

# MULTI-LEVEL VIDEO ANALYTICS FOR EFFICIENT DEDUPLICATION IN CLOUD STORAGE

APARNA RAMALINGIAH<sup>1</sup>, AND SHILPA CHAUDHARI<sup>1</sup>

<sup>1</sup>DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, M S RAMAIAH INSTITUTE OF TECHNOLOGY, BENGALURU, INDIA (AFFILIATED TO VTU)

**CORRESPONDING AUTHOR:**

APARNA RAMALINGIAH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, M S RAMAIAH INSTITUTE OF TECHNOLOGY, BENGALURU, INDIA (AFFILIATED TO VTU)

email: [aparnasurya.js@gmail.com](mailto:aparnasurya.js@gmail.com)

---

## ABSTRACT

A video search query resulting in several similar videos is a common search problem users encounter these days while they expect few relevant videos matching their search context. Also, redundant/duplicate videos lead to inefficient utilization of cloud storage which is the cost-effective storage now a day. Extensive exploration of video deduplication techniques has good scope for research. Hence, the proposed multi-level video analytics aims to identify if a given input video is a possible duplicate of existing videos. Three-level similarity check namely summary\_check, transcript\_check and video\_frame\_check is proposed for duplicate determination. Summary\_check level uses cosine similarity metric for determining semantic likeness of a set of documents. Transcript\_check level uses text-to-speech service provider called AssemblyAI which provides the transcripts of uploaded video or audio files for comparison. Video\_frame\_check level based on deep learning algorithm compares videos based on the search relevant objects present in the video. The threshold defined for each level decides the exactness of duplicates and metadata generated in each level is stored in a database for future use. The input video is stored in cloud storage if found to be unique. The results obtained prove to be significantly better compared to existing approaches of deduplication check.

**Keywords:** Video deduplication, Video analysis, Cloud storage, Object Detection, Deep Learning

---

## 1. INTRODUCTION

Information explosion is being witnessed due to wide range of generators such as social networks, mobile devices, and streaming media. On a positive note, this led to the emergence of several new applications like media streaming, cloud data storage, digital libraries etc., These applications expect to ingest data of high quality to provide reliable services. With massive amounts of data being generated every day, it is critical that data be managed efficiently. For instance, Google Photos is a cloud storage platform that contains a massive number of media. To handle storage efficiently, Google Photos has identical duplicate detection, which means that Google Photos will not allow re-uploading such photos that were already uploaded whenever user tries an upload action. The unique hash that is maintained for every uploaded photo file will be used to detect duplicates. Currently, they can only identify duplication of photos (images) while video deduplication is yet to be explored extensively.

Usage of the internet today has grown to an extent that e-learning resources have become the first choice to learn about any topic. There are a lot of platforms on the internet that provide tutorials, explanatory videos on different subjects. These platforms essentially act as a database of videos from which consumers can access required videos. YouTube as a search platform is a great example for this purpose. However, many a times users/consumers face difficulty in selecting videos to watch especially if the videos are longer in duration, selecting any one of similar videos gives them a feeling of missing out on some topics. We might find relevant

videos while searching for videos on the internet by providing the keywords, yet those videos might be redundant for most of the part.

One of the biggest problems in cloud data is inefficient utilization of storage when videos are similar on redundant topics and do not provide additional information. Therefore, a solution to this issue is elimination of duplicates. When the user who is uploading a video would know how much of the video is already existing in the database or is redundant, then the user can revise or adjust the video accordingly. This process of duplicate elimination namely, deduplication can be used to achieve storage optimization and thus more storage capacity could be used to achieve longer-term data retention, provide with larger backup capacity, and achieve continuous verification of backup data, improve the level of data recovery services, and aids in effective data disaster recovery.

This paper discusses one possible way to address this situation. Analysis of videos in the initial levels of our method using their transcripts and summary gives initial confidence about possible duplication considering a certain threshold. Later, when the deduplication result is greater than specified threshold in transcripts, then video frame check is applied for duplication check. We have coined the term multi-level video deduplication since the video deduplication is performed through multiple stages. In addition to video deduplication check, a database is maintained for future deduplication checks when users upload videos for storage. Our specific contributions for effective management of video data in the cloud environment are as follows. (1) design and development of multi-level video deduplication check model (2) design and development of summary check level using cosine similarity check (3) design and development of transcript check level using AssemblyAI (4) design and development of video frame check level using deep learning techniques. (5) Maintenance of video data with the deduplication facility.

The outline of the paper is presented as follows. Section 2 presents the review of the literature. The multi-level model for video deduplication is explained in Section 3. Each level design and development details are also elaborated in Section 3. Comparison of the proposed model performance is discussed on Section 4. The paper's conclusions are discussed in Section 5.

