Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 6, No. 2, December 2020, pp. 47-57
ISSN: 2338-3070, DOI: 10.26555/jiteki.v6i2.18978

A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work

Md Shofiqul Islam[1,2], Shanjida Sultana[3], Uttam kumar Roy[4], Jubayer Al Mahmud[5]

[1]Faculty of Computing, Universiti Malaysia Pahang, 26300, Kuantan, Pahang, Malaysia
[2]IBM Centre of Excellence (Universiti Malaysia Pahang), Cybercentre, Pahang Technology Park, 26300 Kuantan, Pahang, Malaysia.
[3]Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh.
[4]Assistant Programmer at Bangladesh Bank-The Central Bank of Bangladesh, Head Office, Motijheel, Dhaka 1000.
[5]Senior Software Engineer at Charja Solutions Limited,129-Kha/1, Elephant Road, New Market, Dhaka-1205.

ARTICLE INFO

Article history's:

Received 10 December 2020,
Revised 12 December 2020,
Accepted 23 December 2020.

Keywords:

Video Classification,
Machine learning,
Deep learning,
Video,
Video classification.

ABSTRACT

In recent years, there has been a rapid development in web users and sufficient bandwidth. Internet connectivity, which is so low cost, makes the sharing of information (text, audio and videos) more common and faster. This video content needs to be analyzed for prediction it class in different purpose for the users. Many machines learning approach has been developed for the classification of video to save people time and energy. There are a lot existing review papers on video classification, but they have some limitations such as limitation of analysis, badly structured, not mention research gaps or findings, not clearly describe advantages, disadvantages, and future work. But our review paper almost overcomes these limitations. This study attempts to review existing video-classification procedures and to examine the existing methods of video-classification comparatively and critically and to recommend the most effective and productive process. First of all, our analysis examines the classification of videos with taxonomical details, latest application, process and datasets information. Secondly, overall inconvenience, difficulties, shortcomings and potential work, data, performance measurements with the related recent relation in science, deep learning and the model of machine learning. Study on video classification systems using their tools, benefits, drawbacks, as well as other features to compare the techniques they have used also constitutes a key task of this review. Lastly, we also present a quick summary table based on selected features. In terms of precision and independence extraction functions, the RNN(Recurrent Neural Network), CNN(Convolutional Neural Network) and combination approach performs better than the CNN dependent method.

Md Shofiqul Islam,
Faculty of Computing, Universiti Malaysia Pekan, 26600, Kuantan, Pahang, Malaysia.
Email: shafiqcseiu07@gmail.com

1. INTRODUCTION

The internet is currently commonly used by the people worldwide. Social media have an essential role to play of content distribution (audio, video, text, image) sharing [1]. About the same period, they also share their emotions in social media about a certain aspect so that those users can quickly find out exactly what is happening and with this reason, user views are used to estimate the public opinion on certain issues. However, if consumer employ a person to evaluate the views of people through multitudes of content it is very difficult and time consuming. In order to evaluate public attitudes, the researchers present a machine learning approach to data mining. Video classification is part of mining which analyzes text through natural language processing, video by machine linguistics in order to find views of people by gathering and analyzing social and other resources of subjective knowledge. Deep learning methodology is more reliable and effective than other approaches.

This paper reviews different approaches for video classification. There are a number of review and survey paper in video classification. Some recent review papers are listed here with their works and limitations. A nice review on deep learning based on video classification and captioning task [2]. This review is on only deep learning-based approach for video classification with good description on deep model, data and feature extraction tools but does not able to mention research gaps, advantages and performance. A simple review on video classification technique proposed in 2019 by Q. Ren [3]. This is a very simple review because it just presents video classification approach with a short description. This method does not describe method, dataset, performance metrics, research gaps, limitations of existing methods. In 2020, A very simple review has given by Anusya for video classification [4]. This review simply gives introduction and state some recent existing method in video classification for tagging. There are many lacking this review like has limited information, does not provide information about the research limitations, used tools in existing method. A recent review on video classification in 2020 by Rani [5]. This review states video classification approach and summary-based description of recent works. The limitations of this work are short description, not properly analyzed on recent task to find research output, gaps and finding. Another systematic, recent, and good review [6] on live sport video classification has done by s in 2020. This review properly presents recent works in live sport video classification with tools, feature extraction, video interaction features etc. This is a longer review and has no summarized table for research gaps, finding, advantages and disadvantages of existing methods.

The explanation above indicates that most of reviewers have historically reviewed existing research. Current survey articles usually describe the techniques and related studies with a basic introduction of method. In the traditional survey paper for video research, we usually see similar classification trends of the comparative research or associated study. But with numerous forms of critical research our research study paper is unique. The following can be mentioned as our key contribution to this review paper:

1. Discussion by concerning the taxonomy of video classification, recent usage, methods and databases.
2. State general disadvantages, difficulties, challenges including future tasks, data, Performance measures with relevant recent references to science, deep learning as well as a model of machine learning.
3. Comparison of the techniques used objectively analyzes video classification systems utilizing their tools, benefits, disadvantages as well as other features. Show a quick overview table based on the feature selected.

The rest of this paper is arranged as follows. Section 1 gives background knowledge of the research for of video classification technique. Section 2 states critical analysis on recent research with their advantages, disadvantages, features based quick summary, quick summary, drawback, challenges, limitations, and future works followed by the conclusions. Overall methodology of this review is shown in Figure 1.

