International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Restaurant Based Emotion Detection Of Images From Social Media Sites Using Deep Learning Model

Mamatha M.[1]*, Shilpa ShivaKumar[2], Thriveni J.[3], Venogopal K. R.[4]

Submitted: 27/05/2023 **Revised**: 05/07/2023 **Accepted**: 24/07/2023

Abstract: Nowadays, online reviews and rating has become a powerful means of sharing opinions among the customers. From past few years people have started reacting about restaurants through pictures as reviews. The facial expression of the person or food in the image is reviewed along with a short text added to it and then classified into positive or negative category. This gives a clear understanding of the image. Because analysis of text sentiment is carried over enormously in natural language processing(NLP), this work focuses on visual sentiment analysis of restaurant review image dataset to find if the image belongs to positive or a negative category. An image prediction model is built using deep learning method such as Convolutional Neural Networks(CNN) to identify the sentiment. Image classification is performed and accuracy is enhanced using images posted on social media sites. The proposed system performs comparatively better than the machine learning methods in analyzing the opinions in images.

Keywords: *Sentiment Analysis, Feature Extraction, Adam Optimizer, CNN, Softmax*

1. Introduction

Consumers share their reviews on a product on the internet. Nowadays, opinion on a restaurant is also being written on their websites. For ex, before a purchase decision of a product is made from a particular website, we need to check for the reviews and ratings made by the consumers who bought it previously. The choice of a restaurant is also made by reviewing its ambience, food and service details posted by people on the little app or eatigo app which could be in text or image format.

Sentiment analysis on text data is performed previously and best results are obtained. Limited research is done on images and videos posted on social media. Classification of images is a difficult task as the emotions need to be extracted and it is done using deep learning techniques.

Recently, deep neural networks methods have been used to obtain high performance in classification of images and feature extraction[1,2]. In this work, we obtain and classify the images of restaurants and facial expressions of customers into positive or negative. CNN is a tool used mostly for image segmentation to extract the deep features from images. By doing this the sentiment classification is also improved[3,4]. To do this a sentiment prediction model using CNN is proposed. The performance of the proposed work is shown experimentally by comparing it with machine learning techniques.

1,2,3,4 Department of Computer Science and Engineering, Bangalore University, Karnataka, India*
**Email: mmtha.s@gmail.com*

1.1 Motivation:

Sharing of customer feedback of a restaurant visited in the form text on the internet sources helps in identifying the sentiment and make decisions. It is in the form of text data which is analyzed and divided into positive, negative and neutral. This is done using machine learning methods. In the recent years, people share reviews on restaurant they visited in the form of images. These are classified further by performing feature extraction on those images which is a difficult task. The size of the dataset is increased by performing data augmentation such as rotation, flipping, scaling, etc. By this the class label of the image can be figured out correctly.

1.2 Contribution:

A prediction model is constructed on neural networks to analyze the restaurant image dataset and recognize the facial expression. The difficulties in understanding the emotion behind the images is overcome using these methods. These are the pictures shared on the internet by customers who visited the restaurant to express their views.

The contributions are:

- Performing image identification on extracted images.
- Obtaining a good classification accuracy on dataset. This is different from classification of labels on text.
- Class labels of the images in the dataset that are put up by users are identified.

1.3 Organization:

The remaining parts of this paper is structured as follows. In section II literature review is analyzed. Section III explains the background work. Section IV discusses about the proposed system in detail. Section V specifies the implementation details. Analysis of performance on the proposed work is explained in section VI. And, in section VII the paper is concluded briefly.

2. Related Work

Restaurant review images on social media are used in the training and testing process. In the recent years, food identification has become a crucial role as it is difficult to find the sentiment in an image by extracting few features from it. There are also variations in the shape of food, color and texture. [5]Preprocessing and segmentation is used in food identification on the images. This work uses CNN based deep learning techniques for food image classification to identify the objects in images.

In the modern era, digital technologies are replacing the job of humans in most of the industries. In the same way resorts and restaurants have also been taken away by posting many pictures to attract customers. With this images of variety of foods also have been added. These images are classified using convolutional neural network. The author [6]proposed CNN for classification task by removing the spatial features in the posted pictures. Transfer learning with Inception V3 is also used in this work.

In this work[7]recognition of food using deep learning technique to estimate the calories in various Malaysian foods is performed. Malaysian diet datasets are used for analysis. Alexnet which is based on Convolutional Neural Networks and transfer learning is used in classification. An accuracy of 91.4% is achieved using large Alexnet network.

