# Deep Learning Approach for COVID-19 Meme Categorization

S.Rajasree, R.Chinmaya, K.Dharsini Prabha, B. Indujapriya

**Abstract:- Meme is an image, which contains both text and images. Now-a-days memes are very popular and it becomes viral by sharing the memes in Instagram, Twitter and other social media. Some information from memes may be fake. By sharing the unwanted memes may harm others or may cause some other social issues. COVID-19 dataset has been considered for meme categorization in the proposed work. By classifying the memes, it is easy to find whether the meme is positive or negative or covid related. The dataset used for this project is memes from social media. The memes can be classified by using OCR technique and YOLO technique. Proposed methodology is Text Sentimental Analysis and Image Analysis to categorize the emotion of the memes related to COVID-19. YoloV5 is used for object detection and OCR Technique is used for Text extraction, LSTM is used for Sentimental Analysis from text. OCR is used to recognize and extract the text in the meme and yolov5 is used for object detection and label it. The proposed method can be evaluated using measures such as accuracy and precision.**

*Keywords:- Covid-19 Memes, Text Analysis, Image Processing, Sentimental Analysis.*

## I. INTRODUCTION

The new set of communication technologies have previlage of expressing the people opinion in different forms like text, image, videos etc. one of the distinct cultural participation is internet meme. Memes are images in which it describes the author feelings in the form of image. Usually, the sentiment analysis is performed from the collections of texts or sentences. The sentimental analysis can be extracted from both text and object in the meme images. By classification of Covid-19 memes helps people to understand which category it belongs to, such as funny, sad, or awareness. The text analysis is done using LSTM and object detection from the meme image is done using YoloV5 algorithm.

EasyOCR is a tool which can have multiple languages at the time but they have to be compatible with each other. English is compatible with all languages and converting the images into machine readable code. It is one of the fields in pattern recognition, artificial intelligence and computer vision. This technique of extracting text from images is carried out in different work environments where it contains images with text data. The function takes an image object and returns the recognized text.

By Using YOLOV5, YOLOV4,YOLOV5 can achieve 140FPS on Tesla P100 rapid detection, YOLOV4 is only of 50 FPS. Meanwhile, the size of YOLOV5 is only of 27mb, while the size of yolov4 is of 244mb. YOLOV5 also has approximately the same accuracy as yolov4. Yolov5 has inherited the advantage of yolov4, namely adding SPP-Netmodifying the SOTA method and putting forwardnew data enhancement methods such as Mosaic training, self-adversary training(SAT),and multi-channel feature replacing FPN fusion with PANet In Image processing, classification ,which is mainly used to structure the image into certain type of information ,to respective category or instance ID to describe the nature of the image. Detection, the classification task is concerned with the whole. Compared with classification, detection gives the understanding of picture foreground and background. The output of the detection model is a list, and each item uses a data group for detecting the target.ie, rectangular detection box. Then the label is fed to the classification module to find whether the meme is related to covid-19.

## II. LITERATURE SURVEY

The literature survey is based on the text classification using various machine learning algorithms, it also covers the detailed explanation to image processing and sentimental analysis using multimodal approach.

Akshi Kumar & Geetanjali Garg [1] used the concept of multimodal approach to determine the Image sentiment scoring using SentiBank and SentiStrength with R-CNN and Text sentiment scoring using Novel context aware hybrid technique. Amalia Amalia, Amer Sharif, Fikri Haisar, Dani Gunawan, Benny B Nasution [2] classifies the memes by image processing and OCR Tesseract combined with Naive Bayer Algorithm. The accuracy is almost depending on the OCR result which utilizes tesseract engine. Zhao Jianqiang , Gui Xiaolin , Zhang Xuejun [3] used word embeddings method obtained by unsupervised learning based on large twitter corpora, and also combined with n-grams features and word sentiment polarity, integrated into deep convolution neural network for training and predicting sentiment classification labels. Savitha Choudhary, Nikhil Kumar Singh, Sanjay Chichadwani [4] recognized the text using self-trained Neural Network and using Maximally Stable Extremal Regions (MSER) . The text was individually isolated as single characters by simple algorithms on the binary image. Shahriar Shakir Sumit, Junzo Watada , Anurava Roy, DRA Rambli1 [5] compares the performance of YOLO and Mask R-CNN. YOLO detect the real time object by improving computer vision task that requires both