## 2. RELATED WORKS

Many of the studies conducted as part of our research fit into multiple themes, we decide to include each study to fit the most representative theme of the study as a whole. Videos have become the default medium to share various content among the people and thus has resulted in huge number of videos on different topics being generated and stored in the web. Such videos are easily accessible by just using a web browser over the internet. So, anyone can access and view the videos using a browser or on the widely used YouTube platform. We will aim at providing the recent research in the themes that are discussed in this paper. Each theme in described in detail next.

### 2.1 Theme 1: Analyzing Transcripts and Deduplication

In [1], authors have conducted their research studies in the areas of data privacy, data integrity, and data deduplication. Their analysis classifies existing Data Deduplication Schemes based on different dimensions like data popularity, homomorphic validation, target-based deduplication, static/dynamic method, proof of storage, digital signature scheme and convergent encryption. Authors have also discussed open research challenges in these abovementioned areas.

In [2], authors were able to reduce user labeling efforts in large-scale deduplication tasks by developing a two-stage sampling selection strategy (T3S). The strategy involves selecting small random subsamples from different datasets and incrementally analyzing them to eliminate redundancy. The authors plan to explore genetic programming's ability to combine similarity functions to assess the closeness of Minimum True Pair (MTP) and Maximum False Pair (MFP) boundary estimates to the ideal values.

In [3], authors use Burrows - Wheeler Data Transform (BWT) encoding to compress plain data files, resulting in superior performance and significant storage and bandwidth savings. The authors are exploring extending this approach to other multimedia data, such as audio and video files, to address the issue of redundant data copies in cloud computing.

In [4], authors depicted the overall performance of the video duplication detection using the precision-recall curves. The larger areas under the curve indicate a better performance. Authors fused the search results from multiple tables by developing an efficient similarity ranking method based on the index. The similarity ranking method incorporated both the Hamming distances and the temporal order between the videos.

CSPs use hashing to avoid redundant copies, but in [5], a Nature-Inspired, Genetic Programming Approach is proposed. This approach uses biological evolution ideologies, Sequence Matching Algorithms,

and Levenshtein's Algorithms for text comparison. Online education is popular for its high-quality content, but students may struggle with time constraints. So, avoiding duplicates will result in highly relevant search results.

In [6], authors worked on obtaining a ranked list of video transcripts for a particular query for the e-learning resources from a repository. The work is carried out on 16,012 available video transcripts from the Media Website at Universitat Politecnica de Valencia. The concept is to generate a bag-of-words, compute TF-IDF scores, cluster the transcripts and build a Latent Dirichlet Allocation (LDA) model for each cluster. The results were on par in terms of comparing with the currently existing search mechanism.

Authors of [7] worked on the text data and used Keyword Extraction as a prime task for judging the important parts of text rather than going through the entire text. Such a method of keyword extraction was used exclusively for educational video transcripts from MOOCs in this research work. Keyword Extraction helps the reader to easily analyze the semantics of the transcripts rather than reading through the whole text saving a lot of time. The technique devised by authors identifies the Noun Chunks in the text of the transcript by employing an approach called Regular Expression Grammar Rule. Extracting keywords help in finding out the specifically important part of the educational material.

The authors of [8] developed a Python module using Natural Language Processing (NLP) algorithms to summarize online class videos. The model uses TF-IDF and Gensim for information retrieval and topic modeling. Cosine similarity and ROUGE score methods are used, with efficiency ranging from 40-50%. The model is designed for cloud storage optimization, allowing for efficient summary creation and reference.

Authors of [9] have used cross-user deduplication to reduce the amount of usage of server storage. They present a single-server protocol based on secure LSH (SLSH) for identifying cross-user nearly identical duplicates. Many authors propose using multiple servers to maintain privacy during deduplication, but they have used a single server to define ideal security and have proved using simulation that their protocol is secure against colluding and malicious adversaries. They have arrived at experimental proofs to indicate that the individual parts of the protocol are computationally feasible.

The authors of [10] used Natural Language Processing and Machine Learning Learning to summarize the YouTube video transcripts without losing the key elements. Thus, they proposed a video summarization system that makes it easy to know the content of a video. This system eases the job of the user by not having to watch the entire video. Their proposed method retrieves transcript text from the given video link and then performs transcript text summary by using Hugging Face Transformers and Pipelining. Authors aimed at shortening the length of the transcript text extracted from input videos. The model built by the authors accepts video link, retrieve transcript and then summarize the transcript text Summarized transcript is the output of the proposed model. Experimental results showed that the time taken to obtain final translated text was comparatively less than other proposed techniques.

## 2.2 Theme 2: Video analysis for sample comparisons

Now a days, extensive research is being carried out to improve analysis of video content as many applications are generating or consuming video data. Neural networks are currently achieving things that no other machine algorithm can achieve [11]. They use similarity measures on CNN features extracted from large and diverse datasets". CNN features based on object detection network using one query image is proposed for image and video face retrieval in [12]. Cosine similarity metric and Manhattan similarity metric are used. Human face recognition in video sequence using improved Viola Jones' face detection algorithm determines the lossless facial area [13]. Usage of fuzzy logic in face detection method for comparison check of query image with the existing images gives 99.3%.