This author in this paper[8] proposed convolutional block attention module(CBAM) with ResNet50, Mobile NetV2 and VGG16 to improve the classification accuracy of Asian food images. A mixed-data enhancement algorithm(Mixup) is also used to obtain a smooth discrimination ability. Accuracy of 87.3% is achieved by combining CBAM and Mixup.

The author[9]uses machine learning techniques to identify the profound information in restaurant images that suggests customers to make a decision to dine in. The dataset is trained on ResNet-101 and ResNet-50. The work is carried over on the Yelp dataset using various descriptive labels such as:

0: good_for_lunch

1:good_for_dinner

3:outdoor_seating

5:has_alcohol

The author[10] works on a combination of both text and visual data together where an image has both in it. A deep fusion convolutional neural network is proposed to learn the sentiment of both text and visual representations. The information from the two types are put together in the pooling layer of CNN and forwarded to 3 fully-connected layer. Then, the sentiment polarity is predicted in the softmax layer.

Authors[11] train the model using deep learning techniques. This model performs both classification and recommendation of food to customers based on diet. Convolutional neural network algorithms are used to improve the accuracy to 86.33%. The data collected from google chrome, kaggle and UCI is classified into 12 classes namely, butter, dosa, idly, upma, bisibelebath, chapati etc. This work contributes to diabetic patents healthy diet. In this paper[12] the author applied CNN to identify the facial expressions in the restaurant and food review system. Food type preferences in the restaurant, and quality and taste can be known by reviewing the food images. This method is applied on real-time images and the performance is tested.

Image-based dietary assessment has had a lot of improvement in the recent years. CNN plays an important role in image classification and recognition of variety of food or non-food items. [13]GoogLeNet model on deep convolutional neural networks is used to work on recognition and classification of food images collected through social media, existing image datasets and smart phones. Food/non-food category achieved an accuracy of 99.2% and 83.6% on food recognition.

The problem that arises during recognition process is it needs a lot of details to perform the task. CNN in the existing techniques identify the sentiment in the image by analyzing the features of the entire picture through different regions and can have distinct effects on the invoked emotions. [14]A weakly supervised coupled convolutional neural network(WSCNet) is introduced to select soft proposals by providing the weak comments. This reduces the burden of annotation and performs the following task: Firstly, the convolutional network which includes the cross spatial pooling strategy is trained to identify a sentiment-specific soft map. Secondly, a semantic vector is formed by combining the sentiment map with deep features during classification, both holistic and localized information are utilized. Experiments show that WSCNet performs better than the other existing models on the seven datasets.

This paper[15] works on deep learning techniques to predict the sentiment of unlabelled images. Sentiment analysis of images has been a major area of research lately.

Image recognition and prediction, an area under neural networks is used to examine the performance of convolutional neural network and region-based CNN in image sentiment analysis. Image analysis is a combination of noun adjective pair that are helpful in determining the exact thoughts behind the image.

In this paper the author[16] works on classification of food item images such as restaurants without waiter for service and diet calculators for intake of food. Pre-trained deep convolutional neural networks is used to extract the features. Transfer learning is applied on the images for better analysis of performance. A new Australian dietary guidelines database on food images is used for classification. The results obtained are comparable to the existing techniques, but the training time taken to train the model is less than the other methods.

This paper is a corpus with number of different photos uploaded with simple information such as, whether the restaurant has outdoor seating, good for kids, good for lunch, good for dinner, etc. [17]The images are labelled accordingly to help the customer decide which restaurant they want to dine in. Three different machine learning models, specific to image classification are used to train and test the dataset. Overfitting problem is solved using CNN and ResNet, an activation function which is variant of CNN. A total of 234842 images is used for training and these are divided under each of the 9 categories.

There are number of challenges in sentiment analysis because of the increase in social media and people sharing their expression in text, video or image form. The author[18] proposed a deep visual sentiment analyzer for images related to disaster in social media. Different aspects such as data collection, model selection, implementation etc. Using many participants world-wide, a crowdsourcing study has been conducted to collect the data and analyze it.

In this paper[19] the author solves the problem of visual sentiment analysis by introducing two branches to detect the sentiment using supervised convolutional network. It is a challenging task because of high level abstraction in the process of recognition. Relevant soft maps are automatically selected from the weak explanations written. The work contributes to detect soft map by training a fully convolution network in WSCNet together with pooling technique. Next, for classification, the comprehensive and localized features are used and combined with the deep features of sentiment map. These are integrated to form a unified deep framework. The advantage of this method is the reduction in annotation and time taken for sentiment classification and detection is less. WSCNet performs better than the state-of-the-art methods on all the datasets.

