Computers, Materials & Continua
DOI:10.32604/cmc.2022.019284
Article

Tech Science Press

AI Cannot Understand Memes: Experiments with OCR and Facial Emotions

Ishaani Priyadarshini* **and Chase Cotton**

Department of Electrical and Computer Engineering, University of Delaware, Newark, 19716, DE, United States
*Corresponding Author: Ishaani Priyadarshini. Email: ishaani@udel.edu
Received: 08 April 2021; Accepted: 11 May 2021

Abstract: The increasing capabilities of Artificial Intelligence (AI), has led researchers and visionaries to think in the direction of machines outperforming humans by gaining intelligence equal to or greater than humans, which may not always have a positive impact on the society. AI gone rogue, and Technological Singularity are major concerns in academia as well as the industry. It is necessary to identify the limitations of machines and analyze their incompetence, which could draw a line between human and machine intelligence. Internet memes are an amalgam of pictures, videos, underlying messages, ideas, sentiments, humor, and experiences, hence the way an internet meme is perceived by a human may not be entirely how a machine comprehends it. In this paper, we present experimental evidence on how comprehending Internet Memes is a challenge for AI. We use a combination of Optical Character Recognition techniques like Tesseract, Pixel Link, and East Detector to extract text from the memes, and machine learning algorithms like Convolutional Neural Networks (CNN), Region-based Convolutional Neural Networks (RCNN), and Transfer Learning with pre-trained denseNet for assessing the textual and facial emotions combined. We evaluate the performance using Sensitivity and Specificity. Our results show that comprehending memes is indeed a challenging task, and hence a major limitation of AI. This research would be of utmost interest to researchers working in the areas of Artificial General Intelligence and Technological Singularity.

Keywords: Technological singularity; optical character recognition; transfer learning; convolutional neural networks (CNN); region-based convolutional neural networks (RCNN)

1 Introduction

Artificial intelligence (AI) plays a major role in the constantly evolving technology. Machines are becoming more and more sophisticated with increased training and data. While machine intelligence may have served humanity generously in many ways it may not always have desirable outcomes. Past research works have also been known to highlight how AI systems go rogue. Examples like Self-driving cars jumping red lights [1], Image Recognition software labeling black people as gorillas [2], AI systems performing racial discrimination [3,4], robots killing humans [5,6] are all wake up calls for setting boundaries on feeding intelligence to machines. Machines imitating

humans in cyberspace have also shown undesirable outcomes [7]. This led to many researchers and visionaries thinking of a point in time when machine intelligence becomes equal to or surpasses human intelligence [8] commonly referred to as Technological Singularity [9,10]. Moreover, if a system starts making decisions of its own, and improvising itself, it may not be possible for humans to control it until the capabilities of humans exceed it in some way or the other. It is also likely that a machine may not make correct decisions in situations that call for humanity, compassion, psychology, critical thinking, etc. since many of these cannot be easily fed to a system. This makes it imperative to identify the incapabilities or limitations of AI with respect to humans, i.e., specific tasks that may be performed by humans but not by machines. Over the last few decades, AI has proven its indispensability and intelligent machines have made themselves proficient in various spheres [11,12]. They can detect humor, identify sentiments, generate poetry, automate complex processes, beat humans in games [13,14]. The list goes on, which makes it even harder to find techniques that could demarcate human and machine intelligence. Not only this, but machines also try to improvise themselves constantly and become superior in areas that are yet to be perfected [15,16]. However, one of the areas that is abundantly researched and still poses a challenge for AI is emotion detection. Emotion detection may be performed on texts [17] as well as facial expressions [18]. Languages in texts may be diverse and include infinite grammatical variations, misspellings, slang, complicated structures, which could make emotion detection from texts challenging [19]. Emotion detection from facial expressions is a bigger challenge because expressions may vary across many billion human faces in the world. Surprisingly, Internet Memes incorporate texts, pictures, emotions, and cognizance, understanding which would be a significantly challenging task for AI, as multiple emotions need to be assessed simultaneously [20,21]. Internet memes may be defined as humor traveling virally through the internet. Usually, an image with a concept or a catchphrase may be used to express emotions in various situations. Since the situations and emotions are expressed in limited words and objects, it may not be easy for AI to fully comprehend the underlying message of a meme. Moreover, understanding Internet Memes completely could be forever a challenge for AI and remain an unsolvable problem. There are several reasons that highlight why AI cannot comprehend memes with 100% efficiency. There is no end to the number of memes, which grows at an exponential rate. This makes it difficult to train since there can be infinite memes. Since memes can be created from old as well as new data, meme prediction problems are extremely challenging unlike many other problems [22,23]. Moreover, one single meme can be used in different contexts, and may also manifest different emotions depending on the situation. The variations may be in the form of texts, images, fonts, colors, etc. They may convey the same message or different messages, either way, it will be new training data and may lead to ambiguity issues for AI. Memes may support intentional misspellings and grammatical mistakes, which is yet again a challenge for AI to understand. There is an issue of language and culture barriers. Memes vary across cultures, hence similar memes can be expressed in various languages. The meaning of a meme in one culture may not be the same in the other culture. Either way it is new training data, and much easier for humans to comprehend.

As we know, machine learning methods improvise according to past experiences and patterns. However, this may not work in the case of internet memes. This is because if all the previous instances n, are dissimilar, there is practically no pattern and the machine may not learn anything concrete from previous data. It may not be able to interpret what $(n+1)$th meme means even after adequate training. Therefore even after the machines are taught adequately, they may still fail to accomplish certain tasks. It may not be an easy task for AI to comprehend Internet Memes given the plethora and variety of memes found on the internet, along with the different patterns each of the memes can generate. In this paper, we will be validating our claims using emotion

detection (text and facial recognition). We use a combination of Optical Character Recognition (OCR), techniques like Tesseract, Pixel Link, and East Detector for text extraction, and machine learning algorithms like CNN, R-CNN, and Transfer Learning using dense networks to adequately train the system. The combined emotion from the text and facial recognition can assess the underlying idea of the meme. Failure to understand the sentiment behind a meme would unveil the limitations of AI and machine learning capabilities and draw a line between human and machine intelligence. The novelty and main contributions of the paper are as follows:

(a) We rely on Internet Memes for distinguishing human and machine intelligence followed by an empirical study. This is the first paper to analyze human and machine intelligence from the perspective of comprehending Internet Memes.
(b) We perform Emotion Detection for Texts as well as Facial Expressions in the memes. This is the first paper to detect emotions extensively in memes.
(c) We use three OCR techniques for extracting text from memes, that are Tesseract, Pixel Link, and East Detector.
(d) Machine Learning Techniques used in the study are CNN, R-CNN, and Transfer Learning.
(e) We propose Internet memes as a potential approach to combat Technological Singularity, should it happen in the future. Comprehending the underlying idea of Internet memes draws a line between human intelligence and machine intelligence, and may assist in identifying the limitations of AI.

The rest of the paper is organized as follows. Section 2 highlights the materials and methods used in this study. In this section, we present some related works that have been done in the past followed by our proposed work. The results and evaluation based on the experimental analysis are described in Section 3. Section 4 presents a discussion along with a comparative analysis. Section 5, highlights the conclusions and future work.

