SURF model to object feature selection.

3) Object recognition: Recognizing objects inside the bounding box region and determining the accuracy of the object recognition. The region that has a high score of object recognition indicates the affective object.

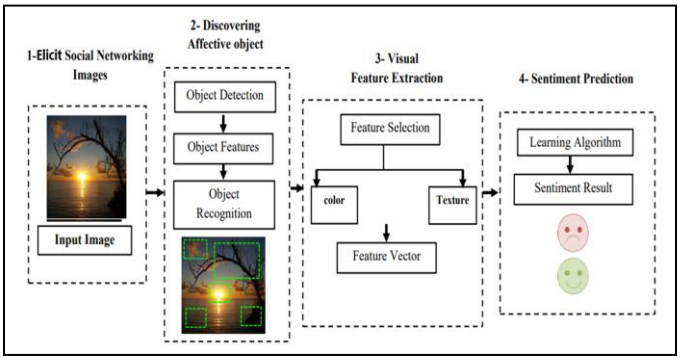

Fig. 1. The proposed approach is divided into four stages.

C. Visual Feature Extraction

The affective object region was identified in the previous phase. This phase [18] seeks to extract visual elements (such as color, texture, and form) from this region that reflect image appearance and can also predict image sentiment. A single feature cannot discriminate among a homogeneous group of photos, hence a vector including all retrieved image features is required to describe the image. Color, texture, and shape are examples of mid-level visual elements retrieved. Color is the most effective feature; color retrieval methods include color

histograms and color correlograms. There are also various image retrieval methods based on Block Truncation Coding (BTC), which use encoding data in RGB color using BTC to recover image features. The texture of an image [19] is an important quality that describes not only the properties of an object's surface but also its relationship to its surroundings. The wavelet transforms and Gaborfilters are two of the most prominent tools for detecting picture texture; these measures attempt to capture elements of the image relating to changes in specific directions and the scale of these changes. Following that, employing feature [20] selection approaches to choose the most significant, best, most important, and most optimal subset of features collected from the region. Also used to remove irrelevant or redundant properties without changing the data, resulting [21] in increased efficiency, improved accuracy, and reduced data complexity. The output features are then combined to form a feature vector that represents the sentiment features for the affective object region.

D. Sentiment Prediction

Classify images and predict sentiment from those images (positive or negative). Sentiment prediction models can be easily trained by using learning algorithms. There are various algorithms used for learning models, such as the logistic regression model, which leads to better performance than SVM classifiers with sentiment features.

IV. Experiments

The images dataset is mentioned in this section, and the study aims to generate sentiment labels for the dataset as well as an experimental setup to predict event photo sentiments on the test set.

A. Dataset

To conduct the experiment, using eight event categories as search queries, retrieve public pictures from social media engines. These events are broad, encompassing both planned and unexpected events and including both personal and community-based activities, getting about 11,000 photos labelled as (1) positive and (2) negative. A dataset consisting of annotated event images. The dataset was divided into different sentiment classes.

B. Experimental Setup

For each class, 70% of the images were randomly selected as positive training data, while an equal number of images from the remaining mood classes were chosen as negative training data. The remaining 30% of images from each class were reserved for testing purposes. During testing, an equal number of negative training data points from sentiment classes other than those being tested were included. This ensured that the baseline accuracy for sentiment prediction was always 50%. The experiment was repeated five times, and the sentiment prediction accuracy for each class was averaged to obtain the final accuracy. In this stage, different techniques are employed to effectively detect objects within the bounding box regions. utilized EdgeBoxes for object detection. Once the objects were detected, extracted their features within the bounding box region using HOG feature extraction in combination with an SVM machine learning model. Additionally, employing a bag-of-words model with features like SURF to further enhance object recognition. Furthermore, the Viola-Jones algorithm, known for its capability to identify various objects like faces and upper bodies, was also utilized. By recognizing objects within the bounding box region and assessing the accuracy of object recognition, we were able to determine the affective object. The region with a high score of object recognition indicated the presence of the affective object. After that, select features such as color and texture. Color is regarded as the most effective feature; numerous approaches for picture retrieval based on Block Truncation Coding (BTC) extract image features from BTC that store data in RGB color. The texture of an image is an important attribute that describes the surface properties of an object and their relationship to its surroundings. The wavelet transforms and Gaborfilters are two common methods for identifying picture texture and the model used Gaborfilters. These methods attempt to capture image components with respect to changes in particular directions and the scale of the changes. Following that, employing feature selection techniques to identify the most significant, best, most important, and most optimal subset of characteristics collected from the region also used to remove irrelevant or redundant features without transforming the data, resulting in increased efficiency, improved accuracy, and reduced data complexity. The output features are then combined to form a feature vector that represents the sentiment features for the affective object region. To calculate event scores on the images, we utilized the Caffe deep learning framework. Specifically, we extracted features using the activations from the seventh layer ('fc7') of a CNN (Convolutional Neural Network) known as AlexNet. This CNN architecture was pre-trained on HybridCNN, which incorporates knowledge from a pre-training phase on 978 object categories from the ImageNet database and 205 scene categories from the Places dataset. The extracted features from the 'fc7' layer were 4096-dimensional, capturing rich representations of the images. The last step is to classify images and predict sentiment from those images (positive, negative). Sentiment prediction models can be easily trained using a logistic regression model.

