

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA
COMPUTAÇÃO**

Bernadete Aparecida da Rocha Aquino

Classificação Automática de Estilos de Videoaulas

Juiz de Fora

2023

Bernadete Aparecida da Rocha Aquino

Classificação Automática de Estilos de Videoaulas

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Dr. Jairo Francisco de Souza

Coorientador: Dr. Eduardo Barrére

Juiz de Fora

2023

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Aquino, Bernadete.

Classificação Automática de Estilos de Videoaulas / Bernadete Aparecida da Rocha Aquino. – 2023.

53 f. : il.

Orientador: Jairo Francisco de Souza

Coorientador: Eduardo Barrére

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciências da Computação, 2023.

1. Videoaula. 2. Estilos de Videoaula. 3. Classificação automática. I. Souza, J. F., orient. II. Barrere, E., coorient. III. Título.

Bernadete Aparecida da Rocha Aquino

Classificação Automática de Estilos de Videoaulas

Dissertação
apresentada ao
Programa de Pós-
graduação em
Ciência da
Computação
da Universidade
Federal de Juiz de
Fora como requisito
parcial à obtenção do
título de Mestre
em Ciência da
Computação. Área de
concentração: Ciência
da Computação.

Aprovada em 19 de janeiro de 2023.

BANCA EXAMINADORA

Prof. Dr. Jairo Francisco de Souza - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Eduardo Barrére - Coorientador

Universidade Federal de Juiz de Fora

Prof. Dr. Victor Stroële de Andrade Menezes

Universidade Federal de Juiz de Fora

Prof. Dr. Carlos de Salles Soares Neto

Universidade Federal do Maranhão

Juiz de Fora, 14/12/2022.



Documento assinado eletronicamente por **Jairo Francisco de Souza, Professor(a)**, em 19/01/2023, às 11:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Victor Stroele de Andrade Menezes, Professor(a)**, em 07/02/2023, às 14:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Barrere, Professor(a)**, em 09/02/2023, às 11:29, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos de Salles Soares Neto, Usuário Externo**, em 07/03/2023, às 17:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1078606** e o código CRC **7942BF8F**.

Dedico este trabalho às minhas sobrinhas e sobrinhos!

AGRADECIMENTOS

Aos meus pais Manoel Aquino e Aparecida Rocha por sempre estarem presentes e incentivarem meus estudos.

Ao meu irmão Baltazar Aquino por todo suporte e inspiração.

Ao meu marido Marco Pereira por me dar muita força desde o momento que decidi participar do processo seletivo, sempre me encorajando a seguir em frente.

Aos coordenadores Fernanda Campos e Mario Dantas pelas conversas e direção quando tive problema de saúde mental. Aos professores Victor Stroële, Regina Braga, Stênio Sã e Luciana Brugiolo por reorganizarem as agendas das disciplinas durante esse período. Por meio desses apoios, pude seguir no mestrado.

Aos professores do PGCC pela contribuição em minha formação. Em especial, aos professores Jairo Souza e Eduardo Barrère por toda orientação, ajuda e conselhos durante todo o período desse trabalho.

Aos meus colegas do mestrado pelo apoio e companheirismo durante o tempo em que estivemos juntos, em especial, Gilson Fonseca e João Paulo Radd que acompanharam todo o processo de desenvolvimento dessa dissertação.

À banca avaliadora, professores Victor Stroële e Carlos Salles, pela disponibilidade e pertinentes apontamentos.