2 Materials and Methods

This section highlights two components of the study. In the first part, we highlight the related works specific to the study, and in the second part, we describe the proposed work.

2.1 Related Works

A study [24] performed text-based emotion analysis using feature selection. The research mentions only three approaches for the study and highlights sentence ambiguity as one of the biggest issues for emotion detection in sentences. Using Naive Bayes Classifier, Support Vector Machines, Bag of words, N-gram, and WordNetAffect, the accuracy achieved is 81.16%. In [25] the authors proposed the use of deep learning and big data for emotion detection in texts. The emotions detected are happy, sad, and angry in textual dialogues and the study is based on the combination of both semantic and sentiment-based representation. The limitation of the research work is the emotion detection from only textual data and the limited number of emotions used for training purposes. Another study [26] is based on detecting emotions in Hindi-English mixed social media texts incorporating 12000 texts. The emotions considered for the study are happy, sad and anger, and it is observed that CNN-BiLSTM (Long Short Term Memory) performs daily with an accuracy of 83.21%. Likewise [27] recommended a multitask framework for detecting depression and sentiment from suicide notes. The study uses Corpus of Emotion Annotated Suicide notes in English (CEASE) dataset, and IsaCore and AffectiveSpace vector-spaces for analysis. The results show that the proposed system achieves the highest cross validation of 56.47%. Reference [28]

suggested GloVe embeddings for decoding emotions in text. The embeddings relied on LSTM for handling long term dependencies and the model achieved an F-1 score of 0.93. Reference [29] performed a study to analyze human emotions at various tourist locations. Spatial clustering of user-generated footprints was used for the construction of places and online cognitive services were utilized to extract human emotions from facial expressions. Although the dataset incorporates 80 tourist attractions and emotions from over 2 million faces from 6 million photos, the correlation coefficient for the emotions considered is not too good (0.28 and 0.30). Again this hints at how challenging facial emotion detection is [30] presented a novel method for facial emotion detection, using a combination of modified eyemap mouth map algorithms. Overall accuracy of 30% indicates that facial emotion detection is a highly challenging problem in AI. Yet another study [31] was conducted to analyze facial emotion recognition using deep learning methods. The study is based on pre-trained networks like Resnet50, vgg19, Inception V3, and Mobile Net, and deploys transfer learning for analysis. An accuracy of 96% was achieved. In [32] a face-sensitive convolutional neural network was suggested to recognize facial emotions using deep learning. In the first stage, faces in high resolution images are detected and cropped. This is followed by prediction of expressions using CNNs. The proposed model is 95% accurate. Likewise [33] proposed a CNN-based approach for facial emotion recognition, focussing on the removal of background and feature vector extraction. Although the proposed model achieves an accuracy of 96%, the algorithm works with different orientations (less than 30°) due to the unique 24 digit long EV feature matrix. Based on the past research works, we made certain observations. Emotion detection from text is slightly less complicated with respect to emotion detection from facial expressions. We observe that none of the research papers on facial emotion recognition achieve 100% accuracy if more than one emotion is considered. According to [34], there are as many as 27 different types of emotions, and not all of the emotions are considered in many research works. Also, it is interesting to observe that no study has conducted emotion detection from texts and images simultaneously. Since memes may incorporate facial expressions as well as texts, this will be the first article to carry out emotion detection extensively. We have also observed that since emotion detection is not perfected yet, systems often misidentify human emotions. This approach may be essential in validating that comprehending internet memes is a challenge for machines, and may potentially distinguish machine intelligence and human intelligence to some extent.

2.2 Proposed Work

The proposed work is divided into three sections. The first one is the overall system architecture. This is followed by text extraction techniques and emotion detection techniques. Finally, the proposed combined approach is presented.

2.2.1 System Architecture

Internet Memes mostly incorporate emotions and textual data. Our approach is based on analyzing both the components. The architecture of the system is based on two different approaches. First, text extraction using OCR techniques, and second, Emotion detection from facial expressions (Fig. 1). The meme dataset incorporates a variety of popular internet memes like Chubby Bubbles, Distracted Boyfriend, Left Exit, Overly Attached Girlfriend, and Left Exit. There are at least 150 variations of each of these templates. All the memes include texts and faces. The OCR approach is used to extract text from these images, using techniques like the tesseract, pixel link, and east detector. Emotion detection from facial expression has been performed using machine learning techniques like CNN, RCNN, and Transfer Learning using pre-trained DenseNet. A combination of text extraction and emotion detection from facial expression has been considered

for evaluating the capability of AI. The results are evaluated using validation parameters. For facial emotion detection, we rely on validation parameters like validation losses, training losses, validation accuracy and training accuracy. For evaluating the performance of OCR for the memes, we rely on sensitivity and specificity. The combined result determines how capable the system is at comprehending memes.

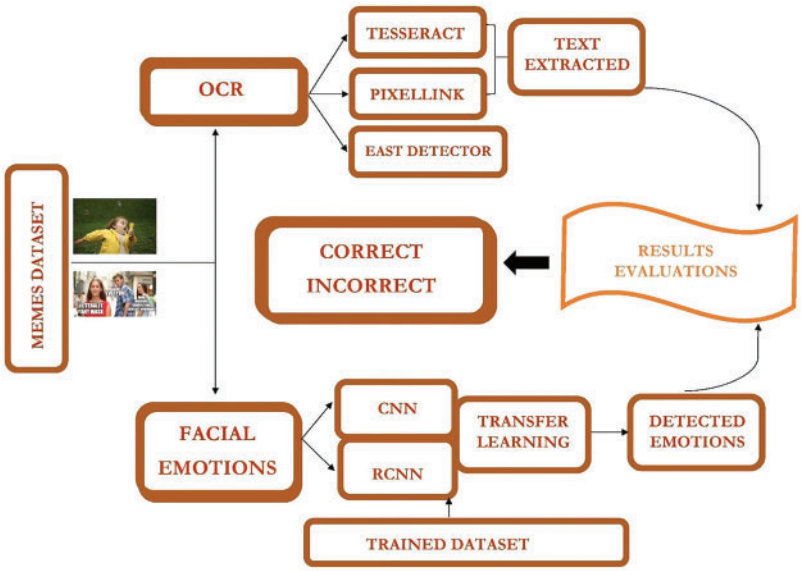

Figure 1: The architecture of the proposed work

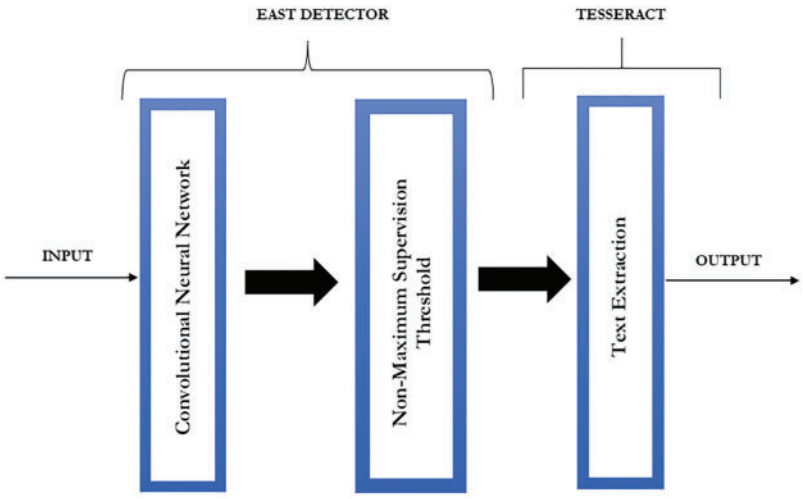

Figure 2: EAST Detector with Tesseract

2.2.2 Text Extraction Using OCR

OCR, or Optical Character Recognition, is a method of identifying text inside pictures and converting it into an electronic structure [35]. These pictures could be of manually written content,

printed text like documents, receipts, name cards, etc. OCR has two sections to it. The initial section is text detection where the literary part inside the picture is identified. This localization of text inside the picture is significant for the second segment of OCR, text recognition, where the content is separated and extracted from the picture. Utilizing these methods together is the means by which one can extricate textual data from any picture.

