Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 8
August 2024

Preserving Copyright Integrity: A Comprehensive Approach to Combat YouTube Video Infringement and Privacy-Sensitive Video Analysis

M. Thanga Subha Devi [1,2,*] ,Dr. R. Suji Pramila [3]

[1]Research Scholar (PT), Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu 629180, India.

[2]Assistant Professor, Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India.

[3]Associate Professor, Computer Science and Engineering, Mar Baselios Institute of Technology and Science Kothamangalam, Kerala 686693, India.

Abstract

Addressing the pressing issue of YouTube video infringement, which has become increasingly prevalent in the digital age, this study aims to uncover its underlying causes and explore potential solutions. Through a comprehensive analysis, explore the challenges posed by copyright violations on YouTube and propose strategies to detect and combat infringement effectively. By leveraging advanced techniques in multimedia forensics and legal frameworks, our approach aims to protect the rights of content creators and uphold copyright integrity in online video content. Propose a novel approach for video representation and classification, focusing on privacy-preserving techniques and robust performance. Our methodology begins with pre-processing using Gabor filters and key frame selection to enhance feature extraction. Feature extraction is performed using Petri Net and Horse Herd optimization (HHO) algorithms to capture intricate patterns in the data. For video representation and classification, employ a BiLSTM network with an attention mechanism, followed by a softmax classifier for accurate classification. This comprehensive framework offers a robust solution for privacy-sensitive video analysis across various domains. The proposed method leveraging Python tools achieves a remarkable accuracy of 99% when applied to a dataset extracted from YouTube videos. This study underscores the efficacy of the approach in effectively analyzing and classifying video content, offering promising prospects for various applications in video analysis and understanding.

Keywords: copyright integrity; petrinet; BiLSTM; horse herd optimization; softmax

1. Introduction

Large-scale internet video sharing services like YouTube and Twitch have made it possible for regular people to broadcast and watch real-time footage to a worldwide audience [1]. The ability to broadcast live videos to an international audience on online social media platforms like Twitch and YouTube is a relatively new phenomenon. Although streams are created and consumed in real-time, these video sharing services are essentially distinct from "static" media distribution systems like Netflix and Hulu [2]. Internet surfers may violate the copyright of the original work's proprietors when they mix, remix, mash-up, or parody preexisting cultural resources. The accessible nature of such user-generated material on Internet platforms is determined by the interaction of two sets of legal rules [3]. But with the increased awareness of intellectual property in society in recent years, altering videos and capturing screenshots is equally risky when it comes to infringement [4].

Many aspects of our daily lives have been altered as a result of the new digital environment. Along with fresh possibilities, computers and the internet also brought unforeseen challenges and responsibilities [5]. The amount of data shared on numerous websites and platforms for exchanging information has increased exponentially. But on a network as large as the worldwide web, this data which is available in a variety of formats, including text, audio, video, and image is frequently exposed to theft and alteration [6] [7] [8] [9]. With the quick advancement of information and communication technology, may now communicate with each other far removed from places that are inaccessible due to huge distances by using electronic media [10].

In Taipei City, the synchronous video-observed therapy (SVOT) program that permitted reciprocal video conversations was launched in 2014 [11]. Educational movies are valuable teaching resources that are frequently utilized by instructors in flipped or student-focused classrooms as well as by students to close knowledge gaps [12]. Studies on social cybersecurity has shown how propaganda efforts utilize social media's multi-platform usage, particularly YouTube [13]. Many video producers post their videos online due to the widespread use of social media. But a lot of the video clips found in these uploaded films are really clips from other longer video that have been edited in numerous manners.

Key contribution of the proposed work is,

- Conduct a thorough examination of the prevalent issue of YouTube video infringement, identifying the challenges posed by copyright violations in the digital era.
- Through meticulous analysis, propose effective strategies for detecting and combatting infringement on YouTube. These strategies leverage advanced techniques in multimedia forensics and legal frameworks to protect the rights of content creators and uphold copyright integrity in online video content.
- Introduce a pioneering methodology for video representation and classification, prioritizing privacy-preserving techniques and robust performance.
- Our approach begins with pre-processing using Gabor filters and key frame selection. This step enhances feature extraction by identifying key frames and extracting relevant visual information using Gabor filters.
- Employ Petri Net and Horse Herd optimization algorithms for feature extraction, capturing intricate patterns in the data. These techniques enable comprehensive representation of video content, facilitating accurate classification.
- For video representation and classification, utilize a BiLSTM network with an attention mechanism, followed by a softmax classifier. This architecture ensures accurate classification while preserving privacy-sensitive information.

The article is arranged in the following manner: Section 2 investigates previous studies on prediction problems various optimization approaches are used. Section 3 discussed the problematic statement. Section 4 discusses about proposed method. Section 5 experimental evaluation comprises mathematically developed system models. The paper is concluded in Section 6.

2. Related Works

Mandelli et al. [14] proposed a method to address the device identification problem by matching a stabilised video pattern with a specified fingerprint; a method to extract a device's distinctive fingerprint from either a group of images or steady video footage. The suggested approach is evaluated using movies from a variety of devices that were extracted from the current, publicly accessible Vision Dataset. The carried out tests also offer an intriguing perspective on how contemporary smartphone image stabilization algorithms affect certain video frames.

Mareen et al. [15] proposed a unique method of video watermarking that modifies only one encoder choice manually. The watermark's shape is then automatically represented by a vast set of implicit distortions that are a result of the explicit modification. Hidden distortions are invisible because they mimic common compression artifacts produced by encoders. They also show themselves to be resistant to video tampering. Moreover, the suggested plan doesn't call for changing any of the current consumer electronics. As a result, the suggested watermarking strategy may be used to aid in the fight against piracy without upsetting innocent users with strange irregularities.

Zheng et al. [16] Proposed the technique of extracting video keyframes using the image association coefficient. Additionally, integrate the SIFT method, QR breakdown, and Contourlet Transform domain to strengthen the watermark's resistance against geometric assaults based on the idea of obscurity. Next, in order to encrypt the strong watermark and bolster security, employ Arnold Transformation (Cat Map), which is based on the Maximum Entropy and Resistant Segment. Furthermore, effectively wrap up the integrity verification of the watermarked video and pinpoint the precise tamper position of the attacked video based on the properties of actual fragile watermarking. Furthermore, SM2 signs and uploads to the blockchain the hash digest of the

video watermark and the rightful owner's user ID. Once the personal authentication is successful, an individual can register for the copyright. They execute security evaluations and evaluations on the watermark system's protection, the system's blockchain effectiveness, and the business-grade cryptography algorithm's effectiveness.

Kumar et al. [17] proposed a blockchain-powered distributed peer-to-peer picture and video sharing network based on IPFS (InterPlanetary File System). They identify copyright breaches of multimedia using a perceptual hash (pHash) approach. The pHash of multimedia material that is to be uploaded to the IPFS is calculated and compared to the current pHash numbers in the blockchain network. If the multimedia were to match the current pHash values, it would be identified as tampered with. One benefit of blockchain technology is that it eliminates the need for a third party, which helps to prevent single points of failure.

Kong et al. [18] proposed a novel paradigm that uses the combined knowledge of the face and voice to show the real face concealed behind the phony one. More precisely, given the audio segment and the fake face, the cross-modality transferring capacity is utilized by learning to construct the authentic face feature dependent on the audio segments's underlying indications and the false face look. A series of evaluations are conducted to verify the effectiveness of the proposed scheme. The experimental results demonstrate that the proposed approach achieves a promising face reorganization performance in revealing the hidden faces, both in terms of identity and face characteristic inference accuracy and reconstructed images quality.

The existing research presents various innovative techniques for multimedia identification, watermarking, and authentication. However, there remains a notable research gap in addressing the comprehensive integration of multiple modalities for robust multimedia authentication and copyright protection. While individual studies focus on specific aspects such as device identification, watermarking techniques, blockchain-based copyright management, and face recognition, there is a lack of research that combines these diverse approaches into a unified framework. Specifically, there is a need for a holistic solution that leverages the synergies between device fingerprinting, watermark embedding, blockchain authentication, and multimodal fusion techniques to provide a more robust and comprehensive approach to multimedia content protection and copyright enforcement. Such a comprehensive framework would enhance the security and integrity of multimedia content across various platforms and mitigate the challenges posed by evolving piracy and tampering techniques.

