

# **A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos**

## **ABSTRACT:**

The exponential growth of videos on YouTube has attracted billions of viewers among which the majority belongs to a young demographic. Malicious uploaders also find this platform as an opportunity to spread upsetting visual content, such as using animated cartoon videos to share inappropriate content with children. Therefore, an automatic real-time video content filtering mechanism is highly suggested to be integrated into social media platforms. In this study, a novel deep learning-based architecture is proposed for the detection and classification of inappropriate content in videos. For this, the proposed framework employs an ImageNet pre-trained convolutional neural network (CNN) model known as EfficientNet-B7 to extract video descriptors, which are then fed to bidirectional long short-term memory (BiLSTM) network to learn effective video representations and perform multiclass video classification. An attention mechanism is also integrated after BiLSTM to apply attention probability distribution in the network. These models are evaluated on a manually annotated dataset of 111,156 cartoon clips collected from YouTube videos. Experimental results demonstrated that EfficientNet-BiLSTM (accuracy 95.66%) performs better than attention mechanism based EfficientNet-BiLSTM (accuracy 95.30%) framework. Secondly, the traditional machine learning classifiers perform relatively poor than deep learning classifiers. Overall, the architecture of EfficientNet and BiLSTM with 128 hidden units yielded state-of-the-art performance (f1 score 0.9267). Furthermore, the performance comparison against existing state-of-the-art approaches verified that BiLSTM on top of CNN captures better contextual information of video descriptors in network architecture, and hence achieved better results in child inappropriate video content detection and classification.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

## EXISTING SYSTEM:

Rea *et al.* [37] proposed a periodicity-based audio feature extraction method which was later combined with visual features for illicit content detection in videos.

The machine learning algorithms are usually employed as classifiers Liu *et al.* [38] classified the periodicity-based audio and visual segmentation features through support vector machine (SVM) algorithm with Gaussian radial basis function (RBF) kernel. Later on, they extended the framework [39] by applying the energy envelope (EE) and bag-of-words (BoW)-based audio representations and visual features.

Ulges *et al.* [23] used MPEG motion vectors and Mel-frequency cepstral coefficient (MFCC) audio features with skin color and visual words. Each feature representation is processed through an individual SVM classifier and combined in a weighted sum of late fusion. Ochoa *et al.* [40] performed binary video genre classification for adult content detection by processing the spatiotemporal features with two types of SVM algorithms: sequential minimal optimization (SMO) and LibSVM.

Jung *et al.* [41] worked with the one dimensional signal of spatiotemporal motion trajectory and skin color. Tang *et al.* [42] proposed a pornography detection system\_PornProbe, based on a hierarchical latent Dirichlet allocation (LDA) and SVM algorithm. This system combined an unsupervised clustering in LDA and supervised learning in SVM, and achieved high efficiency than a single SVM classifier. Lee *et al.* [43] presented a multilevel hierarchical framework by taking the multiple features of different temporal domains. Lopes *et al.* [44] worked with the bag-of-visual features (BoVF) for obscenity detection.

Kaushal *et al.* [21] performed supervised learning to identify the child unsafe content and content uploaders by feeding the machine learning classifiers (i.e., random forest, K-nearest neighbor, and decision tree) with video-level, user-level and comment-level metadata of YouTube Reddy *et al.* [45] handled the explicit content problem of videos through text classification of YouTube comments. They applied bigram collocation and fed the features to the naïve Bayes classifier for final classification.

## DISADVANTAGES:

- An existing system doesn't ANALYSIS OF PRE-TRAINED CNN MODEL VARIANTS.
- An existing system doesn't ANALYSIS OF EFFICIENT-NET FEATURES WITH DIFFERENT CLASSIFIER VARIANTS.

## **PROPOSED SYSTEM:**

1. The system proposes a novel CNN (EfficientNet-B7) and BiLSTM-based deep learning framework for inappropriate video content detection and classification.
2. The system presents a manually annotated ground truth video dataset of 1860 minutes (111,561 seconds) of cartoon videos for young children (under the age of 13). All videos are collected from YouTube using famous cartoon names as search keywords. Each video clip is annotated for either safe or unsafe class. For the unsafe category, fantasy violence and sexual-nudity explicit content are monitored in videos. We also intend to make this dataset publicly available for the research community.
3. The system evaluates the performance of our proposed CNN-BiLSTM framework. Our multiclass video classifier achieved the validation accuracy of 95.66%. Several other state-of-the-art machine learning and deep learning architectures are also evaluated and compared for the task of inappropriate video content detection.

## **ADVANTAGES:**

- The most frequent applications of image/video classification employed the convolutional neural networks.
- The EfficientNet model is a convolutional neural network model and scaling method that uniformly scales network depth, width and resolution through compound co efficient.

## **System Requirements:**

### **Hardware system configuration:**

- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

## Software Requirements:

- Operating system : Windows 7 Ultimate
- Coding Language : Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : Html, css, javascript.
- Data Base : MySQL (WAMP Server).

## Project Team:

ROSHINI SINGH	20W91A05J9
S.SANJANA	20W91A05K5
SIMRAN SINGH	20W91A05L3
SUBHADIP MUDI	20W91A05L8

Internal Guide Signature

Dr.MD ASHFAKUL HASAN