(a) Tesseract: Tesseract is an open-source OCR engine initially developed by HP (Hewlett-Packard) as proprietary software [36]. Tesseract presumes that the input text picture is genuinely clean, however many input pictures may contain plenty of objects and not only a clean preprocessed text. Consequently, it becomes necessary to have a satisfactory text recognition framework that can distinguish text which could then be easily extradited. There are a reasonable number of ways for text detection like the conventional method of utilizing OpenCV, contemporary method of utilizing Deep Learning models and building your own tailored model. Tesseract is not 100% accurate. It performs ineffectively when the picture is noisy or when the text style of the language is one on which Tesseract OCR is not trained. Different conditions like skewness or brightness of text will likewise influence the capability of Tesseract.

(b) Pixel Link and East Detector: When text instances lie very close to one another, separating them is very hard. PixelLink separates text locations conveniently from an instance segmentation result, rather than from bounding box regression [37]. In PixelLink, a Deep Neural Network (DNN) is prepared to complete two sorts of pixel-wise forecasts, text/non-text prediction, and link prediction. Pixels inside the text instances are marked as positive or text pixels. Likewise, pixels outside the text instance are termed as negative or non-text pixels. Each pixel has 8 neighbors. For a given pixel and one of its neighbors, in the event that they exist in a similar instance, the link between them is marked as positive, else negative. Predicted positive pixels are combined into Connected Components (CC) by predicted positive links. When each CC represents a detected text, instance segmentation is achieved. The EAST detector or the Efficient and Accuracy Scene Text detection pipeline is based on a neural network model. It is trained to instantly predict text instances along with their geometries from full images [38]. The EAST detector model is a fully convolutional neural network specifically introduced for text detection that outputs dense per-pixel predictions of words or text lines. This eliminates several unnecessary intermediary steps like candidate proposal, text region formation, and word partition.

(c) Text Extraction from the dataset: In the proposed approach, two different integrations are used such as the pixel link with the tesseract, and the east detector with the tesseract. The aim is to deploy the bounding box around the detected text and extract the detected text from the bounding box via tesseract. Tesseract is based on finding words from the text lines, chopped into characters, recognizing the characters and joining the chopped characters to make words extracted from the text line. East detector and pixel link are used with tesseract to enhance the accuracy of the system.

East detector and pixel link both are based on the CNN with non maximum supervision threshold and instance segmentation respectively. In the East Detector, pre-trained model is based on the ICDAR 2015 [39] and 2013 [40] training images with F1 score of 80.83. Pixel link also uses the pre-trained dataset of ICDAR 2015. These two models are integrated with the google tesseract for the extraction of words in the lines. Fig. 2 depicts the architecture of the EAST detector with Tesseract. The input is fed into the CNNs which undergoes a non maximum supervision threshold leading to text extraction. While EAST detector is responsible for the CNN operations,

tesseract performs the actual text extraction. In Pixel Link with Tesseract (Fig. 3), the Pixel Link is responsible for CNN operations, predictions and instance segmentation. The result undergoes text extraction, which is performed by Tesseract.

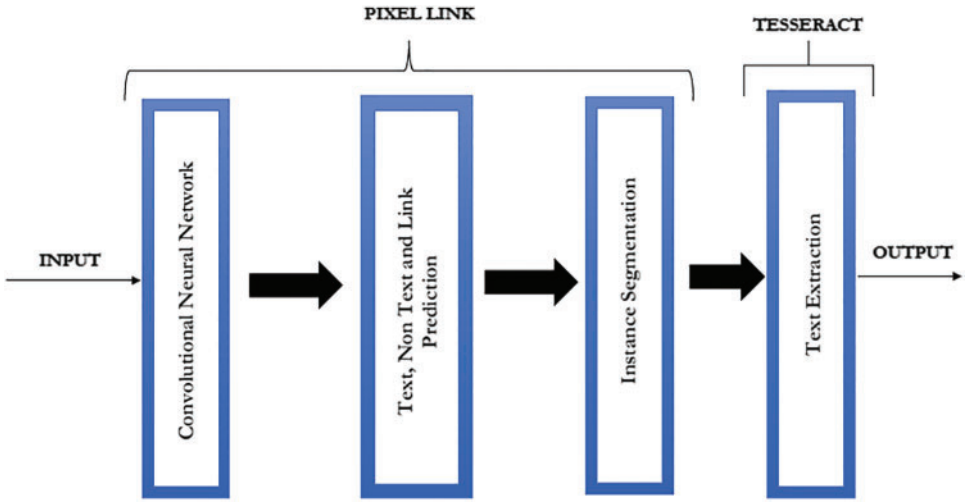

Figure 3: Pixel Link with Tesseract

2.2.3 Emotion Detection From Facial Expressions

The second approach in the proposed work is to detect the emotions from the untrained images. There are three different techniques such as CNN, RCNN and transfer learning with the pre-trained DenseNet applied to recognize the different kinds of emotion in the memes dataset. FER2013 [41] is a dataset which contains 35,685 examples of 48 * 48 pixel grayscale image of faces and is further categorized into 6 different facial expressions like anger, happy, sad, surprise, fear, neutral.

(a) Convolutional Neural Networks (CNN): Convolution neural networks or CNN are basically neural networks with some mathematical operations (usually involving matrices) in the middle of their layers referred to as convolution. Yann LeCun proposed the concept in 1999. Convolution neural networks consist of the, Input layer, Convolutional layer, and Output layer [42]. Input layers are associated with convolutional layers which perform numerous operations, for example, padding, striding, working of kernels for many operations of this layer, this layer is considered as the basic framework of CNN. Basically, the convolution represents the mathematical operation that is used to integrate the two functions and the producible third function is the result outcome. Convolution is applied on the input data combined through the aid of filters to generate the feature map. Afterwards, a pooling layer is added for dimensionality reduction. The FER2013 dataset contains emotions for different images in training, testing folders as well as x, y labels from the pixels and emotion columns respectively. We load the features and label into x and y variables respectively followed by standardizing x by subtracting and dividing by mean and standard deviation respectively. Then emotions images dataset is split into training and testing sets to save the test features and labels. Then, CNN model is created with different functions such as sequential model, two-dimensional convolution layer, batch normalization, max pooling

layer, dropout layer, flatten, dense layer and model fitting is done with batch size 64. The performance is evaluated on the proposed meme dataset.