3. Problem Statement

Detecting the capturing device of a video is crucial for tracing copyright infringement and combating distribution of illicit content. However, advancements in video stabilization hinder traditional sensor pattern noise-based techniques, making device attribution challenging. This paper addresses the problem of attributing stabilized videos to their recording devices by proposing methodologies for fingerprint extraction and matching, tested on a dataset of smartphone videos [14]. With the proliferation of digital content online, copyright infringement issues have become rampant due to challenges in confirming ownership and gathering evidence. Traditional copyright protection systems struggle with monitoring and providing robust evidence for rights protection. To address these challenges, present a novel video copyright protection scheme leveraging blockchain and double watermarking technologies, aiming to enhance copyright confirmation, infringement monitoring, and evidence gathering [16].

4. Proposed Method

In this paper, propose a comprehensive methodology to address the pressing issue of YouTube video infringement, aiming to protect content creators' rights and uphold copyright integrity while preserving privacy. Our approach combines advanced techniques in multimedia forensics with privacy-preserving methodologies to detect and combat infringement effectively. Begin with pre-processing using Gabor filters and key frame selection for feature enhancement and computational efficiency. Feature extraction employs Petri Net and Horse Herd optimization algorithms to capture intricate patterns, while video representation and classification utilize a BiLSTM network with an attention mechanism and softmax classifier for accurate classification.

Integrate differential privacy and federated learning techniques to ensure privacy preservation during analysis. Evaluation includes standard performance metrics and privacy analysis, while experimental validation involves dataset curation and comparative analysis. Finally, address legal and ethical considerations, ensuring compliance with regulations and mitigating potential biases and unintended consequences. Through this methodology, aim to develop a robust solution for combatting YouTube video infringement while safeguarding privacy and upholding copyright integrity. Figure 1 shows flow diagram of YouTube video infringement.

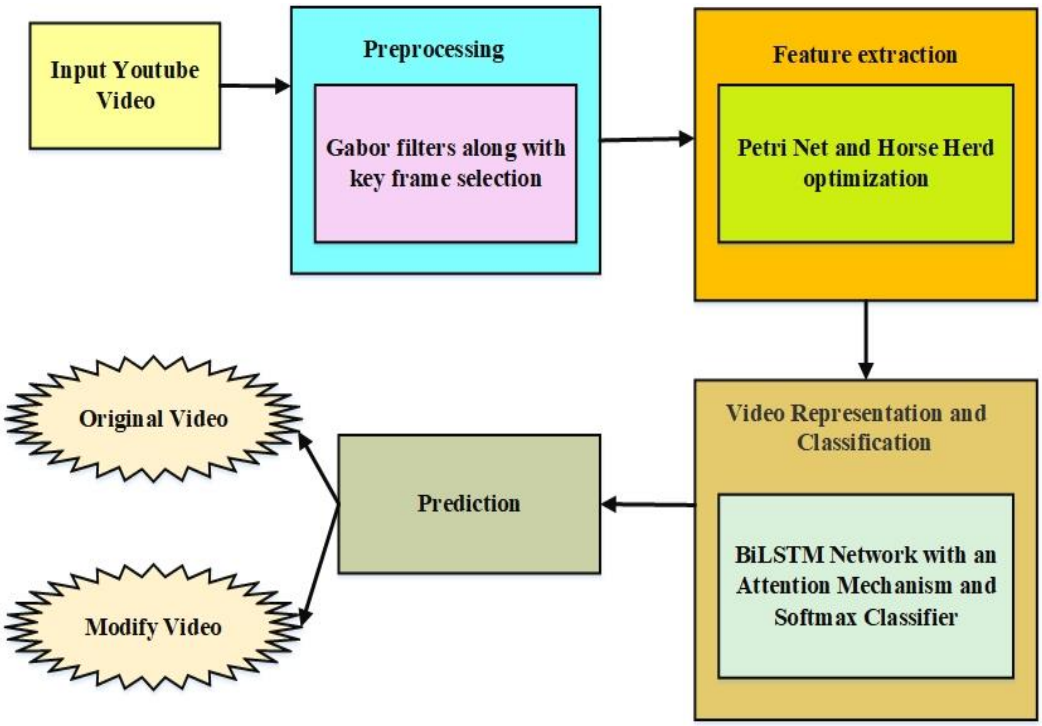


Figure 1: Flow Diagram of YouTube Video Infringement

4.1. Dataset Description

The dataset consists of real-time YouTube cartoons, comprising both modified and original videos. These cartoons span various genres, styles, and durations, sourced from publicly available YouTube channels. The modified videos encompass alterations such as color grading, scene rearrangement, and addition of effects, while the original videos remain unaltered. The dataset is divided into two subsets: an 80% training set and a 20% testing set. This split ensures a sufficient volume of data for model training while reserving a portion for evaluating model performance. Each video is accompanied by metadata, including title, upload date, and channel information, facilitating organization and categorization. Additionally, the dataset includes annotations indicating whether each video is original or modified, enabling supervised learning approaches for classification tasks. Overall, this dataset serves as a valuable resource for researchers and practitioners in multimedia analysis, enabling the development and evaluation of algorithms for detecting and classifying modified videos.

4.2. Pre-processing:

4.2.1. Utilizing Gabor Filters in Conjunction with a Key Frame Selection Scheme

4.2.1.1 Key frame selection

Determine each frame's pixel difference and designate it as a crucial frame if it exceeds a threshold of 10. Key frames from a film containing a ranges are represented as $f(a,b)$ in an input where a (a,b) spatial location is. First, the important frames are sent via gabor filtering, which eliminates duplicated and noisy data. Each frame gains more characteristics as a consequence of this preprocessing.

There are 'n' frames in the video F_i, which are denoted by $F_i = \{v_1, v_2, v_3 \ldots v_n\}$. Using key frame selection, the videos are framed at thirty frames per second on average. To make room for additional processing, the frames are thereafter reduced in size 250×250. The threshold approach is employed to identify key frames, whereby the initial frame of every frame is designated as a key frame.

4.2.1.2. Gabor filters response:

Gabor has shown to be a highly helpful technique in computer vision and image processing because of its features, which include perfect location in both the frequency and spatial domains. Gabor functions are harmonic oscillators made up of sinusoidal plane waves with a particular frequency and direction that are contained in a Gaussian envelope. The definition of the image domain (c, d) at a complicated 2-D gabor filter is

$$G(c,d) = \exp\left(-\frac{(c-c_0)^2}{2\sigma_c^2} - \frac{(d-d_0)^2}{2\sigma_d^2}\right) \times \exp(-2\pi i(e_0(c-c_0) + g_0(d-d_0))) \quad (1)$$

here (c_0, d_0) specifies the direction, the position of the image (e_0, g_0), the spatial frequency $\omega_0 = \sqrt{e_0 + g_0}$ that is used to space out the modulation $\theta_0 = \arctan(g_0 / e_0)$, σ_c and σ_d the Gaussian envelope along the c- and d-axes as determined by the standard deviation. By carefully choosing the aforementioned parameters, equation may be used to create a 2-D Gabor filer with an even-symmetric real component.

$$s(c,d;T,\phi) = \exp\left(-\frac{1}{2}\left[\frac{c_\phi^2}{\sigma_c^{2\phi}} + \frac{d_\phi^2}{\sigma_d^2}\right]\right)\cos\left(\frac{2\pi c_\phi}{T}\right) \quad (2)$$

$$c_\phi = c\cos\phi + d\sin\phi \quad (3)$$

$$d_\phi = -\sin\phi + d\cos\phi \quad (4)$$

there T indicates the sinusoidal plane wave of the period, and there ϕ indicates the gabor generated filter by orientation. By splitting formula into two transverse pieces, one perpendicular to orientation and the other parallel, the following formula may be inferred ϕ :

$$s(c,d,T,\phi) = h_c(c;T,\phi).h_d(d;\phi) \quad (5)$$

$$= \left\{\exp\left(-\frac{c_\phi^2}{2\sigma_c^2}\right)\cos\left(\frac{2\pi c_\phi}{T}\right)\right\}.\left\{\exp\left(-\frac{d_\phi^2}{2\sigma_d^2}\right)\right\} \quad (6)$$

The first portion h_c functions as a Gabor function 1-D band pass filter, while the second part h_d is a Gaussian function low pass filter. Consequently, a ϕ low pass filter is applied along ϕ its direction and a band pass filter is applied orthogonally to it via the 2-D even-symmetric Gabor filter. Low pass and band pass qualities along these two orthogonal orientations are greatly advantageous in boosting face emotion pictures, since ridges and valleys are often alternated oppositely to local parallel and orientation shows local alignment throughout an approximate smoothness.