V. Results and Discussion

Table I shows the sentiment prediction accuracies for our proposed event features and many powerful state-of-the-art baselines, and our proposed approach surpasses the state-of-the-art for not only all the sentiment classes but also the overall average sentiment prediction.

TABLE I. Four Classification Models and the Proposed Event Concepts with Objects Detection Model for Performance Evaluation

Models	Accuracy	F1-score	Recall	Precision
Hybrid CNN	66.38	66.01	65.70	65.38
SentiBank	67.73	67.55	67.40	66.70
Deep SentiBank	70.69	70.40	69.60	69.50
Event concepts	73.06	73.01	72.60	72.10
Event Concepts with Object Detection	74	74.02	73.59	73.20

As seen in Table I, the events with objects model outperformed all other models, achieving an accuracy of 74%, an f1-score of 74.02%, a recall of 73.59%, and a precision of 73.20%. Fig. 2 displays the accuracy of the five approaches and the proposed events with the object model.

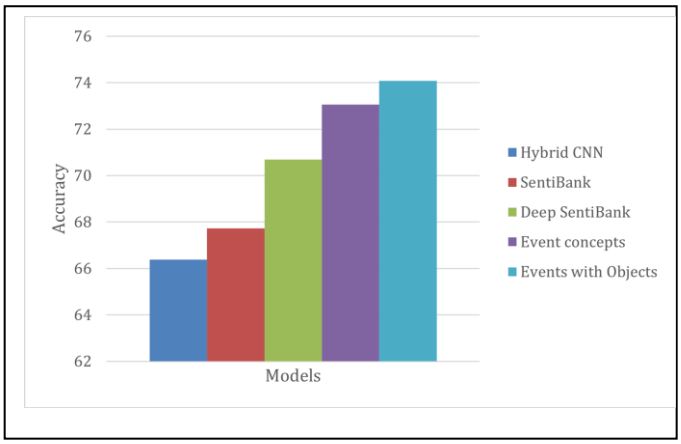

Fig. 2. Comparison between the proposed events with objects model and different models in the term of accuracy.

VI. Limitations and Future Work

The proposed approach has limitations that have been identified. While the taught model can accurately recognize positive images when the visual cues are strong, it tends to make mistakes when differentiating between positive and negative sentiments. In summary, the observation of a disparity between human perception of events (e.g., assuming all photographs of the Nepal disaster should be negative) and the actual images that exhibit diverse emotions influenced by those events. However, we believe that the proposed method adequately captures the nuanced nature of how an event impacts the emotional content of an image. In future developments, expanding the richness of social event data by incorporating more test data and richer labels into the sentiment recognition pipeline could potentially enhance the classifier's performance and address confusion between the three sentiments.

VII. Conclusion

Through event concepts and object detection approaches, we present a system for predicting complicated image sentiment using visual content, presenting an annotated social event dataset and demonstrating that suggested event concepts and object detection approaches can be efficiently mapped to sentiment. Comparing this method to state-of-the-art approaches, it outperforms them by a wide margin. Also investigated the proposed method's generalizability and validity by testing its performance on an unseen dataset of photos encompassing events not covered in model training.

References

[1] Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. Social media analytics–Challenges in topic discovery, data collection, and data preparation. International Journal of Information Management, 39, 156-168, 2018.