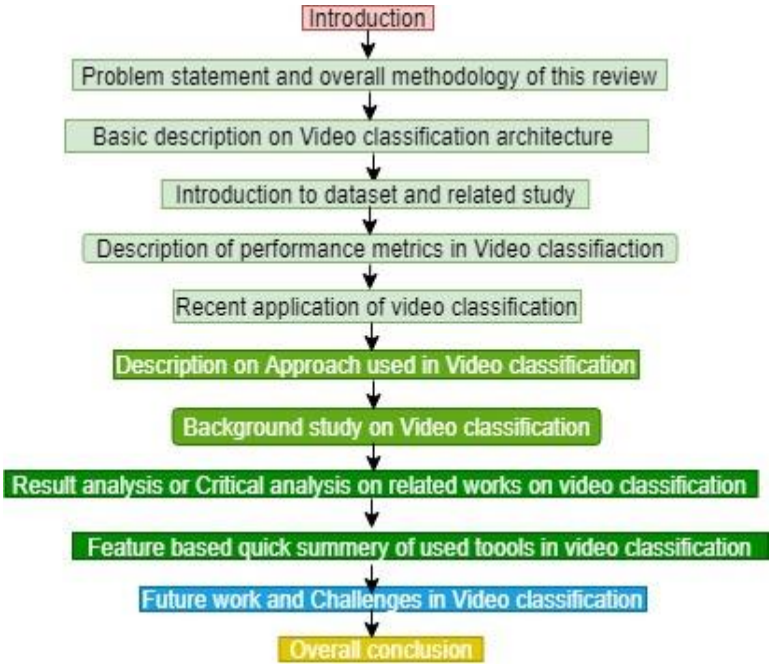
Introduction
Problem statement and overall methodology of this review
Basic description on Video classification architecture
Introduction to dataset and related study
Description of performance metrics in Video classifiaction
Recent application of video classification
Description on Approach used in Video classification
Background study on Video classification
Result analysis or Critical analysis on related works on video classification
Feature based quick summery of used toools in video classification
Future work and Challenges in Video classification
Overall conclusion

Fig. 1. Overall methodology of this review

1.2 Video classification architecture

Video classification technique have some basic steps and those steps should be done in sequential order. Figure 2 shows basic steps in video classification. First step is data collection, then preprocessing of data for feature extraction, then method execution for feature matching and classification. In data collection section data can be in form of video, text, speech, and image on the review of video. Preprocessing part deals an important task in video processing, its play the role of video conversion, segmentation and analysis for further feature or information extraction. Feature extraction, feature matching and feature classification with algorithm is the main part of the video classification process.

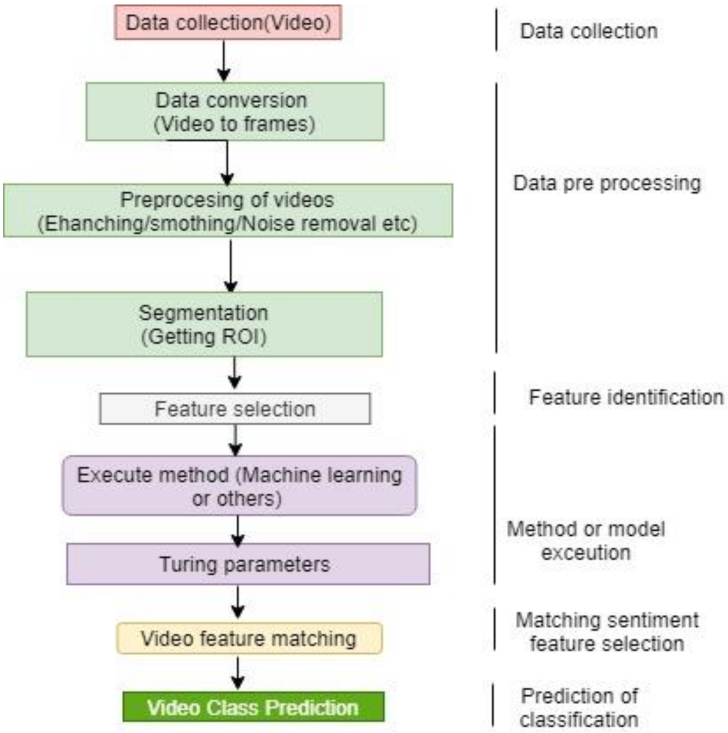
Data collection(Video)
Data conversion (Video to frames)
Preprocecing of videos (Ehanching/smothing/Noise removal etc)
Segmentation (Getting ROI)
Feature selection
Execute method (Machine learning or others)
Turing parameters
Video feature matching
Video Class Prediction
Data collection
Data pre processing
Feature identification
Method or model exceution
Matching sentiment feature selection
Prediction of classification

Fig. 2. Basic steps in video classification

1.3 Data set used in video classification

Many science and research organizations have invested a lot of time gathering and marking video data sets in media-related fields of research. YouTube-8M, HCF-50, HCF-101, HMDB51 and many more are the widely used datasets. The small sets of data include Weizmann, KTH, and Hollywood, with smaller, but very well-labeled overall quantity and video forms. And over 50 images, such as UCF101, Thumbos'14 and HMDB51, are included in medium set info. The big data collection such as the YouTube 8M (Google collects), Sports-1M, ActivityNet, Kinetics and others. More detailed information is summarized in Table 1. Here Weizmann and KTH dataset are static, but all other dataset are dynamic.

