

# Classificação Automática de Estilos de Videoaulas

Bernadete Aquino<sup>1</sup>, Eduardo Barrére<sup>1</sup>, Jairo Francisco de Souza<sup>1</sup>

<sup>1</sup>Programa de Pós Graduação em Ciência da Computação  
Universidade Federal de Juiz de Fora (UFJF)  
Juiz de Fora – MG – Brasil

{baquino, eduardo.barrere, jairo.souza}@ice.ufjf.br

**Abstract.** *Much has been done to investigate the effects of video lesson styles on student engagement and learning outcome. However, few studies have sought to automatically classify these styles. Thus, the current study used visual characteristics of the video lessons' styles (presence of people and texts) and different classifiers to evaluate the proposed classification method. This automatic classification can be used by recommendation systems to suggest styles that are more adhering to student preferences and the intended learning outcome. The experiments carried out showed that the classification method used is adequate to the problem, reaching values up to 92% of accuracy.*

**Resumo.** *Muito tem sido feito para investigar os efeitos dos estilos das videoaulas no envolvimento do aluno e no resultado de aprendizagem. Porém, poucos estudos buscaram classificar automaticamente esses estilos. Assim, o estudo atual utilizou características visuais dos estilos das videoaulas (presença de pessoas e textos) e diferentes classificadores para a avaliação do método de classificação proposto. Essa classificação automática poderá ser utilizada por sistemas de recomendação para sugestão de estilos mais aderentes a preferências dos alunos e ao resultado de aprendizagem pretendido. Os experimentos realizados mostraram que o método de classificação utilizado é adequado ao problema, atingindo valores de até 92% de acurácia.*

## 1. Introdução

As videoaulas estão se tornando uma das mídias de aprendizagem mais poderosas que captura e distribui informações, além de fornecer um ambiente de aprendizagem estimulante, onde os alunos podem melhor entender e reter informações [Sablić et al. 2021]. Vídeos educacionais podem ser organizados em relação a diferentes características, como aspectos físicos (*codec* utilizado, tamanho, resolução, *aspect ratio*, etc), conteudistas (tópicos abordados, profundidade/abragência de cada tópico, etc) e pedagógicos (objetivo do vídeo, tipo de conteúdo – exposição de tópicos, simulação, exercício, etc –, estilo projetado para o vídeo, etc). Neste trabalho, investiga-se o estilo de vídeos educacionais. O estilo do vídeo educacional se refere ao principal método de organização visual que é empregado para realizar os objetivos do vídeo e alcançar resultados específicos quando o vídeo é visualizado [Hansch et al. 2015].

Estudos indicam que a escolha do estilo correto da videoaula impulsiona o envolvimento do aluno e os resultados de aprendizagem, sugerindo que, para cada cenário específico, alunos aprendem melhor através de um determinado estilo de

videoaula [Lackmann et al. 2021, Rosenthal and Walker 2020, Ng and Przybyłek 2021, Rahim and Shamsudin 2019]. Já [Chen and Thomas 2020] verificou que os estilos da videoaula afetaram os estudantes de maneira diferente com base em seu nível de conhecimento prévio. Tais pesquisas geralmente se baseiam em evidências de que o aprendizado requer atenção, e certos estilos de videoaula são mais propícios do que outros para ganhar e manter a atenção [Rosenthal and Walker 2020]. Além do desempenho acadêmico, os estilos também podem afetar os alunos psicologicamente [Rahim and Shamsudin 2019] e esses têm fortes preferências por certos estilos [Choe et al. 2019]. Não há necessariamente um estilo ideal e pouca pesquisa foi realizada para analisar todos os diferentes estilos [Lackmann et al. 2021].

Com base nos estudos relatados, há a necessidade de investigar o estilo das videoaulas disponíveis para os alunos. Em [Rahim and Shamsudin 2019], os autores realizaram uma investigação em relação à caracterização dos estilos de vídeo aulas existentes, criando manualmente uma classificação de estilos. Outros autores apresentaram diferentes listagens e definições dos tipos de estilos [Hansch et al. 2015, Crook and Schofield 2017, Chorianopoulos 2018]. Embora não exista um consenso entre os tipos existentes e suas características, existem um subconjunto de estilos de vídeos educacionais que estão presentes na maioria desses estudos e que são mais populares em repositórios de objetos de aprendizado. Pouco esforço, contudo, tem-se dedicado à identificação automática desses estilos. A capacidade de classificar automaticamente os estilos das videoaulas pode servir como uma base sólida para os sistemas que visam fornecer recomendações personalizadas de objetos de aprendizagem com base nos estilos de aprendizagem de um aluno, conforme ilustrado em [de Oliveira et al. 2018, Abreu et al. 2020]. Ainda, a identificação de estilos permite fornecer uma nova dimensão de busca em repositórios de objetos de aprendizado.

