

## Automatic Classification of Learning Material Styles

*Bernadete Aquino*

*Programa de Pós-Graduação em  
Ciência da Computação  
Universidade Federal de Juiz de Fora  
ORCID: [0000-0002-6206-7003](https://orcid.org/0000-0002-6206-7003)  
<beteaquino.jf@gmail.com>*

*Jairo Francisco de Souza*

*Departamento de Ciência da  
Computação  
Universidade Federal de Juiz de Fora  
ORCID: [0000-0002-0911-7980](https://orcid.org/0000-0002-0911-7980)  
<jairo.souza@ice.ufjf.br>*

*Eduardo Barrére*

*Departamento de Ciência da  
Computação  
Universidade Federal de Juiz de  
Fora  
ORCID: [0000-0002-1598-5362](https://orcid.org/0000-0002-1598-5362)  
<eduardo.barrere@ice.ufjf.br>*

### Abstract

*Although video lessons are often used in diverse areas, the lack of a common approach to defining and classifying their styles results in using many different models for these purposes. There is a need to build a framework through which these styles can be defined and classified. Much has been done to investigate the effects of these styles on student engagement and learning outcomes. These studies suggest that video lesson styles affect academic performance and that students learn better through a certain video lesson style. Based on this, we propose a unified model for classifying video lesson styles based on the nomenclatures and definitions used in the literature. Furthermore, we present an approach for automatically classifying four popular video lesson styles. The automatic classification is useful for recommendation systems to suggest materials more consistent with student preferences and their intended learning outcomes.*

**Keywords:** *Video Lesson Styles, Learning Material Styles, Instructional Media Design, Automatic Classification.*

## 1 Introduction

The current century is sometimes referred to as the Internet video era (Mayer et al., 2020). Video lessons have become an important form of learning, especially during the pandemic period where the number of students has significantly increased on educational platforms where they are frequently used (Lu et al., 2020). Video lessons are becoming one of the most powerful learning media that capture and distribute information, as well as provide a stimulating learning environment where students can better understand and retain information (Sablic et al., 2020). In (Deng & Benckendorff, 2021), the authors have identified video lessons as one of the themes that most contribute to positive learning experiences. Its use increases student satisfaction and performance due to its potential to increase active and student-centered learning (Bordes et al., 2021), allowing students to review the content taught and follow their own learning rhythms.

Video lessons are produced in a variety of ways and their production styles vary substantially. They can be organized according to different characteristics, such as physical aspects (codecs, size, resolution, aspect ratio, etc), content (topics, depth/coverage of each topic, etc), and pedagogical (goals, type of content – exposition of topics, simulation, exercise, etc –, instructional design, etc). In this work, video lesson styles are investigated. Video lesson style refers to the main method of visual organization that is employed to accomplish the objectives of the video and achieve specific results (Hansch et al., 2014).

Studies indicate that choosing the right video lesson style is crucial to student engagement and learning outcomes, suggesting that students learn better through certain video lesson styles (Lackmann et al., 2021; Rosenthal & Walker, 2020; Ng & Przybylek, 2021; Rahim & Shamsudin, 2019a). Such research is generally based on evidence that learning requires attention, and certain video lesson styles are more conducive than others to gaining and maintaining attention (Rosenthal & Walker, 2020). In addition to academic performance, styles can also affect students psychologically (Rahim & Shamsudin, 2019a) and students have strong preferences for certain styles (Choe et al., 2019). In (Chen & Thomas, 2020), the authors show that video lesson styles affect students differently based on their level of prior knowledge. There is not necessarily an ideal style and there is room for more research to analyze all the different styles in relation to engagement and performance (Lackmann et al., 2021).

In an investigation of the impact of using a video lesson style, it is first necessary to identify and characterize the existing styles. There are different proposals for classifying video lesson styles where the authors have presented different definitions of styles (Hansch et al., 2015; Santos Espino et al., 2016; Crook & Schofield, 2017; Chorianopoulos, 2018; Köse et al., 2021). Although there are different theoretical classifications of video lesson styles in the literature, in reality, none was unanimously accepted. Also, style names used for different authors often do not match, even if they describe the same video lesson style (Arruabarrena et al., 2021). Because of that, these theoretical classifications are usually not used in studies on preferences, benefits, and impacts of the use or production of different styles. In addition, those classifications do not name, describe, or present screenshots that help in the correct characterization of styles. Thus, the results found in these works are difficult to interpret because they use often conflicting definitions. Achieving better knowledge about the classification model can allow new functionalities in search and recommendation mechanisms for students.

In (Rahm & Shamsudin, 2019a), the authors performed a manual classification of video lesson styles available in a private repository, using the classification proposed by Crook & Schofield (2017). Although there is no consensus between the existing types and their characteristics, there is a subset of educational video styles that are present in most of these studies and that are more popular in learning object repositories. Little effort, however, has been devoted to automatically identifying these styles. The ability to automatically classify video lesson styles

can serve as a solid foundation for systems that aim to provide personalized learning object recommendations based on a student's learning styles, as illustrated in (de Oliveira et al., 2018). The identification of styles allows for providing a new dimension of search in repositories of learning objects.

This article aims to (i) propose a unified classification model of video lessons styles; (ii) an approach to automatically classifying four video lesson styles (Talking Head, Voice Over Slides, Presentation Style, and Khan Style), which have been found to be some of the most popular styles. The ability to automatically classify video lesson styles can serve as a solid foundation for systems that aim to provide personalized learning object recommendations based on a student's learning styles. Furthermore, the identification of styles provides a new dimension for searching within learning object repositories.

## 2 Related Work

Creating classifications for any field of knowledge has benefits such as establishing a set of unifying constructs that provide a common terminology for communication, understanding interrelationships, and identifying knowledge gaps (Vegas et al., 2009). Some classification proposals on video lesson styles have already been presented in the literature.

The work of Hansch et al. (2015) and Crook & Schofield (2017) presented methods for classifying video lessons as secondary results of their work. In (Hansch et al., 2015), the authors sought to analyze the standardization of the video production process and its effectiveness as a pedagogical tool. In this process, 18 styles were cataloged. This result was based on a literature review, observation of online courses, and 12 semi-structured interviews with professionals in the field of educational video production. The work of Crook & Schofield (2017) sought to identify how variations in styles can result in different experiences for students. Sixteen styles were identified through a sample of 200 videos from different disciplines and from different virtual learning environments.

In (Santos Espino et al., 2016; Chorianopoulos, 2018), the authors aimed to organize the styles of video classes in two dimensions. In (Santos Espino et al., 2016), the authors classified the communication strategy into two opposite categories: centered on the speaker or centered on the board. The seven most frequent styles were identified through a mapping of the catalog of styles presented in (Hansch et al., 2015). On the other hand, the work of Chorianopoulos (2018) generalized the two categories presented in (Santos Espino et al., 2016) by considering that they are not conflicting. Such categories were presented in two dimensions with more common notions in the learning sciences: human incorporation and instructional media, both have the same limits from digital to physical, for example, from human animation to the human body. The authors present thirteen styles discovered through a literature review and examination of educational videos. However, these styles were not named, but represented by symbols. In (Köse et al., 2021), the authors have as main references the articles by Hansch et al. (2015) and Crook & Schofield (2017), and they sought to expand the classification of video lessons beyond their style, identifying new characteristics such as interaction (ability to stop, advance, or rewind the video); connection (internet access requirement for video playback); sequence (transitions between parts of the video are structured); components (image, moving image, audio, and text); image format (2D, 3D); instant (live video); content (the subject of the video has individual and independent parts or not).