Table 1. Summary of Video task dataset

Name of the dataset	Year	Number of video categories	Amount of video
Weizmann	2005	9	81
KTH	2004	6	2361
Hollywood	2008	8	430
UCF50	2012	50	6676
HMDB51	2013	51	6474
UCF101	2012	101	13320
Thumos’14	2014	101	18394
Youtube-8M	2016	4800	8264650
Sports-1M	2014	87	1133158
ActivityNet	2015	203	27901
Kinetics	2017	400	306245

1.4 Performance metrics in Video classification

Throughout this section we describe most common video classification efficiency metrics. Using performance metrics demonstrate how well a dataset approach works. In the scope of the video classification, there are several performance analysis steps called Precession (Precision measures are conducting positive meaning determination), Recall (Precision tests are percentage tests for productive detection of positive result of the classifier). However, a Table 2 presents some of the performance measures used for the assessment of research on video classification from the latest work on video classification. Here Table 2 is given for related research with performance metrics.

Table 2. List of Performance Metrics used in video classification

Performance metrics	Reference and Year
Accuracy	[7]-2020, [8]-2020, [9]-2020
Precession	[8] [9]-2020
Recall	[8] [9]-2020
F1 Score	[8] [9]-2020
Micro F1	[10] [11]-2020
K-Fold:3,5,10-fold	[12]-2019

1.5 Applications of video classification

There are many numbers of applications of video classification. Here, I have mentioned some of them with recent work reference in Table 3**.** For the application of video in firewall task, user must sure the specification of the types of videos that allowed to login. Live streaming prediction, action recognition, violence detection, character recognition, traffic control, social media analysis, emotion analysis, movie review, event prediction is also the application of video analysis.

Table 3. Summary of Application of Video classification

Name of Applications of Video classification	Authors and Years
Violence detection from video of real time game	[7] - 2020
Video Scene classification	[13] - 2020
Event prediction	[9] - 2020
Animation movie video classification	[8] - 2020
Sport player action recognition	[14] - 2020
Twitter video classification	[15] - 2020
Stock Market prediction	[16] - 2020
Movie video trailer classification	[17] - 2020

2. APPROACH USED IN VIDEO CLASSIFICATION

Because many videos are present in the real world, an effective way to classify those videos is important. The main aim of the video classification method is to classify whether the video is used for athletics, films, amusing videos, school, etc. There are three different ways of classifying video called audio, video and text. Apart from these three methods we can also use hybrid approach (using one more method combined approach) to categories the videos. Figure 3 shows Taxonomy of Video Classification Approaches.

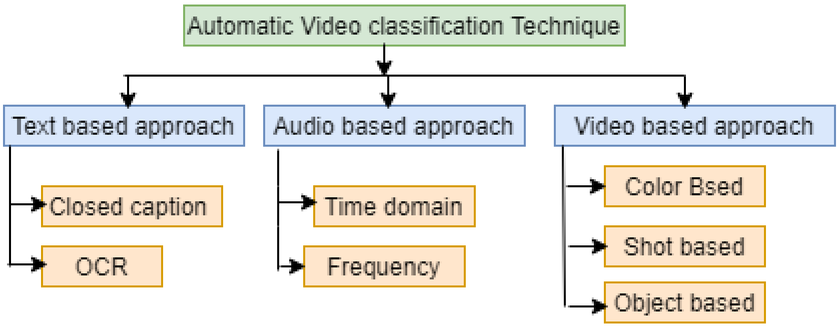

Fig. 3. Taxonomy of Video Classification Approaches

2.1 Text based Approach

We generate video texts and evaluate them for classification in this process. Might be a visible text or text from speech extracted. The text on the computer is derived in the first category. For example, the playing board, number in the player's jersey, subtitles on the display, etc. The text of that sort could be extracted with OCR [18][19]. The text is derived from voice through voice recognition throughout the second category. This technique is used primarily for subtitles and closed subtitles. Closed subtitles are also used for other sound forms such as pet sound or songs. In order to make it clear, subtitles are put on video.

2.2 Audio Based Approach

This technique is being used more that text based on analysis which is ascribed to the fact that audio processing takes fewer time and energy. Audio as well as its characteristics need less space to be stored than video and text. For audio processing, a single signal is sampled, and some characteristics are retrieved for inspection of each sample. In certain instances, these samples may be overlapped. The time domain as well as the frequency domain could be used as the functions.

2.3 Video Based Approach

Most scholars used the approach as most knowledge dependent on the vision is interpreted by human beings. Some authors have also where necessary coupled these visual aspects including audio and text. Visual capability is primarily derived from image sequences or video files. Video's basic structure is like a combination of pictures is a fundamental part of video. Video may also be named as a collection of frames. Visual characteristics are typically dependent on color, motion or shot time. These features must convey lighting, movement, background or video speed detail.

2.4 Comparison among video classification approach

From the description we see that each approach works has some own way of working and success outcomes, based on the suitability to application of existing approach we present a comparison table below. The Table 4 explains the advantages and disadvantages of each method in detail.