(b) Region with Convolutional Neural Network (R-CNN): Region with Convolutional Neural Network (R-CNN) was put forward by [43]. R-CNN models initially select a few proposed areas from a picture (like anchor boxes) and afterward label their categories and bounding boxes. A CNN is considered for performing forward computation for extracting features from each proposed zone. Thereafter, the features of each proposed area are considered for predicting their categories and bounding boxes. A selective search is performed on the input image for choosing numerous high-quality proposed areas. A pre-trained CNN is chosen and set, in a shortened structure, before the output layer. It considers forward computation to yield the features selected from the proposed areas. The features and labeled category of each proposed area is incorporated for training multiple support vector machines (SVM) for object classification. Every SVM is utilized for determining whether an instance belongs to any specific category. The labeled bounding box along with features for every proposed region are integrated for training a linear regression model for ground-truth bounding box prediction. To enhance the CNN based emotion extraction technique, a RCNN model is applied. The main purpose behind this model is to bypass the number of selective regions in the image and extract the particular region of emotion.

(c) DenseNet: [44] introduced yet another enhanced convolutional neural network architecture, known as the Dense Convolutional Network (DenseNet). This architecture is based on densely connected convolutional networks such that each layer is connected to every other layer. For every m layer, there will be $m(m + 1)/2$ direct connections. In every layer, the feature map for previous layers is taken as the input, and the current feature maps are used for input in successive layers. This preserves the feed forward nature making DenseNet highly scalable. The scalability can and reach up to hundreds of layers, without the need to address any kind of optimization problems [m]. As each layer receives more supervision from the loss function, due to shorter connections, there is deep supervision. A dense block is made of layers which incorporate batch normalization, ReLu activation and 3×3 convolution. Moreover, a transition layer between two dense blocks includes batch normalization, 1×1 convolution and average pooling. The fulfilling architecture improvises consistently in terms of accuracy as the number of parameters increases without any issues related to overfitting and performance degradation. DenseNet relies on lesser parameters and less computational resources. Since the internal representations of DenseNet are compact, feature redundancy is reduced. DenseNet is designed to allow layers access to feature maps from all other foregoing layers. They can integrate the properties of deep supervision, diversified depth and identity mappings. Their simple connectivity rule allows use of features throughout the network, and leads to better learning and compactness resulting in better accuracy.

(d) Transfer Learning: Transfer learning takes into account the knowledge gained with respect to solving a specific problem and then applying it to a different but related problem [45]. Basically it refers to the knowledge a model has learned from a task performed with an appreciable amount of available and labeled training data that is supposed to be applied for a related task that does not have sufficient data. In this situation, there is no need to start the learning process from scratch. Rather, the methodology analyzed the patterns learned from solving a related task. In transfer learning, the early and center layers are utilized and we just retrain the last layers. It helps influence the labeled data of the task it was at first trained on. Moreover, in transfer learning, we attempt to transfer however

much information as could be expected from the previous task the model was trained on to the new task needing to be done. In the Pre-Training stage, we train the network on a large dataset considering all the parameters of the neural network. Thus the model may take time for learning. In the Fine Tuning stage, we provide the new dataset to fine-tune the pre-trained CNN. The new dataset is considered almost similar to the previous dataset that was used in the pre-training stage. Due to datasets being similar, the same weights can be used for extracting the features from the new dataset. Pre-trained networks incorporate DenseNet which is a part of the convolution neural network (CNN) family which has trained million pictures with thousand classes on the ImageNet (imagenet reference). The principle intention behind utilizing the pre-trained network is to transfer bias, features and weight to the imaging dataset for extracting emotions. Transfer learning does not rely on high computational power, particularly if the emotions dataset does not include many classes (Fig. 4).

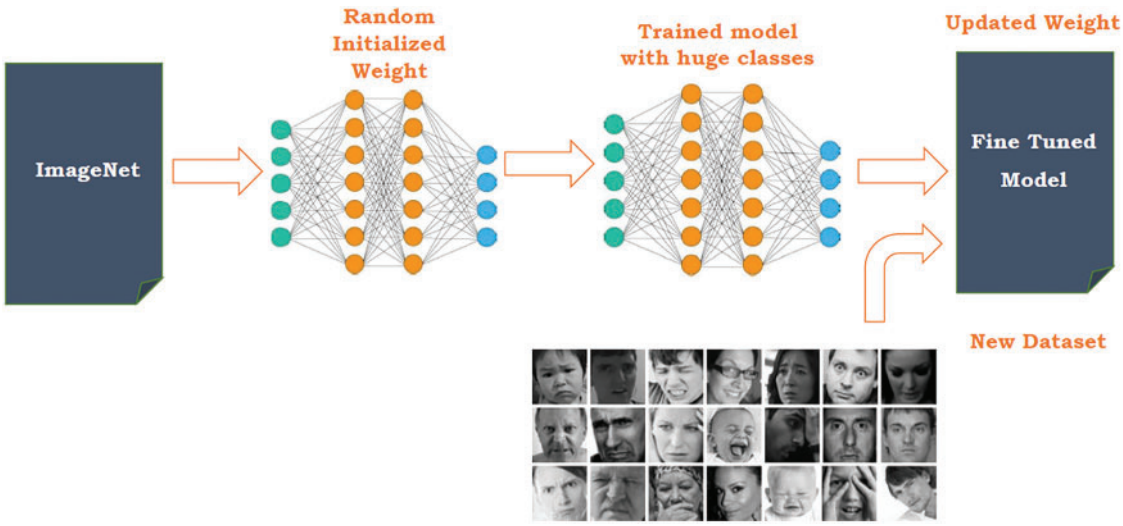

Figure 4: Transfer learning for the proposed work

Given a source emotion-image field $E_s = \{(x_i^s, y_i^s)\}_{i=1}^n$ and, a target emotion-image field $E_t = \{(x_i^t, y_i^t)\}_{i=1}^n$, such that $n_s >> n_t$ our aim is implementing the process of transfer learning for optimizing a CNN with E_s and E_t, for improving the classification performance in E_t. The methodology involves reducing the domain distribution discrepancy at the fully connected layers when CNN is being trained with E_s and E_t simultaneously. There are two reasons for choosing CNN as our base transfer transfer learning model. First, CNN is most appropriate for facial emotion detection with respect to conventional manually crafted features. Second, CNN is capable of learning more transferable image features at the initial layers.

We can use E_s to improvise the performance of a CNN on E_t, therefore we can engage E_s and E_t for training the same CNN simultaneously. Nonetheless, for facial emotion detection, there may be a discrepancy between field distributions $A(\mathbf{X}^s)$ and $B(\mathbf{X}^t)$. The image features transit from general to field specific with CNN, hence at the fully connected layers, there will be a decrease in transferability. The transfer learning approach used in this study can minimize the domain shift at these fully connected layers because of two things. Firstly, there will be a reduction in

marginal distribution discrepancy $\{A(\mathbf{K}^{si}), B(\mathbf{K}^{ti})\}_{i \in G}$ from layer to layer. Secondly, it is possible to reduce joint distribution discrepancy $A(\mathbf{K}^{s1},\ldots, \mathbf{K}^{s|G|})$ and $A(\mathbf{K}^{t1},\ldots, \mathbf{K}^{t|G|})$. $\{\mathbf{K}^{si}\}_{i \in G}$, and $\{\mathbf{K}^{ti}\}_{i \in G}$ are features at the fully connected layers. G may be defined as a set of chosen fully connected layers positioned for joint distribution. It may also encompass fully connected layers of CNN.