O objetivo do presente trabalho é propor uma abordagem para classificação automática de quatro estilos de videoaula (*Talking Head*, *Voice Over Slides*, *Presentation Style* e *Khan-Style*), os quais são alguns dos estilos mais populares segundo a literatura. Para isso, utilizou-se de extração de características visuais desses estilos como a presença de imagens de pessoas e textos ao longo do vídeo, e diferentes métodos de classificação foram empregados para a avaliação do método proposto. Ainda, uma base de dados foi criada e disponibilizada para a comunidade científica, uma vez que não existem bases de dados públicas projetadas especificamente para a classificação de estilos de vídeos educacionais.

O trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. O método utilizado e a análise dos resultados são descritos na Seção 3 e 4, respectivamente. Por fim, a Seção 5 apresenta as considerações finais.

## 2. Trabalhos Relacionados

Diversos artigos tem como foco a extração automática de características das videoaulas [Lin et al. 2019, Shanmukhaa et al. 2020, Barrére et al. 2020, Sonia et al. 2021], porém poucos deles estão relacionados à classificação automática quanto aos estilos da videoaula. Apesar de não tratarem diretamente da classificação dos estilos das videoaulas, os artigos a seguir identificam características presentes em diferentes estilos e que poderiam ser utilizadas para tal propósito.

Tendo como foco a identificação do professor na videoaula, o trabalho de

[Rawat et al. 2014] identificou conceitos como professor falando, professor escrevendo no quadro-negro e professor explicando o slide. Foram identificadas imagens de professor, quadro-negro e slide por meio de recursos quanto as cores dessas imagens. Já [Yousaf et al. 2015] identifica atividades do instrutor usando recursos de reconhecimento de face e estimativa de pose do professor. Enquanto que outros artigos focam na extração do conteúdo escrito na lousa, [Lee et al. 2017] identifica e melhora a qualidade do *frame* que melhor representa o conteúdo, [Davila and Zanibbi 2018] reconhece fórmulas matemáticas presentes na lousa, [Kota et al. 2021, Urala Kota et al. 2018] e [Davila et al. 2021] buscaram resumir o conteúdo escrito na lousa por meio da identificação de conteúdos e *frames* chaves. Já [Ciurez et al. 2019] apresentou um método para classificar os vídeos em diferentes estilos de aprendizagem, com técnicas para calcular a quantidade de texto e imagem de diferentes *frames* do vídeo.

Por outro lado, o trabalho de [Aryal et al. 2018] focou diretamente na classificação automaticamente quanto os estilos de videoaulas, concentra-se em três estilos: *talking head*, *slide* e *code* que é o estilo *screencast* mas apresentando conteúdo de programação de computadores. A abordagem de classificação baseada em imagem foi adotada realizando-se um comparativo entre os modelos de Redes Neurais Convolucionais (CNN), incluindo VGG16, InceptionV3 e ResNet50, onde cada *frame* é classificado em um dos estilos.

O diferencial deste trabalho, em relação às propostas existentes na literatura, está na utilização de diferentes características visuais como a presença de pessoas e textos para identificação automática desses estilos, a realização de testes com diferentes modelos de classificação e a utilização de um conjunto mais abrangente de estilos de videoaulas.

### 3. Materiais e método

O objetivo do sistema é classificar automaticamente o estilo da videoaula. Para isso, foi utilizado um sistema de detecção de objetos para identificar pessoas nas videoaulas e uma ferramenta de reconhecimento óptico de caracteres (OCR) que irá reconhecer e extrair o texto embutido nas imagens. Com esses dados foram extraídos algumas características das videoaulas e então essas características foram utilizadas por modelos de classificação. A Figura 1 mostra a arquitetura em alto nível dessa solução.