Table 4. Comparison among video classification approach.

Classification Method	Feature list	Advantages	Disadvantages
Text Based Video Classification	Optical Character Recognition Closed Captions features Speech Recognition	Higher accuracy Higher dimensionality	Expensive in computation Higher error rate Works of text-based format only.
Audio Based Video Classification	Physical Features as well as Perceptual Features	Short length, Computationally cheaper	Difficult to differentiate multiple and similar sounds
Video Based Video Classification	Color Based Features, Shot-Based Features and Object-Based Features	Easily implemented. Not in converted representation.	Large size Computation is expensive Pre-processing is needed Identification of shots, track is difficult

2.5 Background study on video classification method

Some commonly used methods are supervised, such as SVM, CNN, and also unregulated. There also also a variety of solutions throughout the video classification (LSTM, GRU etc.). This section demonstrates the most widely used method of video classification including their working technique, application advantages and disadvantages. A way to identify video using Naïve Bayes and dictionary for the video classification [8]. If the statement of independent predictors is valid, a classifier from Naive Bayes functions works better than other models. Naive Bayes' primary imitation is autonomous predictors' inference. SVM is also a method of detection that is commonly in video classification [20]. Another approach is used to identify hateful speech from the world wide web of video classification [14]. Another job to classify Twitter videos [9] was to work with SVM tool to classify the pilot and to weight production in order to improve classification precision. SVM method does not performs well for noisy data and when target class are overlapped.

K means are used in various ways for video labeling. Another video classification task performed recently by Peng [13]. This approach is used to retrieve the visual features from video and share resources of visual features by segmenting the video. Original clustering levels of labialized video samples are enhanced with the standard k-means aggregation algorithm. K-means computation is faster than hierarchical clusters most of the time when we hold k smalls. The drawbacks are that K-value is hard to estimate, the K-proximate neighbor (KNN) method is simple and easy to apply for classification as well as extraction purposes, HMM (Hidden Markov Model) is used. A new approach to the study of child face speech for the R-CNN and HMM method of real-time video surveillance [21]. HMM approach gains Solid theoretical base, fast learning algorithms through raw sequence information may take place explicitly and different-length inputs are the simplest generalization for sequence data.

HMM's drawbacks include a. HMMs also have a multitude of unstructured criteria and cannot rely on hidden states to rely on them. A paper shows that 3D CNN is best suited to the classification of the video, and also to analyze its success with the title of an effective deep pipeline template-based architectures to accelerate the whole 2-D and 3-D CNNs on FPGA [22]. Action recognition was used with 3D Deep convolutional Neural Networks [23]. 3D convolutions combine spatial information as well as motion information successfully. Long-term model RNNs maps time dynamics explicitly to variable length video frames. To accomplish this the RNN produces networks with loops that cause knowledge to survive [24]. The neural network will use this loop form to record the input series. The RNN functions like this. RNN assists from the previous feedback anywhere we need meaning.

RNN has two types of LSTM, as well as the other type GRU. RNN with neurons in long short-term memory (LSTM) is being trained in sports video sequences with SIFT features [25]. Baccouche work has been very much respected for its consistency. The function extraction is automatic with the creation of deep learning techniques and architecture. Through back propagation, RNN could be optimized. A new piece of videos with higher fidelity, using the 2D Gated Bidirectional of Neural Networks for the identification of aggression at the end of the day. Kyunghyun developed Gated Return Units (GRUs) as an existing Neural Networking Feature (CNN) [26]. Deep literacy is more reliable and efficient than other techniques [27]. The approach to learning is more effective. Table 5 offers a detailed comparison of the deep-learning video classification system.

Table 5. Overall comparison among deep learning-based method for video classification.

Model	Advantage	Drawback
2D CNN	Can capture spatial feature.	Can capture spatial feature from video data
3D CNN	Can capture both spatial and temporal feature.	Expensive for its 3D structure for working
RNN	Can capture both spatial and temporal feature from sequence data.	Has short memory ability, could not be able in real situation.
LSTM	Can capture both spatial and temporal feature from sequence data.	Gradient explosion, Takes more training time.
GRU	Can capture both spatial and temporal feature from sequence data in a faster time	The reset gate of GRU controls if the previous hidden state needs to be ignored.

2.6 Result discussion with critical analysis on related works on Video Classification

Many deep learning approaches may use a large-scale data collection and working capability resolve the limitation of current or usable methods with increased precision and accurateness. This segment presents analyses of recent progress on the classification of videos. Table 6 contains analytical style columns having data, methods, model, type, advantage, and disadvantage of most recent video classification methods in 2020. Throughout the field of video classification research, there are several similar works. Traditional methods are subdivided into following types: traditional machine learning and deep learning. The SVM is the classification tools generally used in video classification, a system that uses Naïve Bayes and Dictionary [8]. K means used in various ways for video classification. A video classification work performed recently by Peng [14]. Latest analysis focused on the you tube video content Classification with Random Forest algorithm [28]. End-to-End Information Diagrams video classification and K next to neighbor classification [29].