These works aimed to define different video lesson styles or organize them in some dimensions. However, there is no consistent definition of each style with a detailed description and the aspects that characterize them. The lack of a sufficient definition for the styles prevents an accurate evaluation of research results. The creation of a detailed classification with a common

and comprehensible language that facilitates the definition of styles would make possible the correct characterization of the study objects of future investigations. The present work intends to fill this gap by raising the aspects that characterize each style and identifying new styles based on a Systematic Literature Review. Then, we propose a unified classification, contributing to the formation of a shared understanding of defining video lesson styles based on existing scientific studies.

Several articles focus on the automatic extraction of features from video lessons (Lin et al., 2019; Shanmukhaa et al., 2020; Sonia et al. (2021), but few studies are related to the automatic classification of video lesson styles. Although they do not deal directly with the classification of video lesson styles, the articles below identify characteristics present in different styles that could be used for this purpose.

In Rawat et al. (2014), the authors focused on identifying the teacher in the video lesson by finding scenes such as the teacher speaking, the teacher writing on the blackboard, and the teacher explaining the slide. Teacher, blackboard, and slide images are identified considering the colors of these shapes. Yousaf et al. (2015) identifies instructor activities using face recognition and teacher pose estimation. Other articles focus on extracting the content written on the blackboard. Lee et al. (2017) have identified and improved the quality of the frame that best represents the content. Davila & Zanibbi (2018) recognize mathematical formulas present on the blackboard. The works of Kota et al. (2021), Urala et al. (2018), and Davila et al. (2021) sought to summarize the content written on the blackboard by identifying contents and keyframes. Ciurez et al. (2019) presented a method to classify videos in different learning styles, with techniques to calculate the amount of text and image of different frames of the video.

The work of Aryal et al. (2018) focused directly on automatic video style classification, focusing on three styles: talking head, slide, and code (code is screencast style but featuring computer programming content). The image-based classification approach was adopted by performing a comparison between Convolutional Neural Networks models, including VGG16, InceptionV3, and ResNet50, where each frame is classified in one of the styles.

Several articles focus on the extraction of features from video lessons, but few of them focus on the automatic classification of video lesson styles. Thus, the contribution of this work is the use of different visual characteristics of the video lessons, such as the presence of people and texts for automatic identification of these styles, the conduction of tests with different classification models, and the use of a broader set of video lesson styles. The result of this research can help provide personalized course recommendations for a student based on video lesson style, as different styles serve different purposes.

### **3 A unified classification of video lessons styles**

The style of a video lesson can affect student engagement and learning outcomes. There are different proposals in the literature for classifying styles of video lessons but with differences between them. Consolidating production styles not only enhances the grasp of material production possibilities but also streamlines communication among researchers. This section presents a survey of these works based on a Systematic Literature Review (SLR) in order to identify video lesson styles and which aspects characterize them as well as propose a unified classification of these styles. Furthermore, it presents an investigation of how automatic classification systems have been used to identify educational video styles.

The SLR was used to ensure adequate coverage of existing works, encompassing different views and revealing points of consensus and disagreement in nomenclatures, as well as showing gaps and open questions that remain. The research questions identified in this study were: (i) What

aspects have been used to characterize each style of video lesson? (ii) How have works that use AI to automatically classify video lesson styles characterized these styles?

Searches were carried out for works presented in journals or annals of relevant online events in the repositories: ACM Digital Library, ERIC Institute of Education Sciences, IEEE Digital Library, and Scopus. The query string used in all these digital repositories was ( ( "video lecture" OR "lecture video" OR "video lesson" OR "Educational video" ) AND ( " video styles" OR "automatic video classification" OR "Instructional design" OR "learning styles" OR "E-Learning" OR "Online learning" ) ).

The following inclusion criteria were used: (1) The article discusses or presents video lesson styles; (2) The article is in English or Portuguese language. The defined exclusion criteria were: (1) The article was unavailable. (2) Paper was duplicated or is a short version of a more recent study.

The search string was executed on May 31, 2022, resulting in a total of 35 papers after using the exclusion criteria, where 24 papers were used to answer the first research question and 11 papers to the second research question.

### 3.1 Answering the research questions

The two research questions of this systematic review are answered by presenting the aspects that characterize each style and the aspects used in the automatic classification of video lesson styles.

#### 3.1.1 *What aspects have been used to characterize each video lesson style?*

It was possible to verify that different terms are used to identify the same video lesson style. Distinct dimensions of a style are presented in the same list, without hierarchy, making them very long and allowing the realization of some groupings. For instance, in (Hansch et al., 2015), the styles Talking Head, Webcam Capture, On Location, and Green Screen are ambiguous and share some similarities. These styles focus on the instructor and differ only in the configuration of the background image of the video. In (Crook & Schofield, (2017), the styles Fixed Frame Outside, Mobile Frame Outside, Fixed but Overlapping, Mobile Frame and Overlapping, Presence in Split Screen, and Presence in Picture focus on the instructor and the slides, differing only in the position of them on the screen.

It was also identified that the name Talking Head was used to describe different styles. While it was used in (Ilioudi et al., 2013) to represent Classroom Lecture, in (Choe et al., 2019) it was used to describe Presentation Style. However, the vast majority of articles describe it as being the close-up of a lone narrator.

We performed two types of grouping: selecting different terms representing the same style or considering subdimensions of the same style. To create these two groups, a table was created with the descriptions of each style per article. From the reading of new descriptions, these were compared with the previous ones to identify if they did not represent the same style. Finally, new readings of the table were carried out to identify styles with many characteristics in common.

To choose the name of each group, we used the most frequent term found in the literature, coined new terms when needed, and considered the work of Hansch et al. (2015), which presents a considerably clear definition of some styles, when possible. Figure 1 presents examples of each group.

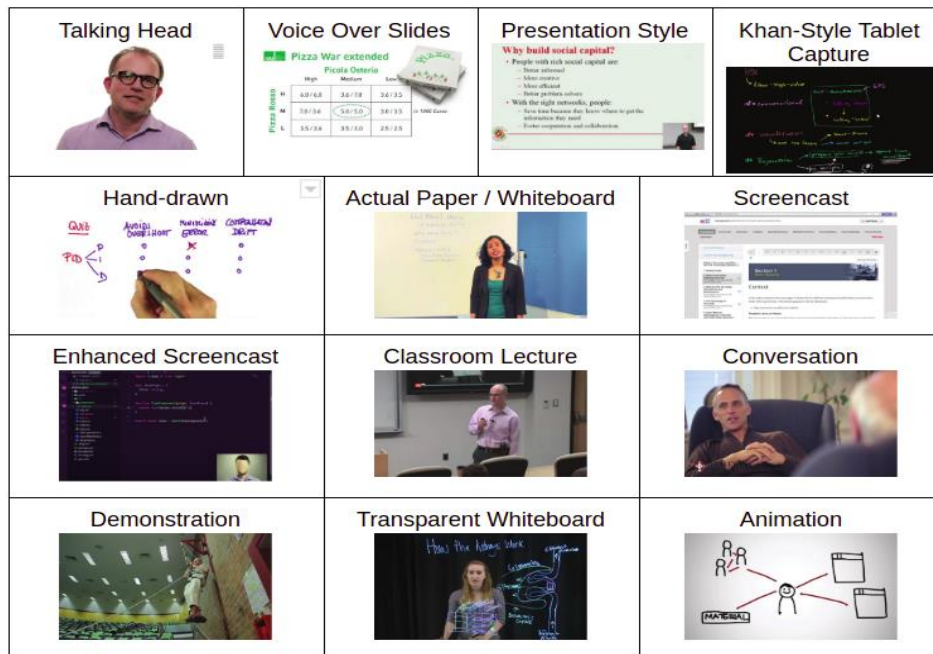


Figure 1: Example frames of each video lessons style.