3. Related Work

With the emergence of mobile phones people have started expressing their opinions in the reviews by including photos or even text. Deep learning techniques such as CNN, DNN, Region R-CNN, and fast R-CNN perform efficiently in identifying the polarity of image and classification. To determine whether the image expresses positive or negative polarity deep learning techniques/models are used.

Six emotion categories defined are happiness, disgust, fear, surprise, anger and sadness. These emotions are used by other authors in the same domain[20,21]. With these the features such as, color and texture are analyzed to make classification easier. These are classified using the dimensional model and categorical model.

In the dimensional model emotions in images are represented as points in dimensional space. They are valence, arousal and control represented in two or three dimensions. The categorical model labels the basic human expressions of images which are assigned to regions. There are a number of emotional classes such as sad, happy, anger, excited,

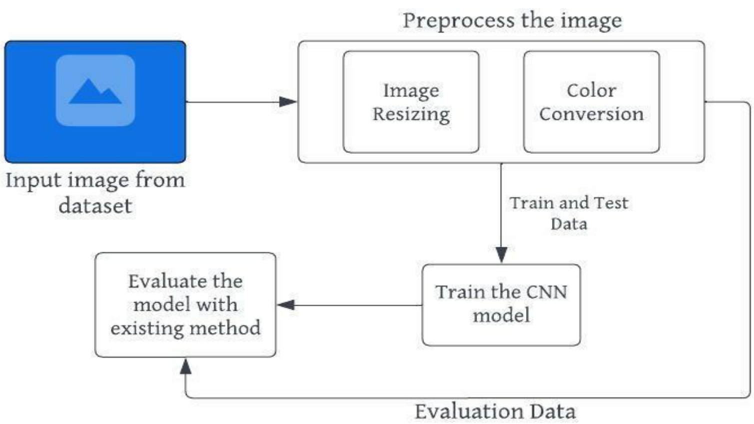

Fig. 1. Process diagram of emotion detection model

disgust etc. associated during classification. The categorical model is basically used in face recognition. The former measures the distance on three dimensional space(valence, arousal, control) and the latter assumes that people who speak the same language think the same way about different distinct emotions.

The architecture of CNN is multilayered with each layer performing transformation to study the features from raw-pixels. Image analysis is a process of finding the emotion in pictures. This is done by extracting the detailed features and applying the necessary tests with respect to colors. The deeper the network, better is the performance obtained.

To represent visual sentiment using the elements in images, multiple instance learning(MIL) is used. The author[22] proposed a deep multiple clustering instance learning network(DMCILN) using deep multiple clustered instance learning method for visual analysis. The picture is transformed into a number of samples using visual instance generation module. This contains pre-trained CNN and two adaptation layers. An attention-based MIL pooling layer is constructed to identify the relation between clustered instances and bags. It is also used to represent bag features.

4. Proposed System Model

CNN is used specifically to process structured data in grid within the dimensional space. Linear algebra is used to process the image and classify them. Convolution and pooling are the important layers used to extract average or maximum value of the pixel from image. The width, depth and height are represented by differing volumes in the input layer. The convolution process is performed on two real-valued arguments and feature map is obtained as output from the input image. Fig. 1 shows the process followed in the proposed system. CNN is a subset of neural networks designed especially for image processing. The raw images in the dataset are pre-processed before it is divided into train and test data. They are then trained on the model and evaluated.

The dataset contains raw restaurant images that are processed and trained using CNN model. Fig. 2 shows some of the samples related to restaurants that are randomly picked from the internet sources.

Fig. 2. Sample set of raw positive and negative images

4.1 Pre-processing:

The input image is converted into pixel value of width and height with an array between 0 and 255. In order to extract features and learn the weights, CNN uses filters across the hidden layers and the input image. Unprocessed images are resized to 150×150 pixels with width and height set to 30 pixels and R, G, B color channels. Thus, dimensionality of the image is reduced to maintain uniform shape. Image scaling is performed by applying a scaling factor of 0.2(30/150) on the images. A color picture consists of three primary colors applied in combination to generate new colors. It is a mix of R, G and B.