Additionally, there are studies that use YOLO to detect fruits and vegetables. A model to detect and classify three different vegetables was implemented in [14] with the model efficiently detecting and classifying 60-70% of the harvest ready vegetables in an image. In this work, the YOLO network was trained using the Coco dataset and deployed on the Google CoLab platform. The primary benefit of using Google CoLab is that a machine with a GPU is not required. CoLab provides extremely quick processing for the model we create. Fruit and vegetable classification using CNN and image saliency outperforms vast methods [14]. Image saliency helps in selecting main saliency regions as per the saliency map. VGG model was used to training for the purpose of classification of fruits and vegetables. An overall top-1 accuracy of 95.6% was achieved using a dataset of 26 categories of fruits and vegetables. Fruit and vegetable intraclass classification is also crucial, and it may be managed by offering a quality dataset with a variety of fruit and vegetable varieties. Although it takes a lot of effort, label photos can be used to create custom datasets in which particular fruits can be annotated.

In [15], experiments were conducted on TensorFlow platform to evaluate image processing algorithms on different parallel processing units. The results were indicative of potential speed gains using parallelization on GPU that the increase in processing speed can be expected from 3.6 times to 15 times using a GPU. CNN model with a transfer learning approach and different regularization techniques classifies and reviews the online video media based on frame extracted [16]. It rates the appropriateness of Thai television programs as per the announcement of national broadcasting and telecommunications commission (NBTC). Python is widely used in [17] to develop scalable applications that can implement parallelism due to its readability and availability in various academic libraries. Using YouTube data collection, processing time was reduced by 400%, allowing more data collection in less time. These performance improvements reduce processing time and allow for broader exploration of YouTube data. Anaphora resolutions are commonly used as an integral part in [18], they want to approach other multimedia. There are many different versions developed in YOLO, and YOLOv3 is the fastest of them but in [19] found that it's not great on the COCO average AP between 0.5 and 0.95 IOU metric.

The authors of [20] presented a large Convolutional Neural Network (CNN) that is trained using a single step model, You Only Look Once version 3 (YOLOv3) on Google CoLaboratory to process the images within a database and to accurately locate people within the images. After dividing the image into areas, YOLOv3 forecasted the probability and bounding boxes for each area. The projected probabilities were used to weight these bounding boxes, and the model then used the final weights to make its detection. A bespoke dataset of 500 high-resolution photos from Google's Open Images was used for experimentation with this model. After training, the neural network could accurately identify the individuals in the photos, generate the test data, and obtain a mean average precision (mAP) of 78.3% and a final average loss of 0.6. It also emphasized how difficult it was to put such a system into production. First and foremost, the system needs to be able to make the most of the GPU. Experiments also revealed that by integrating a faster and more powerful GPU, the training could be performed much faster than training the model using the built-in Tesla K80 GPU on Google CoLaboratory platform.

Authors in [21] use a triplet loss deep learning network for computing robust features from video and a scalable hashing solution was developed based on Fisher Vector aggregation of the convolutional features from the Triplet loss network. They proposed such a content-based video segmentation identification scheme that can be used irrespective of the codec used. Their simulation results exhibit improvement in terms of large-scale video repository de-duplication compared with other existing methods. Maze [22] video deduplication system at web scale exploits visual and the acoustic features of the videos locate potential matching clips based on the Smith-Waterman algorithm. High denominational vector of visual features is obtained using approximate nearest neighboring search (ANNS based on the quantization-based indexing while acoustic features are obtained using spectrograms-based CNN. The authors claim that the design is scalable as the running cost gradually decreases to approximately 250K standard cores with the dataset growing.

A robust and highly discriminative deduplication feature generated using 32-d fisher vector and a thumbnail feature based on principal component analysis (PCA) in [23]. Fisher vector aggregation applied on Scale-Invariant Feature Transform (SIFT) keypoints extracted from frames with Gaussian Mixture Model (GMM) as the generative model to generate fisher vector. A 12x12 resolution frame used for the thumbnail feature. Results obtained are highly accurate and a query video is processed within a few milliseconds. A GOP-level deduplication system [24] uses an adaptive GOP structure. The proposed technique and the fixed-size GOP structure is compared with the GOP sizes of 8, 10, 12, and 15. The proposed technique achieved a 2.18% PSNR gain. Authors of [25] propose a video abstraction approach using multi-modal video data, segmenting input into scenes and obtaining textual and visual summaries. A hybrid features method improves detection shot performance. A hybrid deep learning model is used for abstractive text summarization. The tests used BBC Learning English and BBC News videos, and a news summary dataset. The performance was assessed using metrics like Rouge for textual summarization. Authors of [26] proposes video abstraction approach based on segmentaion of audio and visual data to summarize the events. Hybrid deep learning model summarizes text, detects shot boundaries and scene boundaries, extracts keyframe efficiently.