A new approach to the study of child face speech for the R-CNN and HMM method of real-time video surveillance [21]. A deep learning structure automatically operates in order to learn then represent data across different processing layers through specifically classifying specific input data or vine frames [30]. Unlike a typical designed architectural design, no identifiers or practical extractors are required. For example, in deep learning model, local characteristics are immediately learned from an image rather than through a whole picture [31]. Deep learning techniques that are able to identify high-level or complicated behavior that attract enormous research [32]. The widespread examples of profound learning models are the CNN, repetitive neural network (RNN) as well as a long-term memory (LSTM). The use of deep learning to video data analysis was motivated by an outstanding success with a high accuracy of deep learning method in such a visual work. At first, CNN operates separately for data extraction from still pictures [23]. Although in video streams 2D-CNN cannot retrieve temporary information. For the massive video classification, the paper [33] uses coevolutionary neural networks (CNN) and reveals that the slow melting system performs better than the usually early fusion model [34] evaluated CNN with LSTM-RNN and identified the potential for a stronger creation of Recurrent Convolutional Neural Networks. In [35] the CNN two-stream structure is being used, one for spatial and another one for temporal functionality.

The [36] study uses description and CNN in activities to recognize activity and behavior. The grade bundling codes period details by grouping video frames in sequential order. A Bi-level optimization approach be used for the learning algorithm by convolutions of neural networks. The CNN extractor and batch standardization LSTM function extractor could also be used to optimize performance [37]. Non-linear context gating was introduced in [38] to model interdependencies between features and it was the used to classify videos. 3D-CNN then has designed to retrieve both spatial and temporal knowledge from video frames in able to fix this problem for 2D CNN [39]. The behavior identification was followed by this RNN. The RNN-based approach efficiently records time knowledge based from both current and past observations [40]. This forecast is based on current measurements. However, RNN architecture does indeed have a short-term memory that cannot be extended in the real-world case. The LSTM model was suggested to mitigate this problem. This model will extract time from sequential video files. The LSTM model has a memory device which determines when secret states are to be remembered and forgotten [41]. The LSTM model is primarily used in computer vision applications including action recognition, owing to its excellence. Table 6 given represent some recent approaches for video classification.

Table 6. Video classification recent approach

Author and Year	Type	Task	Lexicon or dataset	Performance and Data domain	Approach	Video analysis Features	Type of data	Advantages or findings	Disadvantages or limitations
[17] - 2020	Deep learning	Video movie analysis	LMTD-9, MMTF-14, and ML-25M	The combined accuracy of ILDNet for LMTD-9, MMTF-14, and ML-25M 86.15%, 83.06%, 85.3% respectively	Bi-LSTM, and LSTM	Can acquire discriminative and comprehensive higher level features with a unique combination of Inception V4, Bi-LSTM, and also LSTM layers.	Video	Can recognize six types of emotion from video trailers	Performance limited predefined and compared EmoGDB dataset content.
[7]- 2020	Deep learning	Video violence detection	Hockey game dataset , Violent Flow dataset and Video Real Life Violence Situations dataset	Accuracy:98%	2D CNN, BiGRU	A simple end-to-end deep learning method to find violence in video sequences with CNN and RNN.	Video	Can detect violence in video sequences.	Works with small dataset, Higher training time.
[9]- 2020	Machine learning	Video analysis	Chilean earthquake and Catalan independence referendum	Dataset 1: Highest accuracy with SVM 0.812±0.067, BAF TAN got highest Precession 0.898±0.003 F1 Score with SVM of 0.899±0.042, BFTAN got highest Recall 0.898±0.003. Dataset-2 :highest : RF got highest precession 0.922±0.002, RF highest accuracy 0.858±0.008, Highest Recall 0.985±0.008 by SVM, RF got highest F1 Score 0.908±0.00.	Bayesian network classifiers, Uses TAN and BF TAN	BOW, Term document matrix (TDM) for tweet to vector representation	Video	Provides good result by using Bayes classifier with Directed acyclic graph (DAG) for twitter comments emotion.	Biased to hashtag, Heavily time and event dependent
[13] - 2020	Hybrid	Video analysis	Haberman sub-library data, German Credit Data sub-library, Heart sub-library	Accuracy: Haberman sub-library 72%, German Credit Data sub-library 75%, Heart sub-library data 92%	CNN, K means	Clustering video using k means algorithm.	Video	Handle multi-visual features	Dataset is small, Accuracy is not high.

Author and Year	Type	Task	Lexicon or dataset	Performance and Data domain	Approach	Video analysis Features	Type of data	Advantages or findings	Disadvantages or limitations
			dataset, UCI data.						
[8] - 2020	Hybrid	Emotion analysis from Video	Danmaku reviews	The F1 scores by SD-NB is 82.3% for positive class and also 93.6% for the negative class. Also shows good result for precession and recall.	Sentiment dictionary and naïve Bayes	Can classify video sentiment and opinion	Text and Video	Can classify seven video sentiment	Domain depended
[21] - 2020	Deep learning	Video analysis of infant	Clinical dataset for infant expression	Mean average precision is 81.9% and also 84.8% for four infant expressions and three states evaluated with both clinical and daily datasets. Precision for discomfort detection 90%.	HMM, CNN	Does infant expressions and also the states detection, object tracking and detection compensation with HMM and R-CNN.	Image	Nicely do expression detection, Utilization of HMM with CNN is good.	Unable to handle temporal data.

The literature is having different methods of video classification based on text, audio and video feature extraction. Different algorithms HMM, ANN, SVM and RNN, all have their own advantages and disadvantages. If it is possible to combine any of these two or more approaches, then there are advantages of both the methods in one scheme.