1) Normalization:

It is a part of pre-processing performed to achieve dynamic range consistency. The output feature dimensions of the image to the model is 900 features for each picture. This is equal to 30×30 features in an image. During model training the pixel values of images are within the range [-1, 1]. The color spectrum is maintained consistently using “(1)”.

$$x_i(m,n)=\frac{2\times x(m,n)}{x_j}-1 \qquad (1)$$

where,

$x_i(m,n)$- normalized output image,

x_j- maximum pixel value, and

$x(m,n)$- input image before normalization

As the dataset is small, the time taken is less and the accuracy is good. The more number of features extracted, training the model is easy as more data is available.

2) Data Augmentation:

It is a part of the image classification. It directly impacts the effect of classification. The importance of adding data is that with a limited amount of image data, the training dataset can be sufficiently and reasonably expanded to improve the generalization ability of the model. Relevant noisy data is added to improve the system robustness and remove the dimensional effects between data.

The data is increased artificially through rotation, contrast jittering, flipping or translation. This improves the models performance and helps to recognize if the image is positive or negative. Rotation is performed by rotating the image on an axis between 1° to 359°.

4.2 CNN Modeling:

CNN is largely used in the branch of computer vision including face recognition and food review images. It performs accurately with respect to scale invariance when compared to machine learning classifiers.

CNN follows the feed-forward artificial neural network pattern. The idea of using CNN is it extracts the features in images accurately and classifies them. The images after preprocessing are passed as input to the model to classify the image expression. The proposed network consists of three convolution layers along with max-pooling, fully connected layer and then softmax output.

1) Feature Engineering

- Convolution Layer: The neurons in the convolution layer are arranged in such a way that it forms a filter of length and height. It is in the form of pixels. Fig. 3 shows how the pattern in the image is identified.

Example: First layer in feature extraction is 5×5 convolution layer and filter size is 3×3. All 3 of these filters keep shifting around the image by performing dot product with input data and filter value and matrix output referred to as feature map or activation map is obtained.

A parameter called stride is considered to determine the filter movement over input. If a stride of 1 is fixed, the filter shifts one pixel at a time over the input. The feature map shows the reduction in image size. The smaller the stride, detailed information is obtained from input. The number of steps increases if the stride is large, but achieves very good performance. In the system non-linearity is present. This is introduced using Rectified Linear Units(ReLU) activation function. Several layers of convolutions along with activation function makes it possible to model non-linear relationships.

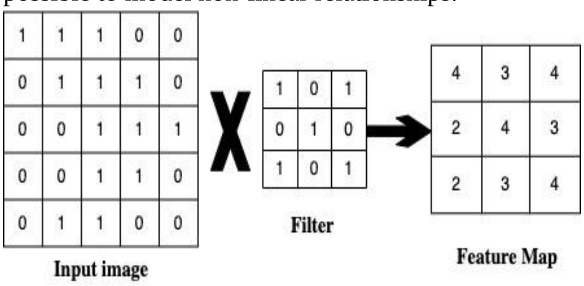

Fig. 3. Convolution process

Padding is performed to find the pixels containing zero value that is to be added on each side of the input. The output dimensions of the feature map is manipulated according to the input using this technique. The output dimension is calculated using (2).

$$\text{output} = \frac{L-W+2p}{S}+1 \qquad (2)$$

where,

L- length of input,

W- length of the filter,

p- padding, and

S- stride

- Pooling Layer: After applying the activation function to the feature map in convolution layer, pooling layer is added. In our work both max-pooling and average pooling is used to reduce the size or number of pixels in each feature map and also noise removal task. Max-pooling extracts the maximum value from the input by selecting the filter size and stride. A 2×2 filter is added in this model to reduce the feature maps dimension with stride of 2. Next, more convolution and pooling layers are added for better feature extraction.

The output dimension obtained after performing in the pooling layer with dimensions $n_r \times n_s \times n_p$ shown in eq. (3)

$$(n_r - f + 1)/L \times (n_s - f + 1)/L \times n_p \qquad (3)$$

where,

n_r - height of feature map,

n_s - width of feature map,

n_p - number of color channels in the map,

f - size of filter, and

L - stride length

By using average-pooling, an average of all features in the patch are obtained. A downsampled feature map is created as output. Smooth features are selected using this technique.

Our system works on three convolution, pooling layers and two fully-connected layers for model 1 and 2. For model 3 another fully-connected layer is added to compare the performance with other systems. 32, 64 and 128 filter size is added.

The filters are selected according to the pattern to be captured. Simple patterns from images are captured using smaller filter size and it combines further to make complex patterns in the successive layers. The size of the filters selected are 3× 3 and 2× 2.