4.3. Feature extraction using Petri Net

In this work, they provide Petri Nets (PNs) for activity detection using YouTube datasets for MLB, Vlog, Atomic Visual Actions, and TV interaction between people. Graph-based methods such as Petri Nets are used to describe and depict many kinds of conduct, such as resource sharing, concurrency, synchronization, and parallelism. A classical finite state machine is a Petri Net because every change is limited to having precisely just one

result and one input. A Petri Net, on the other hand, is a machine with finite states that permits multiple inputs and numerous outputs. Destinations are depicted as circles in a visual way, and transitions are represented as squares or rectangles. Place nodes and transition nodes are connected by arcs that are shown as input or output arcs, respectively. Arrow heads are used to draw regular loops. Dot head are used to illustrate exhibit arcs. Arcs have a multiplication, also known as a weight, attached to them. If arc multiplicity is not stated, it is assumed to be one. The input places (also known as the input set) of a transition are locations that are related to it by input arcs. In the same way, locations linked to a transition by output arcs are referred to as the output places (or output set) of the changeover. A location node can have many tokens (an additional graph element) in it. Within the location node that houses them, tokens are seen as black dots. The suggested approach is based on a rudimentary enhanced Petri Net, also known as the PN with area-velocity tokens, since PN is introduced to detect the velocity distribution of area representations and nodes. First, define the following model and logic [19].

The fundamental place/transition PN can be formatted as a five-tuple, according to this: A bidirectional divided graph with two distinct kinds of nodes the places P, which are depicted as circles, and the points of transition T, which can be seen as bars or boxes can be used to visually convey PN = {L, H, Q,K,N}.

$L = \{l_1, l_2, l_3, \dots., l_n\}$ is limited to some locations

$H = \{h_1, h_2, h_3, \dots., h_n\}$ is a limited collection of changes

The input to the arc operation, denoted as Q: (L × H) ⟶M, can be expressed as the input matrix$Q_{n\times m}$. $Q(L_i, h_j) = k$; Assuming that there was an arc with weight k from the location $L_1$to the point at h_j; else, $Q(L_i, h_j) = 0$.

K: (L × H) ⟶M is results in an arc function, what the output matrix $K_{n\times m}$ may use for representation. $K(h_j, L_k) = c$Where there is an arc of weight c connecting the change in position h_j to the locationL_k; alternatively, $K(h_j, L_k) = 0$.

$N: L \to M$, the net's existing labeling, can be expressed by the vector $N_{1\times n}$. The first marking, or N_0, indicates the net's initial condition.

following a new function transformation, behavioral recognition looks like this:

$$D^i_{new} = D^i_{old} + G_j . H_j \qquad (7)$$

Wherein G_j is the area-velocity efficiency of the process of transition j, H_j old is the previous rating of the token i that is traveling across the changeover j, and G_j is the detection reaction for the action j.

Algorithm 1: Petri Net

Initial State: Represent the initial state where a video is uploaded to YouTube.

Transitions:

a. Copyright Detection: When a video is uploaded, it undergoes a copyright detection process. If copyrighted material is detected, it transitions to a flagged state.

b. Dispute Resolution: If the uploader disputes the copyright claim, the system transitions to a resolution state.

c. Removal: If the copyright claim is upheld or not disputed within a certain timeframe, the video transitions to a removal state.

Flagged State: Videos identified as potentially infringing copyrights are placed in this state pending further action.

Resolution State: Represents the process of dispute resolution where the uploader can provide justification or counter-claim the copyright infringement.

Removal State: Indicates that the video has been removed from YouTube due to copyright infringement.

Final State: Represents the end of the process.

Characters from the input locations of a particular transition (rule) are eliminated once it fires. It is straightforward to understand the firing of a specific transition (rule) and the removal of tokens as the application of this rule to a particular reasoning process. As a result, identifying the input

locations of a fired rule is already not required in the next phases. One way to conceptualize this type of thinking is called forward reasoning. Below figure 2 shows the architecture of Petri Net.

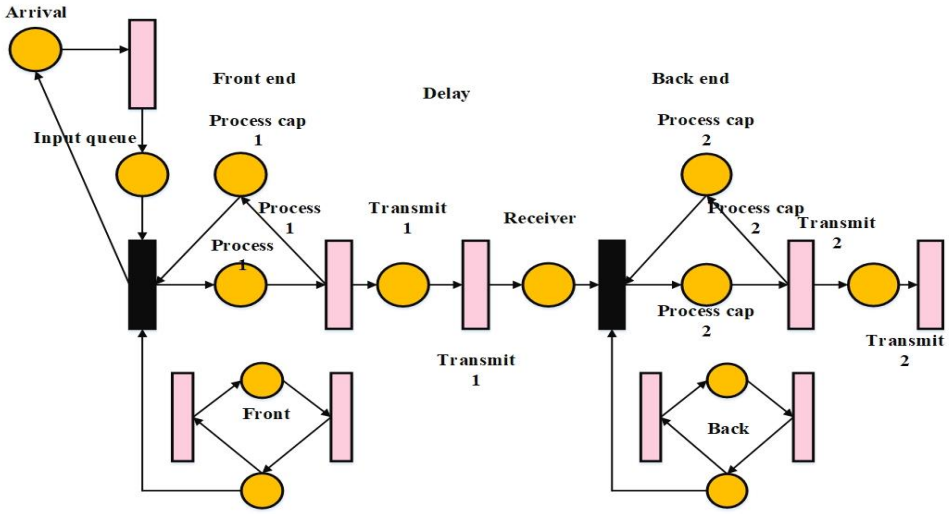


Figure 2: Petri Net [20]

Horse Herd Optimization (HHO) is a metaheuristic algorithm inspired by the behavior of horse herds, where individual horses follow the lead of the dominant ones to find optimal paths. In the context of Petri Nets, which are mathematical models used for describing concurrent systems, HHO can be utilized to reduce the weight of the Petri Net, thereby enhancing its efficiency and performance. By mimicking the collaborative behavior of horses, HHO facilitates the identification and removal of redundant or unnecessary elements within the Petri Net structure. This optimization process helps streamline the Petri Net, leading to improved resource utilization, reduced complexity, and enhanced overall system reliability. By leveraging the collective intelligence inherent in the HHO algorithm, Petri Nets can be fine-tuned to better reflect the underlying system dynamics, ultimately resulting in more effective modeling and analysis capabilities.

Fitness function to assess all chromosomes is stated as the reciprocal. The fitness functions given as below:

$$FIT(d_l) = \frac{1}{\sum_{q=1}^{w}\sum_{o}^{m} e_{qo}^{n}\left\|k_o - v\right\|^2} \quad (8)$$

The above equation store f_{avg} and $f_{\max}$ during iterative process; Then, it is simple demonstrate that lower value in terms of the goal function, greater influence, and the greater the fitness value $FIT(d_l)$. As a result, when the function reaches its minimum, i.e., $FIT(d_l)$ best partition, highest value, and optimal value are obtained.