2.2.4 Combined Approach for OCR and Facial Emotions Extraction

In this paper, a hybrid approach is proposed for the text detection and facial emotion in the memes picture. Algorithm 1 represents the hybrid approach for the OCR and facial emotion extraction.

Algorithm 1: Hybrid approach for OCR and facial emotion extraction.

1: Algorithm: Detection of Facial Emotions and Text in the Memes Pictures
2: Result: Combined Facial Emotions and Text Detection
3: While begin **do**
4: Conditions for putting the Boxes on the text and Extraction
5: If t*ext detection in the picture* **then**
6: put the boxes on the text images:
7: else
8: No Text
9: If *words or character detection in the image* **then**
10: Text Extraction
11: else
12: No Text or wrong characters;
13: If Emotions are detected in the image **then**
14: applying the boxes around the image;
15: else
16: No emotions or wrong emotions detected;
17: If Matching the emotions with the trained model emotion then
18: Emotion Successful matched:
19: else
20: No emotion or wrong emotions detected:
21: If matching text and emotion combined **then**
22: Emotions and Text successful extracted:
23: else
24: System failed;
25: end
26: end
27: end
28: end
29: end

The algorithm is based on a hybrid approach, i.e., text extraction and facial emotion detection. The text extraction from the memes images is based on the pixel link and tesseract. Pixel link model and tesseract assist in extracting the text from the bounding boxes. With the help of pixel link, text can be easily recognized and extracted through the use of tesseract. There are

two possibilities: either text is fully extracted or wrong/partially extracted. In the next step, the algorithm moves towards facial emotion recognition. For facial recognition, transfer learning with DenseNet is used because validation accuracy of the model is high as compared to the other models. For facial emotions, the DenseNet model is applied via transfer learning. As per the trained model, the proposed system is able to detect the six kinds of emotions. The first step of this model is to detect the person's face in the image and then apply the bounding box. In the next step, the bounding box face will match with the trained model information and detect the correct emotion for the detected face. If the emotion and text are successfully matched with the system, the system provides the correct results. Otherwise, the system fails.

3 Results and Evaluations

In this section, we have highlighted the statistical parameters used to evaluate the proposed system, along with evaluation datasets and experimental analysis

3.1 Statistical Parameters

We rely on two statistical parameters for this study, i.e., Sensitivity and Specificity.

3.1.1 Sensitivity

Sensitivity is defined as the quantity or percentage of positives that are identified correctly. For a classifier, Sensitivity is given by the ratio between how many instances were correctly identified as positive with respect to how many were actually positive. It is also known as True Positive Rate (TPR) and is given by the formula:

$$\text{Sensitivity} = \text{True Positives}/(\text{False Negatives} + \text{True Positives})$$

3.1.2 Specificity

Specificity is defined as the quantity or percentage of samples that test negative using the test concerned that are actually negative. For a classifier, Specificity is given by the ratio between how many instances were correctly classified as negative with respect to how many were actually negative. It is also known as True Negative rate (TNR), and is given by the formula:

$$\text{Specificity} = \text{True Negatives}/(\text{False Positives} + \text{True Negatives})$$

3.2 Evaluation Datasets

For the performance evaluation of the OCR approach and Emotion based approach, five different datasets are used, i.e., Chubby Bubbles Girl, Distracted Boyfriend, Left Exit, Overly Attached Girlfriend, and Roll Safe. All these datasets have 150 images of meme variation based on a single template. Hence, although the template is the same, each meme has a different underlying meaning. Most of the memes use a combination of texts and emotions. Tab. 1 represents the details of the evaluation datasets.

Every dataset contains 150 images embedded with the text and emotions except the left exit dataset. There is a variation of text or emotion for every image used in the particular dataset.

3.3 Experimental Analysis

To evaluate the performance of facial emotion-based approach, four parameters are used such as training accuracy, validation accuracy, training losses and validation losses for the CNN, RCNN and transfer learning with pre-trained DenseNet. Tab. 2 represents the performance evaluation of model training and validation.

Table 1: Details of evaluation dataset

Dataset	Features	Number of images	Details
Chubby bubbles	Text and emotions	150	Text variation in every image, face expression in left and right side
Distracted boyfriend	Text and emotions	150	Text variation in every image three different facial expressions
Left exit	Text	150	Variation in text
Overly attached girlfriend	Text and emotions	150	Variation in facial expression, variation in text
Roll safe	Text and emotions	150	Variation in facial expression, variation in text

Table 2: Performance evaluation of model training and validation

Parameters	CNN	RCNN	Transfer learning with pre-trained DenseNet
Training accuracy	0.8697	0.8921	0.9319
Validation accuracy	0.6179	0.6524	0.7920
Training losses	0.4288	0.2987	0.1987
Validation losses	1.2321	1.2492	1.2292
Epochs	50	50	50

For the CNN model, training accuracy, validation accuracy, training losses, validation losses are 0.8697, 0.6179, 0.4288 and 1.2321 respectively. The performance evaluation of the RCNN and transfer learning with DenseNet are 0.8921:0.9319, 0.6524:0.7920, 0.2987:0.1987, 1.2492:1.2292 respectively. The performance for Transfer learning with DenseNet is relatively higher as compared to the CNN and RCNN. Tab. 3 represents the performance evaluation of the models on the evaluation dataset (memes dataset). For the performance evaluation of the meme's dataset on trained models, Sensitivity and Specificity are used. True positive refers to the ability of the system to detect the same emotion correctly as in the picture. True negative is for no emotion is in the picture and also that the system does not detect the emotions as depicted in the pictures. False positive refers to incorrect emotion detection in the picture and false negative means that there is no emotion in the picture but the system detects emotion in the picture. In this study, there are five different datasets that are considered and evaluated through the above-mentioned formula. In most cases sensitivity and specificity values are either 1 or 0 except the overly attached girlfriend. In the chubby bubbles' dataset, no emotion is detected so it falls under the false negative category. In the distracted boyfriend dataset, the system detected 30 correct emotions (true positive) and 140 wrong emotions (false positive). No emotion is available in the left exit dataset and the system could not detect any emotions on the dataset (true negative). Roll safe dataset has emotions in the pictures but the system does not detect emotions, so it falls under false negative category. The system works better for the overly attached girlfriend, as compared to the other dataset since 136 emotions are detected as true positive, 12 as false negative and 2 as false positive.