2) Classification:

- Fully-connected Layer: The output map from feature learning layer is still multi-dimensional. It is reduced into one-dimensional array by performing flattening on the feature map. This is used as input to fully-connected layer. Activation neurons from the previous layer are first flattened and then attached to every neuron in the next layer. The neurons are fully connected in this layer unlike the convolution layer where it connects only to specific regions of the input.
- Softmax Layer: This is the final layer in the CNN model. A softmax activation function is used to determine the output among the classes(positive and negative). The class formed by the system will correspond to the vector probabilities derived from the images. Furthermore, the input images are labelled.

Mathematically, softmax performs the following transform on 'n' number of images x_1........x_n . Mathematical formula of the softmax function is shown in eq. (4) as:

$$\mathrm{Soft}(x_i)=\frac{\exp(x_i)}{\sum_{j=1}^{N}\exp(x_j)} \qquad (4)$$

where,

x - input vector to ’Soft’. It is made up ofx0,...xN.

x_i- these values are element of the input vector.

$\exp(x_i)$ - standard exponential function applied oneach element of x_i. The result value is alwayspositive, close to 0.

N - Number of classes, and $\sum_{j=1}^{N}\exp(x_j)$ - It is a normalization term whichensures the values of output vector sums to 1.

These values are in the range of (0-1) and hence called a valid probability distribution.

- Model Tuning: The model is trained to fit to the dataset by optimizing the loss using Adam optimizer. The performance of the system is enhanced further by fine-tuning various parameters in the different CNN layers. Learning rate parameter is varied and it is fixed to 0.0002 and the batch size is 50. The system is set to execute for 30 epochs. The best performance accuracy is reached at 10 epochs.

5. Implementation

System implementation details of the proposed work is explained in this section.

5.1 Dataset Specification:

The proposed system is evaluated for detection of emotions in images. It is implemented using CNN a deep neural network method to obtain better accuracy.

The different images from restaurants of various websites are collected from social media platform. These images are divided into positive and negative sentiment categories. 800 raw images that show customers sentiment about a restaurant are downloaded and the images with low resolution and those that are labelled neutral are discarded. 103 images are eliminated and the analysis is done using 697 images.

5.2 Model Comparison:

In this work, the parameters on the different models of the CNN are varied and the accuracy between them are compared to find the best model that fits into this work.

Table 1 shows the various parameters and the number of hidden layers to obtain good performance in the different CNN models. In model 1, 3 convolution and 3 pooling layers are added. Max-pooling is used in the pooling layer with a pool size of 2×2. 32 and 64 filters are added along with a filter size of 3×3. Then, the output from last pooling layer is sent to the first fully-connected layer where the neurons are connected to the others in the next FC layer and to the softmax output with 2 possibilities for positive and negative.

In model 2 the kernel filter is decreasing from high to low and the pool size remains the same as in model 1. Two fully-connected layers are used with 2 neurons in the output.

Model 3 uses three fully-connected layers for dimensionality reduction. ReLu is the activation function used in all the FC layers. It is used to

Table I. Parameter comparison

Layers	**Model 1**		**Model 2**		**Model 3**		**Activation Function**
	Parameters	***Size***	***Parameters***	***Size***	***Parameters***	***Size***	
Convolution 1	32 filter	3*3	64 filter	3*3	128 filter	3*3	ReLu
Pooling 1	Max-pool	2*2	Average-pool	2*2	Max-pool	2*2	ReLu
Convolution 2	32 filter	3*3	32 filter	3*3	128 filter	3*3	ReLu
Pooling 2	Max-pool	2*2	Average-pool	2*2	Max-pool	2*2	ReLu
Convolution 3	64 filter	3*3	32 filter	3*3	64 filter	3*3	ReLu

Pooling 3	Max-pool	2*2	Average-pool	2*2	Max-pool	2*2	ReLu
Fully-connected layer	256 Neuron		256 Neuron		512 Neuron		ReLu
Fully-connected layer	128 Neuron		128 Neuron		128 Neuron		
Fully-connected layer	-		-		64 Neuron		ReLu
Output	2 Neurons		2 Neurons		2 Neurons		Softmax

Table II. Model accuracy

Epoch	**Model 1**		**Model 2**		**Model 3**	
	Train1 (%)	***Test1 (%)***	***Train2 (%)***	***Test2 (%)***	***Train3 (%)***	***Test3 (%)***
1	62	74	69	67	64	78
2	71	70	69	70	65	78
3	74	83	72	69	69	75
4	75	90	76	74	72	81
5	78	90	74	78	73	84
6	80	89	77	79	81	87
7	81	92	80	85	81	92
8	84	91	81	86	86	92
9	86	91	81	88	87	94
10	87	93	83	88	89	95

reduce the classification errors in the network and train the network is faster. 128, 128 and 64 filters are used in the convolution layers.