4.3.1. Horse Herd Algorithm for Weight Optimization

Inspired by the collective behaviors observed in horse herds, the Horse Herd Algorithm (HHA) is a fresh and promising technique in the field of weight optimization. Emulating the adaptable and collaborative characteristics of horses in a herd, the HHA utilizes a dynamic optimization approach that repeatedly fine-tunes individual weights to achieve optimal outcomes. This algorithm uses the natural hierarchical structure of a herd to balance local and global search, introducing a novel interaction between exploration and exploitation. This research study highlights the potential of the Horse Herd Algorithm as a useful tool in the field of weight optimization approaches by demonstrating its effectiveness in solving complicated optimization issues across several domains via a series of simulated tests and comparison assessments [21]. The first population was created at random using k input features.

$$SW=\begin{bmatrix} WR_{1,1} & WR_{1,2} & \cdots & \cdots & WR_{1,d} \\ WR_{2,1} & WR_{2,2} & \cdots & \cdots & WR_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ WR_{m,1} & WR_{m,2} & \cdots & \cdots & WR_{m,d} \end{bmatrix} \tag{9}$$

Solution-based fitness analysis; algorithmic inspiration drawn from six common horse motions of different ages; the horse is moved in each iteration in line with the formula

$$R_m^{iter,age}=\vec{v}_m^{iter,age}+WR_m^{(iter-1),age},\quad age=\alpha,\beta,\gamma,\delta \tag{10}$$

their $WR_m^{iter,age}$ shows the location of m^{th} horse, indicating that they are cognizant of the age spectrum of horses, $iter$ indicates the current version, $\vec{v}_m^{iter,age}$ show off the horse's speed.

In light of the above indicated behavioral patterns, the horses' movement vector throughout each algorithm cycle may be written as follows:

$$\begin{aligned} \vec{v}_m^{iter,\alpha} &= \vec{g}_m^{iter,\alpha}+\vec{d}_m^{iter,\alpha} \\ \vec{v}_m^{iter,\beta} &= \vec{g}_m^{iter,\beta}+\vec{h}_m^{iter,\beta}+\vec{s}_m^{iter,\beta}+\vec{d}_m^{iter,\beta} \\ \vec{v}_m^{iter,\gamma} &= \vec{g}_m^{iter,\gamma}+\vec{h}_m^{iter,\gamma}+\vec{s}_m^{iter,\gamma}+\vec{i}_m^{iter,\gamma}+\vec{d}_m^{iter,\gamma}+\vec{r}_m^{iter,\gamma} \\ \vec{v}_m^{iter,\delta} &= \vec{g}_m^{iter,\delta}+\vec{i}_m^{iter,\delta}+\vec{r}_m^{iter,\delta} \end{aligned} \tag{11}$$

The content of this piece covers the stages of social and individual intelligence in horses.

Grazing (g)

They spend 16 to 20 hours a day grazing on the pasture with little chance of relaxation. The HHA method uses the coefficient to represent the grazing land surrounding each horse. Grazing may be represented mathematically by the following equation.

$$\vec{g}_m^{iter,age}=G_{iter}(\breve{U}+P\breve{L})[(DR)_m^{(iter-1)}],\quad age=\alpha,\beta,\gamma,\delta \tag{12}$$

$$G_m^{iter,age}=G_m^{(iter-1),age}\times\omega_G \tag{13}$$

here $\vec{g}_m^{iter,age}$ represents the motion's parameter of i^{th} hours, indicating the horse's inclination to graze. This factor decreases linearity by ω_G in each cycle. The upper and lower borders of the grazing space were identified by $\breve{L}$ and $\breve{U}$, Accordingly, P stands for a random integer between 0 and 1. For every age bracket, it suggested $\breve{L}$ and $\breve{U}$ be set at 0.95 and 1.05, correspondingly, and that coefficient G be set at 1.5.

Hierarchy (h)

Coefficient h represents the tendency of a herd of horses to follow the strongest and most experienced horse, according to HOA. Research indicates that horses in the median age range—five to fifteen years old—complied with the rule of hierarchy. It is summarized in the following few formulas.

$$\vec{h}_m^{iter,age}=H_m^{iter,age}\left[(WR)_*^{(iter-1)}-(WR)_m^{(iter-1)}\right],\quad age=\alpha,\beta,and\gamma \tag{14}$$

$$H_m^{iter,age}=H_m^{(iter-1),age}\times\omega_h \tag{15}$$

their $\vec{h}_m^{iter,age}$ highlights the consequences of the velocity parameters' optimal horse positioning, and $(WR)_*^{(iter-1)}$ identify the finest horse's placement.

Sociability (s)

Horses can live with other animals and need human relationships. Certain horses even seem to be happy alongside sheep and cattle, although they rarely seem to enjoy being by themselves. Factor indicates the tendency toward a typical place of the other horses s .

$$\vec{s}_m^{iter,age} = S_m^{iter,age}\left[\left(\frac{1}{n}\sum_{j=1}^{n}(WR)_j^{(iter-1)}\right)-(WR)_m^{(iter-1)}\right], \quad age=\beta,\gamma \tag{16}$$

$$S_m^{iter,age} = S_m^{(iter-1),age}\times\omega_S \tag{17}$$

In the formula above $\vec{s}_m^{iter,age}$ designates the i^{th} equine social mobility vector and $S_m^{iter,age}$ indicates the direction in which the relevant horse is heading toward the herd in $iter^{th}$ iteration. $S_m^{iter,age}$ reduction in each cycle with a ω_S factor, the total number of horses is n, age range in which every horse is age. The analyses of parameters, S coefficient for horses β and γ is determined.

Imitation (i)

These days, emulating equine behavior is also taken into consideration I. Equation following illustrates how horses try to imitate other horses until they reach maturity.

$$\vec{i}_m^{iter,age} = I_m^{iter,age}\left[\left(\frac{1}{Pn}\sum_{j=1}^{Pn}(WR)_j^{'(iter-1)}\right)-(WR)^{(iter-1)}\right], \quad age=\gamma \tag{18}$$

$$i_m^{iter,age} = i_m^{(iter-1),age}\times\omega_i \tag{19}$$

$i_m^{iter,age}$ shows the path taken by i^{th}. The progression of an example $i_m^{iter,age}$ horse at various locations $(DR)'$ in the path of the best i^{th} horses on average is shown in the formula that follows. The number of horses in the best places P is displayed Pn. It is recommended that 10% of horses be used for. ω_i is a reducing factor for as well I_{iter}, as was shown before.

Defense mechanism (d)

Horses must defend themselves from the onslaught or fight. This sort of protection system is incorporated wherever feasible throughout the whole life of a horse. Horses use a negative coefficient as a safety feature to keep themselves out of harm's way, as demonstrated by the following equation.

$$\vec{d}_m^{iter,age} = -W_m^{iter,age}\left[\left(\frac{1}{Qn}\sum_{j=1}^{Qn}(WR)_j^{'(iter-1)}\right)-(WR)^{(iter-1)}\right] \tag{20}$$

$$age=\alpha,\beta\ and\ \gamma \tag{21}$$

$$W_m^{iter,age} = W_m^{(iter-1),age}\times\omega_D \tag{22}$$

where escape vector of i^{th} horse from a sample of some of the worst-located horses, which are assigned $(WR)'$ is shown by $\vec{d}_m^{iter,age}$. Additionally, Qn shows how many horses are in the worst place. It calculated that Q is equivalent to 20% of all horses. The previously mentioned reduction factor per cycle for D_{iter} shown by the symbol ω_D .

Roam (r)

Throughout the vast pastures, horses graze and roam, seeking food from one pasture to another. A horse could opt to graze someplace else out of the blue. Horses are quite curious creatures, and they frequently go to new places to familiarize themselves with their surroundings and find fresh pastures. When a horse is young, wandering is almost always present and becomes less common as the horse becomes older. The portion of the equation r duplicates this inclination as a random movement and represents it as such. This method is also demonstrated by the equation below.

$$\vec{r}_m^{iter,age} = R_m^{iter,age}P(WR)^{(iter-1)}, \quad age=\gamma,\delta \tag{23}$$

$$R_m^{iter,age} = R_m^{(iter-1),age}\times\omega_R \tag{24}$$

Out of a local minimum, in an escape and quest, $\vec{r}_m^{iter,age}$ depicts the velocity vector of the i^{th} horse, and R symbolizes the component of decrease of $R_m^{iter,age}$ per cycle.