Transcripts are an excellent way of comparing the topic, intent and content of a video with other videos. They have a way of categorization based on keywords, summary and transcript itself. There are various APIs that will be helpful for this work. For the text similarity to match cosine similarity using vectors is a great way of finding duplication. Once the similarity results are obtained, frame by frame analysis of video is an additional layer of deduplication. YOLO v3 and ImageAI are appropriate tools for our work.

### 3. MULTI-LEVEL VIDEO ANALYTICS FOR DEDUPLICATION DETECTION

Main objective of our work is to design and develop a model for video deduplication followed by designing and developing a deduplicated video database with video transcripts and topic detections. The video frame-based deduplication algorithm is designed and developed using deep learning multi-core architecture tools such as CUDA and update the developed deduplicated video database with video frame-based deduplication. Then integrate the video transcript retrieval and video frame-based deduplication as multi-level video analytics for achieving higher deduplication accuracy and performance. The video analysis process takes place in multiple phases as shown in Figure. 1, where a user first submits a video file. This video file is used to generate a transcript using AssemblyAI.

Once the Transcription and topic extraction are done and a separate transcript file is generated, these files are compared with the existing topics in the database. If the topics don't exist in the database, the video is uploaded or else the transcripts are compared. If transcript similarity is less than the threshold, then video is uploaded, or else video analysis is done. Threshold is decided based on the category of the video dynamically. The high-level architecture for video deduplication is depicted as class diagram in Figure. 2.

Transcribe\_upload class generates the summary using upload\_summary method, transcript using upload\_transcript method and object using upload\_video method taking input video file through read\_file method to generate database for future. Separate classes are designed for each level. The get\_file\_match class checks the summary of video in summary\_check level. The Text\_matching class checks transcript of video in transcript\_check level. The Video\_object\_detection class detects similar objects in the frame in video\_frame\_check level.

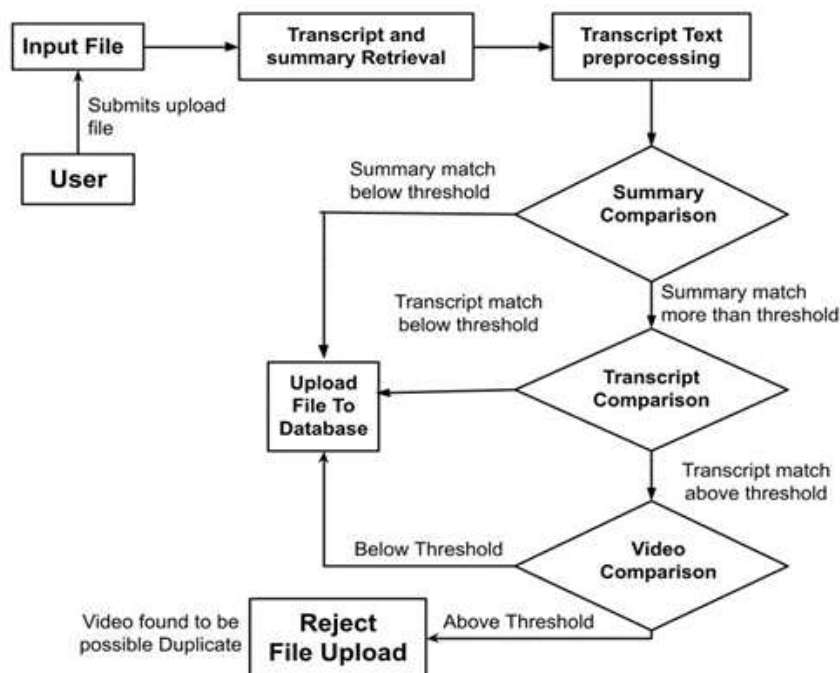


Figure 1. Flowchart describing Multi-level Video Analysis Process

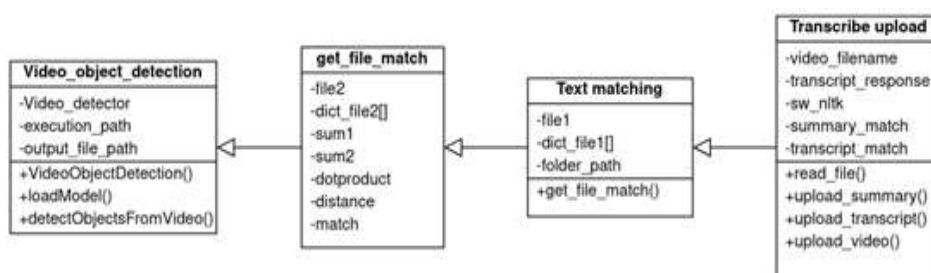


Figure 2. High level Architecture as Class Diagram for Multi-level Video Deduplication



The transcribed upload will get the input video and start the video analysis process. The `get_file_match` class has the document distancing method used where vectors are obtained for transcripts and the cosine similarity using their distance is found out. Using `get_file_match`, the highest similarity among all the files is computed by iterating over it. Finally, the video object detection class containing a `detectObjectsFromVideo` method is used to get the details of objects involved in the video.