[2] Ji R, Cao D, Zhou Y et al Survey of visual sentiment prediction for social media analysis. Front Comput Sci 10(4):602611, 2016.

[3] J. Jia, S.Wu, X.Wang, P. Hu, L. Cai, and J. Tang. Can we understand van goghs mood?: learning to infer affects from images in social networks. In Proceedings of the 20th ACM international conference on Multimedia, pages 857860. ACM, 2012.

[4] Al-onazi, B.B., Nauman, M.A., Jahangir, R., Malik, M.M., Alkhammash, E.H. and Elshewey, A.M.. Transformer-based multilingual speech emotion recognition using data augmentation and feature fusion. Applied Sciences, 12(18), p.9188, 2022.

[5] Elshewey, A.M., Shams, M.Y., El-Rashidy, N., Elhady, A.M., Shohieb, S.M. and Tarek, Z.,. Bayesian optimization with support vector machine model for parkinson disease classification. Sensors, 23(4), p.2085, 2023.

[6] Alkhammash, E.H., Kamel, A.F., Al-Fattah, S.M. and Elshewey, A.M., 2022. Optimized multivariate adaptive regression splines for predicting crude oil demand in Saudi arabia. Discrete Dynamics in Nature and Society, pp.1-9, 2022.

[7] Ravi K, Ravi V, A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl-Based Syst 89(C):1446, 2015.

[8] Alkhammash, E.H., Hadjouni, M. and Elshewey, A.M., A Hybrid Ensemble Stacking Model for Gender Voice Recognition Approach. Electronics, 11(11), p.1750, 2022.

[9] Shams, M.Y., El-kenawy, E.S.M., Ibrahim, A. and Elshewey, A.M., A hybrid dipper throated optimization algorithm and particle swarm optimization (DTPSO) model for hepatocellular carcinoma (HCC) prediction. Biomedical Signal Processing and Control, 85, p.104908, 2023.

[10] Tarek, Z., Shams, M.Y., Elshewey, A.M., El-kenawy, E.S.M., Ibrahim, A., Abdelhamid, A.A. and Mohamed, A., Wind Power Prediction Based on Machine Learning and Deep Learning Models. CMC-COMPUTERS MATERIALS & CONTINUA, 74(1), pp.715-732, 2023.

[11] Yang, J., She, D., Sun, M., Cheng, M. M., Rosin, P., & Wang, L. Visual sentiment prediction based on automatic discovery of affective regions. IEEE Transactions on Multimedia, 2018.

[12] Campos, V., Jou, B., & Giro-i-Nieto, X. From pixels to sentiment: Finetuning CNNs for visual sentiment prediction. Image and Vision Computing, 65, 15-22, 2017.

[13] Ahsan, U., De Choudhury, M., & Essa, I. Towards using visual attributes to infer image sentiment of social events. In Neural Networks (IJCNN), 2017 International Joint Conference on (pp. 1372-1379). IEEE, 2017.

[14] Islam, J., & Zhang, Y. Visual sentiment analysis for social images using transfer learning approach. In Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialComSustainCom), 2016 IEEE International Conferences on (pp. 124-130). IEEE, 2016.

[15] Wang, Y., Hu, Y., Kambhampati, S., & Li, B. Inferring sentiment from web images with joint inference on visual and social cues: A regulated matrix factorization approach. In Ninth international AAAI conference on web and social media, 2015.

[16] You, Q., Luo, J., Jin, H., & Yang, J. Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks. In AAAI (pp. 381-388), 2015.

[17] Chen, T., Yu, F. X., Chen, J., Cui, Y., Chen, Y. Y., & Chang, S. F. Object-based visual sentiment concept analysis and application. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 367- 376). ACM, 2014.

[18] Singh B, AhmadW, Content based image retrieval: a review paper. Int J Comput Sci Mob Comput3 (5):769–775, 2014.

[19] Sidhu S, Saxena J, Content based image retrieval a review. Int J Res Comput Appl Robot 3(5):84–88, 2015.

[20] Esmel ME, A novel image retrieval model based on the most relevant features. Knowl Based Syst 24(1):23–32, 2011.

[21] Hancer E, Xue B, Karaboga D, Zhang M , A binary ABC algorithm based on advanced similarity scheme for feature selection. Appl Soft Comput 36:334–348, 2015.