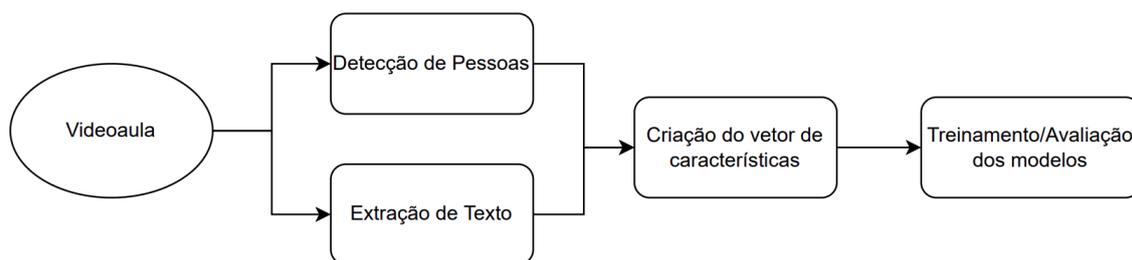


Figura 1. Arquitetura em Alto Nível da Solução

#### 3.1. Escolha dos estilos para o estudo

Dentre os vários estilos de videoaulas definidos na literatura, focou-se apenas em 4 estilos diferentes que são *Talking Head*, *Voice Over Slides*, *Presentation Style* e *Khan-style*.

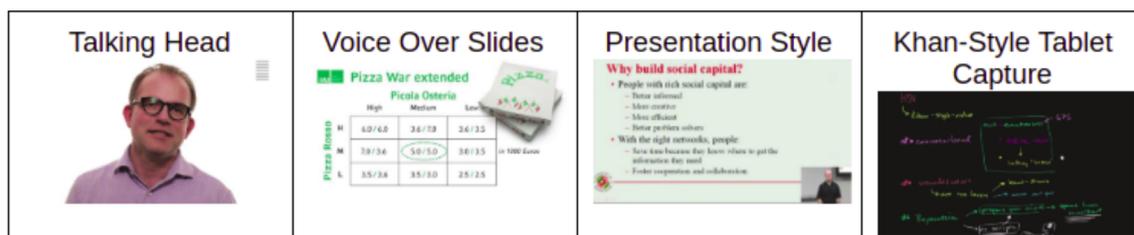


Figura 2. Exemplos dos estilos de videoaula

[Hansch et al. 2015] cita *Talking Head* e *Khan-style* como os estilos mais comumente usados, [Ozan and Ozarslan 2016] cita também a popularidade dos estilos *Voice Over Slides* e *Presentation Style*. A Figura 2 apresenta exemplos desses estilos.

No estilo *Talking Head*, o foco principal é o instrutor, o qual ocupa uma grande área do quadro e fala diretamente para a câmera a maior parte do tempo [Ozan and Ozarslan 2016]. Não está rodeado por *slides* ou outros elementos ricos em texto. Às vezes, o texto sobreposto é colocado para reforçar as ideias-chave da narração ou há mudanças de cena para mostrar outro tipo de material. Essas inserções representam uma quantidade relativamente pequena de tempo de vídeo [Santos Espino et al. 2016]. Por sua vez, o estilo *Voice Over Slides* consiste de slides em PowerPoint ou em outro formato de apresentação visíveis em tela inteira com a narração do instrutor [Inman and Myers 2018]. Em *Presentation Style*, o estilo combina uma apresentação em tela cheia do conteúdo do *slide* com uma projeção menor do vídeo do instrutor (por exemplo, como uma cabeça falante no canto ou lateral do slide) [Rosenthal and Walker 2020]. Por fim, no estilo *Khan-Style* tem-se vídeos em tela cheia de um instrutor escrevendo/desenhando à mão livre em uma lousa digital. É um estilo popularizado pelos vídeos da Khan Academy [Guo et al. 2014].

### 3.2. Construção do Dataset

Não há um dataset público para avaliar estilos de videoaulas. Assim, um conjunto de vídeos do mundo real do domínio da educação foram coletados para conduzir os experimentos e validar a abordagem de classificação. O dataset consiste de 175 vídeos em língua inglesa coletados de plataformas como Khan Academy<sup>1</sup>, Ted Talks<sup>2</sup>, VideoLectures.NET<sup>3</sup> e University of Oxford<sup>4</sup> por estarem licenciados sob uma licença *Creative Commons* e permitem uso não comerciais. A duração total da base é de 40,39 horas. O vídeo mais curto tem 4 minutos enquanto o mais longo tem 52 minutos, com média de 13 minutos de duração. Todos os vídeos mantêm um único estilo durante toda sua duração.