6.Performance Analysis

In this section, the results of different models of CNN are analyzed. The dataset is divided into 80(train):20(test) ratio and evaluated. Deeplearning methods results are compared with machine learning methods. Adam optimizer is used in both the processes and learning rate is fixed to 0.0002 and number of epochs is 10. The images are saved in jpeg format and the implementation is done using python programming. Precision, recall, F1-score and accuracy are the metrics used for analysis. Table 2 shows the training and test accuracy of different models.

In Fig. 4, the model reaches the highest accuracy of 93% at epoch 10 because of the effectiveness of max-pooling. The time taken to process the model is 154 seconds(sec). The test data performs better compared to validation. The loss is comparatively less for the test data.

Fig. 5 shows the training process of model 2. Optimum accuracy of 86% is achieved at epoch 8 and it reaches 88% at epoch 9.

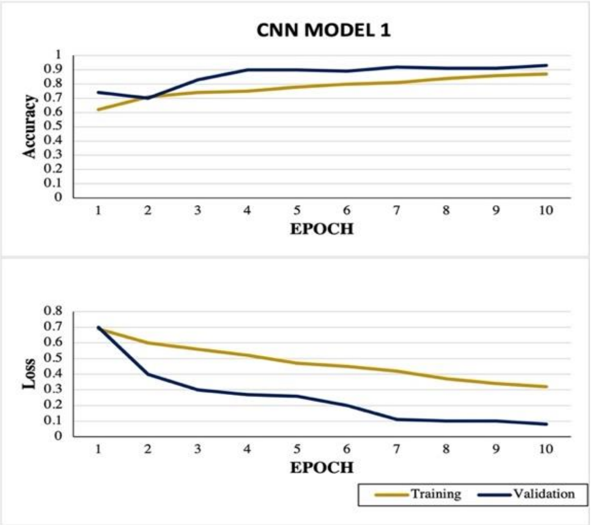

Fig. 4. Accuracy and Loss performance of model 1

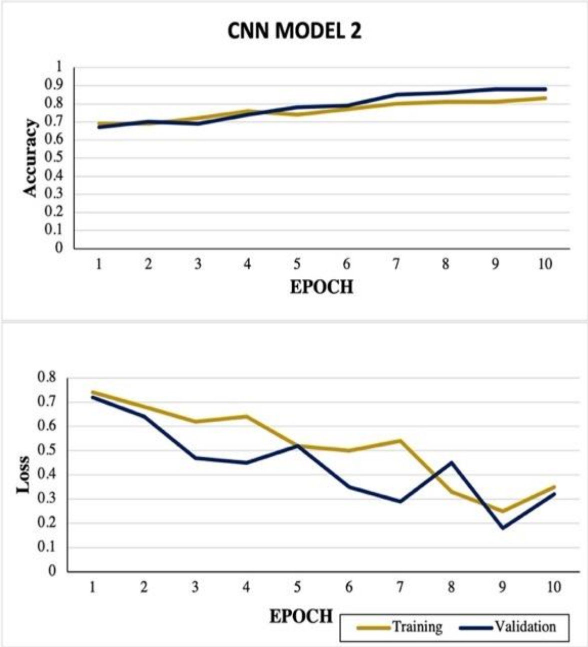

Fig. 5. Accuracy and Loss performance of model 2

The performance of this model reduces since average pooling is used to capture the spatial features instead of max-pooling. As the loss is high in this model, the performance also decreases. And, the training process takes 221 sec to train the model.

The progress of the third model is shown in Fig. 6. It performs better than the first and second system in terms of accuracy by achieving 95%. Max-pooling is added and another fully-connected layer is added to learn the complex representations. This increases non-linearity and prevents overfitting. It takes more time in training than the testing process.

The performance of CNN is better than the machine learning classifiers as feature extraction is handled by CNN itself. Overfitting does not occur in our models as the dataset is small and the number of iterations are less.