Table 3: Performance evaluation for the emotions on the memes Dataset

Dataset	CNN	RCNN	Transfer learning with pre-trained DenseNet
Chubby bubbles	Sensitivity = 0 Specificity = 0	Sensitivity = 0 Specificity = 0	Sensitivity = 0 Specificity = 0
Distracted boyfriend	Sensitivity = 1 Specificity = 0	Sensitivity = 1 Specificity = 0	Sensitivity = 1 Specificity = 0
Left exit	Sensitivity = 0 Specificity = 1	Sensitivity = 0 Specificity = 1	Sensitivity = 1 Specificity = 0
Overly attached girlfriend	Sensitivity = 0.918 Specificity = 0	Sensitivity = 0.905 Specificity = 0	Sensitivity = 0.959 Specificity = 0
Roll safe	Sensitivity = 0 Specificity = 0	Sensitivity = 0 Specificity = 0	Sensitivity = 0 Specificity = 0

Tab. 4 represents the performance evaluation for the text extraction from memes dataset using Tesseract, East Detector and Tesseract, and Pixel Link and Tesseract. The evaluation of these models is done on the basis of the full text extraction from the dataset. In the Chubby Bubbles dataset, out of 150 images Tesseract, East Detector and Tesseract, and Pixel Link and Tesseract extracted texts completely for 16, 20 and 20 images respectively. For the Overly Attached Girlfriend dataset, performance of all models is not satisfactory. In the Left Exit dataset, performance of the models is better as compared to the Overly Attached Girlfriend dataset but not for others. The highest number of texts extracted from images is for the Roll Safe dataset by means of Tesseract, East Detector and Tesseract, and Pixel Link and Tesseract, i.e., 32, 35 and 35 respectively. Fig. 5 depicts the overall performance of different models with respect to training and text extraction.

Table 4: Performance evaluation for the text on the memes Dataset

Dataset	Tesseract (full text extracted)	East detector + tesseract (full text extracted)	Pixel link + tesseract (full text extracted)
Chubby bubbles	16	20	20
Distracted boyfriend	12	23	23
Left exit	2	8	10
Overly attached girlfriend	0	6	10
Roll safe	32	35	35

4 Discussions

In this study we have explored yet another limitation of Artificial Intelligence and successfully demonstrated how AI is incompetent for understanding Internet Memes. In this section we present observations and comparative analysis of the existing work.

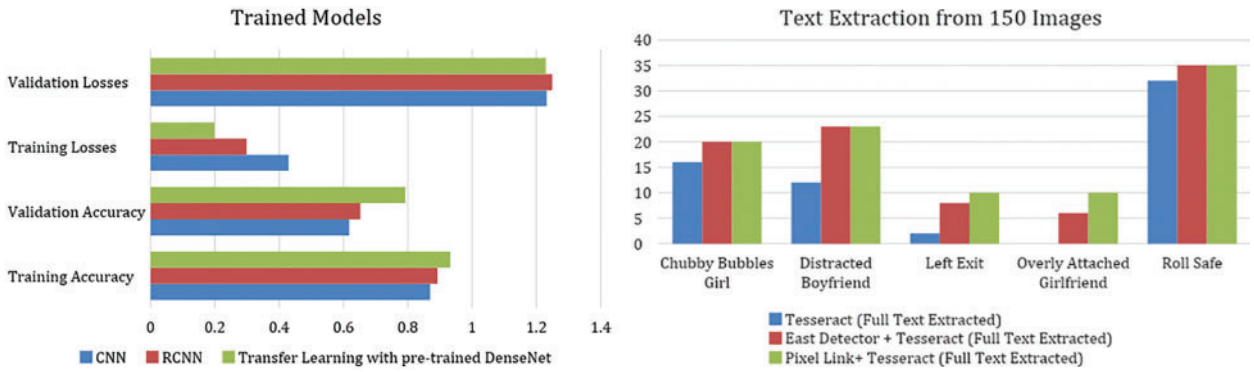

Figure 5: Performance of the different models trained and text extraction

4.1 Observations

The approach used for the study involved a combination of Text Extraction techniques using OCR and Facial Emotion Detection using RÇNN and Transfer Learning. While we were able to conclude that AI cannot interpret memes even using a combination of the two techniques, we made some observations with respect to Text Extraction and Facial Emotion Detection. These observations may be the reasons why AI still fails at understanding something as complex as Internet Memes which is easily understood by humans and is a shared culture in society. These observations also explain why understanding Internet Memes will always remain an unsolvable problem for AI.

4.1.1 Text Extraction Challenges

The study uses multiple Optical Character Detection techniques, which is sufficient for extracting texts from the data. However, despite using three authentic and powerful OCR techniques like the Tesseract, Pixel Link and East Detector, there were still issues with respect to text extraction.

(a) Tilted Texts: Tilted Texts are not easy for extraction. A large portion of internet memes is dominated by tilted texts. Since text extraction and identification is based on artificial intelligence techniques, identifying tilted texts may be another challenge in itself. Many alphabets of English language may be difficult for the software to recognize if they look similar and are tilted. For example, (M and W), (O and Q), (n and u), (d, b, p and q) etc.

(b) Hidden Text: While performing the OCR for text extraction, it was observed that text extracted for few memes was completely different from the text seen in the images. These texts looked like names or messages that were hidden within the image. This kind of extraction will make it further complicated for AI to understand Internet Memes.

(c) Sentences on the Left and Right: Many times, Internet Memes incorporate dialogues that exist in the form of multiple sentences. They may exist as sentences in the left hand side and right hand side of the meme image. It is a challenging task for AI to interpret that the sentences on the left and right are different. OCR techniques read sentences in a single line, and when this text is extracted it is a combination of half of the sentences from each side in one line. This keeps on repeating until the entire text extraction is done, and the resulting output is a combination of words that does not make sense.

(d) Font and Size Issues: Since Internet Memes are continuously generated manually as well as automatically, there is a variation in the font type and its size. Due to the variation of characters in different fonts, not all characters are extracted perfectly. Some characters seem indistinguishable across various fonts. Hence even if there is an AI that identifies all characters perfectly, there will be some kind of confusion for the AI for identifying characters that bear striking resemblance. For normal text, it is easy to judge the character from the American Standard Code for Information Interchange (ASCII) values, however text extraction from image, makes the task utterly challenging, since text is also treated as an image.

(e) Image with text within the image: Many internet memes include logos that incorporate some amount of text. OCR techniques fail to recognize texts within logos that are a part of the meme image.

(f) Similar looking characters: Similar looking characters across different fonts (including symbols) are challenging for text extraction. Since AI techniques recognize and classify images (texts) based on previous learning, it is not surprising to detect a large number of misclassifications or incorrectly extracted texts.

(g) Color Contrast: Color contrast seems to play a major role in OCR text extraction. If the contrast between the text color and the background color is not significant, the OCR may find text extraction difficult. On the other hand if there is less contrast between texts and the background image, yet the texts are enclosed in a boundary, or there are some lines around the text that make it easily distinguishable from the background, text extraction becomes simple.

(h) Blurred Text: Many internet memes encompass a combination of blurred and clear texts in order to put emphasis on clear text. While performing the OCR, it was observed that text extraction fails to detect blurred texts. Moreover, even if AI becomes sophisticated to somehow figure that the blurred image is indeed a text, there are many similar looking texts that may produce similar blurred images, thus making it difficult for AI to extract the text and interpret the meme.

(i) Additional objects in the same image: meme creativity does not end with only adding new texts to existing templates. Many Internet memes are modified to incorporate images that give a different meaning to it. We observed that adding multiple images in the meme template makes text extraction difficult.

(j) Overlapping texts: The concept of overlapping images is not new, and is often employed as a technique in CAPTCHAs for the process of authentication. The underlying idea is that overlapping texts is hard to identify by automated bots that break into authentication systems [46,47]. Similarly, extracting overlapping texts and identifying them may be a challenge for OCR. Often these texts are tilted, and stretched in different angles. Some Internet Memes have overlapping texts within them.