Horses' δ speed from 0 to 5 years old:

$$\begin{aligned}\vec{v}_m^{iter,\delta} &= \left[G_m^{(iter-1),\delta}\omega_G\left(\breve{U}+\breve{P}L\right)\left[(WR)_m^{(iter-1)}\right]\right]\\ &+\left[i_m^{iter-1,\delta}\omega_i\left[\left(\frac{1}{Pn}\sum_{j=1}^{Pn}(WR)_j^{'iter-1}\right)-(WR)^{iter-1}\right]\right]\\ &+\left[R_m^{iter-1,\delta}\omega_R P(WR)^{iter-1}\right]\end{aligned} \tag{25}$$

Horses' γ speeds from five to ten years old:

$$\begin{aligned}\vec{v}_m^{iter,\gamma} &= \left[G_m^{iter-1,\gamma}\omega_G(\breve{U}+P\breve{L})\left[(WR)_m^{iter-1}\right]\right]\\ &+\left[H_m^{iter-1,\gamma}\omega_H\left[(WR)_*^{iter-1}-(WR)_m^{iter-1}\right]\right]\\ &+\left[S_m^{iter-1,\gamma}\omega_S\left[\left(\frac{1}{n}\sum_{j=1}^{n}(WR)_j^{iter-1}\right)-(WR)_m^{iter-1}\right]\right]\\ &+\left[i_m^{iter-1,\gamma}\omega_i\left[\left(\frac{1}{Pn}\sum_{j=1}^{Pn}(WR)_j^{'iter-1}\right)-(WR)^{iter-1}\right]\right]\\ &-\left[W_m^{iter-1,\gamma}\omega_D\left[\left(\frac{1}{Qn}\sum_{j=1}^{Qn}(WR)_j^{'iter-1}\right)-(WR)^{iter-1}\right]\right]\\ &+\left[R_m^{iter-1,age}\omega_R P(WR)^{iter-1}\right]\end{aligned} \tag{26}$$

Horses β aged between ten and fifteen years old:

$$\begin{aligned}\vec{v}_m^{iter,\beta} &= \left[G_m^{iter-1,\beta}\omega_G\left(\breve{U}+P\breve{L}\right)\left[(WR)_m^{iter-1}\right]\right]\\ &+\left[H_m^{iter-1,\beta}\omega_H\left[(WR)_*^{iter-1}-(WR)_m^{iter-1}\right]\right]\\ &+\left[S_m^{iter-1,\beta}\omega_S\left[\left(\frac{1}{n}\sum_{j=1}^{n}(WR)_j^{iter-1}\right)-(WR)_m^{iter-1}\right]\right]\\ &-\left[W_m^{iter-1,\beta}\omega_D\left[\left(\frac{1}{Qn}\sum_{j=1}^{Qn}(WR)_j^{'iter-1}\right)-(WR)^{iter-1}\right]\right]\end{aligned} \tag{27}$$

Horses older than fifteen years old:

$$\begin{aligned}\vec{v}_m^{iter,\alpha} &= \left[G_m^{iter-1,\alpha}\omega_G(\breve{U}+P\breve{L})\left[(WR)_m^{iter-1}\right]\right]\\ &-\left[W_m^{iter-1,\alpha}\omega_D\left[\left(\frac{1}{Qn}\sum_{j=1}^{Qn}(WR)_j^{'iter-1}\right)-(WR)^{iter-1}\right]\right]\end{aligned} \tag{28}$$

The horse value's total velocity is provided as,

$$V=\left\{\vec{v}_m^{iter,\delta},\ \vec{v}_m^{iter,\gamma},\vec{v}_m^{iter,\beta},\vec{v}_m^{iter,\alpha}\right\} \tag{29}$$

4.4. Video Representation and Classification

Although our first topic modeling was helpful for exploration, they decided to apply a supervised classification approach to provide more trustworthy examples for analysis after we had determined which groups were relevant. The state of the art in text classification has significantly improved recently thanks to developments in natural language processing, which also increase the usefulness of deep neural networks for

classification applications. Utilizing "transfer learning," new methods train generic models on massively current data sets after which they fine-tune them for particular uses on considerably smaller labeled datasets [22].

4.4.1. Implementing a BiLSTM Network Enhanced with Attention Mechanism

The attention system in a neural network architecture determines when to examine data (or in this example, video segments) by autonomously prioritizing feature vectors containing the most useful information over those with less value knowledge [23].

Using the i-th BiLSTM's final hidden state as t_{ih}, that is calculated as,

$$t_{ih} = [t_h^f, t_h^z] \quad (30)$$

Next, the formulas that follow are used to calculate the attention mechanism.

$$u_{ih} = tant(C_y t_{ih} + z_y) \quad (31)$$

$$y_{ih} = \frac{\exp(u_{ih})}{\sum_{r=1}^{H} \exp(u_{rh})} \quad (32)$$

$$f_h = \sum_{h=1}^{H} y_{ih} . t_{ih} \quad (33)$$

Equation computes the attention weight and the mechanism for attention applies to the i-th BiLSTM result vector at time h. The attention layer's bias and intensity are represented by C_y and z_y. Ultimately, the output of the attention layers produces an attention vectorf_h, which is computed as the weighted sum of the multiplications of the i-th BiLSTM output vector at time h plus the weight of attention y_{ih}. Below figure 2 show the structure of the BiLSTM model incorporating an attention mechanism.

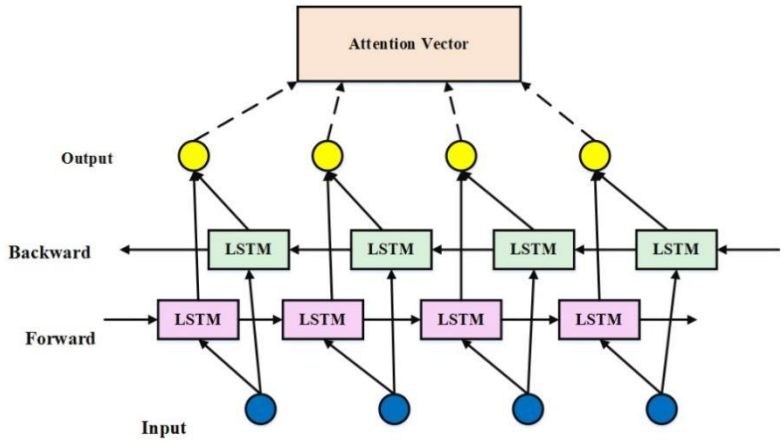


Figure 3: Depicts the Structure of the BiLSTM Model Incorporating an Attention Mechanism

4.4.2. Softmax Classifier

Each deep learning model's output layer's softmax activating function is regarded as a softmax classifier. Another softmax activation function was included into the final fully connected layer of the proposed model in order to estimate the relative probability of the three output units, which allowed each video clip to be classified into one of classes. One calculates the softmax activation function (σ) as,

$$\sigma(b_i) = \frac{\exp(b_i)}{\sum_{w=0}^{M-1} \exp(b_w)} \quad (34)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (35)$$

5. Result and discussion

This section describes the experimental setup, performance measurement, and experimental results. The YouTube vedio dataset, which is accessible to the public. The suggested method will be put into practice using the Python programming language, and the outcomes of the experiments will be assessed and contrasted with previous models in terms of statistical metrics like accuracy, precision, F1-Score, and MCC.

To evaluate these findings, compute the accuracy, recall, precision, and F1-score, FPR, FNR, MCC and NPV indicators.

$$recall = \frac{TP}{TP + FN} \quad (36)$$

$$precision = \frac{TP}{TP + FP} \quad (37)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (38)$$

$$FPR = \frac{F_P}{T_N + F_P} \quad (39)$$

$$FNR = \frac{F_N}{T_P + F_N} \quad (40)$$

$$MCC = \frac{T_P T_N - F_P F_N}{\sqrt{(T_P + F_P) - (T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad (41)$$

$$NPV = \frac{T_N}{T_N + F_N} \quad (42)$$

TP signifies the true positive, FP the false positive, TN the true negative, and FN the false negative.
Assess the viability of the proposed initial-phase within a simulated environment by employing Python software and utilizing Youtube datasets. An Intel(R) Core(TM) i5-3470-equipped computer is used to conduct the test. Additionally, the OS maker micro software 10 pro has inserted physical memory (RAM) 12.7 GB.