2.7 Quick Summery of video classification techniques based on used features

A lot of methods are used to estimate the outcome in the video classification. Deep learning and machine learning based video classification system works with the learning as well as tuning numerous parameters, normalization and support of a layered neural network. A short overview table for the details on the methods used in the video classification can be found in this section. Machine learning as well as underlying neural network models are used to produce a successful outcome in mixed or deep learning. There are several methods and functions that are used for video classification. Here we have included a quick overview of the used for video classification tools and functions. We choose twenty features shown here in Table 7 based on the methods used, tools, algorithms and so forth.

Table 7. Features used in video classification method.

F1:3D CNN	F8:CNN	F15:Dependency tree
F2:Random Forest	F9:Naive Bayes	F16:Machine learning
F3:HMM	F10:Postag	F17: Deep learning
F4:GMM	F11: KNN	F18: Hybrid
F5:2D CNN	F12: Capsule network	F19:Tfidf
F6:RNN	F13:LSTM	F20:K means
F7:Support Vector Machine	F14:GRU	

Table 8 gives a short review of recent techniques for classification tasks based on machine learning. This table is focused on the 20 different features selected throughout Table 7. Here we are introducing latest work and using a sign ✓ to show the marching features with twenty characteristics.

Table 8. Overview of feature-based video classification.

Authors and Year	F$_1$	F2	F3	F4	F5	F6	F7	F8	F9	F9	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
[7]-2020	✓				✓		✓		✓							✓	✓			
[9]-2020							✓	✓								✓				
[13]-2020								✓								✓	✓	✓		✓

[8]-2020									✓							✓	✓	✓		
[14]-2020							✓									✓	✓	✓		
[21]-2020			✓						✓							✓				
[15]-2020									✓	✓		✓				✓				
[28]-2020		✓														✓				
[42]-2020				✓												✓				
[39]-2017	✓															✓				
[40]-2016						✓										✓				
[41]-2015						✓							✓	✓		✓				
[43]-2009																✓			✓	
[29]-2017											✓					✓				
[16]-2018							✓									✓				
[17]-2020						✓							✓			✓				

2.8 Future Work and challenges in Video classification

This research may be generalized to incorporate new methods in the future. However, the characteristics of frames as well as frame retrieval are key to the effective video classification. The role of patterns is also to boost the quality of the classification tasks. Another potential challenge of incorporating larger types of video into the dataset with more efficient and generic functionality, to research methods that expressly clarify camera movement. To classify longer video, to recognize multiple action in video, to find correlation among different videos, classification of multiple objects action in the video. Live steaming game video prediction is the trends work in video classification.

3. CONCLUSION

This article critically reviews on different approach and method in video classification with their advantage, finding, limitations, challenges, data summary, research gaps, and performance. From the analysis of this paper. It is concluded that video-based approach for video classification works better over text and audio. The least employed process of video classification becomes text extraction. In different applications audio and video features extractions are used, but as we can appreciate, also the performance of the classification tasks can be more enhanced if the extraction both of visual and audio features is taken with same importance in the collection of video features. Audio-based solution needs little computing source. We also have the chance to identify videos in multiple ways to overcome the limitations of existing methods. By first segmenting images, we will identify them and then use the threshold and afterwards classify them in order to create new techniques. We may use movie and game or event forecasting classification algorithms. In order to identify the aspect of the movie and also the songs, fighting scene, funny scene, here is also a chance to classify videos. There are many existing methods for video classification and already have shown their performance. This review article also shows limitations of existing methods like unable to handle multiple features at a time, higher training time of deep learning, less adaptability of traditional machine learning, low accuracy to handle multilevel video. To overcome the limitations of the video classification is the trends and opportunity for the researcher. To classify longer video, to recognize multiple action in video, to find correlation among different videos, classification of multiple objects action in the video. Live steaming game video prediction is also trending and future work in video classification.

REFERENCES

[1] Brezeale, D. and D.J. Cook, *Automatic video classification: A survey of the literature.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 3, p. 416-430, 2008. DOI: https://doi.org/10.1109/TSMCC.2008.919173

[2] Wu, Z., et al., *Deep learning for video classification and captioning*, in *Frontiers of multimedia research*, 3122867 p. 3-29, 2017. DOI: https://doi.org/10.1145/3122865.3122867

[3] Ren, Q., et al., *A Survey on Video Classification Methods Based on Deep Learning.* DEStech Transactions on Computer Science and Engineering, cisnrc**,** 33301 .p. 1-7, 2019. DOI: https://doi.org/10.12783/dtcse/cisnrc2019/33301

[4] Anushya, A., *VIDEO TAGGING USING DEEP LEARNING: A SURVEY, International Journal of Computer Science and Mobile Computing,Vol.9 Issue.2,pg. 49-55,2020.*

[5] Rani, P., J. Kaur, and S. Kaswan, *Automatic Video Classification: A Review.* EAI Endorsed Transactions on Creative Technologies, *,7(24), p. 163996,2020).* DOI: https://doi.org/10.4108/eai.13-7-2018.163996

[6] Li, Y., C. Wang, and J. Liu, *A Systematic Review of Literature on User Behavior in Video Game Live Streaming.* International Journal of Environmental Research and Public Health, vol. 17, no. 9, p. 3328,2020. DOI: https://doi.org/10.3390/ijerph17093328