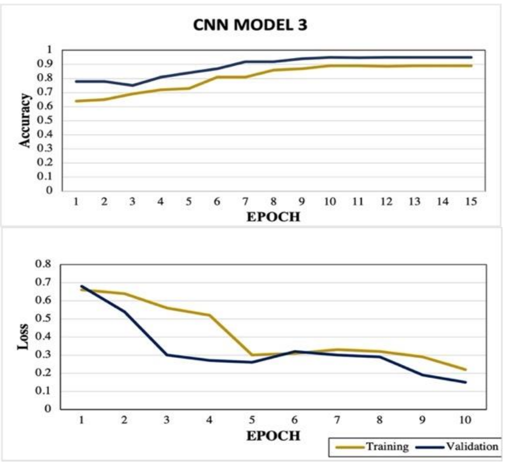

Fig. 6. Train and Test Accuracy and Loss performance of model 3

Table III. Processing time comparison

Classifier	Accuracy	Process Time(s)
GaussianNB	57.1	92
BernoulliNB	57	106
MultinomialNB	52.2	84
CNN 1	93.4	154
CNN 2	88	221
CNN 3	95	197

Table 3 shows the time taken by the different classifiers to predict the emotion of images in the dataset to reach the best accuracy. Multinomial naive bayes takes 84 sec to process the data but the performance is low. The proposed CNN 3 model performs best with accuracy of 95% and the time taken is 197 sec, 3.2 min with 10 epoches.

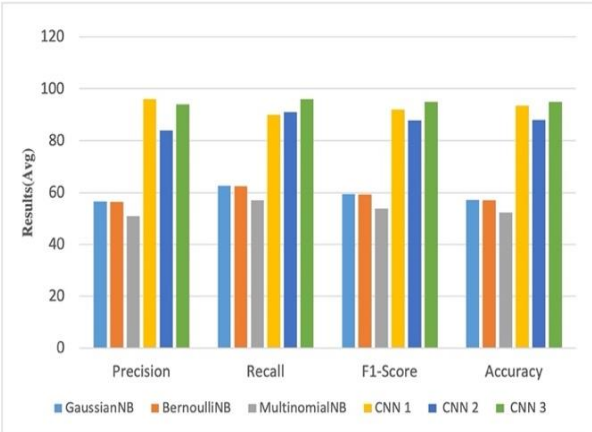

Fig. 7. Accuracy Comparison of Proposed Model with Machine Learning Model

A comparison of the performance of all the classifiers are analyzed on the basis of various metrics as shown in Fig. 7. This shows that the proposed model performs better in analogy with machine learning algorithms in all the metrics. The recall value is higher than the precision which shows that most of the relevant results are obtained and false negative values are less.

7. Conclusion

Image analysis is one of the most important area of research for study, as people are more exposed to social media and converse through visual content. A novel framework is proposed in this paper to identify the customers emotion/sentiment from the restaurant review images and it is compared with machine learning methods. Machine learning classifiers performs extremely well on text data. The neural network models perform better on food images than NLP techniques. Fine tuning of the parameters improve the classification accuracy. Further, the work can be carried over on summarizing the sentiment of video clips. Graph neural networks which is another deep learning method can be applied to review images of restaurants obtained from social media sites. An application for food review can be developed for restaurants that immediately labels the images posted by customers as good, better, best.

Conflicts of Interest

Authors declare no conflict of interest

Author Contributions

Methodology, conceptualization, original draft preparation, dataset collection, implementation, result analysis and comparison, have been done by first and second author. Supervision, review of work and project administration, have been done by third and fourth author.

References

[1] Krizhevsky, I. Sutskever, and G. E. Hinton,"Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1–9, 2012.

[2] Y. Lu, Y. Chen, D. Zhao, B. Liu, Z. Lai and J. Chen, "CNN-G: Convolutional Neural Network Combined with Graph for Image Segmentation with Theoretical Analysis," *Journal of Transac- tions on Cognitive and Developmental Systems*, pp. 1–13, 2020.

[3] M. Mamatha, S. Shilpa, J. Thriveni and K. R. Venugopal, "Visual Sentiment Classification of Restaurant Review Images using Deep Convolutional Neural Networks," *In: Proc. of IEEE International Conf. on Electronics, Computing and Communication Technologies*, pp. 1–6, 2022.

[4] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2481–2495, 2017.

[5] H. Li, Q. Zhang, and D. Zhao, "Deep Reinforcement Learning-based Automatic Exploration for Navigation in unknown Environment," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 2064–2076, 2020.

[6] Shah and H. Bhavsar, "Overview of Deep Learning in Food Image Classification for Dietary Assessment System," *Intelligent Systems, Technologies and Applications*, vol. 1353, pp. 265–285, 2021.

[7] Alex M. Goh and Xiaoyu. L. Yann, "Food Image Classification using Neural Network Model," *International Journal of Electronics Engineering and Applications*, vol. 9, pp. 12–22, 2021.