Table 1: Simulation parameters of dataset.

Simulation parameters	Values
Batch size	32
Epoch	18
Input size	256,256,3
Activation function	Softmax
Dropout	0. 3
Dense	32 and 2

Confusion matrix provides a comprehensive overview of the model's accuracy, highlighting areas of both success and potential improvement. Through a thorough analysis of the matrix, stakeholders can gain a nuanced understanding of the strengths and weaknesses inherent in the classification system employed within the original video, thus facilitating informed decision-making and further refinement of the model.

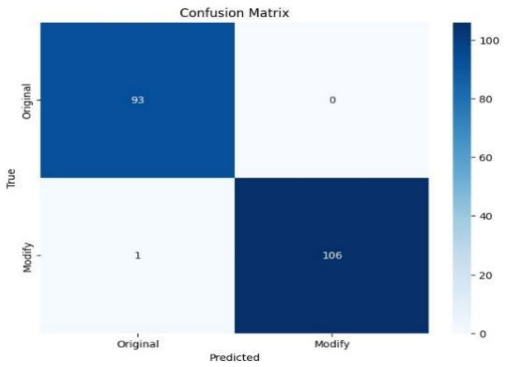
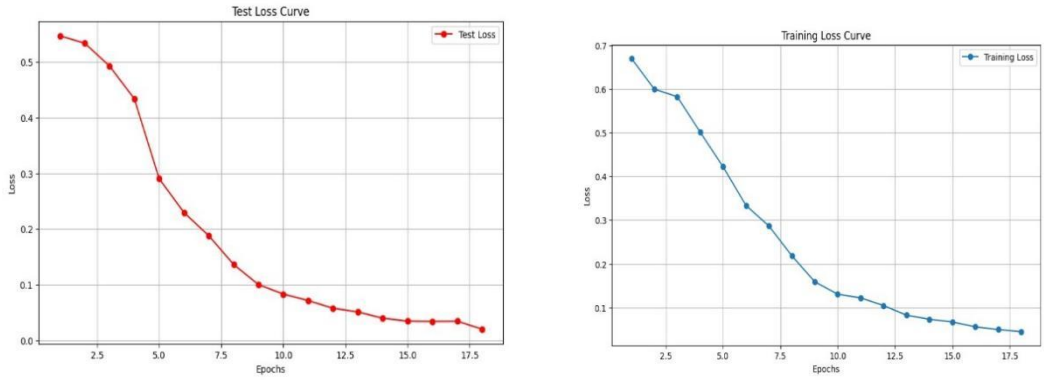


Figure 4: Confusion Matrix

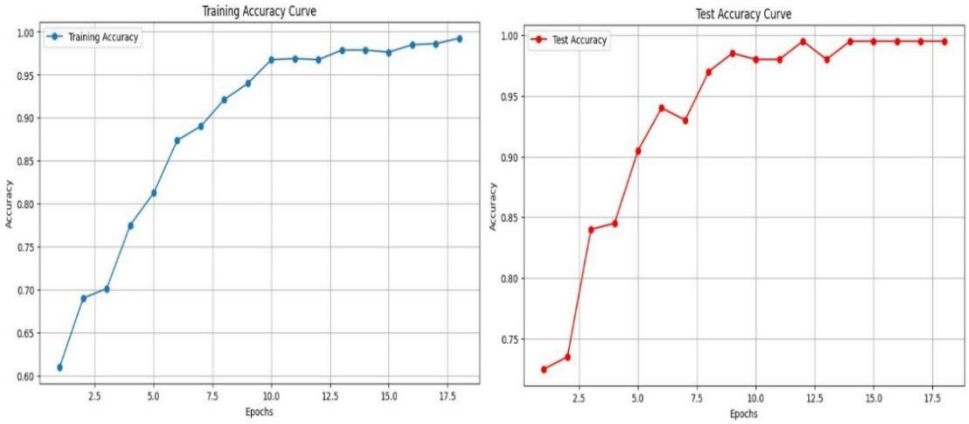
The confusion matrix displayed in Figure 4 is derived from the original video dataset, serving as a pivotal component in assessing the classification performance of the model. By meticulously detailing the distribution of true positive, true negative, false positive, and false negative classifications, this matrix offers invaluable insights into the efficacy of the model's predictive capabilities.



(a) **(b)**

Figure 5: (a) Training and (b) Testing Loss Curve

In Figure 5, both the training and testing loss curves are presented, providing critical insights into the performance of the YouTube video infringement model. The training loss curve, depicted in subplot (a), showcases the gradual decrease in loss over successive epochs, indicative of the model's learning process and improved predictive capabilities during training. Subplot (b) illustrates the testing loss curve, reflecting the model's ability to generalize to unseen data. Discrepancies or similarities between the two curves offer valuable information regarding potential overfitting or underfitting tendencies. These loss curves collectively serve as instrumental tools for assessing the model's efficacy in combatting YouTube video infringement, guiding further refinement and optimization efforts.



(a) **(b)**

Figure 6: (a) Training and (b) Testing Accuracy Curve

In Figure 6, the training and testing accuracy curves are depicted in subplots (a) and (b) respectively, providing crucial insights into the performance of the model for combating YouTube video infringement. The training accuracy curve showcases the model's improvement in correctly predicting labels over successive epochs, highlighting its learning process. Conversely, the testing accuracy curve demonstrates how well the model generalizes to unseen data, serving as a key metric for evaluating its performance in real-world scenarios. Discrepancies or similarities between the training and testing accuracy curves offer valuable insights into potential overfitting or underfitting tendencies, guiding further optimization efforts. These accuracy curves collectively contribute to the comprehensive assessment of the model's effectiveness in addressing YouTube video infringement.

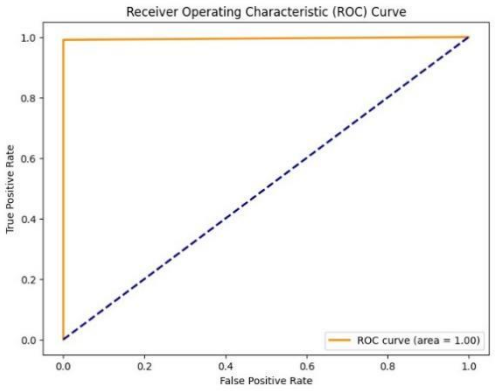


Figure 7: ROC Curve

In Figure 7, the ROC curve is presented, showcasing the performance of the model in terms of sensitivity and specificity. Notably, the area under the ROC curve is reported as 1.00, indicating excellent discriminatory power and a high degree of accuracy in distinguishing between positive and negative instances. This suggests that the model achieves near-perfect classification performance across various thresholds. Such a high ROC curve area underscores the robustness and effectiveness of the model in accurately detecting YouTube video infringement. This result is indicative of the model's reliability and suitability for real-world deployment in combating copyright infringement on the platform.

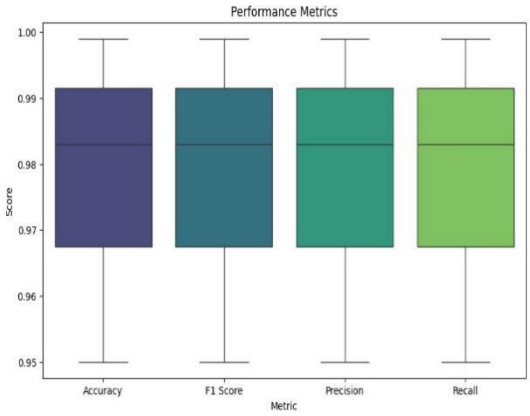


Figure 8: performance Metrics Graph

In Figure 8, the classification graph displays precision, recall, and F1 score metrics, with an accompanying accuracy of 99%. This high accuracy suggests the model's exceptional performance in correctly classifying instances of YouTube video infringement. Precision, recall, and F1 score provide additional insights into the model's ability to precisely identify positive instances while minimizing false positives and false negatives. The high accuracy, combined with strong precision, recall, and F1 scores, underscores the reliability and effectiveness of the model in accurately detecting and addressing copyright infringement on the platform. This robust performance enhances trust in the model's capabilities for real-world deployment, contributing to more effective enforcement of copyright policies on YouTube.