### 3.1. Level-1: Summary Check

In the first level of duplicate detection, input videos are summarized and the summarization stored as text documents are used to find the similarity of video content. Document distancing is used in the design and development of text comparison. The content overlap between documents is used to calculate document similarities. Numerous intricate algorithms like Euclidean Distance, Jaccard Similarity, Manhattan Distance, Pearson Correlation Coefficient etc., are available to address this issue. Cosine similarity, a vector-based similarity metric, is another popular document similarity approach. Words are expressed in vector form to represent the text documents in n-dimensional vector space. The angle between two documents' feature vectors—in this case, word frequency vectors—defines the cosine distance between them. A mapping from words to their frequency count is called the word frequency distribution of a manuscript. This distribution is a useful tool for identifying document content overlap. Hence, Cosine similarity is a useful measure of similarity between documents, especially in text mining and information retrieval tasks.

The concept of document distance is the angle between two supplied document vectors, where words contained in the documents are considered as vectors. For instance, if we need to compare similarity of two documents, say A and B be the document vectors of those two documents being compared. By determining the frequency of word occurrences in each document, document vectors are produced for the two documents under comparison.

The content overlap between documents is used to calculate document similarities.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{||A|| \ ||B||} \quad (1)$$

As per equation (1), Cosine similarity metric measures the cosine of the angle between two n-dimensional vectors projected in a multi-dimensional space and Similarity between the files is computed as cosine of (dot product of both files/ product of mod of both files). This value ranges from 0 to 1. The Cosine similarity of two documents will range from 0 to 1. If the Cosine similarity score is 1, it means two vectors have the same orientation. The value closer to 0 indicates that the two documents have less similarity.

### 3.2. Level-2: Transcripts from AssemblyAI

This level tries to check for the similarity of videos by comparing their transcripts. The transcripts are retrieved using a text-to-speech service provider called AssemblyAI as shown in Figure 3. This AssemblyAI service provides better accuracy than its counterparts namely Google Cloud Speech-to-Text, and AWS Transcribe. Topics are detected from the transcripts which are used for duplicate detection. Transcripts and topics thus generated are stored so that future duplicate checks can be quickly performed. We have stored transcripts and the topic detection files in separate directories of their own so that it helps in easy navigation instead of storing all files in a single directory. We planned to extend this work to classify videos into different categories based on the detected topics so that newer videos can be added to a specific category and indexing also becomes quicker and easy.



Figure 3. Input video transcript generation using AssemblyAI

### 3.3 Level-3: Uploading video files to video database

We are using Python DriveSDK to perform our uploads to our video database after the video passes the three levels of similarity check namely `summary_check`, `transcript_check` and `video_frame_check`. If the video file crosses the threshold at any level, we will alert the user about the same and the video upload is halted with an error message stating that the video is similar to some other video already available in our database. We are using ImageAI for our video duplicate detections where we divide each second of the video into frames and detect objects in each frame to store it in a file for further comparisons. This work uses YOLOv3 deep learning algorithm for object detection in the video. YOLOv3 is considered to be a single neural network architecture that is proven to be more accurate and faster in real-time object detection. Since, this level is computationally intensive, we could get object detection done quickly in real-time using YOLOv3.

### 3.4 Integration of the components of the system

We have developed a simple web-based user interface in our application for the users to upload the videos, which is not included in the paper due to restriction on page limits. The user first tries to upload the video file to the system. The file is then sent to the server for validation. Once the transcripts and topic are identified, the match percentages of the topic are displayed to the user if there is a match else the file gets uploaded. If the match crosses the topic match threshold, then the user is prompted to accept to proceed to level 2 and transcript match is carried out and the percentage match of transcript is also displayed to the user. If the video passes level 2 validation, the file is uploaded to the database else level 3 validation of video frame matching is carried out again by prompting the user to give his/her consent to enter level 3 check. If passed, then the file is uploaded to the database, else the user is alerted of a possible match with another file and the file is discarded. Such a multi-level duplicate detection system helps users quickly identify unique videos during their search operation on various topics and save a lot time avoiding browsing similar videos. Figure 4 depicts the time required to detect duplicates and the accuracy of duplicate detection in the form of a cone for the 3 levels of our proposed approach.