4.1.2 Facial Emotion Detection Challenges

The study uses three artificial intelligence techniques, namely CNN, RCNN, DenseNet, and Transfer Learning, which are sufficient for facial emotion detection. However, despite using these robust techniques, it was observed that facial emotion detection in Internet memes is not an easy task.

(a) Tilted Faces: Many images deal with tilted faces of humans, which makes emotion detection a challenge. At certain angles, faces may not only be tilted, but partially or almost fully hidden. This makes facial image detection hard and emotion detection even harder.
(b) Animal Faces: As the Internet meme culture evolves, it is interesting to see pictures of animals, birds, cartoons etc. being a part of more and more Internet Memes. There is no way to analyze the facial emotion detection for these, which makes our claims even more plausible.
(c) Blurred images: Facial emotion detection is hard for blurred images. Many internet memes have deliberately blurred images which are tied down to the underlying meanings. Unless emotions are detected by AI, understanding the meme is a challenge.

The Internet Meme culture has been rapidly progressing over the Internet using social media platforms specifically. It is humor that travels over the web and connects individuals at the level of their conscience. Everyday, hundreds of thousands of memes are generated, some tweaked from the previous templates, some created from the scratch. Hence there is a plethora of Internet Memes travelling over the network, and this kind of data is difficult to manage. The creativity of humans has made a single meme template be represented in millions of different forms, emanating a million meanings. This is a challenge for AI to understand, since AI learns from patterns and features of data. These patterns and features vary largely when it comes to Internet memes. As more and more Internet Memes will be generated in the future, understanding Internet Memes will become a much harder problem for AI, since it has to learn what is already there and make decisions for the ones that it has never seen. Hence understanding Internet Memes completely may always remain an unsolvable problem for AI.

Table 5: Comparative analysis of our proposed work

Author and year	Proposed work	Methodology/parameters	Results
Ma et al., 2021 [48]	Scene text recognition	Position information enhanced encoder-decoder framework	Accuracy between 85% to 94%
Pandey et al., 2021 [49]	Text Extraction from scene images	Hybrid deep neural network with adaptive galactic swarm optimization	F1- score = 95.2, precision = 93.79
Awan et al., 2021 [50]	Extraction and Classification of Key Phrases in Text	TopicalPostionRank techniques incorporating both topical and positional information	F1- score = 0.73
Li et al., 2021 [51]	Facial Emotion Detection via ResNet-50	Convolutional Neural Networks	Accuracy $\sim$ 95%
Said et al., 2021 [32]	Facial Emotion Recognition	Face-sensitive convolutional neural network (FS-CNN)	Average mean precision $\sim$ 95%
Our proposed work, 2021	Establishing the inability of AI in comprehending Internet Memes (experimental evidences)	Hybrid Text Extraction and Facial Emotion Detection (OCR techniques, CNN, RCNN and Transfer Learning)	Models accuracy: CNN, RCNN and transfer learning are 0.6179, 0.6524 and 0.7920 respectively

4.2 Comparative Analysis

The Comparative Analysis of our work with some previously established research works in similar area with is depicted in Tab. 5

Comparative analysis depicts that text extraction and facial emotion detection are yet to achieve 100% accuracy, and are still a challenge for AI. Moreover the studies conducted in the past reveal how text extraction and facial emotion detection are different areas of study, and have been explored widely, but separately. Most of the studies conducted deploy already existing datasets for texts and facial emotion detection. However, our datasets have been assembled by accumulating memes that have several variations in the form of texts and images. Since Internet Memes incorporate texts as well as emotions this is the first study to use hybrid approaches by combining text extraction and emotion detection for interpreting Internet Memes. As stated in the introduction section, we used experimental evidence to establish how Internet memes can be used to distinguish human and machine intelligence. A number of techniques were used for OCR (Tesseract, Pixel Link, and East Detector) as well as facial emotion detection (CNN, R-CNN, and Transfer Learning), and it can be asserted based on the experimental analysis and comparative analysis that understanding memes is challenging for AI.

5 Conclusion and Future Work

Over the last few decades learning has become relatively easy for AI, and AI has exhibited quick decision making and efficiency. However, there are outliers that have cost humanity in different ways, and this has made researchers think in the direction of technological singularity and overcoming it using limitations of AI. In this study we explore yet another limitation of AI i.e., comprehending Internet Memes. Incapability of AI to understand Internet memes certainly draws a line between human intelligence and machine intelligence, and based on the evidence, it may be asserted that understanding Internet memes is a big challenge for AI. Since internet memes incorporate texts and images, we used the hybrid technique of text extraction and facial emotion detection. Text extraction relies on OCR techniques like tesseract, pixel link and east detector, while facial emotion detection involves algorithms like CNN, RCNN, and Transfer Learning. Despite using a combination of all these techniques, we observe that based on the experimental analysis, interpreting Internet Memes could be a difficult task for AI. There are certain concerns with respect to text extraction and facial emotion detection that need addressing and AI has a long way to go before it can comprehend memes. Moreover, with an increase in the number of memes on a daily basis, training a system to understand Internet Memes will always pose a challenge for AI. The approach may also serve as one of the potential classes of Turing Tests. In the future, we would like to establish more such experimental evidence on how AI cannot comprehend Internet Memes and explore more algorithms which can assist us in the same.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. Brown, "The social life of autonomous cars," *Computer*, vol. 50, no. 2, pp. 92–96, 2017.
[2] K. Crawford, "Artificial intelligence's white guy problem," *The New York Times*, vol. 25, no. 6, 2016.
[3] M. Upchurch, "Robots and AI at work: The prospects for singularity," *New Technology, Work and Employment*, vol. 33, no. 3, pp. 205–218, 2018.

