ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/109-116
Singavarapu Lakshmi Nvs Sai Rajesh *et. al.*, / International Journal of Engineering & Science Research

A Deep Learning-Based Approach For Inappropriate Content Detection And Classification Of Youtube Videos

Singavarapu Lakshmi Nvs Sai Rajesh, K.Sri Devi
Pg Scholar, Department Of Mca, Dnr College, Bhimavaram, Andhra Pradesh.
(Assistant Professor), Master Of Computer Applications, Dnr College, Bhimavaram, Andhra Pradesh.

Abstract

The bulk of YouTube's viewers are young, and the site has attracted billions of them as a result of the site's videos' exponential growth. Now-a-days YouTube content are access by all age group of peoples as this provide digital entertainment content on various topics such as Sports, Religious, Movies and many more. This channel provides vast amount cartoon entertainment for KIDS. Some malicious users are taking advantage of this digital content to spread inappropriate content for kids the in the form of cartoons. Such inappropriate content may put bad influence on growing kids and need a technique to prevent such content before showing to kids. Additionally, malicious up-loaders use this network as a distribution channel for disturbing visual information, such as sharing inappropriate material with kids through animated cartoon videos. This paper proposes a unique deep learning-based architecture for identifying and categorising objectionable information in videos. . Overall, cutting-edge performance was obtained by the Efficient-Net and BiLSTM design with 128 hidden units (f1 score = 0.9267).

Keywords: EfficientNet, CNN, bidirectional LSTM, video classification, social media analysis, deep learning.

I. INTRODUCTION

Over the past few years, social media platforms have seen a significant increase in both the production and consumption of films.

YouTube is the most popular social media website for sharing videos, with a vast selection of content from many genres. More than 2 billion people are registered users of YouTube worldwide, and more than 500 hours of video content are posted every minute, according to YouTube statistics [1].

As a result, billions of hours of video are available for users of all ages to discover both generic and personalised content [2]. Given the size of the crowd sourced database, it is challenging to keep track of and control the contributed content by platform policies. This gives nasty individuals a chance to spam by deceiving audiences with falsely marketed content.

Malicious users' most annoying behaviour is exposing young audiences to upsetting content, especially when it is made to appear as though it is appropriate for them.

Younger audiences prefer this social media site over audiences of other ages, and the reason for this high level of approval is that there are fewer restrictions [6,] according to a press release from YouTube [5]. Due to a lack of rules, children can be exposed to any content online, unlike on television.

One of the risks to children's internet safety, along with cyber hate, cyber predators, cyber bullying, etc., is exposure to upsetting information. [7]. numerous publications [9]–[12] noted the widespread practice of disseminating objectionable material in kid-friendly videos.

The Elsagate controversy, which involved videos of well-known childhood cartoon characters depicted in disturbing situations, such as engaging in drinking alcohol, stealing, mild violence, or

engaging in nudity or sexual activities, caught the attention of the general public.

II. LITERATURE SURVEY

YouTube has one of the most sophisticated and complete industrial recommendation systems in use. In this paper, we provide a high-level overview of the system and focus on the profound performance gains from deep learning. The two components of the work are separated into the conventional two information retrieval steps. Before describing an alternative deep ranking methodology, we discuss a deep candidate generation model. We also provide valuable advice and insight gleaned from designing, enhancing, and running a large-scale recommendation system with significant user-facing implications. [1]

These days, young kids routinely visit the well-known digital media site YouTube. However, concerns about inappropriate video content and poor quality have been raised. Currently, no study or theoretical argument can be used to determine the quality of YouTube videos made for youngsters. We conducted a literature search for this study. We developed a set of design criteria to evaluate the level of YouTube videos produced for young children aged 0 to 8. The four critical evaluation criteria are age appropriateness, content quality, design components, and learning objectives. Educators can assess the effectiveness of early learning videos using this evaluation tool, and YouTube creators can use it as guidance when producing educational films for young children. [2]