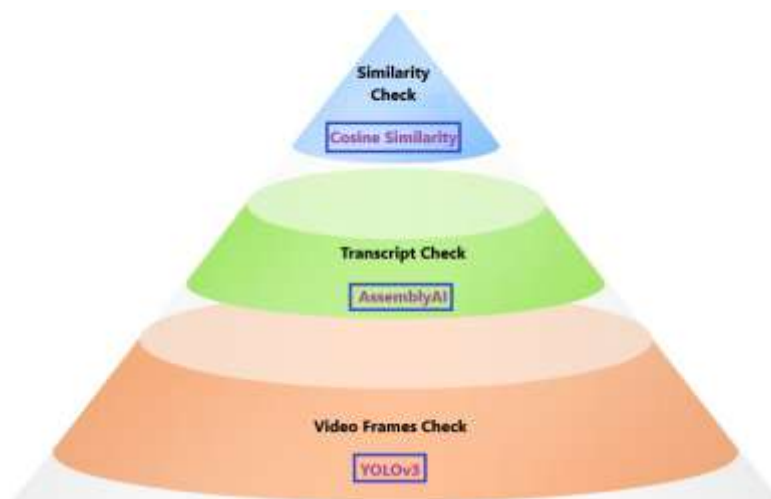


Figure. 4. Overall system of proposed approach

## 4. EXPERIMENTAL RESULTS

Experiments were conducted to check transcription accuracy between AssemblyAI, Google Cloud Speech-to-Text, and AWS Transcribe indicates AssemblyAI gives better accuracy for different input videos of different media type, which is shown in Table 1. We have also conducted experiments to find the deduplication accuracy and time taken as we take the input video through the three levels of our proposed system. The performance of the three levels of validations of video files of different durations is observed as shown in Figure 5. The depicted graphs suggest tests under a GPU environment with a video file to be converted to 20 frames per second and speed mode set to flash from ImageAI's `forFrame` function. From the graph, it is evident that, we can perform video deduplication in phases of transcripts detection for better efficiency and highly reduced time. Level 1 and Level 2 checks take very less time for deduplication checks in cloud as

compared to level 3. If the video transcript is same, then a lot of time is saved. We could conclude from our experiments that depending on the extent to which the duplicate detection is required, we can tune our application to stop the detection at a specific level and inform the users with a quick turnaround time. For instance, if housewives are searching for recipe videos, we can check for duplicates using summary level and quickly ask them to watch unique videos. If students are searching educational videos on a certain topic, we could extend to level 2 and determine duplicates involving transcript comparison as well. If the use case is storing videos on cloud storage, we can perform all 3 levels of duplication check and then accept only unique videos to be stored to achieve storage optimization.

Table 1. Transcription Accuracy of Various Transcript Generation Platforms

Video Media type	AssemblyAI	AWS	Google video model
Virtual demo	92.11%	86.81%	86.93%
Tutorial	97.74%	95.73%	96.47%
Classroom	89.92%	84.82%	85.08%
webinar	89.83%	82.45%	85.23%
Fitness class	87.67%	83.48%	84.45%
Document series	90.15%	85.11%	85.00%
Podcast	93.76%	90.34%	89.99%
News	96.11%	94.02%	93.48%
Sportcast	87.95%	86.02%	80.08%
Phonecall	91.59%	85.91%	94.20%

We have done a performance comparison with [6]. Time Complexity of building Bag of Words and TF-IDF matrices, then forming clusters and running LDA algorithm to identify the closest centroid for the new video is mostly quadratic or cubic. In our proposed method, the first two levels operate quickly to determine the possibility of a duplicate with an acceptable accuracy at sub-quadratic complexity. If the threshold of first two levels indicate the possibility of a duplicate, we can run third level of deep learning to determine duplicate with high precision and accuracy. Hence, we found that our method is computationally much faster and stable. Accuracy is also high compared to [6]. The state-of-the-art approaches is compared with the proposed work in Table 2 wherein the comparison parameters include summary extraction, similarity ranking, multi-level threshold based, and block level deduplication. It clearly indicates that the proposed methods support all parameters unlike the state-of-the-art approaches.

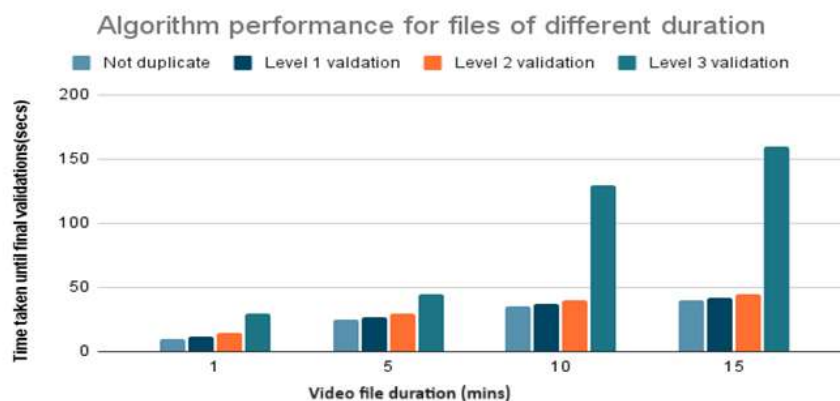


Figure. 5. Performance Analysis



Table 2. Comparison of Features supported with the State-of-the-art Approaches

Methods	Summary Extraction	Similarity ranking	Multi-level threshold based	Block level deduplication
[6]	no	yes	no	no
[8]	yes	yes	no	no
[10]	yes	no	no	no
Proposed method	yes	yes	yes	yes