[4] L. Sweeney, "Discrimination in online ad delivery," *Queue*, vol. 11, no. 3, pp. 10–29, 2013.
[5] K. Koruyan and B. Bedir, "A legal and administrative evaluation of robots and autonomous vehicles," in *The 11th Int. Conf. 'Economies of the Balkan and Eastern European Countries*, Romania, pp. 53, 2019.
[6] I. Priyadarshini, "Cyber security risks in robotics," in *Cyber Security and Threats: Concepts, Methodologies, Tools, and Applications*. IGI Global, pp. 1235–1250, 2018.
[7] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.
[8] R. Yampolskiy, "The singularity may be near," *Information: An International Interdisciplinary Journal*, vol. 9, no. 8, pp. 190, 2018.
[9] I. Priyadarshini and C. Cotton, "Intelligence in cyberspace: The road to cyber singularity," *Journal of Experimental and Theoretical Artificial Intelligence*, pp. 1–35, 2020.
[10] I. Priyadarshini, P. R. Mohanty and C. Cotton, "Analyzing some elements of technological singularity using regression methods," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3229–3247, 2021.
[11] V. Puri, S. Jha, R. Kumar, I. Priyadarshini, L. Son *et al.,* "A hybrid artificial intelligence and internet of things model for generation of renewable resource of energy," *IEEE Access*, vol. 7, pp. 111181–111191, 2019.
[12] S. Quek, G. Selvachandran, M. Munir, T. Mahmood, K. Ullah *et al.,* "Multi-attribute multi-perception decision-making based on generalized T-spherical fuzzy weighted aggregation operators on neutrosophic sets," *Mathematics*, vol. 7, no. 9, pp. 780, 2019.
[13] I. Priyadarshini and C. Cotton, "A novel LSTM-CNN–grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing*, pp. 1–22, 2021.
[14] S. Patro, B. K. Mishra, S. K. Panda, R. Kumar, H. V. Long *et al.,* "A hybrid action-related k-nearest neighbour (HAR-KNN) approach for recommendation systems," *IEEE Access*, vol. 8, pp. 90978–90991, 2020.
[15] I. Priyadarshini, R. Kumar, L. Tuan, L. Son, H. Long *et al.,* "A new enhanced cyber security framework for medical cyber physical systems," *SICS Software-Intensive Cyber-Physical Systems*, pp. 1–25, 2021.
[16] N. Rokbani, R. Kumar, A. Abraham, A. Alimi, H. Long *et al.,* "Bi-heuristic ant colony optimization based approaches for traveling salesman problem," *Soft Computing*, pp. 1–20, 2020.
[17] K. Sailunaz and R. Alhajj, "Emotion and sentiment analysis from Twitter text," *Journal of Computational Science*, vol. 36, pp. 101003, 2019.
[18] N. Sebe, M. Lew, Y. Sun, I. Cohen, T. Gevers *et al.,* "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.
[19] T. Vo, R. Sharma, R. Kumar, L. H. Son, B. T. Pham *et al.,* "Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering," *Journal of Intelligent & Fuzzy Systems*, pp. 1–13, 2020.
[20] I. Priyadarshini, H. Wang and C. Cotton, "Some cyberpsychology techniques to distinguish humans and bots for authentication," in *Proc. of the Future Technologies Conf.*, USA, Springer, pp. 306–323, 2019.
[21] I. Priyadarshini and C. Cotton, "Internet memes: A novel approach to distinguish humans and bots for authentication," in *Proc. of the Future Technologies Conf.*, USA, Springer, pp. 204–222, 2019.
[22] S. Jha, R. Kumar, L. Son, M. Abdel-Basset, I. Priyadarshini *et al.,* "Deep learning approach for software maintainability metrics prediction," *IEEE Access*, vol. 7, pp. 61840–61855, 2019.
[23] D. Dansana, R. Kumar, J. Adhikari, M. Mohapatra, R. Sharma *et al.,* "Global forecasting confirmed and fatal cases of COVID-19 outbreak using autoregressive integrated moving average model," *Frontiers in Public Health*, pp. 8, 2020.
[24] P. Apte and S. Khetawat, "Text-based emotion analysis: Feature selection techniques and approaches," in *Emerging Technologies in Data Mining and Information Security*. Springer, pp. 837–847, 2019.
[25] A. Chatterjee, U. Gupta, M. Chinnakotla, R. Srikanth, M. Galley *et al.,* "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, pp. 309–317, 2019.

[26] T. Sasidhar, B. Premjith and K. Soman, "Emotion detection in Hinglish (Hindi+English) code-mixed social media text," in *Third Int. Conf. on Computing and Network Communications*, India, Elsevier, vol. 171, pp. 1346–1352, 2020.
[27] S. Ghosh, A. Ekbal and P. Bhattacharya, "A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes," *Cognitive Computation*, pp. 1–20, 2021.
[28] P. Gupta, I. Roy, G. Batra and A. Dubey, "Decoding emotions in text Using GloVe embeddings," in *2021 Int. Conf. on Computing, Communication, and Intelligent Systems*, India, IEEE, pp. 36–40, 2021.
[29] Y. Kang, Q. Jia, S. Gao, X. Zeng, Y. Wang *et al.,* "Extracting human emotions at different places based on facial expressions and spatial clustering analysis," *Transactions in GIS*, vol. 23, no. 3, pp. 450–480, 2019.
[30] A. Joseph and P. Geetha, "Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow," *The Visual Computer*, vol. 36, no. 3, pp. 529–539, 2020.
[31] K. Chaudhary, T. Nguyen and D. Hemanth, "Deep learning-based facial emotion recognition for human-computer interaction applications," in *Neural Computing and Application*. Springer, pp. 1–18, 2021.
[32] Y. Said and M. Barr, "Human emotion recognition based on facial expressions via deep learning on high-resolution images," in *Multimedia Tools and Applications*. Springer, pp. 1–13, 2021.
[33] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Applied Sciences*, vol. 2, no. 3, pp. 1–8, 2020.
[34] A. Cowen and D. Keltner, "Self-report captures 27 distinct categories of emotion bridged by continuous gradients," *Proc. of the National Academy of Sciences of the United States of America*, vol. 114, no. 38, pp. E7900–E7909, 2017.
[35] R. Mithe, S. Indalkar and N. Divekar, "Optical character recognition," *International Journal of Recent Technology and Engineering*, 2013.
[36] R. Smith, "An overview of the Tesseract OCR engine," in *IEEE Ninth Int. Conf. on Document Analysis and Recognition*, Brazil, vol. 2, pp. 629–633, 2007.
[37] D. Deng, H. Liu, X. Li and D. Cai, "Pixellink: Detecting scene text via instance segmentation," *arXiv preprint* arXiv: 1801.01315, 2018.
[38] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou *et al.,* "East: An efficient and accurate scene text detector," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, USA, pp. 5551–5560, 2017.
[39] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, S. A. Bagdanov *et al.,* "competition on robust reading," in *2015 13th Int. Conf. on Document Analysis and Recognition*, USA, pp. 1156–1160, 2015.
[40] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, M. L. Bigorda *et al.,* "Robust reading competition," in *2013 12th Int. Conf. on Document Analysis and Recognition*, USA, pp. 1484–1493, 2013.
[41] P. Carrier, A. Courville, I. Goodfellow, M. Mirza and Y. Bengio, *FER-2013 face database*. Universit de Montral, Canada, 2013.
[42] I. Priyadarshini and V. Puri, A convolutional neural network (CNN) based ensemble model for exoplanet detection. In: *Earth Science Informatics*. Springer, 2021.
[43] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, USA, pp. 580–587, 2014.
[44] G. Huang, Z. Liu, L. Van der Maaten and K. Weinberger, "Densely connected convolutional networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, USA, pp. 4700–4708, 2017.
[45] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global, pp. 242–264, 2010.
[46] T. Tuan, H. Long, L. Son, R. Kumar, I. Priyadarshini *et al.,* "Performance evaluation of Botnet DDoS attack detection using machine learning," *Evolutionary Intelligence*, pp. 1–12, 2019.
[47] V. Puri, I. Priyadarshini, R. Kumar and C. Le, "Smart contract based policies for the Internet of Things," *Cluster Computing*, pp. 1–20, 2021.

[48] X. Ma, K. He, D. Zhang and D. Li, "PIEED: Position information enhanced encoder-decoder framework for scene text recognition," *Applied Intelligence*, pp. 1–10, 2021.

[49] D. Pandey, B. Pandey and S. Wairya, "Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images," *Soft Computing*, vol. 25, no. 2, pp. 1563–1580, 2021.

[50] M. Awan and M. Beg, "TOP-rank: A TopicalPostionRank for extraction and classification of keyphrases in text," *Computer Speech & Language*, vol. 65, pp. 101116, 2021.

[51] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–34, 2021.