successful object localization in order to draw a boundary box and Mask R-CNN benefit from extra data even if the data is unlabelled. Ashraf Elnagar, Ridhwan Al-Debsi, Omar Einea [6] used the concept of automatically tagging a textual document with most relevant labels or categories. Technique used are Learning Vector Quantization(LVQ) classifier and self-organization maps under neural Network. Wenchi Ma, Yuanwei Wu, Feng Cen, Guanghui Wang [7] proposed an innovative object detector by the leveraging of deep features learned in highlevel layers. Single shot multi-box detector(SSD) is used for multi-box matching and sifting. MDFN utilizes the VGG-16 as its base network for the features of extraction , and adopted the transfer learning as the basic learning style for object detection. LA szlo Nemes & Attila Kiss [8] trained a model to sentiment prediction by looking correlations between words using Natural Language Processing and with Sentiment Classification using recurrent Neural Network.Kashfia Sailunaz, Reda Alhajj [9] implemented using Naïve bayes and support vector Machine(SVM) and Random Forest. It focuses on the emotions from text in their tweets and use them for generating recommendations. In most of the existing system, memes were classified using SVM classifier or Naïve Bayes, but in our proposed modal covid -19 memes are classified using multimodal approach.

## III. PROPOSED SYSTEM

The proposed system is used to identify the emotions of covid-19 memes using multimodal approach. Our goal is to train the model with the training dataset and finally evaluate model's performance with the test dataset. The dataset for image consists of images which are related to covid-19, and dataset for text contains sentences which are extracted from the covid-19 memes. The training and testing sets are splitted up for image dataset. After completing the training process, the testing will be done in both image and text module using lstm and yolov5. Finally with the help of accuracy, the memes will be classified.
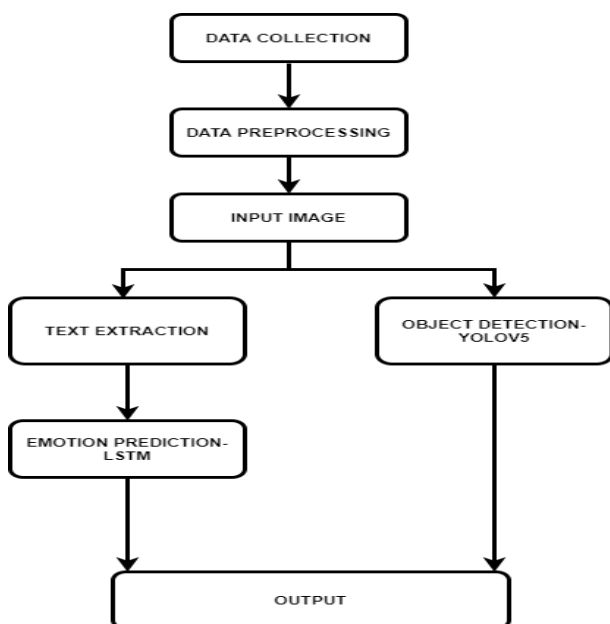


**Fig .1 Architecture Diagram**

### A. Collection of meme datasets:

The data used for the implementation of this paper is the meme dataset which consists of text, objects like mask, vaccine are considered as an input.



a. Example 1



b. Example 2

**Fig 2: Examples of COVID-19 memes dataset.**

### B. Text Extraction:

Optical Character Recognition is a tool used for extracting the text from memes. Since most OCR algorithms recognize text from black and white background, so it is challenging to detect text from coloured background. EasyOCR is a python package that allows the image to be converted to text.EasyOCR works in all background. It is an open-source tool and ready-to-use OCR which supports more than 70 languages. The memes and images related to covid-29 have been collected and text has been extracted using EasyOCR. After extracting the text, we classify the text into corresponding emotion with the help of LSTM where LSTM stands for Long Short-Term Memory. It is an RNN architecture used in the field of deep learning. The dataset has been trained by LSTM algorithm, which is mainly used for classifying ,processing and making predictions.

### C. Pre-Processing:

The raw image from the dataset is cleaned and normalized for further processing. The object in the image is boxed and the unwanted objects in the image will be removed for image processing, and for text classification, The HTML entities in the memes are decoded, URLs were removed, expressions corresponding to the tweet at the beginning of the

tweet are removed, Contractions present in the tweet are replaced by their extended words, punctuations present including hash-tag are removed. Extra spaces in the tweets are removed and all the characters of the tweets were changed to lowercase . Additionally , all non-ASCII-English character were removed, to keep the domain of the data specific to the English language.