[8] R. A. Rahmat and S. B. Kutty, "Malaysian Food Recognition using Alexnet CNN and Transfer Learning," *In: IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 59–64, 2021.

[9] Xu, X. He and Z. Qu, "Asian Food Image Classification based on Deep Learning," *Journal of Computer and Communications*, vol. 9, pp. 10–28, 2021.

[10] H. Peng,Y. Ding, and B. Huang,"Yelp Restaurant Photo Classification," pp. 1–5, 2020.

[11] X. Chen, Y. Wang, and Q. Liu, "Visual and Textual Sentiment Analysis using Deep Fusion Convolutional Neural Networks," *In: Proc. Of IEEE International Conf. on Image Processing(ICIP)*, pp. 402–413, 2017.

[12] Manjula G, Sukanya U, Suryashree T. S, and Shubha M, "Food Classification and Recommendation for Health and Diet Tracking using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*,vol. 9,pp. 2587–2589,2022.

[13] N. Mahrab, S. A. Salim, A. I. Ali, I. J. Mimand R. Khan, "Facial Expression based Automated Restaurant Food Review System using CNN," *In: Proc. of IEEE International Conf. on Artificial Intelligence in Engineering and Technology (IICAIET)*, pp. 1181–1187, 2021.

[14] Singla, S. A. Salim, L. Yuan, and T. Ebrahami, "Food/non-food Image Classification and Food Categorization using pre-trained GoogLeNet Model," *In: Proc. of International Workshop on Multimedia Assisted Dietary Management*, pp. 3–11, 2016.

[15] She, J. Yang, M. Cheng, Y. K. Lai, P. L.Rosin and L. Wang, "WSCNet: Weakly Supervised Coupled Networks for Visual Sentiment Classification and Detection," *IEEE Transactions on Multimedia*, pp. 1358–1371, 2020.

[16] N. Mittal, D. Sharma and M. Lata Joshi, "Image Sentiment Analysis using Deep Learning," *In: Proc. of IEEE/WIC/ACM International Conference on Web Intelligence(WI)*, pp. 684–687, 2018.

[17] Rahul Subhash Gaikwad and SharanabasappaC. Gandage, "An Exploration of Deep Transfer Learning for Food Image Classification," *International Journal of Systems and Applications in Engineering*, pp. 1–5, 2018.

[18] K. T. Islam, S. Wijewickrema, M. Pervez, and S. O'Leary, "MCNN: Visual Sentiment Analysis using various Deep Learning Framework with Deep CNN,"

2018 Digital Image Computing: Techniques and Applications (DICTA), vol. 11, pp. 265–278, 2023.

[19] Y. Wang and B. Li, “Sentiment Analysis for Social Media Images,” *In: Proc. of International Conf. on Data Mining Workshop*, pp. 1584–1591, 2015.

[20] J. Yang, D. She, Y. K. Lai, P. L. Rosin, and M. H. Yang, “Weakly Supervised Coupled Networks for Visual Sentiment Classification and Detection*,” In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 7584–7592, 2018.

[21] M. Dellagiacoma, P. Zontone, G. Boato and L. Albertazzi, “Emotion based Classification of Natural Images,” *In:. DETECT’11: Proc. of International Workshop on DETecting and Exploiting Cultural DiversiTy on the Social Web*, pp. 17–22, 2011.

[22] M. Ulinski, V. Soto, and J. Hirschberg, “Finding Emotion in Image Descriptions,” *In: Proc. of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, pp. 1–7, 2012.

[23] Q.T. Truong and H. W. Lauw, “Visual Sentiment Analysis for Review Images with Item-oriented and User-oriented CNN,” *In: Proc. of ACM International Conf. on Multimedia*, pp. 1274–1282, 2017.

[24]

[25] Alexei Ivanov, Machine Learning for Traffic Prediction and Optimization in Smart Cities , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[26]

[27] Garcia, P., Martin, I., Garcia, J., Herrera, J., & Fernández, M. Enhancing Cyber security with Machine Learning-Based Intrusion Detection. Kuwait Journal of Machine Learning, 1(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/157

[28] Rohokale, M. S., Dhabliya, D., Sathish, T., Vijayan, V., & Senthilkumar, N. (2021). A novel two-step co-precipitation approach of CuS/NiMn2O4 heterostructured nanocatalyst for enhanced visible light driven photocatalytic activity via efficient photo-induced charge separation properties. Physica B: Condensed Matter, 610 doi:10.1016/j.physb.2021.412902