"Quem olha para fora sonha, quem olha para dentro desperta".
Carl G. Jung

RESUMO

Embora as videoaulas sejam frequentemente utilizadas em diversas áreas que abrangem uma ampla gama de estudos e aplicações, a falta de uma abordagem comum para a definição e classificação de seus estilos resulta na utilização de vários modelos diferentes para esses fins. Existem diferentes propostas de classificação desses estilos mas com divergências entre si. Consolidar os estilos de produção permitiria entender as possibilidades de produção de materiais e facilitaria a comunicação entre pesquisadores ou produtores de conteúdo. Para desenvolver essa abordagem comum, existe a necessidade de construir uma estrutura através da qual esses estilos possam ser definidos e classificados. Além disso, muito tem sido feito para investigar os efeitos desses estilos no envolvimento do aluno e no resultado de aprendizagem. Esses estudos sugerem que os estilos de videoaula afetam o desempenho acadêmico e que os alunos aprendam melhor através de um determinado estilo de videoaula em comparação com outros. Com base nisso, os objetivos deste trabalho são propor um modelo unificado para classificação de estilos de videoaulas com base nas nomenclaturas e definições usadas na literatura e propor uma abordagem para classificação automática de 4 estilos de videoaula (*Talking Head*, *Voice Over Slides*, *Presentation Style* e *Khan-Style*), utilizando características visuais desses estilos. O modelo de classificação proposto permite classificar os estilos de vídeo com base em 2 dimensões visuais, incorporação humana e mídia instrucional. Acredita-se que com esse modelo tornará possível a correta caracterização e formação de um entendimento comum de definições de estilos de videoaulas com base em estudos científicos existentes. Já a classificação automática dos estilos poderá ser utilizada por sistemas de recomendação para sugestão de estilos mais aderentes a preferências dos alunos e ao resultado de aprendizagem pretendido. A abordagem apresentada para a classificação automática mostra que as *features* extraídas dos vídeos são capazes de classificar de forma acurada o conjunto de estilos apresentados neste estudo, sendo essa classificação realizada por meio de *features* simples e fáceis de serem extraídas.

Palavras-chave: Videoaula. Estilos de Videoaula. Classificação Automática de vídeos.

ABSTRACT

Although video lessons are often used in diverse areas that cover a wide range of studies and applications, the lack of a common approach to defining and classifying their styles results in the use of many different models for these purposes. There are different proposals for classifying these styles, but with differences between them. Consolidating production styles makes it possible to understand the possibilities of producing materials and facilitates communication between researchers or content producers. To develop this common approach, there is a need to build a framework through which these styles can be defined and classified. In addition, much has been done to investigate the effects of these styles on student engagement and learning outcome. These studies suggest that video lessons styles affect academic performance and that students learn better through a certain video lesson style compared to others. Based on this, the objectives of this work are to propose a unified model for classifying styles of video lessons based on the nomenclatures and definitions used in the literature also, to propose an approach for automatic classification of 4 video lesson styles (Talking Head, Voice Over Slides, Presentation Style and Khan-Style) using visual characteristics of these styles. The proposed classification model allows classifying video styles based on 2 visual dimensions which are human embedding and instructional media. It is believed that this model will make possible the correct characterization and formation of a common understanding of definitions of video lessons styles based on existing scientific studies. The automatic classification of styles can be used by recommendation systems to suggest styles that are more consistent with student preferences and the intended learning outcome. The approach presented for the automatic classification shows that the features extracted from the videos are capable of accurately classifying the set of styles presented in this study and this classification is carried out through simple and easy-to-extract features.

Keywords: Video Lessons. Video Lessons Styles. Video Automatic Classification.

LISTA DE ILUSTRAÇÕES

Figura 1	– Quantidade de artigos por ano de publicação	23
Figura 2	– <i>Frames</i> de exemplos de cada estilo de videoaulas	28
Figura 3	– Modelo de Classificação dos Estilos de Videoaulas	32
Figura 4	– Arquitetura em Alto Nível da Solução	35
Figura 5	– Exemplos dos estilos de videoaula	35
Figura 6	– Matriz de confusão do algoritmo <i>Logistic Regression</i>	41
Figura 7	– Importância das <i>features</i> no modelo <i>Logistic Regression</i>	44
Figura 8	– Matriz de confusão dos modelos de classificação	53