In terms of popularity, YouTube has grown to become a worldwide phenomenon. Kids use YouTube for pleasure and education. According to several reports, YouTube has many advantages, including fostering children's learning, skill development, and world understanding. On the other hand, children's exposure to inappropriate and upsetting YouTube content, screen time, and poor video quality has generated concerns. This article describes YouTube, examines recent YouTube research, and investigates how young kids use YouTube. It also offers suggestions and tactics for using YouTube to promote young children's early learning and development at home and school for parents, teachers, and other carers. **[3]**

To assess if the theories put up to explain the impacts are in line with the results of the cumulative research on aggressive behaviour and media violence. We looked into the short- and long-term effects of violent behaviour. The theory-based premise that adults should have more significant short-term impacts and children should experience more robust long-term consequences was also examined. The results also showed that exposure to media violence generally had a small but significant impact on helpful conduct, arousal levels, angry feelings, violent thoughts, and aggressive behaviour. The findings are consistent with the concept that short-term outcomes are caused mainly by priming well-established beliefs, schemas, or scripts, which adults have had more time to encode. **[4]**

With the rapid development of Internet technology, the prevalence of graphic films has increased and negatively impacted society. This work built a database of gory films using web crawlers and data augmentation methods without open bloody video material. After that, it extracted the spatiotemporal characteristics of the visual channels using CNN and LSTM techniques. [5]

The proliferation of graphic videos has gotten worse with the quick advancement of Internet technology and has had a significant negative impact on society. This work built a database of gory films using web crawlers and data augmentation methods without open bloody video material. After that, it extracted the spatiotemporal

characteristics of the visual channels using CNN and LSTM techniques. In this paper, many designs of hidden layers and nodes in DNN have been built using the try-error methodology and equation-based method to examine the effect of the number of hidden layers and nodes on the classification performance. According to the findings, a try-and-error strategy has a 53% accuracy rate, while an equation-based system has a 51% accuracy rate. [6]

III. PROPOSED METHOD

A brand-new deep learning-based architecture is suggested to identify and categorise objectionable film information. An attention mechanism is also included after using the attention probability distribution on the web and following the BiLSTM. Two algorithms are trained on YouTube video annotated images such as Attention based BI-LSTM and EfficientNetB7 based BI-LSTM and in both algorithms EfficientNetB7 is giving better accuracy. On same dataset we have experiment with existing SVM algorithm but its accuracy is less than propose EfficientNetB7-BILSTM algorithm. In propose paper author has used vulgar videos like nudity and sex so we cannot used such dataset to test above algorithms so we have used 'Normal and Violence (fight)' type of dataset to train above algorithms. For kids violence and fighting videos are also consider as Inappropriate Content.