- *Talking Head*: The main focus is on the instructor, who occupies a large board area and often speaks directly to the camera (Ozan & Ozarslan, 2016). It is not surrounded by slides or other rich text elements. Sometimes overlay text is placed to reinforce key ideas, or there are scene changes to show another type of material. These insertions represent a relatively small amount of video time (Santos Espino et al., 2016). The hallmark of this style is that the narrator appears on the video (Arruabarrena et al., 2021). Some groups were created in the works of Hansch et al. (2015) and Guo et al. (2014) because the classes differ only in the set of the scene or background.
- *Voice Over Slides*: consists of slides in PowerPoint or another delivery format displayed in full screen with instructor narration (Inman & Myers, 2018). A subclass of this style is Writing Over Slides, which also includes instructor writing.
- *Presentation Style*: combines a full-screen presentation of the content on the slide with a smaller projection of the instructor's camera (for example, a talking head in the corner or side of the slide) (Rosenthal & Walker, 2020) or with the instructor being physically superimposed on the slide at the time of recording (slides and presenter appear on screen as a single unit) (Gilardi et al., 2015). Some classes of Crook & Schofield (2017) and Rosenthal & Walker (2020) were grouped because they contain differences only in the instructor's position about the slide.
- *Khan Style*: this is a full-screen video of an instructor writing/drawing freehand on a digital whiteboard. It's a style popularized by the Khan Academy videos (Guo et al., 2014).
- *Hand-drawn*: includes hand-drawn writings where the instructor's hand becomes visible while taking notes (Chen & Thomas, 2020). In Udacity's proposed style, the instructor's hand is presented transparently so that the writing is not obscured (Hansch et al., 2015). Thus, the Udacity style was considered a subclass of Hand-drawn.
- *Actual Paper / Whiteboard*: is a monologue presentation while the instructor moves in front of the whiteboard content and acts on it (Crook & Schofield, 2017). The instructor looks directly at the camera, giving the impression that they are talking directly to the viewer.
- *Screencast*: is the visual recording of the screen output of a computer session. It usually includes a voice narration with a description of the actions being performed (Santos Espino et al., 2016).

- *Enhanced Screencast*: The presentation is recorded as a screencast with the addition of the instructor. This usually occurs with the instructor's face in a corner or on the side of slides (Gilardi et al., 2015).
- *Classroom Lecture*: It is the recording of a lecture in the classroom or at a conference with the audience in the room being visible or implied (Santos Espino et al., 2016). It involves video recording of a physical lecture (Rosenthal & Walker, 2020). It always feels like being recorded in a single take and the instructor speaks directly to the audience.
- *Conversation*: It is a recording of a conversation containing a speaker or expert on the field who answers questions or discusses a topic (Ozan & Ozarslan, 2016). Some groups were created using the styles described at (Hansch et al., 2015) and (Crook & Schofield, 2017) because all these styles are dialog-centric.
- *Demonstration*: It is the capture of the instructor performing orchestrated experiments to illustrate some concept (Choe et al., 2019). It allows the viewer to see an experiment in action, rather than just someone talking about it (Hansch et al., 2015).
- *Transparent Whiteboard*: The instructor stands behind a large glass panel facing the camera while writing or drawing on the glass. The camera reverses the instructor's writing and drawing so that it is readable for the viewer (Stull et al., 2018).
- *Animation*: It refers to computer-generated moving images showing associations between drawn figures (Mayer & Moreno, 2002).

### 3.1.2 *How video lesson styles have been characterized in the works that use AI to automatically classify video lesson styles?*

During the SLR, several retrieved articles focused on the automatic classification of video lessons, but few of them relate directly to the classification of video lesson styles. The works retrieved based on characteristics of video lesson styles differ in several aspects. Regarding the focus of this review, only the work carried out by Aryal et al. (2018) automatically classified the styles of video lessons.

The work of Aryal et al. (2018) is focused on three styles: talking head, slide, and code which is a screencast style but featuring computer programming content. The image-based classification approach was adopted by performing a comparison between Convolutional Neural Networks (CNN) models, including VGG16, InceptionV3, and ResNet50, where each frame is classified in one of the styles. The work of Rawat et al. (2014) identified visual concepts that can help in the identification of styles, such as teacher speaking, teacher writing on the blackboard and teacher explaining a slide. The authors identified three basic aspects, which are the teacher, blackboard, and slide, identified through resources regarding the colors of the images.

The other retrieved articles, despite not directly mentioning a style, use it to automatically extract different characteristics from the video lessons. Considering the Voice Over Slides style, Ali et al. (2021) presented improvements for indexing video classes, and Balasubramanian et al. (2015) showed a way to summarize the contents presented. Regarding the Writing Over Slides style, Kao et al. (2013) developed a method for detecting annotations made in slide presentations.

Using the Classroom Lecture style, some articles focused on the teacher. For example, Xu et al. (2019) present a methodology for extracting video lesson content using the teacher's writing or delete actions that potentially indicate characteristics directly related to the lecture content. Yousaf et al. (2015) present an approach for evaluating teacher performance and behavior in the classroom using face recognition and teacher pose estimation. Other articles focus on extracting the content written on the board (Lee et al., 2017; Davila & Zaniboni, 2018; Kota et al., 2021; Davila et al., 2021).

It was possible to observe two gaps in the literature. First, the articles that work with the identification of characteristics in a certain style do not make use of the theory of video lesson

styles and the nomenclatures defined in the literature. For the most part, these works only present descriptions that are not very detailed to characterize the style that is the scope of their study. This makes it difficult to compile the results from different studies. The second gap concerns the lack of studies in the area of AI applied to Education to automatically identify some video lesson styles. Only the styles Talking Head, Voice Over Slides, and Screencast were automatically identified in Aryal et al.'s (2018) work. The presentation of solutions for the automatic identification of video lesson styles would contribute to the area of Intelligent Tutors and Adaptive Systems for recommending content that is more adherent to the characteristics of each student and the course.

### 3.1.3 Video lesson style classification model

Addressing the importance of defining video lesson styles, a classification model is proposed that aims to unify the definitions in the literature. As pointed out by Crook & Schofiel (2017), a video lesson can also contain more than one style. To define the model, the taxonomy proposed by Chorianopoulos (2018) was used as a base model. The author presents the main factors (Human Incorporation and Instructional Media) that define the classification of the video lesson style as well as the possible aspects of these factors that range from digital (eg. slides) to physical (eg. blackboard). We extend the work of Chorianopoulos (2018) by naming each video lesson style, including new styles identified in the literature, and mapping the styles with the aspects that define them. The classification was based on visual items present in the video lessons and styles found in this SLR.