LISTA DE TABELAS

Tabela 1 – Quantidade de artigos retornados por repositório de pesquisa	23
Tabela 2 – Estilos de Videoaulas dos Artigos Analisados	24
Tabela 3 – Exemplos de Classificações de Videoaulas	33
Tabela 4 – Distribuição dos estilos de videoaulas por área de conhecimento, onde KS, TH, VO, PS e O representam os estilos <i>Khan-Style</i> , <i>Talking Head</i> , <i>Voice Over Slides</i> , <i>Presentation Style</i> e Outros, respectivamente	37
Tabela 5 – Valores adotados nos experimentos com os métodos de classificação . .	40
Tabela 6 – Resultados das avaliações dos diferentes modelos	40

LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Networks</i>
FPS	<i>Frame por segundo</i>
kNN	<i>K-Nearest Neighbor</i>
OCR	<i>Optical Character Recognition</i>
RSL	Revisão Sistemática da Literatura
SVM	<i>Support Vector Machine</i>
YOLO	<i>You Only Look Once</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	15
1.2	CONTRIBUIÇÕES	16
1.3	VISÃO GERAL DOS CAPÍTULOS	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	ESTILOS DE VIDEOAULAS	17
2.2	EFEITOS DOS ESTILOS DE VIDEOAULAS NA APRENDIZAGEM	17
2.3	TRABALHOS RELACIONADOS	19
2.3.1	Classificação de estilos de videoaulas	19
2.3.2	Classificação automática de estilos de videoaulas	20
2.3.3	Relevância deste estudo	21
3	UMA PROPOSTA DE CLASSIFICAÇÃO UNIFICADA DE ESTILOS DE VIDEOAULAS	22
3.1	METODOLOGIA	22
3.2	ANÁLISE DAS QUESTÕES DE PESQUISA	24
3.2.1	Q1 - Quais aspectos têm sido utilizados para caracterizar cada estilo de videoaula?	24
3.2.2	Q2 - Como trabalhos que utilizam IA para classificar automaticamente estilos de videoaulas têm caracterizado estes estilos?	30
3.3	MODELO DE CLASSIFICAÇÃO DE ESTILOS DE VIDEOAULAS	31
3.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO	32
4	CLASSIFICAÇÃO AUTOMÁTICA DE ESTILOS DE VIDEOAULAS	34
4.1	MATERIAIS E MÉTODO	34
4.1.1	Escolha dos estilos para o estudo	35
4.1.2	Construção do dataset	36
4.1.3	Extração de características	36
4.1.4	Modelos de Classificação	39
4.2	ANÁLISE DOS RESULTADOS	39
4.3	ANÁLISE DA IMPORTÂNCIA DAS <i>FEATURES</i>	42
4.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO	43
5	CONSIDERAÇÕES FINAIS	45
5.1	CONTRIBUIÇÕES	46
5.2	LIMITAÇÕES	46
5.3	TRABALHOS FUTUROS	47
	REFERÊNCIAS	48

APÊNDICE A – Matriz de Confusão dos Modelos de Classificação	53
--	----

1 INTRODUÇÃO

O século atual é por vezes referido como a era do vídeo na Internet (MAYER et al., 2020). Quando os vídeos possuem foco educacional/instrucional, é comum o uso do termo videoaula, apesar do mesmo não ter sido sugerido por pesquisadores. As videoaulas tornaram-se uma forma importante de aprendizado, especialmente durante o período da pandemia onde o número de alunos aumentou significativamente nas plataformas educacionais onde elas são frequentemente usadas (LU et al., 2020). As videoaulas vêm 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 (SABLIC et al., 2020). O trabalho de (DENG; BENCKENDORFF, 2021) identificou as videoaulas como um dos temas que mais contribuem para experiências positivas de aprendizagem. Sua utilização aumenta a satisfação e o desempenho do aluno devido ao seu potencial de aumentar o aprendizado ativo e centrado no aluno ([BORDES et al., 2021), possibilitam que os alunos revisem o conteúdo ministrado e permitem que sigam seu próprio ritmos de aprendizado, pausando e navegando (BARRÉRE et al., 2020).