Foram recuperados 35 vídeos para cada estilo (*Khan-Style*, *Talking Head*, *Voice over Slides*, *presentation Style* e Outros). O estilo Outros foi acrescentado para verificar se as soluções de classificação são capazes de identificar vídeos de estilos diversos que não foram rotulados nos quatro estilos anteriores. A classe de Outros, então, representa uma classe de confusão para o modelo, a qual auxilia na verificação da qualidade da solução

<sup>1</sup><http://ko.mujica.org/>

<sup>2</sup><https://www.ted.com/talks>

<sup>3</sup><http://videolectures.net/>

<sup>4</sup><https://podcasts.ox.ac.uk/>

para os quatro estilos escolhidos e permite verificar como as soluções podem se comportar em ambientes reais, quando vídeos de diversos estilos podem estar presentes no repositório. A Tabela 1 mostra a distribuição dos estilos das videoaulas pela área de conhecimento, a grande maioria das videoaulas recuperadas estão no domínio de Ciência da Computação. O dataset utilizado neste estudo está disponível<sup>5</sup> para outros pesquisadores.

**Tabela 1. Distribuição dos estilos de videoaulas por área de conhecimento, onde KS, TH, VO, PS e O representam os estilos *Khan-Style*, *Talking Head*, *Voice Over Slides*, *Presentation Style* e *Outros*, respectivamente.**

Área de Conhecimento	KS	TH	VO	PS	O
Computação	-	2	35	35	17
História	-	1	-	-	5
Jornalismo	-	2	-	-	3
Matemática	10	-	-	-	6
Medicina	18	4	-	-	2
Administração	-	5	-	-	1
Direito	-	6	-	-	-
Psicologia	-	4	-	-	-
Geografia	-	11	-	-	1
Química	6	-	-	-	-
Física	1	-	-	-	-
<b>Total</b>	<b>35</b>	<b>35</b>	<b>35</b>	<b>35</b>	<b>35</b>

### 3.3. Extração de Características

Para a extração das características da videoaula, utilizou-se como base o trabalho de [Chorianopoulos 2018], o qual apresenta os principais fatores (Incorporação Humana e Mídia Instrucional) que definem a classificação do estilo da videoaula bem como os possíveis aspectos para esses fatores que vão do digital (por exemplo, slides) ao físico (por exemplo, lousa). O autor buscou fornecer uma gama de opções de aspectos a serem escolhidos ao se planejar a criação de uma nova videoaula com base em um estilo. Já o presente trabalho utiliza desses aspectos apresentados para realizar a classificação automática. Assim, buscou-se por características que incorporem esses dois fatores, Incorporação Humana através da presença de pessoas e Mídia Instrucional pela presença de textos.

Primeiramente, foi identificado a presença de imagem de pessoas no vídeo. Para isso, utilizou-se a biblioteca do Python ImageAI que fornece uma API para reconhecimento de diferentes objetos usando modelos pré-treinados. *You Only Look Once* (YOLO) é um dos modelos de detecção de objetos mais utilizados [Yilmaz et al. 2021]. O YOLO v3 foi utilizado como modelo de detecção de pessoas neste estudo. Devido a restrições de processamento da máquina, foram analisados 1 *frame* por segundo (FPS) do vídeo. O algoritmo implementado gerou um *json* contendo a quantidade de pessoas identificadas e um array com todos números dos *frames* onde essa quantidade foi identificada.

Em paralelo, ocorreu a extração de texto da videoaula. Para isso, foi utilizado o Python-tesseract, uma ferramenta de reconhecimento óptico de caracteres (OCR) para

<sup>5</sup><https://drive.google.com/drive/folders/168SRUrT0OvtZYKkQndd5dzuGquWvDeSc?usp=sharing>

Python. Essa ferramenta reconhece e extrai o texto embutido nas imagens, é um wrapper para o Tesseract-OCR Engine do Google [Jayoma et al. 2020]. Por meio de sua utilização, foi gerado um segundo *json* contendo as palavras extraídas e o número do *frame* onde foram identificadas. Foram realizados testes com a extração de textos de 1 à 5 *frames* por segundo. Nesse trabalho, foram analisados 5 *frames* por segundo por ser a quantidade que melhor identificou os textos dentro do limite de processamento da máquina utilizada.

À partir da saída dos dois algoritmos, foram extraídas as características tempo de aparição de uma pessoa, tempo de aparição de mais de uma pessoa, tempo de aparição de slides, quantidade de slides, média de palavras inexistentes, média de palavras novas por *frame*. Sendo  $F^{(P1)}$  a quantidade de *frames* do vídeo que apareçam 1 pessoa;  $F^{(P2)}$  a quantidade de *frames* do vídeo que apareçam 2 pessoas ou mais,  $T$  o tempo total do vídeo,  $F$  a quantidade de *frames* analisados do vídeo,  $P^{(In)}$  a quantidade de palavras inexistentes em um *frame*, foram consideradas inexistentes as palavras que não existem em um corpus definido para o idioma. Para isso, foi utilizado o corpus fornecido pela classe *words* da biblioteca *nlk.corpus*,  $F^{(Pal)}$  a quantidade de *frames* com palavras,  $P^{(Novas)}$  a quantidade de palavras novas por *frame*. Assim, as características foram extraídas dos vídeos:

1. Tempo de aparição de uma pessoa, o qual representa o tempo total do vídeo onde é visto uma única pessoa, dado por  $\frac{F^{(P1)} \times T}{F}$ .
2. Tempo de aparição de mais de uma pessoa, o qual retrata o tempo total do vídeo onde é visto mais de uma pessoa, calculado como  $\frac{F^{(P2)} \times T}{F}$ .
3. Tempo de aparição de slides, o qual simboliza o tempo total do vídeo onde se tem slides. Foi considerado slide o *frame* que possui mais de 3 palavras, pois em muitos *frames* apareciam somente a logo da universidade, por isso foram desconsiderados. O tempo é dado por  $\frac{F^{(Pal)} \times T}{F}$ .
4. Quantidade de slides, que corresponde à quantidade de slides apresentados no vídeo. Para essa definição, foi utilizado a similaridade de Jaccard para encontrar a semelhança entre dois conjuntos de termos usados em *frames* subsequentes. Na comparação, foram utilizadas somente as palavras existentes no corpus definido para o idioma. A quantidade foi calculada como  $\sum_{i=1}^F Se(similaridadeDeJaccard() < 70\%, 1, 0)$ .
5. Média de palavras inexistentes: descreve a média da quantidade de palavras identificadas pelo algoritmo de OCR mas que não estão presentes no corpus definido para o idioma. As palavras consideradas inexistentes são as que não estão nesse conjunto. Média calculada como  $\frac{\sum_{i=1}^F P^{(In)}}{F}$ .
6. Média de palavras novas por *frame*: constitui a médias palavras novas acrescentadas por *frame*. A quantidade de palavras novas é a quantidade de palavras de um *frame* removendo-se a quantidade de palavras da interseção com o *frame* anterior. Média dada por  $\frac{\sum_{i=1}^F P^{(Novas)}}{F}$ .

### 3.4. Modelos de Classificação

Utilizou-se diferentes modelos de classificação com o propósito de comparação e descoberta do método que apresenta o melhor resultado. Os modelos adotados foram: *K-Nearest Neighbor* (kNN), *Random Forest*, *Support Vector Machine* (SVM), *Logistic Regression*, *Naive bayes*, *AdaBoost* e uma rede neural. A aplicação e testes dos modelos foram realizadas utilizando o Orange versão 3.32.0.

Foram realizados testes com a normalização dos dados dentro do intervalo de 0 e 1, porém foi a padronização das variáveis com uma média igual a zero que obteve maiores valores de acurácia. A fim de se observar a correlação entre cada par de atributo, foi utilizado o coeficiente de correlação de Spearman. A maior correlação (0,77) ocorreu entre as variáveis “Média de palavras inexistentes” e “Tempo de aparição de slides”, assim nenhuma variável foi desconsiderada para o processo de classificação.

Cada método de classificação tem hiperparâmetros diferentes e alguns desses parâmetros são fundamentais para criar um modelo acurado. Neste sentido, na condução dos testes realizados, alguns experimentos com diferentes valores de parâmetros foram realizados para se obter as melhores configurações para os algoritmos. A Tabela 2 apresenta os valores utilizados no ajuste de parâmetros configuráveis no Orange e o valor selecionado. Os valores dos parâmetros foram escolhidos com base em 5 execuções dos algoritmos, obtendo-se a média da acurácia, a melhor acurácia definiu a escolha do parâmetro. Nos casos em que a acurácia se manteve, deu-se prioridade ao valor menos custoso para o algoritmo. Foi empregue a validação cruzada dividindo-se o conjunto total de dados em 10 subconjuntos.