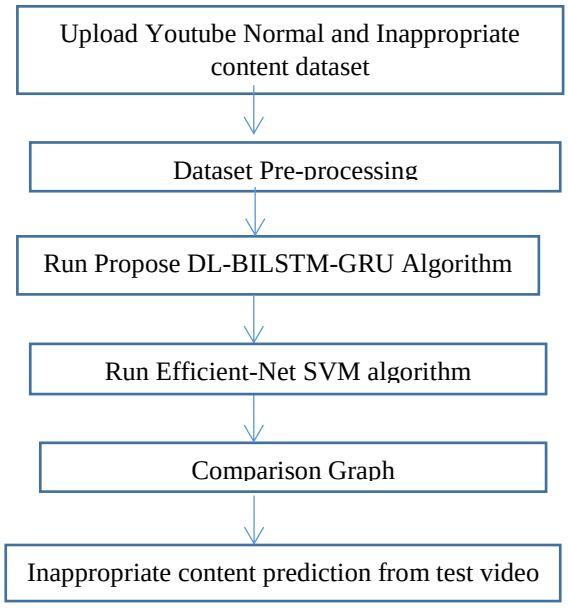


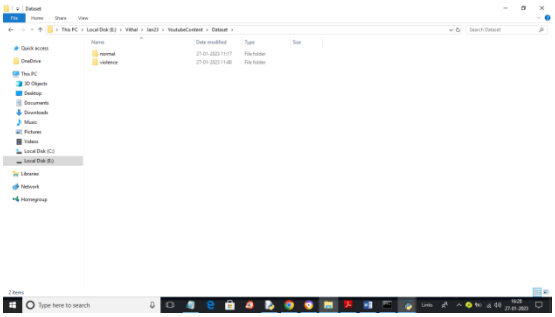
Fig. Flowchart for proposed method

IV. RESULT

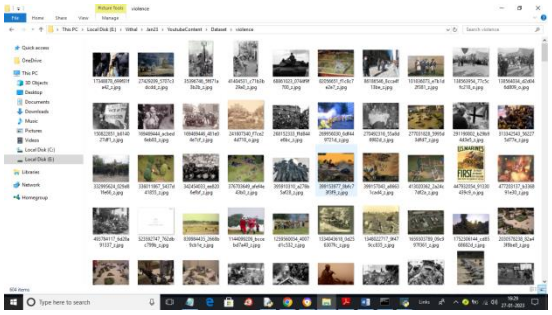
To prevent such content many machine learning and deep learning algorithms are introduced but their inappropriate content detection and classification accuracy is not up to the mark. To overcome from this issue author of this paper employing combination of CNN and BI-LSTM algorithm to detect inappropriate content.

Pre-trained EfficientNetB7 CNN algorithm is employed to extract features from the YouTube video images and then retrained with BI-LSTM algorithm to enhance prediction accuracy.

To train all algorithms we have used below YouTube images consist of two folders called 'Normal and Violence' and below screen showing dataset details



In above screen we have two folders and juts go inside any folder to view training images

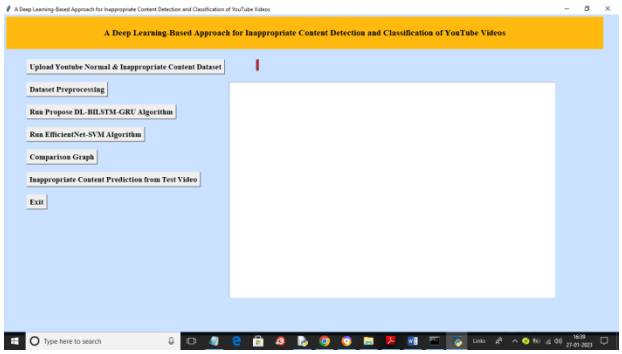


In above screen we can see violence images used to train algorithms

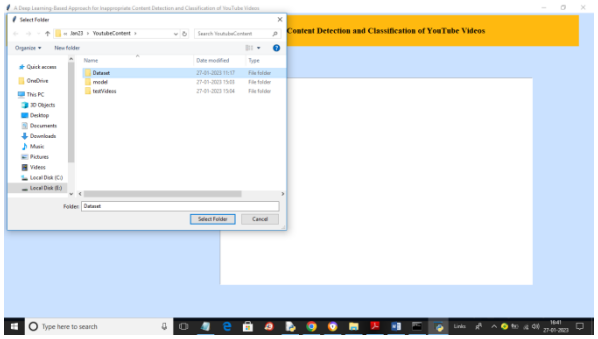
To implement this project we have designed following modules