As videoaulas são produzidas de diversas maneiras e seus estilos de produção variam substancialmente dentro de uma plataforma. Elas podem ser organizadas em relação a diferentes características, como aspectos físicos (*codec* utilizado, tamanho, resolução, *aspect ratio*, etc), conteudistas (tópicos abordados, profundidade/abrangência 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 os estilos das videoaulas. O estilo de uma videoaula 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 é crucial para o envolvimento do aluno e os resultados de aprendizagem, sugerindo que os alunos aprendam melhor através de um determinado estilo de videoaula em comparação com outros (LACKMANN et al., 2021; ROSENTHAL; WALKER, 2020; NG; PRZYBYŁEK, 2021; RAHIM; SHAMSUDIN, 2019a). 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; WALKER, 2020). Além do desempenho acadêmico, os estilos também podem afetar os alunos psicologicamente (RAHIM; SHAMSUDIN, 2019a) e esses tem fortes preferências por certos estilos (CHOE et al., 2019). Em (CHEN; THOMAS, 2020), os autores mostram que os estilos da videoaula afetam os estudantes de maneira diferente com base em seu nível de conhecimento prévio. Não há necessariamente um estilo ideal e pouca pesquisa foi realizada para analisar todos os diferentes estilos em relação a engajamento e performance (LACKMANN et al., 2021).

Atualmente, os profissionais da área justificam o investimento em formatos mais ricos em multimídia porque eles supostamente aumentam o aprendizado e o engajamento mais do que materiais menos ricos em multimídia (KÖSE et al., 2021). É por isso relevante estudar os ganhos, em termos de engajamento ou realização, que estão associados a esses custos.

Em uma investigação acerca do impacto do uso de algum estilo de videoaula, é necessário, primeiramente, identificar e caracterizar os estilos existentes. Existem diferentes propostas de classificação de estilos de videoaulas onde os autores apresentaram diferentes listagens e definições dos estilos (HANSCH et al., 2015; SANTOS ESPINO et al., 2016; CROOK; SCHOFIELD, 2017; CHORIANOPOULOS, 2018; KÖSE et al., 2021). Embora existam na literatura diferentes classificações dos estilos de videoaulas, na realidade, nenhuma foi aceita por unanimidade. Além disso, os nomes dos estilos geralmente não coincidem, mesmo que descrevam o mesmo estilo de vídeo (ARRUABARRENA et al., 2021). Ainda, essas classificações nem sempre são utilizadas em estudos sobre preferências, benefícios e impactos da utilização ou produção dos diferentes estilos, uma vez que essas classificações apresentam divergências entre si, além de não nomearem, descreverem ou apresentarem capturas de telas que auxiliem na correta caracterização dos estilos. Assim, os resultados encontrados nesses trabalhos são de difíceis interpretação por utilizarem definições muitas vezes conflitantes. Alcançar melhor conhecimento sobre o modelo classificatório pode permitir novas funcionalidades em mecanismos de busca e recomendação para os alunos.

Há também a necessidade de investigar o estilo das videoaulas disponíveis para os alunos. Em (RAHIM; SHAMSUDIN, 2019a), os autores realizaram uma classificação manual dos estilos das videoaulas disponíveis em um repositório privado, utilizando a classificação proposta por (CROOK; SCHOFIELD, 2017). 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.

1.1 OBJETIVOS

Esta dissertação de mestrado tem por objetivos (i) propor um modelo de classificação de estilos de videoaulas unificado; (ii) propor uma abordagem para classificação automática

de quatro estilos de videoaula (*Talking Head*, *Voice Over Slides*, *Presentation Style* e *Khan-Style*), os quais verificou-se ser alguns dos estilos mais populares.

1.2 CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

1. Uma visão geral sobre os estilos de videoaulas através de duas questões de pesquisa: Q1 - "Quais aspectos têm sido utilizados para caracterizar cada estilo de videoaula?" e Q2 - "Como trabalhos que utilizam IA para classificar automaticamente estilos de videoaulas têm caracterizado estes estilos?"
2. A proposta de um modelo de classificação de estilos de videoaulas que visa unificar as definições da literatura.
3. Um sistema para classificação automática do estilo das videoaulas.
4. Um novo *dataset* com um conjunto de vídeos do mundo real do domínio da educação contendo os estilos *Khan-Style*, *Talking Head*, *Voice over Slides*, *presentation Style* e Outros.