[7] Zhen, M., et al. *Learning Discriminative Feature with CRF for Unsupervised Video Object Segmentation*. in *European Conference on Computer Vision. Springer, LNCS, volume 12372,pp 445-46,2020.* DOI: https://doi.org/10.1007/978-3-030-58583-9_27

[8] Li, Z., R. Li, and G. Jin, *Sentiment Analysis of Danmaku Videos Based on Naïve Bayes and Sentiment Dictionary.* IEEE Access, vol. 8, p. 75073-75084,2020. DOI: https://doi.org/10.1109/ACCESS.2020.2986582

[9] Ruz, G.A., P.A. Henríquez, and A. Mascareño, *Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers.* Future Generation Computer Systems, 106: p. 92-104,2020. DOI: https://doi.org/10.1016/j.future.2020.01.005

[10] Xu, Q., et al., *Aspect-based sentiment classification with multi-attention network.* Neurocomputing, vol. 388, p. 135-143, 2020. DOI: https://doi.org/10.1016/j.future.2020.01.005

[11] Bibi, M., et al., *A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis.* IEEE Access, Vol. 8, p. 68580 - 68592,2020. DOI: https://doi.org/10.1109/ACCESS.2020.2983859

[12] Sailunaz, K. and R. Alhajj, *Emotion and sentiment analysis from Twitter text.* Journal of Computational Science, vol. 36, p. 101003, 2020. DOI: https://doi.org/10.1016/j.jocs.2019.05.009

[13] Peng, T., et al., *Video Classification Based On the Improved K-Means Clustering Algorithm.* E&ES, vol. 440, no. 3, p. 032060,2020. DOI: https://doi.org/10.1088/1755-1315/440/3/032060

[14] Li, X. and S. Geng, *Research on sports retrieval recognition of action based on feature extraction and SVM classification algorithm.* Journal of Intelligent & Fuzzy Systems, vol. 39, no. 4, pp. 5797-5808, 2020. DOI: https://doi.org/10.3233/JIFS-189056

[15] Alomari, E., R. Mehmood, and I. Katib, *Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection*, in *Smart Infrastructure and Applications*, Springer. p. 37-54, 2020. DOI: https://doi.org/10.1007/978-3-030-13705-2_2

[16] Ren, R., D.D. Wu, and T. Liu, *Forecasting stock market movement direction using sentiment analysis and support vector machine.* IEEE Systems Journal, vol. 13, no. 1, p. 760-770, 2020.DOI: https://doi.org/10.1109/JSYST.2018.2794462

[17] Yadav, A. and D.K. Vishwakarma, *A unified framework of deep networks for genre classification using movie trailer.* Applied Soft Computing, vol. 96: p. 106624, 2020. DOI: https://doi.org/10.1016/j.asoc.2020.106624

[18] Parameswaran, S., et al., *Exploring Various Aspects of Gabor Filter in Classifying Facial Expression*, in *Advances in Communication Systems and Networks*, Springer. p. 487-500, 2020. DOI: https://doi.org/10.1007/978-981-15-3992-3_41

[19] Hauptmann, A., et al., *with the Informedia Digital Video Library System, MULTIMEDIA '94,Pages 480–481*, 1994.

[20] Warner, W. and J. Hirschberg. *Detecting hate speech on the world wide web*. in *Proceedings of the second workshop on language in social media*. 2012. Association for Computational Linguistics. (LSM 2012), pages 19–26, 2012.

[21] Li, C., et al., *Infant Facial Expression Analysis: Towards A Real-time Video Monitoring System Using R-CNN and HMM.* IEEE Journal of Biomedical and Health Informatics, 9254091, pp 1-12, 2020. DOI: https://doi.org/10.1109/JBHI.2020.3037031

[22] Shen, J., et al., *Towards an efficient deep pipelined template-based architecture for accelerating the entire 2D and 3D CNNs on FPGA.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2019. 1442 - 1455,Vol. 39, no. 7, July 2020. DOI: https://doi.org/10.1109/TCAD.2019.2912894

[23] Meng, B., X. Liu, and X. Wang, *Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos.* Multimedia Tools and Applications, vol. 77, no. 20, p. 26901-26918,2018. DOI: https://doi.org/10.1007/s11042-018-5893-9

[24] Yang, H., et al., *Asymmetric 3d convolutional neural networks for action recognition.* Pattern recognition, vol. 85, p. 1-12, 2019. DOI: https://doi.org/10.1016/j.patcog.2018.07.028

[25] Kar, A., et al. *Adascan: Adaptive scan pooling in deep convolutional neural networks for human action recognition in videos*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. (CVPR), pp. 3376-3385,2017. DOI: https://doi.org/10.1109/CVPR.2017.604

[26] Cho, K., et al., *Learning phrase representations using RNN encoder-decoder for statistical machine translation.* arXiv preprint arXiv:1406.1078, p. 1-45, 2014. DOI: https://doi.org/10.3115/v1/D14-1179

[27] Shofiqul, M.S.I., N. Ab Ghani, and M.M. Ahmed, *A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations.* COMPUSOFT: An International Journal of Advanced Computer Technology, vol. 9, no. 7, p. 3768-3776, 2020.