1) Upload YouTube Normal & Inappropriate Content Dataset: using this module we will upload YouTube dataset images to application
2) Dataset Pre-processing: With the help of this module, we'll input every image, scale them to be the same size, normalise their pixel values, and finally shuffle the dataset.
3) Run Propose DL-BILSTM-GRU Algorithm: using this module we will split dataset into train and test and then input 80% training data to Pre-Trained CNN (EfficientNetB7) algorithm to extract digital content from images and then those features will get retrained with BI-LSTM algorithm to train a model. Trained model will be applied on 20% test data to calculate prediction accuracy
4) Run EfficientNet-SVM Algorithm: EfficientNetB7 features will get retrained with existing SVM algorithm and then calculate prediction accuracy
5) Comparison Graph: using this module we will plot accuracy comparison graph between propose EfficientNetB7-BILSTM and EfficientNetB7-SVM.
6) Inappropriate Content Prediction from Test Video: using this module we will upload any YouTube and if video contains fighting or violence then application will predict as 'Inappropriate Content' otherwise will predict SAFE content.

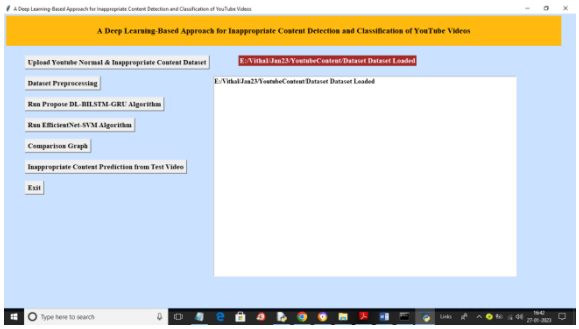
To run project double click on 'run.bat' file to get below screen



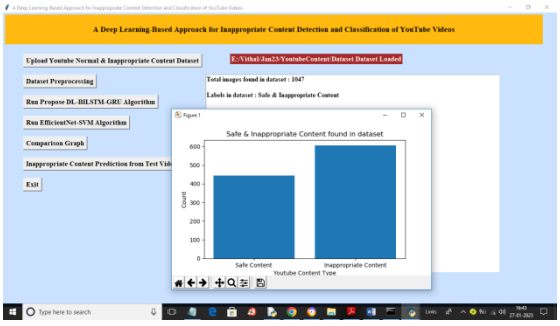
In above screen click on 'Upload YouTube Normal & Inappropriate Content Dataset' button to upload dataset and get below output



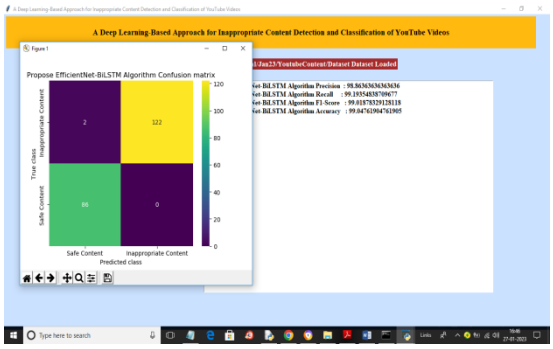
Selecting the complete "Dataset" folder in the screen above, uploading it, and then clicking "Select Folder" to load the dataset will produce the output below.



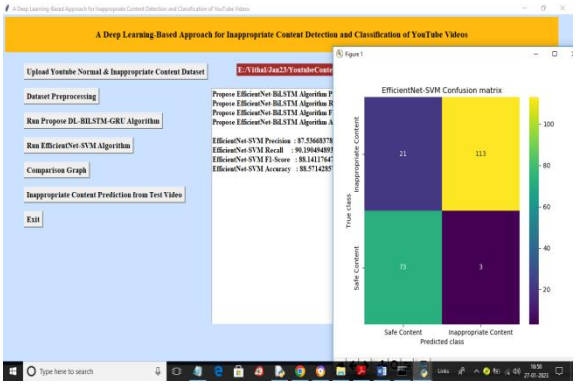
In above screen dataset loaded and now click on 'Dataset Preprocessing' button to read all images and then processes those images for training and get below output