**Tabela 2. Valores adotados nos experimentos com os métodos de classificação.**

Algoritmo	Hiperparâmetros	Valores Testados	Valor Selecionado
kNN	Nº Vizinhos	5 / 10 / 15	5
	Métrica de Distância	Mahalanobis / Chebyshev / Euclidean / Manhattan	Mahalanobis
<i>Random Forest</i>	Número de arvores	10 / 15 / 20	15
	Profundidade	4 / 5 / 6	5
SVM	Kernel	Linear / Polinomial / RBF / Sigmoid	Linear
<i>Logistic Regression</i>	Tipo de Regularização	Ridge / Lasso	Ridge
	Força	1 / 0.9 / 0.8	1
<i>Naive Bayes</i>	Não possui hiperparâmetros		
<i>AdaBoast</i>	Nº Estimadores da Árvore	40 / 50 / 60	50
	Algoritmo	SAMME.R / SAMME	SAMME.R
<i>Rede neural</i>	Neurônios na Camada Escondida	20 / 25 / 30	25
	Função de Ativação	Identity / Logistic / Tanh / ReLu	Tanh
	Solucionador	SGD / Adam / L-BFGS-B	L-BFGS-B
	Nº Iterações	100 / 150 / 200	150

#### 4. Análise dos resultados

A Tabela 3 apresenta os resultados das avaliações dos 7 modelos analisados quanto a acurácia, F1, precisão, *recall* e área sob a curva ROC. Por meio dela, é possível verificar

que os atributos utilizados são relevantes para a classificação, uma vez que os modelos tiveram alta acurácia e alta área sob a curva ROC.

**Tabela 3. Resultados das avaliações dos diferentes modelos.**

<b>Modelo</b>	<b>Acurácia</b>	<b>F1</b>	<b>Precisão</b>	<b>Recall</b>	<b>Área ROC</b>
<i>Logistic Regression</i>	0,926	0,926	0,926	0,926	0,977
SVM	0,914	0,913	0,919	0,914	0,988
<i>Random Forest</i>	0,914	0,914	0,915	0,914	0,982
kNN	0,88	0,879	0,888	0,88	0,964
<i>Naive Bayes</i>	0,88	0,879	0,883	0,88	0,979
Rede neural	0,863	0,862	0,863	0,863	0,97
AdaBoost	0,857	0,857	0,859	0,857	0,911

O modelo de classificação que apresentou maior acurácia foi o *Logistic Regression* com 92%. A Figura 3 apresenta a matriz de confusão desse classificador. Com ela, foi possível analisar os possíveis motivos das videoaulas terem sido classificadas erroneamente. No estilo *Khan-Style*, ocorreram 2 erros, uma videoaula foi classificada como *Presentation Style* por apresentar uma imagem estática de uma pessoa durante toda a videoaula, outra foi classificada como *Voice Over Slides*, essa videoaula já inicia com um volume grande de texto e novos textos vão sendo acrescentados. Já no estilo *Presentation Style*, ocorreram 3 erros. Em todos esses casos, o algoritmo de identificação de pessoas não conseguiu identificar a imagem do professor na videoaula, deixando a quantidade de pessoas como 0 ou identificando-a em pouquíssimos *frames*, 2 casos que possuíam maior quantidade de textos foram classificados como *Voice Over Slides* e 1 caso com pouquíssimo texto foi classificado como *Khan-Style*. Ao passo que o estilo *Talking Head* teve 2 erros, ambos classificações como Outros, por essas videoaulas apresentarem pequena quantidade de textos em alguns frames. Ao mesmo tempo que o estilo *Voice Over Slides* teve 2 erros, um classificado como *Khan-Style* por apresentar muitas fórmulas matemáticas e tabelas o que fez com que o algoritmos de reconhecimento de texto não identificasse corretamente esses textos e outro como *Presentation Style* por existir imagem de pessoas estáticas na tela. Por outro lado, o estilo Outros teve 4 erros, todos classificados como *Talking Head*, 3 deles pertenciam ao estilo *Classroom Lecture* porém o algoritmo não conseguiu identificar corretamente os textos na lousa, o outro pertence ao estilo *Conversation*, como entrevistador e entrevistado não aparecem juntos na tela, foi considerado a presença de somente uma pessoa.

Com base na observação dos erros dos algoritmos é possível identificar algumas melhorias na identificação dos atributos de classificação. Durante a análise da presença de pessoas, passar a analisar se essa pessoa realiza algum tipo de movimento, pois a videoaula pode conter imagens estáticas de uma pessoa e isso não configurará a presença do professor, ainda identificar a presença de mais de uma pessoa utilizando também o tom de voz, pois como aconteceu na videoaula de estilo *Conversation*, entrevistador e entrevistado não apareciam juntos no mesmo *frame*. Além disso, poderia-se realizar novos testes com as configurações do algoritmo de identificação de pessoas utilizado ou com outros algoritmos pois nem todos os casos onde existiam pessoas nas videoaulas foram identificados. Já na técnica de identificação de textos, o algoritmo teve mais dificuldade em identificar textos escritos na lousa, uma possibilidade seria passar a identificar a própria

	Khan-Style	Outros	Presentation Style	Talking Head	Voice Over Slides	$\Sigma$
Khan-Style	33	0	1	0	1	35
Outros	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

Figura 3. Matriz de confusão do algoritmo *Logistic Regression*

lousa no estilo *Classroom Lecture*.