## 5. CONCLUSION











We designed and built a web application with the intention of checking for possible video duplicates, before uploading them to storage. Videos are validated by matching all the criteria and parameters set in the algorithm. Multilevel validation was designed to enable checking at levels like summary, transcripts, and video frames. If a match in similarity for more than the threshold is found, then we conclude that it is a possible duplicate. If it passes all the levels of validations, then video was allowed to be uploaded. Our proposed system has a huge utilization from the storage perspective since we are dealing with transcripts rather than analysing the videos or audios themselves as those methods take a lot of time as well. The uniqueness will immensely optimize the database since we won't have redundancy when working on the same topics as the videos with higher percentage of plagiarism are being rejected and the user is notified about the situation. Many use cases like searching educational videos, recipe videos etc., find our solution a very useful way of filtering redundant videos and watch only the videos of higher relevance and thus customers will not be exhausted while searching for online videos. It has a great scope in the future since we not only just compare and perform deduplication based on syntax, rather we can perform deduplication by taking semantic parameters into consideration as well. We can also look forward to lowering the time constraints even on larger duration of videos by making the retrieval method to be even more robust and efficient.

## REFERENCES

- [1] G., A. K., & P., S. C. (2020), "An extensive research survey on data integrity and deduplication towards privacy in cloud storage", *International Journal of Electrical and Computer Engineering*, 10(2), 2011. <https://doi.org/10.11591/ijece.v10i2.pp2011-2022>.
- [2] G. D. Bianco, R. Galante, M. A. Gonçalves, S. Canuto and C. A. Heuser, "A Practical and Effective Sampling Selection Strategy for Large Scale Deduplication," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2305-2319, 1 Sept. 2015, doi: 10.1109/TKDE.2015.2416734.
- [3] A. Miri and F. Rashid, "Secure Textual Data Deduplication Scheme Based on Data Encoding and Compression," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 0207-0211, doi: 10.1109/IEMCON.2019.8936222.
- [4] Li, Y. and Xia, K., 2016, August, "Fast video deduplication via locality sensitive hashing with similarity ranking", *International Conference on Internet Multimedia Computing and Service* (pp. 94-98). <https://doi.org/10.1145/3007669.3007725>
- [5] G. Madhubala, R. Priyadharshini, P. Ranjitham and S. Baskaran, "Nature-Inspired enhanced data deduplication for efficient cloud storage," 2014 International Conference on Recent Trends in Information Technology, 2014, pp. 1-6, doi: 10.1109/ICRTIT.2014.6996211.
- [6] G. Turcu, M. C. Mihaescu, S. Heras, J. Palanca and V. Julián, "Video Transcript Indexing and Retrieval Procedure," 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2019, pp. 1-6, doi: 10.23919/SOFTCOM.2019.8903790.
- [7] H. Shukla and M. Kakkar, "Keyword extraction from Educational Video transcripts using NLP techniques," 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), 2016, pp. 105-108, doi: 10.1109/CONFLUENCE.2016.7508096.
- [8] K. Kulkarni and R. Padaki, "Video Based Transcript Summarizer for Online Courses using Natural Language Processing," 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2021, pp. 1-5, doi: 10.1109/CSITSS54238.2021.9683609.