1.3 VISÃO GERAL DOS CAPÍTULOS

Esta tese de mestrado está dividida em 5 capítulos:

- **Capítulo 2:** Esse capítulo expõe o conceito sobre os estilos de videoaulas e os possíveis efeitos desses estilos na aprendizagem. Além disso, apresenta os trabalhos encontrados na literatura relacionados aos problemas apresentados nesta dissertação.
- **Capítulo 3:** Uma Revisão Sistemática da Literatura (RSL) foi realizada a fim identificar quais são os estilos de videoaulas já identificados, os aspectos que os caracterizam e como trabalhos que realizam a classificação automática caracterizam esses estilos. Por fim, foi apresentado um modelo de classificação de estilos de videoaulas.
- **Capítulo 4:** Uma classificação automática de estilos de videoaulas é apresentada. O estudo utilizou características visuais dos estilos das videoaulas e diferentes classificadores para a avaliação do método de classificação proposto.
- **Capítulo 5:** Por fim, esse capítulo discute as metodologias e os resultados de cada um dos capítulos anteriores, apresentando as principais contribuições, limitações e perspectivas para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os conceitos fundamentais relacionados à proposta do trabalho. Informações sobre os estilos das videoaulas e seus efeitos na aprendizagem são apresentadas de forma a embasar a proposta. Além disso, estão incluídos os trabalhos relacionados às soluções apresentadas neste trabalho.

2.1 ESTILOS DE VIDEOAULAS

A ampla utilização de videoaulas despertou o interesse de instrutores e acadêmicos em investigar mais sobre o uso desse método. Essa ampla utilização também resultou na criação de diversos estilos de videoaulas (RAHIM;SHAMSUDIN, 2019b). O estilo da videoaula identifica a forma como os aspectos visuais de um vídeo são organizados e apresentados ao público (LACKMANN et al., 2021). A classificação quanto ao estilo concentra-se em fatores visuais da videoaula, como por exemplo a incorporação humana e a inclusão de mídias instrucionais (CHORIANOPOULOS, 2018).

Há a necessidade de uma linguagem comum e compreensível que facilite a definição dos estilos das videoaulas. Essa necessidade é importante não apenas para pesquisadores, mas também para alunos e professores desenvolverem uma visão comum sobre os estilos das videoaulas e serem capazes de selecionar e acessar os materiais de que precisam de forma mais conveniente (KÖSE et al., 2021).

Ao escolher um estilo de videoaula, é importante ter em mente o objetivo do vídeo e os resultados desejados. Diferentes estilos de produção têm diferentes propósitos, por isso é vital que o processo de seleção seja ponderado e intencional (HANSCH et al., 2015). A personalização é importante na criação de vídeos para fins de aprendizado, e é vital que as diferenças cognitivas dos alunos e a modalidade das aulas em vídeo sejam levadas em consideração ao projetar ambientes de aprendizado baseados em vídeo. Investigar a relação entre os estilos de videoaulas e os níveis de atenção dos alunos poderia subsidiar a criação de um modelo de aprendizagem personalizada em videoaulas (KOKOÇ et al., 2020). Os alunos podem ver o vídeo com o estilo de videoaula mais adequado ao seu nível de atenção e conhecimento. Os diferentes estilos de videoaulas são descritos na Subseção **3.2.1**.