Figure 2 presents the aspects that characterize each style of video lesson separated by two factors: Human Embodiment and Instructional Media. In this classification, Human Embodiment may not be visible, be a person alone or in a group, or only a part of the body (hand). Instructional Media refers to animation, writing by pen or pen tip, computer screen, slides, instrument, transparent board, blackboard, or no media. In comparison to the attributes presented by Chorianopoulos (2018), regarding Human Embodiment, we mapped the values of people and audience to people in a group. The presence of an instructor and talking head were mapped to a single person. In the Instructional Media, we added the computer screen and the transparent board because they belong to newly identified styles.

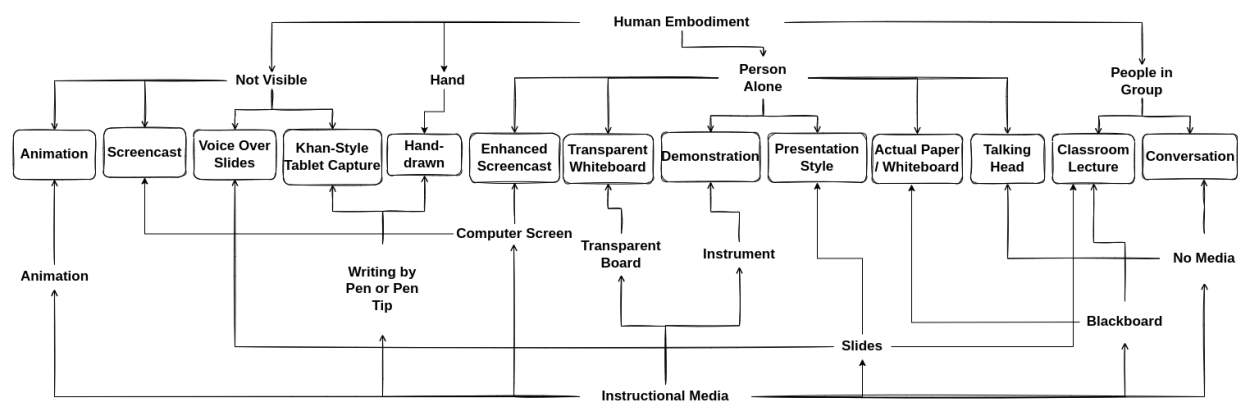
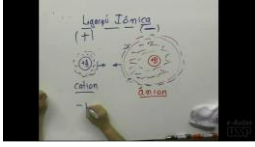
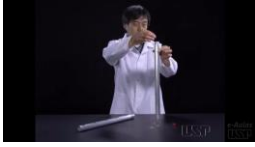




Figure 2: Video lessons styles classification model.

To demonstrate the proposed model, it was randomly retrieved four video lessons from the public repository of the University of São Paulo. Table 1 presents examples of using the theoretical classification model. With a screen capture of the video lesson, it is possible to assign a value to each of the factors, Human Incorporation and Instructional Media, among the values presented by the model. It is possible to use the template to define the video lesson style with this information in hand.



Table 1: Examples of video lessons classification.

Screenshot	Human embodiment	Instructional media	Video lesson style
	Hand	Writing by Pen or Pen Tip	Hand-drawn
	Person Alone	Instrument	Demonstration
	Person Alone	Slides	Presentation Style
	People in Group	No Media	Conversation

### 4 Automatic classification of video lessons styles

Many studies have been conducted to investigate the effects of video lesson styles on student engagement and learning outcomes. Such studies suggest that video lesson styles have a positive impact on academic performance and that students learn better through certain video styles but few studies have tried to automatically classify these styles. Thus, an approach is proposed for the automatic classification of 4 video lesson styles using the classification model proposed in Section 3.2.1, and using different machine-learning models to evaluate the method. This automatic classification can be used by recommendation systems to suggest styles that are closely aligned with student preferences and the learning outcome.

The automatic style classification is carried out by identifying the Human Embodiment and Instructional Media factors described in the classification model (Figure 2). The idea is to verify which attributes representing these factors are present in the video lesson and then perform the classification. The Human Incorporation factor has attributes related to the presence of people in the video lesson, so an object detection system was used to identify people in the video lessons. The Instructional Media factor has attributes related to the types of media present in the video lesson. For its extraction, an Optical Character Recognition (OCR) tool was used to recognize and extract the text embedded in the images. Based on these data, some characteristics of the video lessons were extracted and used by machine-learning models. Figure 3 shows the high-level architecture of this solution.

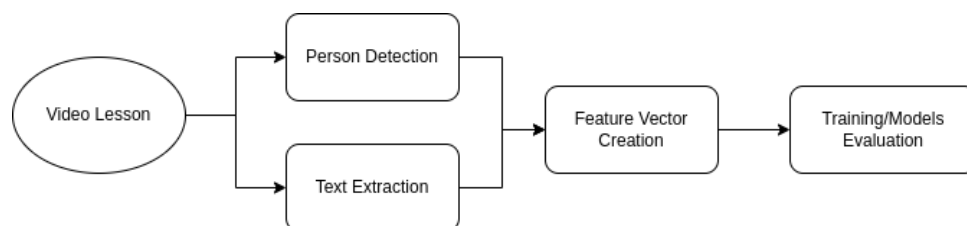


Figure 3: High-Level Solution Architecture.

Among the various styles of video lessons defined in the literature, we classify 4 different styles: Talking Head, Voice Over Slides, Presentation Style, and Khan-Style. Hansch et al. (2015) considered Talking Head and Khan-style the most commonly used styles but Wang et al. (2019) considered the Voice Over Slides style. In addition to the Voice Over Slides style, Ozan & Ozarslan (2016) also mention the popularity of the Presentation Style. Furthermore, these styles appeared in most of the 24 articles retrieved by SLR that were used to answer the first research question (section 3.1.1).

#### 4.1 Dataset

There is no public dataset to evaluate video lesson styles. Thus, a set of real-world videos from the education domain were collected to conduct the experiments and validate the classification approach. The dataset consists of 175 English-language videos collected from platforms such as Khan Academy, Ted Talks, VideoLectures.NET and the University of Oxford that are licensed under a Creative Commons license and permit non-commercial use. The total duration of the base is 40.39 hours. The shortest video is 4 minutes, and the longest is 52 minutes, with an average of 13 minutes in length. All videos maintain a single style throughout their duration. For each style (Khan-Style, Talking Head, Voice Over Slides, presentation Style, and Others), 35 videos were retrieved. The Other class has been added to verify that the classification solutions can identify videos of different styles that were not labeled in the previous four styles. The Other class then represents a confusion class for the model.

#### 4.2 Feature extraction

To extract the features of the video lesson, the classification model proposed in the previous section was used as a basis, which presents the main factors (Human Incorporation and Instructional Media) that define the classification of video lesson style, as well as the possible aspects for these factors ranging from digital (eg. slides) to physical (eg. blackboard). For the Human Incorporation factor, the styles Talking Head and Presentation Style have the attribute "person alone", whereas in the styles Khan-Style and Voice Over Slides, this attribute is "not visible". For the Instructional Media factor, the Voice Over Slides and Presentation Style styles have the "slides" attribute, the Khan-Style has the "writing using a pen or pen tip" attribute, while the Talking Head style has "no media" in this factor. The Human Incorporation factor was identified through the presence of people and the Instructional Media factor through the presence of texts since this is a characteristic of "slides" and "writing using a pen or pen tip".