## 5. Considerações Finais

O estudo de estilos de videoaulas permite a produção de objetos de aprendizagem mais adequados para sua finalidade. A caracterização de estilos de vídeos e sua identificação automática permite que sistemas de recomendação educacionais possam entregar materiais mais adequados aos objetivos do estudante.

Este trabalho apresenta como contribuições: (i) uma abordagem para classificação de estilos de videoaulas, o que permite que novas aplicações em Informática na Educação possam explorar o uso desses vídeos nas suas soluções. Embora outros modelos de classificação possam ser avaliados ou novos experimentos com outros hiperparâmetros possam ser realizados, a abordagem apresentada mostra que as características extraídas dos vídeos são capazes de classificar de forma acurada o conjunto de estilos apresentados neste estudo. Ainda, (ii) o estudo gerou uma base de dados disponível para outros pesquisadores que permite replicar e avançar em pesquisar na área de estilos de videoaulas. Até onde é de conhecimento dos autores, este é o primeiro *dataset* anotado com estilos de videoaulas.

Os resultados demonstram que a abordagem permite identificar de forma satisfatória os estilos de videoaulas selecionados. Contudo, alguns autores [Hansch et al. 2015, Santos Espino et al. 2016, Crook and Schofield 2017, Chorianopoulos 2018, Köse et al. 2021] definem uma variedade de estilos, os quais ainda precisam ser avaliados e novas *features* possivelmente serão necessárias para permitir a classificação de outros estilos. Ainda, experimentos com bases de dados mais volumosas e mais diversificadas em relação ao domínio e características dos vídeos, como duração, tomada de cena, etc, podem mostrar que outros modelos de classificação tenham melhor acurácia que a regressão logística. Embora a abordagem seja independente de idioma, as ferramentas adotadas para produzir texto à partir do vídeo afetam o resultado final da abordagem e precisam ser escolhidas de acordo com o idioma do vídeo. Por fim, uma videoaula pode conter mais de um estilo, porém esses casos não foram tratados no presente estudo.

## Referências

- [Abreu et al. 2020] Abreu, R., Pitangui, C., Andrade, A., Assis, L., and Silva, C. (2020). Detecção automática de estilos de aprendizagem por meio de técnicas de clusterização

- e classificação. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1022–1031, Porto Alegre, RS, Brasil. SBC.
- [Aryal et al. 2018] Aryal, S., Porawagama, A. S., Hasith, M. G. S., Thoradeniya, S. C., Kodagoda, N., and Suriyawansa, K. (2018). Using pre-trained models as feature extractor to classify video styles used in mooc videos. *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 1–5.
- [Barrére et al. 2020] Barrére, E., Souza, J., and Soares, E. (2020). Framework para segmentação temporal de vídeos educacionais. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 972–981, Porto Alegre, RS, Brasil. SBC.
- [Chen and Thomas 2020] Chen, H.-T. M. and Thomas, M. (2020). Effects of lecture video styles on engagement and learning. *Educational Technology Research and Development*, 68(5):2147–2164.
- [Choe et al. 2019] Choe, R., Scuric, Z., Eshkol, E., Cruser, S., Arndt, A., Cox, R., Toma, S., Shapiro, C., Levis-Fitzgerald, M., Barnes, G., and Crosbie, R. (2019). Student satisfaction and learning outcomes in asynchronous online lecture videos. *CBE Life Sciences Education*, 18.
- [Chorianopoulos 2018] Chorianopoulos, K. (2018). A taxonomy of asynchronous instructional video styles. *International Review of Research in Open and Distributed Learning*, 19(1).
- [Ciurez et al. 2019] Ciurez, M. A., Mihaescu, M. C., Giménez, M., Heras, S., Palanca, J., and Julian, V. (2019). Automatic categorization of educational videos according to learning styles. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6.
- [Crook and Schofield 2017] Crook, C. and Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, 34:56–64.
- [Davila et al. 2021] Davila, K., Xu, F., Setlur, S., and Govindaraju, V. (2021). Fcn-lecturenet: Extractive summarization of whiteboard and chalkboard lecture videos. *IEEE Access*, 9:104469–104484.
- [Davila and Zanibbi 2018] Davila, K. and Zanibbi, R. (2018). Visual search engine for handwritten and typeset math in lecture videos and latex notes. *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 50–55.
- [de Oliveira et al. 2018] de Oliveira, E. S., Sales, G. L., de Sousa Pereira, P., and do Nascimento Moreira, R. (2018). Identificação automática de estilos de aprendizagem: Uma revisão sistemática da literatura. In *Anais do XXVI Workshop sobre Educação em Computação*, Porto Alegre, RS, Brasil. SBC.
- [Guo et al. 2014] Guo, P., Kim, J., and Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, pages 41–50.
- [Hansch et al. 2015] Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., and Schmidt, P. (2015). Video and online learning: Critical reflections and findings from the field. *SSRN eLibrary*.