### D. Object detection:

Deep learning, a subfield of Artificial Intelligence (AI), is a popular technique that has been applied in various domains including biology, computer vision, speech recognition and others. The process of object detection from an image is classified into four steps, Data collection and labelling, training, testing, Detection. The memes and images related to Covid-19 have been collected and the annotations are done manually using labelImg tool, the bounding boxes and coordinates will be trained and the weights of the trained dataset. The dataset has been trained by YOLOV5 algorithm, which is mainly used to detect real time object detection from videos, images or webcam. The weights generated while training the dataset have been saved and used for future work..

### E. Sentimental Analysis:

Social networks are a rich platform to learn about people's opinion and sentiment regarding different concepts as they can communicate and by share their opinion actively on social medias including Facebook and Twitter. Sentiment analysis is extremely important because it helps businesses quickly understand the overall opinions of their customers. By automatically sorting the sentiments behind the reviews, social media conversations, and other, can make faster and more accurate decisions. Sentimental analysis aims to detect emotions, like happiness, frustration, anger, sadness, and so on. Many emotion detection systems use lexicons methodology (i.e., lists of words and the emotions they convey) or complex machine learning algorithms.
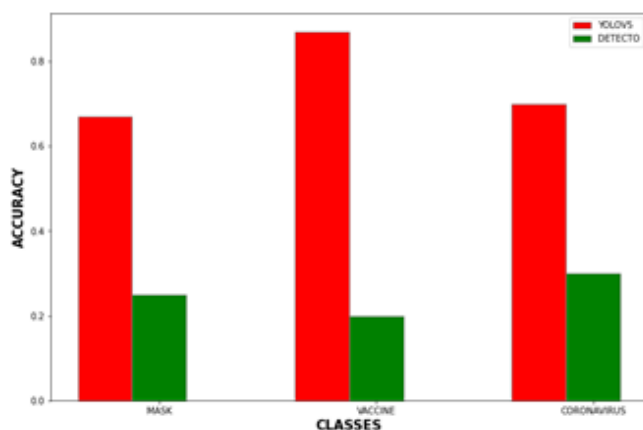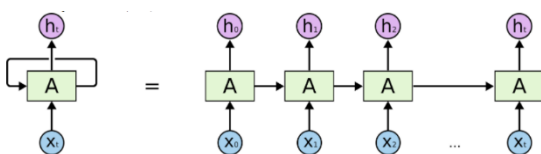


**Fig 3. LSTM Architecture**

### LSTM is a recurrent neural network (RNN)

Architecture that remembers values over arbitrary intervals. LSTM is well- suited to classification , processing and to predict the time series. One of the advantages with LSTM is insensitivity to gap length. RNN and HMM rely on the hidden state before emission/sequence. RNN cell takes two inputs and output from the last hidden state and observation at time =t. Besides the hidden state, there is no information about the past. The long-term memory is usually called the cell state. The looping arrows indicate recursive nature of the cell. This allows the information from previous intervals to be stored with LSTM. Cell state is modified by forgetting the gates placed below the cell state and by adjust the input modulation gate. The remember vector is usually called the forget gate. The output of the forget gate tells the cell state in which the information to forget by multiplying to a position in s the matrix. If the output of the forget gate is 1, then the information is in the cell state.

LSTM model is trained with the help of training datasets where it contains seven emotions and its respective sentences. Word Embedding is represented by the text where words that have the similar meaning have a similar representation. Glove stands for global vectors for word representation. we used glove.6b.100d here as glove file. Deep network takes the sequence of embedding vectors as an input and converts into a compressed representation. The compressed representation effectively captures all the information about the sequence of words in the text. The deep network part is an RNN or some forms of LSTM. The dropout added to overcome tendency, a very common problem with RNN based networks. The fully connected layer takes the deep representation from the RNN/LSTM and transforms it into the final output classes or class scores. This component is comprised of fully connected layers along with batch normalization and optionally dropout layers for regularization. The output layer has both ReLu and Softmax activation functions for multiclassification output.

### F. Training and Testing of dataset:

The proposed system is used to identify the emotions of covid-19 memes using multimodal approach. Our goal is to train the model with the training dataset and finally evaluate model's performance with the test dataset. The dataset for image consists of images which are related to covid-19, and dataset for text contains sentences which are extracted from the covid-19 memes. The training and testing sets are splitted up for image dataset. Finally with the help of accuracy, the memes will be classified .