The Python ImageAI library was used to identify the presence of people in the video. This library provides an API to recognize different objects using pre-trained models. The You Only Look Once (YOLO) is one of the most used object detection models (Yilmaz et al., 2021). YOLO v3 was used as the person detection model in this study. Due to machine processing restrictions, it was processed one video frame per second (FPS). The implemented algorithm generates a JSON containing the number of people identified and an array with the position (frame number) of people present in the video.

At the same time, text was extracted from the video lesson. For this, we used Python-tesseract, an Optical Character Recognition (OCR) tool for Python. This tool recognizes and extracts text embedded in images and is a wrapper for Google's Tesseract-OCR Engine (Jayoma et al., 2020). Through its use, a second JSON is generated containing the extracted words and the frame number where the words were identified. In this work, five frames per second were processed because it is the amount that best identified the texts within the processing limit of the machine used.

From the output of the two algorithms, the following features were extracted: one person's appearance time, more than one person's appearance time, slide appearance time, number of

slides, nonexistent words average, and new words per frame average. Being  $F^{(P1)}$  the amount of frames where a person appears;  $F^{(P2)}$  the number of frames where 2 people or more appear,  $T$  the total time of the video,  $F$  the number of frames analyzed,  $P^{(In)}$  the number of non-existent words in a frame, words that do not exist in a defined corpus for the language were considered non-existent, this work used the corpus provided by the words class of the nltk.corpus library,  $F^{(Pal)}$  the number of frames with words,  $P^{(New)}$  the amount of new words per frame. Thus, the features were extracted from the videos:

- One person's appearance time: which represents the total time of the video where only one person is seen, given by  $(F^{(P1)} \times T) / F$ .
- More than one person's appearance time: which represents the total time of the video where more than one person is seen, calculated as  $(F^{(P2)} \times T) / F$ .
- Slide appearance time: which symbolizes the total time of the video where there are slides. The frame with more than 3 words was considered a slide, as in many frames, the university logo appeared so they were disregarded. The time is given by  $(F^{(Pal)} \times T) / F$ .
- Number of slides: which corresponds to the number of slides presented in the video. For this definition, Jaccard's similarity was used to find the similarity between two sets of terms used in subsequent frames. In the comparison, only the words existing in the corpus defined for the language were used. The amount was calculated as  $\sum \text{If}(\text{Jaccardsimilarity}() < 70\%, 1, 0)$ .
- Nonexistent words average: describes the average number of words identified by the OCR algorithm but are not present in the corpus defined for the language. The words considered non-existent are those that are not in the corpus. Average calculated as  $\sum P^{(In)} / F$ .
- New words per frame average: constituted by the medium of new words added by frame. The number of new words is the number of words in a frame removing the number of words from the intersection with the previous frame. Average is given by  $\sum P^{(New)} / F$ .

Since it is expected that the OCR algorithm produces noisy text when recognizing handwritten words, the feature "Nonexistent words average" was added in order to account for this characteristic.

### 4.3 Classification models

Different classification models were used with the purpose of comparing and discovering the method that presents the best result. The adopted models were: K-Nearest Neighbor (kNN), Random Forest, Support Vector Machine (SVM), Logistic Regression, Naive Bayes, AdaBoost, and a neural network. The training and testing of the models were performed using the Orange version 3.32.0.

Tests were performed with data normalization within the range of 0 and 1, but it was the standardization of variables with a mean equal to zero that obtained the highest accuracy values. In order to observe the correlation between each pair of attributes, Spearman's correlation coefficient was used. The highest correlation (0.77) occurred between the variables "Nonexistent words average" and "Slide appearance time", so no variable was disregarded for the classification process.

Each classification method has different hyperparameters and some of these are fundamental to creating an accurate model. In this sense, experiments with different parameter values were carried out to obtain the best configurations for the algorithms. Table 2 presents the values used to adjust the configurable parameters in Orange and the selected value. The parameter values were chosen based on 5 executions of the algorithms, obtaining the average accuracy, and the best accuracy defined which was the parameter selected. In cases where accuracy was maintained, priority was given to the least costly value for the algorithm. Cross-validation was used by dividing the total dataset into 10 subsets.

Table 2: Values adopted in experiments with classification methods.

Algorithm	Hyperparameters	Values Tested	Selected Value
kNN	No. of neighbors	5 / 10 / 15	5
	Distance metric	Mahalanobis / Chebyshev / Euclidean / Manhattan	Mahalanobis
Random Forest	Number of trees	10 / 15 / 20	15
	Depth	4 / 5 / 6	5
SVM	Kernel	Linear / Polinomial / RBF / Sigmoid	Linear
Logistic Regression	Strength	1 / 0.9 / 0.8	1
Naive Bayes	Does not have hyperparameters		
AdaBoost	No. of Tree Estimators	40 / 50 / 60	50
	Algorithm	SAMME.R / SAMME	SAMME.R
Neural Network	Neurons in the Hidden Layer	20 / 25 / 30	25
	Solver	SGD / Adam / L-BFGS-B	L-BFGS-B
	No. of Iterations	100 / 150 / 200	150
	Activation Function	Identity / Logistic / Tanh / ReLu	Tanh

#### 4.4 Results analysis

Table 3 presents the results of the evaluations of the 7 analyzed models regarding the accuracy, F1, precision, recall, and area under the ROC curve. Through it, it is possible to verify that the features used are relevant for the classification since the models had high accuracy and a high area under the ROC curve.

Table 3: Evaluation results of the different models.

Model	Accuracy	F1	Precision	Recall	ROC Area
Logistic Regression	0.926	0.926	0.926	0.926	0.977
SVM	0.914	0.913	0.919	0.914	0.988
Random Forest	0.914	0.914	0.915	0.914	0.982
kNN	0.88	0.879	0.888	0.88	0.964
Naive Bayes	0.88	0.879	0.883	0.88	0.979
Neural Network	0.863	0.862	0.863	0.863	0.97
AdaBoost	0.857	0.857	0.859	0.857	0.911

The classification model that showed the highest accuracy was Logistic Regression with 92%. Figure 4 presents the confusion matrix of this classifier. From the analysis of the confusion matrix of the Logistic Regression model, it was possible to analyze the possible reasons why the video lessons were misclassified. In the Khan Style, 2 errors occurred, one video lesson was classified as Presentation Style for having a static image of a person throughout the video class, and another was classified as Voice Over Slides because this video lesson already starts with a large volume of text and new texts are being added. In the Presentation Style, 3 errors occurred. In all these cases, the people identification algorithm was unable to identify the teacher's image, leaving the number of people as 0 or identifying it in very few frames, 2 cases that had a greater amount of text were classified as Voice Over Slides and 1 case with very little text was classified as Khan-Style. While the Talking Head style had 2 errors, both classified as Others, because these video lessons present a small amount of text in some frames. At the same time, the Voice Over Slides style had 2 errors, one was classified as Khan-Style for presenting too many mathematical

formulas and tables which made the text recognition algorithms not correctly identify these texts, and another one as Presentation Style because there are images of people on the screen. On the other hand, the Others style had 4 errors, all classified as Talking Head, 3 of them belonged to the Classroom Lecture style but the algorithm could not correctly identify the texts on the blackboard, the other belonged to Conversation style, as the interviewer and interviewee do not appear on the screen together, the presence of only one person was considered.