Table 2: Classification Table

Parameters	Precision (%)	Recall (%)	F1-score (%)
Original	99	100	99
Modify	100	99	100
Accuracy	99		

Table 2 presents the classification table, detailing precision, recall, and F1-score percentages for both the original and modified instances. The results demonstrate consistently high performance metrics across both categories, with precision, recall, and F1-score values ranging between 99% and 100%. These findings indicate the model's ability to accurately classify instances of both original and modified content with minimal errors. The overall accuracy of 99% further corroborates the model's effectiveness in distinguishing between the two classes. Such robust performance metrics underscore the reliability and utility of the model in identifying instances of YouTube video infringement, contributing to improved content moderation and enforcement efforts on the platform.

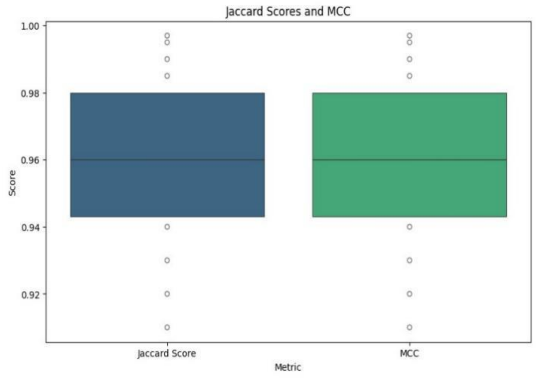


Figure 9: Jaccard Score MCC Graph

Figure 9 illustrates the Jaccard Score and Matthews Correlation Coefficient (MCC) graph for assessing the performance of the proposed method in detecting YouTube video infringement. The graph demonstrates a consistent high Jaccard Score and MCC values, indicating robustness and accuracy in identifying instances of infringement across the dataset. Such results underscore the effectiveness of the method in detecting and potentially mitigating copyright violations in YouTube videos.

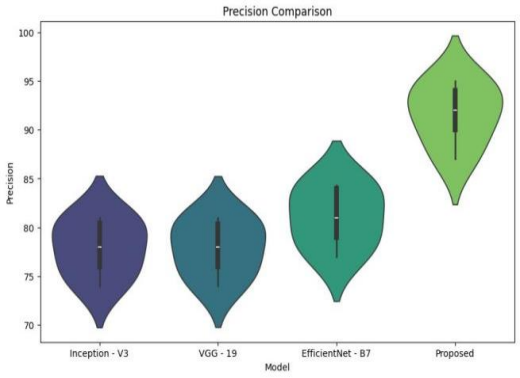


Figure 10: Precision Comparison with Existing Methods

Figure 10 showcases a precision comparison between existing methods, including Inception V3, VGG19, EfficientNet B7 [24], and the proposed method, which achieves a notably high precision of 98% in detecting YouTube video infringement. The comparison highlights the superiority of the proposed approach over established models, emphasizing its potential for accurate and reliable identification of copyright violations in YouTube content. Such findings underscore the significance of leveraging advanced methodologies to enhance infringement detection and enforcement measures on online platforms.

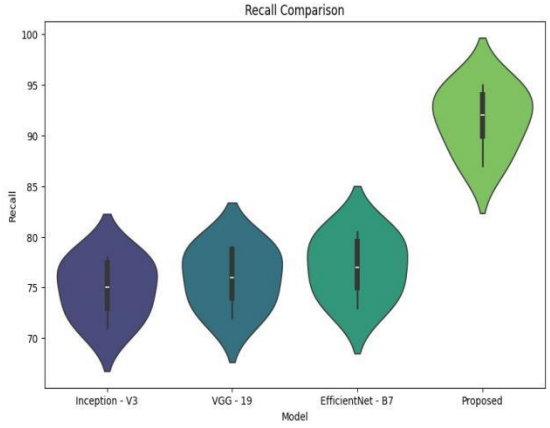


Figure 11: Recall Comparison with Existing Methods

Figure 11 presents a recall comparison between the proposed method and existing techniques, including Inception V3, VGG19, and EfficientNet B7, wherein the proposed method demonstrates a remarkable recall value of 98% in identifying instances of YouTube video infringement. This comparison underscores the efficacy of the proposed approach in efficiently capturing copyright violations, surpassing the recall performance of established models. The findings underscore the significance of adopting advanced methodologies to enhance the detection and mitigation of copyright infringements on online platforms, contributing to a more robust and effective enforcement framework.

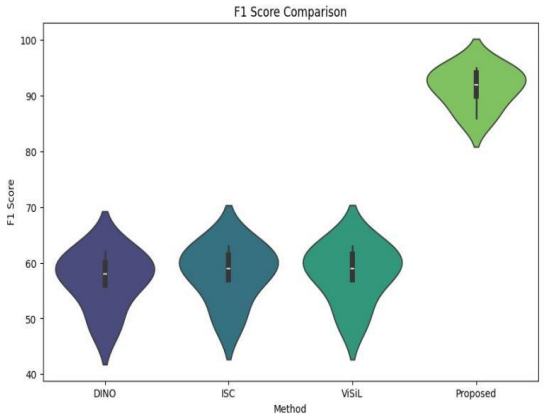


Figure 12: F1 Score Comparison with Existing Methods

In Figure 12, a comparison of F1 scores between the proposed method and existing approaches such as ISC [25], DINO [26], and ViSiL [27] reveals a notably high F1 score value of 98% for the proposed method in detecting YouTube video infringement.

This comparison highlights the superior performance of the proposed approach over established methods, emphasizing its effectiveness in accurately identifying copyright violations in YouTube content. Such results underscore the significance of leveraging advanced methodologies to enhance the detection and enforcement of copyright infringements on online platforms, ultimately fostering a more robust system for protecting intellectual property rights.

Figure 13: Fitness Improvement over Iteration Graph

Figure 13 depicts the fitness improvement over iterations utilizing the HHO for detecting YouTube video infringement. The graph illustrates a steady enhancement in fitness values as iterations progress, indicative of the algorithm's efficacy in refining the infringement detection process. Such results underscore the potential of optimization algorithms like HHO in bolstering the accuracy and efficiency of copyright violation detection mechanisms, thereby contributing to more effective enforcement strategies on online platforms.

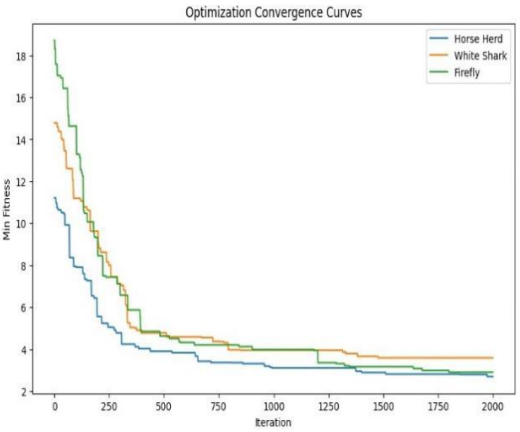


Figure 14: Convergence Curve

Figure 14 displays the convergence curve comparing the performance of the proposed optimization algorithm, HHO, with existing methods White Shark [28] and Firefly [29] in detecting YouTube video infringement. The curve illustrates the convergence behavior of each algorithm, showcasing HHO's superior convergence rate and efficiency in reaching optimal solutions for infringement detection tasks. These findings highlight the potential of HHO as a promising optimization approach for enhancing the accuracy and effectiveness of copyright infringement detection mechanisms on online platforms.

6. Conclusion and Future Scope

In conclusion, this paper has addressed the critical issue of YouTube video infringement by proposing a comprehensive framework that combines advanced techniques in multimedia forensics and legal frameworks. Through our novel approach for video representation and classification, have demonstrated a robust methodology that prioritizes privacy-preserving techniques and achieves high accuracy in infringement detection. By leveraging Gabor filters, key frame selection, Petri Net, and Horse Herd optimization algorithms, along with a BiLSTM network and softmax classifier, our approach offers a sophisticated solution for analyzing and classifying video content while safeguarding the rights of content creators. The achieved accuracy of 99% highlights the effectiveness of our method in detecting copyright violations, thereby emphasizing its potential for widespread applications in video analysis and understanding. Moving forward, continued research and development in this field will be crucial for further improving the efficacy of copyright enforcement measures and ensuring the integrity of online video content. Further research could explore the integration of deep learning models with emerging technologies like blockchain for immutable copyright protection in online video content. Additionally, extending the framework to incorporate real-time monitoring and adaptive learning algorithms could enhance its applicability in dynamic digital environments.