- [Inman and Myers 2018] Inman, J. and Myers, S. (2018). Now streaming: Strategies that improve video lectures. idea paper #68. *IDEA Center, Inc.*
- [Jayoma et al. 2020] Jayoma, J. M., Moyon, E. S., and Morales, E. M. O. (2020). Ocr based document archiving and indexing using pytesseract: A record management system for dswd caraga, philippines. In *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pages 1–6.
- [Kota et al. 2021] Kota, B. U., Stone, A., Davila, K., Setlur, S., and Govindaraju, V. (2021). Automated whiteboard lecture video summarization by content region detection and representation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10704–10711.
- [Köse et al. 2021] Köse, E., Taslibeyaz, E., and Karaman, S. (2021). Classification of instructional videos. *Technology, Knowledge and Learning*, 26.
- [Lackmann et al. 2021] Lackmann, S., Léger, P.-M., Charland, P., Aubé, C., and Talbot, J. (2021). The influence of video format on engagement and performance in online learning. *Brain Sciences*, 11:128.
- [Lee et al. 2017] Lee, G. C., Yeh, F.-H., Chen, Y.-J., and Chang, T.-K. (2017). Robust handwriting extraction and lecture video summarization. *Multimedia Tools and Applications*.
- [Lin et al. 2019] Lin, J., Liu, C., Li, Y., Cui, L., Wang, R., Lu, X., Zhang, Y., and Lian, J. (2019). Automatic knowledge discovery in lecturing videos via deep representation. *IEEE Access*, 7:33957–33963.
- [Ng and Przybyłek 2021] Ng, Y. Y. and Przybyłek, A. (2021). Instructor presence in video lectures: Preliminary findings from an online experiment. *IEEE Access*, 9:36485–36499.
- [Ozan and Ozarslan 2016] Ozan, O. and Ozarslan, Y. (2016). Video lecture watching behaviors of learners in online courses. *Educational Media International*, pages 1–15.
- [Rahim and Shamsudin 2019] Rahim, M. I. and Shamsudin, S. (2019). Video lecture styles in moocs by malaysian polytechnics. In *Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology, ICEMT 2019*, page 64–68. Association for Computing Machinery.
- [Rawat et al. 2014] Rawat, Y., Bhatt, C., and Kankanhalli, M. (2014). Mode of teaching based segmentation and annotation of video lectures. *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–4.
- [Rosenthal and Walker 2020] Rosenthal, S. and 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.
- [Sablić et al. 2021] Sablić, M., Miroslavljević, A., and Škugor, A. (2021). Video-based learning (vbl)—past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning*, 26(4):1061–1077.

- [Santos Espino et al. 2016] Santos Espino, J. M., Afonso-Suárez, M., and Guerra Artal, C. (2016). Speakers and boards: A survey of instructional video styles in moocs. *Technical Communication*, 63:101–115.
- [Shanmukhaa et al. 2020] Shanmukhaa, G. S., Nandita, S. K., and Kiran, M. V. K. (2020). Construction of knowledge graphs for video lectures. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 127–131.
- [Sonia et al. 2021] Sonia, S., Kumar, P., and Saha, A. (2021). Automatic question-answer generation from video lecture using neural machine translation. In *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 661–665.
- [Urala Kota et al. 2018] Urala Kota, B., Davila, K., Stone, A., Setlur, S., and Govindaraju, V. (2018). Automated detection of handwritten whiteboard content in lecture videos for summarization. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 19–24.
- [Yousaf et al. 2015] Yousaf, M. H., Azhar, K., and Sial, H. A. (2015). A novel vision based approach for instructor’s performance and behavior analysis. In *2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA’15)*, pages 1–6.
- [Yılmaz et al. 2021] Yılmaz, A., Nur Uzun, G., Zahid Gürbüz, M., and Kivrak, O. (2021). Detection and breed classification of cattle using yolo v4 algorithm. In *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–4.