	Khan-Style	Others	Presentation Style	Talking Head	Voice Over Slides	$\Sigma$
Khan-Style	33	0	1	0	1	35
Others	0	31	0	4	0	35
Presentation Style	1	0	32	0	2	35
Talking Head	0	2	0	33	0	35
Voice Over Slides	1	0	1	0	33	35
$\Sigma$	35	33	34	37	36	175

Figure 4: Confusion matrix of the Logistic Regression algorithm.

Based on the observation of the algorithm's errors, it is possible to find some improvements in feature identification. The analysis of the presence of people was used to extract the features "One person's appearance time" and "More than one person's appearance time", however, some improvements can be made in this analysis. Some video lessons presented photos of teachers on the cover slide or in video lessons about medicine, images of people were used to explain the content. These images were considered as the teacher's image. Thus, analyzing whether the person performs some movement can guarantee the correct definition of the features, since the video classes can contain static images and this does not mean the teacher's presence. Also, identifying the presence of more than one person using the tone of voice, because as happened in the Conversation style, the interviewer and interviewee did not appear together, so it was considered the appearance of only one person. In addition, in some video lessons, the Presentation Style did not identify the teachers because they were in a small video frame. Therefore, new tests could be carried out with new configurations of the person identification algorithm or other object recognition algorithms, since not all people in the video lessons were identified.

The text identification technique was used to define the features "Slide appearance time", "Number of slides", "Nonexistent words average" and "New words per frame average". In some video lessons of the Classroom Lecture style, the algorithm cannot identify the content written on the blackboard. Other solutions would be to identify the blackboard or use other techniques, for example, those described in the article by Davila & Zanibbi (2018) that recognize mathematical formulas, because in some cases, the formulas were not identified.

## 5 Analysis of Features Importance

To determine whether the predictions made by the Logistic Regression model are reasonable, a qualitative evaluation of the model was used through the importance of each feature for each class. Graphs were generated to indicate which features had a high impact on each prediction. Figure 5 presents 5 charts, where each one presents the most important features for style definition.

The charts in Figure 5 can be analyzed as follows: the values that have a high impact on the prediction of the class are on the right and those that vote against the class are on the left. The dot color represents the feature value (red for high values and blue for low values). For example,

a feature that has more red dots on the left means that high values for that feature contribute against the target value.

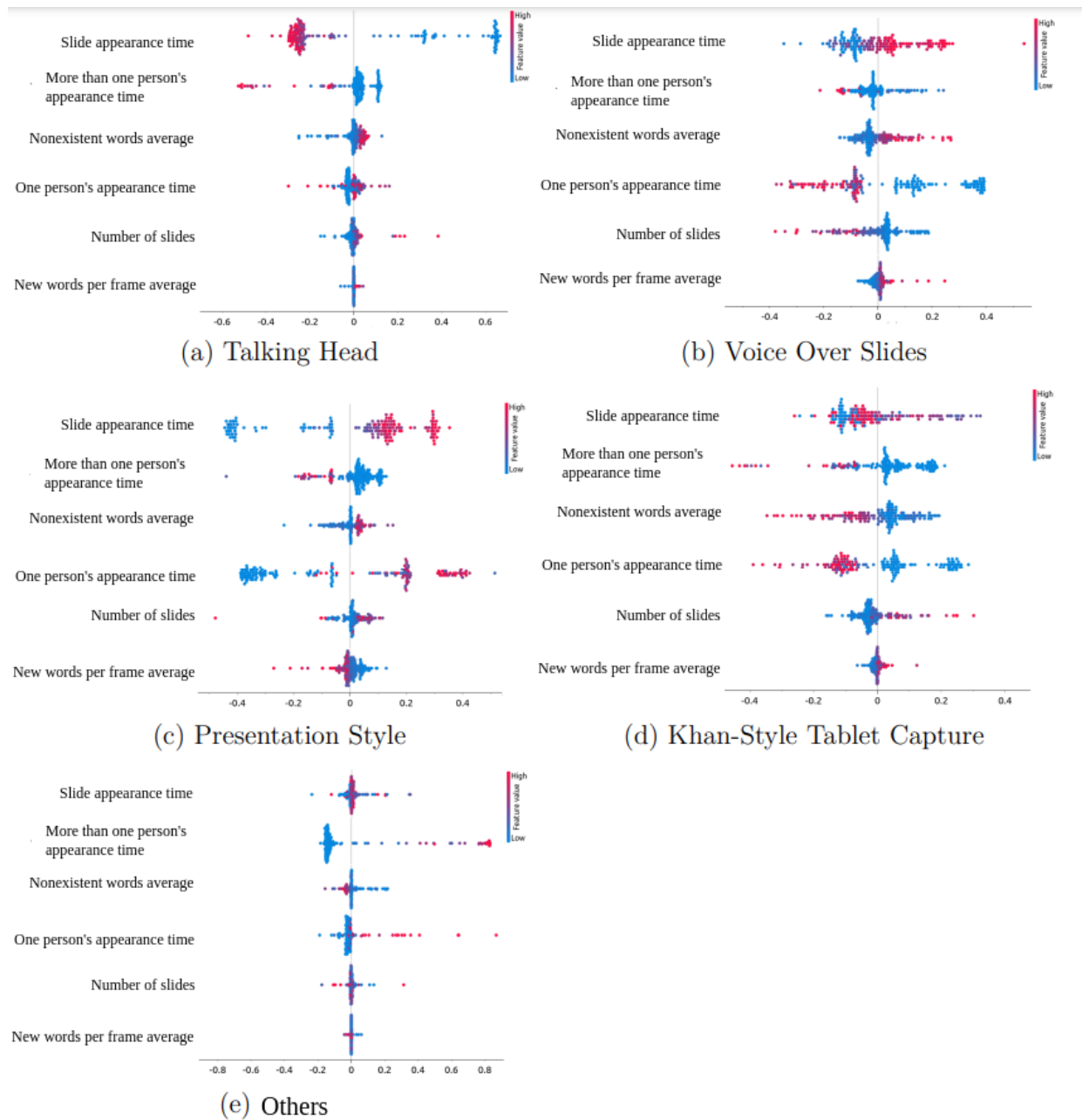


Figure 5: Importance of features in the Logistic Regression model.

Thus, through Figure 5a, it is possible to verify that in the Talking Head style, lower values for the two most important features "Slide appearance time" and "More than one person's appearance time" contribute more to the definition of this style, since this style does not have slides and only one person appears in the video lesson. In the Voice Over Slides style presented in Figure 5b, the main features were "One person's appearance time" with a high contribution of smaller values since this style does not have the teacher presence and "Slide appearance time" with a high contribution of higher values since this style is composed of slides. While in Presentation Style (Figure 5c), higher values for features "One person's appearance time" and "Slide appearance time" cooperate with this classification, since this style is composed of slides and the teacher's presence. In the Khan Style, shown in Figure 5d, smaller values for the variable "One person's appearance time" was more important since this style does not present the instructor's image. It is believed that there was no clear definition of the "Slide appearance time"

values because, in some video lessons, the OCR algorithm can not detect the words. In Figure 5e, which represents the confusion class Others, higher values for the variable "More than one person's appearance time" contributed the most, given that the repository of this style contains many video lessons in the style Classroom Lecture and Conversation where the video lesson has more than one person, the same occurs for the feature "One person's appearance time".

Based on the analysis of the charts, it is possible to conclude that the Logistic Regression model made relevant predictions about the styles of the analyzed video lessons.