Reference

[1] D. Y. Zhang, L. Song, Q. Li, Y. Zhang, and D. Wang, “StreamGuard: A Bayesian Network Approach to Copyright Infringement Detection Problem in Large-scale Live Video Sharing Systems,” in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA: IEEE, Dec. 2018, pp. 901–910. doi: 10.1109/BigData.2018.8622306.

[2] D. Y. Zhang, Q. Li, H. Tong, J. Badilla, Y. Zhang, and D. Wang, “Crowdsourcing-Based Copyright Infringement Detection in Live

Video Streams," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, Spain: IEEE, Aug. 2018, pp. 367–374. doi: 10.1109/ASONAM.2018.8508523.

[3] K. Erickson and M. Kretschmer, "Analyzing Copyright Takedown of User-Generated Content on YouTube," 2018.

[4] "An Empirical Study on Copyright Infringement of Movie Narrative Short Videos in the Age of Self-Media," *Front. Soc. Sci. Technol.*, vol. 5, no. 18, 2023, doi: 10.25236/FSST.2023.051817.

[5] I. Atanasova, "COPYRIGHT INFRINGEMENT IN DIGITAL ENVIRONMENT," *J. Law Econ.*, vol. 1, pp. 13–22, Nov. 2019.

[6] C. Iwendi *et al.*, "KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks," *IEEE Access*, vol. 8, pp. 72650–72660, 2020, doi: 10.1109/ACCESS.2020.2988160.

[7] D. Połap, G. Srivastava, and K. Yu, "Agent architecture of an intelligent medical system based on federated learning and blockchain technology," *J. Inf. Secur. Appl.*, vol. 58, p. 102748, May 2021, doi: 10.1016/j.jisa.2021.102748.

[8] M. Kamal, G. Srivastava, and M. Tariq, "Blockchain-Based Lightweight and Secured V2V Communication in the Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 3997–4004, Jul. 2021, doi: 10.1109/TITS.2020.3002462.

[9] R. Kumar, R. Tripathi, N. Marchang, G. Srivastava, T. R. Gadekallu, and N. N. Xiong, "A secured distributed detection system based on IPFS and blockchain for industrial image and video data security," *J. Parallel Distrib. Comput.*, vol. 152, pp. 128–143, Jun. 2021, doi: 10.1016/j.jpdc.2021.02.022.

[10] N. H. Sharfina, H. Paserangi, F. P. Rasyid, and M. I. N. Fuady, "Copyright Issues on the Prank Video on the Youtube:," presented at the International Conference on Environmental and Energy Policy (ICEEP 2021), Surakarta, Central Java, Indonesia, 2021. doi: 10.2991/assehr.k.211014.021.

[11] S.-H. Chen *et al.*, "Advantage in privacy protection by using synchronous video observed treatment enhances treatment adherence among patients with latent tuberculosis infection," *J. Infect. Public Health*, vol. 13, no. 9, pp. 1354–1359, Sep. 2020, doi: 10.1016/j.jiph.2020.03.013.

[12] S. Pulukuri and B. Abrams, "Incorporating an Online Interactive Video Platform to Optimize Active Learning and Improve Student Accountability through Educational Videos," *J. Chem. Educ.*, vol. 97, no. 12, pp. 4505–4514, Dec. 2020, doi: 10.1021/acs.jchemed.0c00855.

[13] T. Ginossar, I. J. Cruickshank, E. Zheleva, J. Sulskis, and T. Berger-Wolf, "Cross-platform spread: vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations," *Hum. Vaccines Immunother.*, vol. 18, no. 1, pp. 1–13, Jan. 2022, doi: 10.1080/21645515.2021.2003647.

[14] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, "Facing Device Attribution Problem for Stabilized Video Sequences." arXiv, Nov. 05, 2018. Accessed: Feb. 12, 2024. [Online]. Available: http://arxiv.org/abs/1811.01820

[15] H. Mareen, J. D. Praeter, G. V. Wallendael, and P. Lambert, "A Novel Video Watermarking Approach Based on Implicit Distortions".

[16] J. Zheng, S. Teng, P. Li, W. Ou, D. Zhou, and J. Ye, "A Novel Video Copyright Protection Scheme Based on Blockchain and Double Watermarking," *Secur. Commun. Netw.*, vol. 2021, pp. 1–16, Dec. 2021, doi: 10.1155/2021/6493306.

[17] R. Kumar, R. Tripathi, N. Marchang, G. Srivastava, T. R. Gadekallu, and N. N. Xiong, "A secured distributed detection system based on IPFS and blockchain for industrial image and video data security," *J. Parallel Distrib. Comput.*, vol. 152, pp. 128–143, Jun. 2021, doi: 10.1016/j.jpdc.2021.02.022.

[18] C. Kong, B. Chen, W. Yang, H. Li, P. Chen, and S. Wang, "Appearance Matters, So Does Audio: Revealing the Hidden Face via Cross-Modality Transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 423–436, Jan. 2022, doi: 10.1109/TCSVT.2021.3057457.

[19] H. Tu, R. Xu, R. Chi, and Y. Peng, "Multiperson Interactive Activity Recognition Based on Interaction Relation Model," *J. Math.*, vol. 2021, pp. 1–12, Aug. 2021, doi: 10.1155/2021/5576369.

[20] R. D. Sousa *et al.*, "Performability Evaluation and Sensitivity Analysis of a Video Streaming on Demand Architecture," *Appl. Sci.*, vol. 13, no. 2, p. 998, Jan. 2023, doi: 10.3390/app13020998.

[21] M. A. Awadallah, A. I. Hammouri, M. A. Al-Betar, M. S. Braik, and M. A. Elaziz, "Binary Horse herd optimization algorithm with crossover operators for feature selection," *Comput. Biol. Med.*, vol. 141, p. 105152, Feb. 2022, doi: 10.1016/j.compbiomed.2021.105152.

[22] J. E. Gray and N. P. Suzor, "Playing with machines: Using machine learning to understand automated copyright enforcement at scale," *Big Data Soc.*, vol. 7, no. 1, p. 205395172091996, Jan. 2020, doi: 10.1177/2053951720919963.

[23] K. Yousaf and T. Nawaz, "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos," *IEEE Access*, vol. 10, pp. 16283–16298, 2022, doi: 10.1109/ACCESS.2022.3147519.

[24] K. Yousaf and T. Nawaz, "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos," *IEEE Access*, vol. 10, pp. 16283–16298, 2022, doi: 10.1109/ACCESS.2022.3147519.

[25] S. Yokoo, "Contrastive Learning with Large Memory Bank and Negative Embedding Subtraction for Accurate Copy Detection." arXiv, Dec. 08, 2021. Accessed: Feb. 15, 2024. [Online]. Available: http://arxiv.org/abs/2112.04323

[26] M. Caron *et al.*, "Emerging Properties in Self-Supervised Vision Transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9630–9640. doi: 10.1109/ICCV48922.2021.00951.

[27] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras, and Y. Kompatsiaris, "ViSiL: Fine-Grained Spatio-Temporal Video Similarity Learning," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 6350–6359. doi: 10.1109/ICCV.2019.00645.

[28] K. Lee *et al.*, "Oceanographic conditions associated with white shark (Carcharodon carcharias) habitat use along eastern Australia," *Mar. Ecol. Prog. Ser.*, vol. 659, pp. 143–159, Feb. 2021, doi: 10.3354/meps13572.

[29] J. Wu, Y.-G. Wang, K. Burrage, Y.-C. Tian, B. Lawson, and Z. Ding, "An improved firefly algorithm for global continuous optimization problems," *Expert Syst. Appl.*, vol. 149, p. 113340, Jan. 2020, doi: 10.1016/j.eswa.2020.113340.