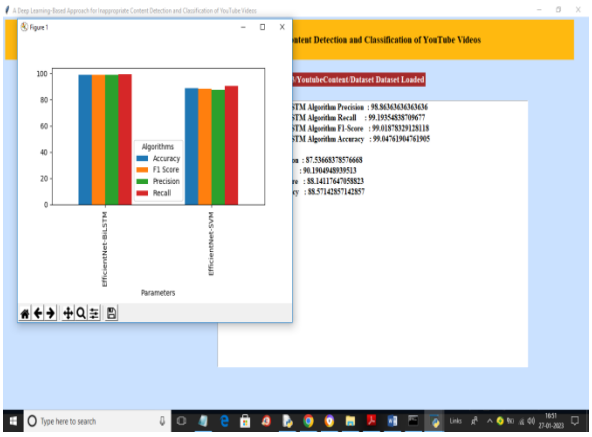
As seen in the screen above, the dataset contains 1047 photographs. The x-axis of the graph shows the type of images, such as "Safe and Inappropriate," while the y-axis indicates the number of such images. Now that the dataset has been processed, click "Run Propose DL-BILSTM-GRU Algorithm" to train the proposed method and obtain the output below.



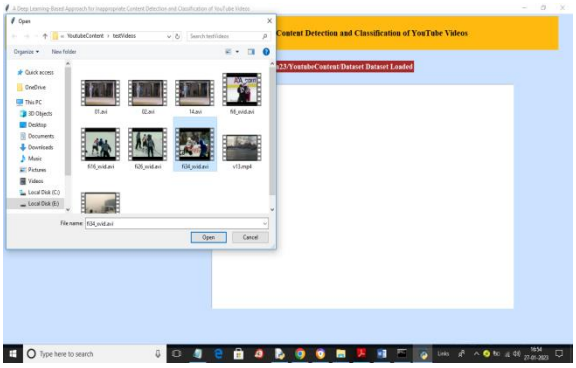
In the image above, we achieved 99.04% accuracy using the proposed EfficientNetB7-BI-LSTM. In the confusion matrix graph, the x-axis represents predicted labels and the y-axis represents true labels. Green and yellow boxes contain the correct prediction count, while blue boxes contain the incorrect prediction count, which is only 2. Close the previous graph now, and then click the "Run Efficient Net-SVM Algorithm" button to obtain the output shown below.



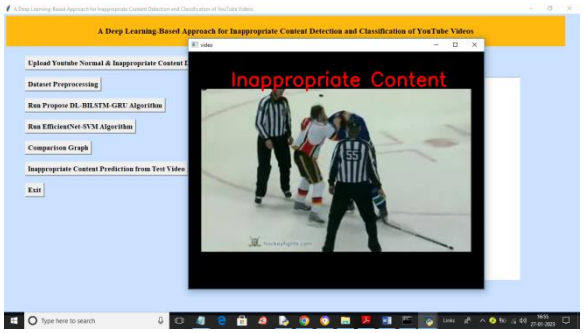
With EfficientNetB7-SVM, we achieved 88% accuracy in the image above, and in the confusion matrix graph, we can see in the blue boxes that SVM predicted a total of 24 wrong predictions, indicating that its accuracy is lower. Close the previous graph now, and then click "Comparison Graph" to obtain the output shown below.



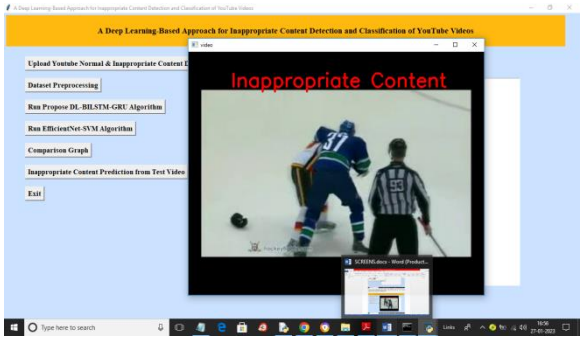
In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. In both algorithms propose EfficientNetB7-BI-LSTM got high accuracy. Now close above graph and then click on ‘Inappropriate Content Prediction from Test Video’ button to upload test video and classify it as Safe or inappropriate.