## 6 Concluding Remarks

Although video lessons are often used in many areas covering a wide range of studies and applications, the lack of a common approach to defining and classifying video lessons results in using many different models for these purposes (Kose et al., 2021). There are different classifications in the literature for video lesson styles, but none was unanimously accepted. Style names often do not match, even if they describe the same video style (Arruabarrena et al., 2021). Furthermore, recent studies indicate that the style of the video lesson influences student engagement and student retention, and some styles seem to have a higher impact on learning than others (Lackmann et al., 2021). Based on this, it becomes important to characterize the video lesson styles and their automatic classification, allowing recommendation systems to select materials compatible with the student's learning objectives.

This work presents an SLR to identify the video lesson styles and which aspects characterize them. Thus, it was possible to propose a unified classification that enables communication through a common terminology. It also surveyed which aspects are used in the automatic classification of video lessons in terms of style. With this survey, it was possible to identify two gaps in the literature: (i) the articles that work with this automatic classification do not use the style classes already proposed in the literature. They usually define a new name for the style or subjectively describe the style, which makes compiling results difficult; and (ii) not all video lesson styles have studies to automatically identify them, which would bring advances in personalized learning environments. Finally, a video lesson style classification model was proposed that unifies the definition of each style.

Based on the proposed classification model, an evaluation of this model was carried out through the use of different classification algorithms. It focused on 4 different styles: Talking Head, Voice Over Slides, Presentation Style, and Khan Style. An Other class has been added as a confusion class for the model. Although other classification models can be evaluated or new experiments with other hyperparameters can be performed, the approach presented shows that the features extracted from the videos can accurately classify the set of styles shown in this study. The results demonstrate that the approach satisfactorily identifies the styles of selected video lessons, and this identification is carried out through simple and easy-to-extract features.

The main contributions of this work are: (i) an SLR on video lesson styles. This study presents an overview of which aspects have been used to characterize each video lesson style and how recent works have tried to apply AI techniques to automatically classify video lesson styles. (ii) a proposal for a video lesson style classification model based on video lessons' visual characteristics that aim to unify literature definitions and map styles based on the aspects that define them. This classification proposal offers practical contributions to researchers by allowing them to classify videos according to the listed aspects, which will facilitate and speed up access to video types. (iii) An approach for automatically classifying video lesson styles based on simple and easy-to-extract features, which allows new applications in the education domain to explore the use of these videos in their solutions, since studies indicate that the students learn better through a certain video lesson style compared to others (Lackmann et al., 2021; Rosenthal &

Walker, 2020; Ng & Przybylek, 2021; Rahim & Shamsudin, 2019a). (iv) a new dataset with a set of real-world videos from the education domain containing the Khan Style, Talking Head, Voice Over Slides, Presentation Style, and Other styles available to researchers to allow replication and advancement in researching in the area of video lesson styles. To the best of our knowledge, this is the first dataset labelled with video lesson styles.

This study has some limitations. The classification model was not evaluated by specialists in the field of Educational Technologies. Possible deficiencies in the proposed classification model can be eliminated through the examination of the work by the scientific community through manual validation or through a classification procedure of a significant number of video lessons based on the proposed model. The performance of classification algorithms is directly related to an adequate selection of design parameters. However, these parameters were selected empirically. Although the approach is language-independent, the tools adopted to produce text from the video affect the final result of the approach and need to be chosen according to the language of the video. In addition, the same video lesson can contain more than one style, but this classification was carried out using video lessons with only one style. Finally, the dataset created does not have video lessons on all areas of knowledge and some of these areas have an unbalanced number of video lessons representing each style, which may cause some bias in the classification.

This work opens up a range of possibilities for future research. The proposed classification model can be better evaluated in future analyses by specialists in the Educational Technologies field, thus ensuring its clear understanding. In addition, include a manual validation of the model through a classification procedure with several video lessons. There are a variety of styles, which were not included in the automatic classification of this study and which still need to be evaluated. New features will be necessary to allow the classification of these other styles. Also, experiments with more voluminous and more diversified databases about the domain and characteristics of the videos, such as duration, scene shot, etc., may show that other classification models have better accuracy than logistic regression. Finally, a video lesson can contain more than one style, but these cases were not addressed in this work.

## References

- Ali, M. M., Qaseem, Mohammad S. & Hussain, Altaf. (2021). Segmenting lecture video into partitions by analyzing the contents of video. *International Conference on Data Analytics for Business and Industry (ICDABI)*, 191–196. <https://doi.org/10.1109/ICDABI53623.2021.9655924> [GS Search]
- Arruabarrena, R., Sánchez, A., Domínguez, C., & Jaime, A. (2021). A novel taxonomy of student-generated video styles. *International Journal of Educational Technology in Higher Education*, 18, 68. <https://doi.org/10.1186/s41239-021-00295-6> [GS Search]
- Aryal, S., et al. (2018). Using pre-trained models as feature extractor to classify video styles used in mooc videos. *IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, 1–5. <https://doi.org/10.1109/ICIAfS.2018.8913347> [GS Search]
- Balasubramanian, V., Sooryanarayan, D., & Kanakarajan, N. (2015). A multimodal approach for extracting content descriptive metadata from lecture videos. *Journal of Intelligent Information Systems*, 46, 121–145. <https://doi.org/10.1007/s10844-015-0356-5> [GS Search]
- Bordes, S. J., et al. (2021). Towards the optimal use of video recordings to support the flipped classroom in medical school basic sciences education. *Medical education online*, 26, 1841406. <https://doi.org/10.1080/10872981.2020.1841406> [GS Search]



- Chen, H. T., & Thomas, M. (2020). Effects of lecture video styles on engagement and learning. *ETRD*, 68, 2147–2164. <https://doi.org/10.1007/s11423-020-09757-6> [GS Search]
- Choe, R., et al. (2019). Student satisfaction and learning outcomes in asynchronous online lecture videos. *CBE Life Sciences Education*, 18. <https://doi.org/10.1187/cbe.18-08-0171> [GS Search]
- Chorianopoulos, K. (2018). A taxonomy of asynchronous instructional video styles. *International Review of Research in Open and Distributed Learning*, 19. <https://doi.org/10.19173/irrodl.v19i1.2920> [GS Search]
- Ciurez, M. A., et al. (2019). Automatic categorization of educational videos according to learning styles. *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 1–6. <https://doi.org/10.23919/SOFTCOM.2019.8903601> [GS Search]
- Crook, C.; Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, 34, 56–64. <https://doi.org/10.1016/j.iheduc.2017.05.003> [GS Search]
- Davila, K. et al. (2021). Fcn-lecturenet: Extractive summarization of whiteboard and chalkboard lecture videos. *IEEE Access*, 9, 104469–104484. <https://doi.org/10.1109/ACCESS.2021.3099427003> [GS Search]
- Davila, K., & Zanibbi, R. (2018). Visual search engine for handwritten and typeset math in lecture videos and latex notes. *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 50–55. <https://doi.org/10.1109/ICFHR-2018.2018.00018427003> [GS Search]
- Oliveira, E. S. et al. (2018). Identificação Automática de Estilos de Aprendizagem: Uma Revisão Sistemática da Literatura. *XXVI Workshop sobre Educação em Computação*. <https://doi.org/10.5753/wei.2018.3488> [GS Search]
- Deng, R., & Benckendorff, P. (2021). What are the key themes associated with the positive learning experience in moocs? an empirical investigation of learners' ratings and reviews. *International Journal of Educational Technology in Higher Education*, 18, 28. <https://doi.org/10.1186/s41239-021-00244-3> [GS Search]
- Gilardi, M., Holroyd, P., Newbury, P., & Watten, P. (2015). The effects of video lecture delivery formats on student engagement. *Science and Information Conference*, 791–796. <https://doi.org/10.1109/SAI.2015.7237234> [GS Search]
- Guo, P., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. *First ACM Conference on Learning @ Scale Conference*, 41–50. <https://doi.org/10.1145/2556325.2566239> [GS Search]
- Hansch, A. et al. (2015). Video and online learning: Critical reflections and findings from the field. *SSRN eLibrary*. <https://doi.org/10.2139/ssrn.2577882> [GS Search]
- Ilioudi, C., Giannakos, M., & Chorianopoulos, K. (2013). Investigating differences among the commonly used video lecture styles. *WAVE Workshop on Analytics on Video-based Learning*.
- Inman, J., & Myers, S. (2018). Now streaming: Strategies that improve video lectures. *IDEA Center, Inc.* [GS Search]
- Jayoma, J. M., Moyon, E. S., & Morales, E. O. (2020). Ocr based document archiving and indexing using pytesseract: A record management system for dswd caraga, philippines. *IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 1–6. <https://doi.org/10.1109/HNICEM51456.2020.9400000> [GS Search]