- [9] J. Takeshita, R. Karl and T. Jung, "Secure Single-Server Nearly-Identical Image Deduplication," 2020 29th International Conference on Computer Communications and Networks (ICCCN), 2020, pp. 1-6, doi: 10.1109/ICCCN49398.2020.9209728.
- [10] A. N. S. S. Vybhavi, L. V. Saroja, J. Duvvuru and J. Bayana, "Video Transcript Summarizer," 2022 International Mobile and Embedded Technology Conference (MECON), 2022, pp. 461-465, doi: 10.1109/MECON53876.2022.9751991.
- [11] Hachchane, I., Badri, A., Sahel, A., & Ruichek, Y. (2020). Large-scale image-to-video face retrieval with convolutional neural network features. *IAES International Journal of Artificial Intelligence* , 9 (1), 40~45, DOI: 10.11591/ijai.v9.i1.pp40-45
- [12] Hachchane, I., Badri, A., Sahel, A., Elmourabit, I., & Ruichek, Y. (2022). Image and video face retrieval with query image using convolutional neural network features. *IAES International Journal of Artificial Intelligence* , 11 (1), 102-109, DOI: 10.11591/ijai.v11.i1.pp102-109
- [13] Younis, AN, & Ramo, FM (2023). Developing Viola Jones' algorithm for detecting and tracking a human face in video file, *IAES International Journal of Artificial Intelligence*, 12 (4), 1603~1610, DOI: 10.11591/ijai.v12.i4.pp1603-1610.
- [14] S. C, N. Manasa, V. Sharma and N. K. A. A., "Vegetable Classification Using You Only Look Once Algorithm," 2019 International Conference on Cutting-edge Technologies in Engineering (Icon-CuTE), Uttar Pradesh, India, 2019, pp. 101-107, doi: 10.1109/Icon CuTE47290.2019.8991457.
- [15] D. Demirović, E. Skejić and A. Šerifović-Trbalić, "Performance of Some Image Processing Algorithms in Tensorflow," 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), Maribor, Slovenia, 2018, pp. 1-4, doi: 10.1109/IWSSIP.2018.8439714.
- [16] Tanantong, T., & Yongwattana, P. (2023). A convolutional neural network framework for classifying inappropriate online video contents. *IAES International Journal of Artificial Intelligence*, 12 (1), 124. DOI: 10.11591/ijai.v12.i1.pp124-136
- [17] J. Kready, S. A. Shimray, M. N. Hussain and N. Agarwal, "YouTube Data Collection Using Parallel Processing," 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2020, pp. 1119-1122, doi: 10.1109/IPDPSW50202.2020.00185.
- [18] A. K. Elmagarmid, P. G. Ipeirotis and V. S. Verykios, "Duplicate Record Detection: A Survey," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 1, pp. 1-16, Jan. 2007, doi: 10.1109/TKDE.2007.250581.
- [19] P. Adarsh, P. Rath and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 687-694,doi: 10.1109/ICACCS48705.2020.9074315.
- [20] N. I. Hassan, N. M. Tahir, F. H. K. Zaman and H. Hashim, "People Detection System Using YOLOv3 Algorithm," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2020, pp. 131-136, doi:10.1109/ICCSCE50387.2020.9204925.
- [21] Jia, Wei, Li Li, Zhu Li, Shuai Zhao, and Shan Liu. "Scalable hash from triplet loss feature aggregation for video de-duplication." *Journal of Visual Communication and Image Representation* 72 (2020): 102908. <https://doi.org/10.1016/j.jvcir.2020.102908>
- [22] An Qin, Mengbai Xiao, Ben Huang, and Xiaodong Zhang. 2022, Maze: A Cost-Efficient Video Deduplication System at Web-scale . In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*, October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3503161.3548145>
- [23] Henry, Chris, Rijun Liao, Ruiyuan Lin, Zhebin Zhang, Hongyu Sun, and Zhu Li. "Fast and Robust Video Deduplication." In *Proceedings of the 2nd Mile-High Video Conference*, pp. 160-160. 2023. <https://doi.org/10.1145/3588444.3591050>
- [24] Sujatha, G., A. Devipriya, D. Brindha, and G. Premalatha, "An Efficient Cloud Storage Model for GOP-Level Video Deduplication using Adaptive GOP Structure." *Cybernetics and Systems* (2023): 1-26 <https://doi.org/10.1080/01969722.2023.2176665>
- [25] Abdulsahib, M. G., & Abdulmunim, M. E. (2023), Multimodal video abstraction into a static document using deep learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(3), 2752-2760. <https://doi.org/10.11591/ijece.v13i3.pp2752-2760>
- [26] Abdulsahib, MG, & Abdulmunim, ME (2023). Multimodal video abstraction into a static document using deep learning. *International Journal of Electrical and Computer Engineering (IJECE)* , 13 (3), 2752-2760. DOI: 10.11591/ijece.v13i3.pp2752-2760

## BIOGRAPHIES OF AUTHORS

	<p><b>Aparna Ramalingaiah</b>     is currently working as Subject Matter Expert at CloudThat Technologies, Bangalore. She received her M. Tech in CSE from NITK Surathkal. She is currently a research candidate at Computer Science and Engineering, MSRIT, Bangalore research centre of VTU, Belgaum. Her research interests secure cloud storage. She is a professional member of IEEE and CSI. She can be contacted at <a href="mailto:aparnasurya.js@gmail.com">aparnasurya.js@gmail.com</a></p>
	<p><b>Shilpa Chaudhari</b>     is currently working as a Professor, CSE, MSRIT, Bangalore. She has completed her Ph.D. in Electronics Engineering from the VTU, Belgaum at the REVA ITM, Bangalore. She has been a technology educator and corporate trainer since 1999. Her research contributions on cloud/network security, and wireless networks exists in various national and international indexed journals and conferences. She is a technical reviewer for Springer (Wireless Personal Communication and Telecommunication Systems). She is a professional member of the CSI and IEEE. She can be contacted at <a href="mailto:shilpasc29@msrit.edu">shilpasc29@msrit.edu</a></p>