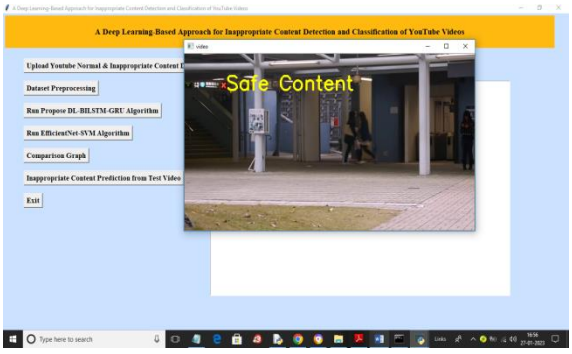
In above screen selecting and uploading video and then click on “Open’ button to play video and perform classification



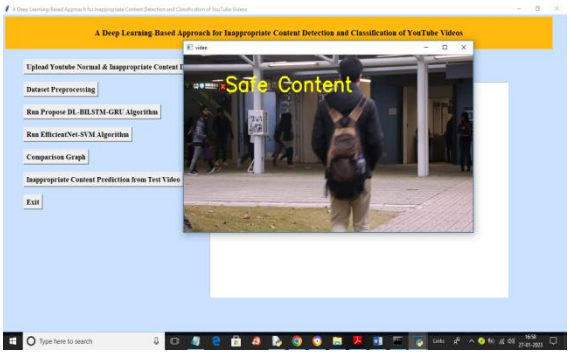
In above screen propose algorithm evaluating playing video and then detecting and classifying it as ‘Inappropriate Content’



In above video also we got classification output



In above we got result as Safe Content



In above screen we got output as Safe Content as peoples are only moving in the video.

Similarly you can upload and test other videos also

V. CONCLUSION

While the model learns effective video representations, the BiLSTM network evaluates the recovered video features and performs multiclass video classification. A dataset of 111,156 carefully annotated cartoon video clips downloaded from YouTube is used in all evaluation experiments. By outperforming previously used models and methods and achieving a maximum recall score of 92.22%, our BiLSTM-based system outperformed previously used models and methodologies. Our approach is independent of the metadata associated with YouTube videos, which malicious uploaders can alter to deceive consumers while looking for inappropriate children's content on the site.

REFERENCES

1. P. Covington, J. Adams, and E. Sargin, ‘‘Deep neural networks for YouTube recommendations,’’ in Proc. 10th ACM Conf. Recommender Syst., Sep. 2016, pp. 191–198, doi: 10.1145/2959100.2959190.

2. M. M. Neumann and C. Herodotou, ‘‘Evaluating YouTube videos for young children,’’ Educ. Inf. Technol., vol. 25, no. 5, pp. 4459–4475, Sep. 2020, doi: 10.1007/s10639-020-10183-7.

3. J. Marsh, L. Law, J. Lahmar, D. Yamada-Rice, B. Parry, and F. Scott, Social Media, Television and Children. Sheffield, U.K.: Univ. Sheffield, 2019. [Online]. Available: https://www.stac-study.org/downloads/ STAC_Full_Report.pdf

4. M. M. Neumann and C. Herodotou, ‘‘Young children and YouTube: A global phenomenon,’’ Childhood Educ., vol. 96, no. 4, pp. 72–77, Jul. 2020, doi: 10.1080/00094056.2020.1796459.

5.C. Hou, X. Wu, and G. Wang, ‘‘End-to-end bloody video recognition by audio-visual feature fusion,’’ in Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV), 2018, pp. 501–510, doi: 10.1007/978-3-030-03398- 9_43.

6. A. Ali and N. Senan, ‘‘Violence video classification performance using deep neural networks,’’ in Proc. Int. Conf. Soft Comput. Data Mining, 2018, pp. 225–233, doi: 10.1007/978-3-319-72550-5_22