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

The environment is set up and the images are loaded into the model for training and testing and it will evaluate the performance, accuracy ,precision of the model. More than 1000 images are used for training and 20% of images for testing. The image is fed into the separate module, i.e., for text extraction and object detection. The result will be predicted by the accuracy. Classification of the meme images belong to which class using multimodal sentimental analysis.

**Fig 4.Input Image**



**Fig 5.Object detection**

The comparison of detecto method and Yolov5 method with accuracy. In detecto method the accuracy is low with that of yolov5 method. For the class labels mask, virus and vaccine the accuracy is compared here.
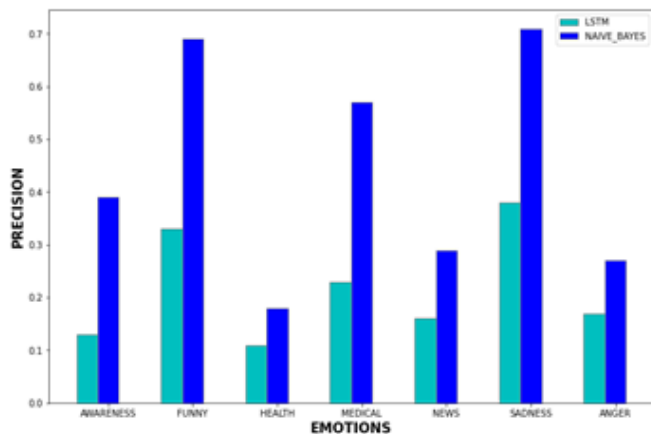


**Fig 6. Comparison of Detecto and Yolov5 method.**

| Model | Precision | Recall | F1-Score | Support |
|-------|-----------|--------|----------|---------|
| NAIVE BAYES | 0.47 | 0.37 | 0.38 | 13 |
| LSTM | 0.21 | 0.15 | 0.16 | 9 |

**Table 1:- Comparison of LSTM With Naive Bayes.**

For LSTM and Naïve Bayes comparison we have taken precision as a main parameter and graph is drawn based on it. From that we have inferred Naïve Bayes is better suitable for text sentimental analysis than LSTM technique.

## V.    CONCLUSION

The proposed model uses the YOLOV5 and LSTM technique for processing the image and text classification. Lstm is well suited for classification for text with different layers and activation functions like ReLU, SoftMax.YOLOV5 model is used to detect the object and classify in one at a time. From the results it can be said that the proposed model is able to detect the emotion of the memes to which class it belonged to. It collects the image from the social media like Instagram, tweeter, process the image, identifies the object and emotion, and finds the output of the image. Deep learning technique provides better result in image processing than in text analysis.

## FUTURE WORK

In the proposed system, COVID-19 meme analysis has been done using both images and text, but for further development of this project includes emotion prediction from videos and to increase the accuracy, dataset will be expanded.

## REFERENCES

[1]. Kumar, Akshi, and Geetanjali Garg. "Sentiment analysis of multimodal twitter data." Multimedia Tools and Applications 78.17 (2019): 24103-24119.

[2]. Amalia, Amalia, et al. "Meme Opinion Categorization by Using Optical Character Recognition (OCR) and Naïve Bayes Algorithm." 2018 Third International Conference on Informatics and Computing (ICIC). IEEE, 2018.

[3]. Jianqiang, Zhao, Gui Xiaolin, and Zhang Xuejun. "Deep convolution neural networks for twitter sentiment analysis." IEEE Access 6 (2018): 23253-23260.

[4]. Choudhary, Savita, Nikhil Kumar Singh, and Sanjay Chichadwani. "Text Detection and Recognition from Scene Images using MSER and CNN." 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC). IEEE, 2018.

[5]. Sumit, Shahriar Shakir, et al. "In object detection deep learning methods, YOLO shows supremum to Mask R-CNN." *Journal of Physics: Conference Series*. Vol. 1529. No. 4. IOP Publishing, 2020.

[6]. Elnagar, Ashraf, Ridhwan Al-Debsi, and Omar Einea. "Arabic text classification using deep learning models." *Information Processing & Management* 57.1 (2020): 102121.

[7]. Ma, Wenchi, et al. "MDFN: Multi-scale deep feature learning network for object detection." *Pattern Recognition* 100 (2020): 107149.

[8]. Nemes, László, and Attila Kiss. "Social media sentiment analysis based on COVID-19." Journal of Information and Telecommunication (2020): 1-15.

[9]. Sailunaz, Kashfia, and Reda Alhajj. "Emotion and sentiment analysis from Twitter text." Journal of Computational Science 36 (2019): 101003.