- Kao, J. L., Chen, S. Y., & Duh, D.J. (2013). Detecting handwritten annotation by synchronization of lecture slides and videos. *Computer Engineering and Applied Computing*, 29. [[GS Search](#)]
- Kota, B. U. et al. (2021). Automated whiteboard lecture video summarization by content region detection and representation. *25th International Conference on Pattern Recognition (ICPR)*, 10704–10711. <https://doi.org/10.1109/ICPR48806.2021.9412386> [[GS Search](#)]
- Köse, E., Taslibeyaz, E., & Karaman, S. (2021). Classification of instructional videos. *Technology, Knowledge and Learning*, 26, 1079-1109. <https://doi.org/10.1007/s10758-021-09530-5> [[GS Search](#)]
- Lackmann, S. et al. (2021). The influence of video format on engagement and performance in online learning. *Brain Sciences*, 11, 128. <https://doi.org/10.3390/brainsci11020128> [[GS Search](#)]
- Lee, G. C. et al. Robust handwriting extraction and lecture video summarization. *Multimedia Tools and Applications*, p. 357-360, 2017. <https://doi.org/10.1007/s11042-016-3353-y> [[GS Search](#)]
- Lin, J. et al. (2019). Automatic knowledge discovery in lecturing videos via deep representation. *IEEE Access*, 7, 33957–33963. <https://doi.org/10.1109/ACCESS.2019.2904046> [[GS Search](#)]
- Lu, X. et al. (2020). Research on the Impacts of Feedback in Instructional Videos on College Students' Attention and Learning Effects. *IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 513-516. <https://doi.org/10.1109/CSCWD49262.2021.9437774> [[GS Search](#)]
- Mayer, Richard E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, 68. <https://doi.org/10.1007/s11423-020-09749-6> [[GS Search](#)]
- Mayer, R., & Moreno, R. (2002). Animation as an Aid to Multimedia Learning. *Educ Psychol Rev.*, 14, 87-99. <https://doi.org/10.1023/A:1013184611077> [[GS Search](#)]
- Ng, Yen Y., & Przybyłek, A. (2021). Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. *IEEE Access*, 9, 36485-36499. <https://doi.org/10.1109/ACCESS.2021.3058735> [[GS Search](#)]
- Ozan, O., & Ozarslan, Y. (2016). Video lecture watching behaviors of learners in online courses. *Educational Media International*, 53, 1-15. <https://doi.org/10.1080/09523987.2016.1189255> [[GS Search](#)]
- Rahim, Muhamad I., & Shamsudin, S. Video Lecture Styles in MOOCs by Malaysian Polytechnics. *3rd International Conference on Education and Multimedia Technology, Association for Computing Machinery*, 64–68. <https://doi.org/10.1145/3345120.3345169> [[GS Search](#)]
- Rawat, Y., Bhatt, C., & Kankanhalli, M. (2014). Mode of teaching based segmentation and annotation of video lectures. *12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, 1-4. <https://doi.org/10.1109/CBMI.2014.6849840> [[GS Search](#)]
- Rosenthal, S., & Walker, Z. (2020). Experiencing Live Composite Video Lectures: Comparisons with Traditional Lectures and Common Video Lecture Methods. *International Journal for the Scholarship of Teaching and Learning*, 14. <https://doi.org/10.20429/ijstol.2020.140108> [[GS Search](#)]
- Sablic, M., Miroslavljević, A., & Škugor, A. (2020). Video-Based Learning (VBL)—Past, Present and Future: an Overview of the Research Published from 2008 to 2019. *Technology, Knowledge and Learning*, 1-17. <https://doi.org/10.1007/s10758-020-09455-5> [[GS Search](#)]

- Santos Espino et al. (2016). Speakers and boards: A survey of instructional video styles in MOOCs. *Technical Communication*, 63, 101-115. [[GS Search](#)]
- Shanmukhaa, G. S., Nandita, S. K., & Kiran, M V. (2020). Construction of knowledge graphs for video lectures. *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 127–131. <https://doi.org/10.1109/ICACCS48705.2020.9074320> [[GS Search](#)]
- Sonia, S., Kumar, P., & Saha, A. (2021). Automatic question-answer generation from video lecture using neural machine translation. *8th International Conference on Signal Processing and Integrated Networks (SPIN)*, 661–665. <https://doi.org/10.1109/SPIN52536.2021.9566139> [[GS Search](#)]
- Stull, T. et al. (2018). Using transparent whiteboards to boost learning from online STEM lectures. *Computers & Education*, 120, 146-159. <https://doi.org/10.1016/j.compedu.2018.02.005> [[GS Search](#)]
- Urala, B. et al. (2018). Automated detection of handwritten whiteboard content in lecture videos for summarization. *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 19–24. <https://doi.org/10.1109/ICFHR-2018.2018.00013> [[GS Search](#)]
- Vegas, S., Juristo, N., & Basili, V. R. (2009). Maturing Software Engineering Knowledge through Classifications: A Case Study on Unit Testing Techniques. *IEEE Transactions on Software Engineering*, 35, 551-565. <https://doi.org/10.1109/TSE.2009.13> [[GS Search](#)]
- Wang, Y. et al. (2019). Research on Learners' Eye Movements for Online Video Courses. *14th International Conference on Computer Science & Education (ICCSE)*, 661-666. <https://doi.org/10.1109/ICCSE.2019.8845375> [[GS Search](#)]
- Xu, F. et al. Content Extraction from Lecture Video via Speaker Action Classification Based on Pose Information. *International Conference on Document Analysis and Recognition (ICDAR)*, 1047-1054. <https://doi.org/10.1109/ICDAR.2019.00171> [[GS Search](#)]
- Yilmaz, A. et al. (2021). Detection and breed classification of cattle using yolo v4 algorithm. *International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 1–4. <https://doi.org/10.1109/INISTA52262.2021.9548440> [[GS Search](#)]
- Yousaf, M. H., Azhar, K., & Sial, H. A. (2015). A novel vision based approach for instructor's performance and behavior analysis. *International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15)*, 1–6. <https://doi.org/10.1109/ICCSPA.2015.7081291> [[GS Search](#)]