[28] Kalra, G.S., R.S. Kathuria, and A. Kumar. *YouTube Video Classification based on Title and Description Text*. in *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. 2019. IEEE. ICCCIS48478,p. 8974514,2019. DOI: https://doi.org/10.1109/ICCCIS48478.2019.8974514

[29] Yuan, F., et al., *End-to-end video classification with knowledge graphs.* arXiv preprint arXiv:1711.01714, 2017. 1711.01714, pp 1-9, 2017.

[30] Voulodimos, A., et al., *Deep learning for computer vision: A brief review.* Computational intelligence and neuroscience, 7068349, pp 1-13, 2019. DOI: https://doi.org/10.1155/2018/7068349

[31] Sargano, A.B., P. Angelov, and Z. Habib, *A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition.* applied sciences, vol. 7, no. 1, p. 110,2017. DOI: https://doi.org/10.3390/app7010110

[32] Elboushaki, A., et al., *MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences.* Expert Systems with Applications, vol. 139: p. 112829, 2020. DOI: https://doi.org/10.1016/j.eswa.2019.112829

[33] Huiqun, Z., W. Hui, and W. Xiaoling. *Application research of video annotation in sports video analysis*. in *2011 International Conference on Future Computer Science and Education*.IEEE, 6041660, p. 1-5, 2011. DOI: https://doi.org/10.1109/ICFCSE.2011.24

[34] Herath, S., M. Harandi, and F. Porikli, *Going deeper into action recognition: A survey.* Image and vision computing, vol. 60, p. 4-21, 2017. DOI: https://doi.org/10.1016/j.imavis.2017.01.010

[35] Chen, H., et al., *Action recognition with temporal scale-invariant deep learning framework.* China Communications, vol. 14, no. 2, p. 163-172, 2017. DOI: https://doi.org/10.1109/CC.2017.7868164

[36] Peng, X., et al. *Action recognition with stacked fisher vectors*. in *European Conference on Computer Vision*, Springer. ECCV,2014,pp 581-595, 2014. DOI: https://doi.org/10.1007/978-3-319-10602-1_38

[37] Lan, Z., et al. *Beyond gaussian pyramid: Multi-skip feature stacking for action recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (CVPR), pp. 204-212, 2015. DOI: 10.1109/CVPR.2015.7298616

[38] Dalal, N., B. Triggs, and C. Schmid. *Human detection using oriented histograms of flow and appearance*. in *European conference on computer vision*, Springer. ECCV, p. 428-441, 2006. DOI: https://doi.org/10.1007/11744047_33

[39] Asadi-Aghbolaghi, M., et al. *A survey on deep learning based approaches for action and gesture recognition in image sequences*. in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, IEEE. 7961779, p. 1-8, 2017. DOI: https://doi.org/10.1109/FG.2017.150

[40] Yang, X., P. Molchanov, and J. Kautz. *Multilayer and multimodal fusion of deep neural networks for video classification*. in *Proceedings of the 24th ACM international conference on Multimedia*, 2964297, p. 978–987. 2016. DOI: https://doi.org/10.1145/2964284.2964297

[41] Yue-Hei Ng, J., et al. *Beyond short snippets: Deep networks for video classification*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*,(CVPR), p. 4694-4702, 2015. DOI: 10.1109/CVPR.2015.7299101

[42 Dvir, A., et al., *Encrypted Video Traffic Clustering Demystified.* Computers & Security, Volume 96, p. 101917, 2020. DOI: https://doi.org/10.1016/j.cose.2020.101917

[43] Yin, D., et al., *Detection of harassment on web 2.0.* Proceedings of the Content Analysis in the WEB, 2: p. 1-7, 2009.

BIOGRAPHY OF AUTHORS

Md Shofiqul Islam, Currently, he is doing Masters (Research based), student at University Malaysia Pahang (UMP), Pahang, Malaysia, He have completed my B. Sc. in 2014 in CSE from Islamic University, Kushtia, Bangladesh. Now he is a research assistant at University Malaysia Pahang (UMP), He is also a teacher at CSE under the faculty of FST at ADUST university, Dhaka. He is also in teaching profession since 2015. His research field are: Deep learning, Machine learning, Natural Language Processing, Image Processing. He has published a lot of papers in his field. Email: shafiqcseiu07@gmail.com

Shanjida Sultana, she is completing master's degree and completed bachelor's degrees from the department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. She is working in the field of image processing, video processing and text processing. Her email is sunjidasultana51984@gmail.com

Uttam Kumar Roy, he has completed bachelor and master's degrees from the department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. Now he is working as Assistant Programmer at Bangladesh Bank-The Central Bank of Bangladesh. Head Office, Motijheel Commercial Area, PO Box 325, Dhaka 1000.He is also doing his research work in the field of Machine learning, image processing, video processing and text processing. His email is cseuttamiu@gmail.com

Jubayer Al Mahmud, he has completed masters and bachelor's degrees from the department of Computer Science and Engineering, Islamic University, Kushtia-7600, Bangladesh. Now he is working as Senior Software Engineer at Charja Solutions Limited,129-Kha/1, Elephant Road, New Market, Dhaka-1205.He is also doing his research work in the field of Machine learning, IOT, image processing, video processing and text processing. His email is jubayear.iu0708@gmail.com