2.2 EFEITOS DOS ESTILOS DE VIDEOAULAS NA APRENDIZAGEM

A escolha do estilo da videoaula pode ter grande impacto na capacidade dessa videoaula afetar os objetivos pedagógicos e os resultados de aprendizagem desejados (HANSCH et al., 2015). Diversos estudos examinam o impacto dos diferentes estilos de videoaulas no engajamento e aprendizado. Essas avaliações geralmente são realizadas por

meio de questionários, análise de movimento dos olhos dos alunos e/ou medidas emocionais e cognitivas registradas com instrumentos neurofisiológicos. O trabalho de (CHEN, 2017) fornece evidências de que diferentes estilos de videoaula produzem diferentes tipos de emoções e como a emoção é um fator importante que afeta a aprendizagem este estudo fornece algumas ideias sobre os estilos de videoaulas que os desenvolvedores de MOOC devem priorizar para produzir ambientes de aprendizagem positivos. Foram analisados 10 estilos de videoaulas e concluiu-se que os estilos de videoaula *Picture in Picture*, *Text Overlay*, *Khan Style Tablet Capture*, *ScreenCast* e *Animation* são capazes de induzir emoções positivas e, portanto, devem ser priorizados. Já em (LACKMANN et al., 2021), os autores compararam os estilos *Infographic* e *Lecture capture* e constataram que *Lecture capture* desencadeia maior envolvimento emocional em um período mais curto, enquanto o estilo *Infographic* mantém maior envolvimento emocional e cognitivo por períodos mais longos. Em relação à aprendizagem dos alunos, as videoaulas no estilo *Infographic* contribuem para melhorar significativamente o desempenho em questões difíceis. Além disso, os resultados sugerem uma relação significativa entre o engajamento e o desempenho dos alunos.

Sobre o desempenho de aprendizagem, em (KOKOÇ et al., 2020), os autores identificaram que os desempenhos de aprendizagem dos participantes diferiram significativamente entre o estilo de videoaula e o nível de atenção. Observou-se que tanto para os alunos com alto nível de atenção quanto para os alunos com baixo nível de atenção, o uso de videoaulas do tipo *Picture in Picture* levou a maiores pontuações. Já os autores de (WANG et al., 2019) concluíram que, em comparação com a videoaula no estilo PPT com o mesmo conteúdo, o vídeo no estilo Khan foi melhor para ajudar os alunos a aumentar o envolvimento e melhorar a pontuação. Uma explicação possível é que a função do desenho no *tablet* digital orienta os alunos a focar sua atenção na área que os professores querem que eles prestem atenção. Com a ajuda de tal orientação, os alunos podem filtrar informações importantes a tempo de evitar distrações. Por outro lado, o trabalho de (ROSENTHAL; WALKER, 2020) identificou que o estilo *Live composite* é superior aos estilos *Voiceover* e *Picture in Picture* em termos de atenção, emoção positiva, qualidade percebida e presença do instrutor. Também, não foi encontrado diferenças no aprendizado de curto prazo entre os estilos das videoaulas, porém o método *Live composite* resultou em uma melhor experiência para os alunos. O estudo de (CHEN; THOMAS, 2020) verificou que o estilo *Hand-drawn* foi avaliado como o mais envolvente e que deu maior suporte aos indivíduos com baixo conhecimento prévio do conteúdo.

Já com relação a presença do professor nas videoaulas, o trabalho de (COLLIOT; JAMET, 2018) mostrou que essa presença melhorou significativamente a retenção dos alunos nas explicações faladas, sem atrapalhar seu desempenho em problemas com diagramas. Enquanto que o estudo de (CAO, 2018) identificou que essa presença é percebida pelos alunos e que a satisfação dos alunos e essa percepção são maiores ao aprender cursos complexos no estilo *Chroma Key*, enquanto que o design *Picture in Picture* foi favorável

para cursos mais fáceis.

2.3 TRABALHOS RELACIONADOS

Nessa seção, são apresentados trabalhos encontrados na literatura relacionados aos problemas apresentados nesta dissertação. Na Subseção **2.3.1**, são apresentados propostas de classificação de estilos de videoaulas. Já na Subseção **2.3.2**, são apresentados trabalhos que buscaram classificar automaticamente os diferentes estilos de videoaulas. Enquanto que a Subseção **2.3.3** apresenta a relevância deste trabalho.

2.3.1 Classificação de estilos de videoaulas

A criação de classificações para qualquer campo do conhecimento traz benefícios, como o fornecimento de um conjunto de construções unificadoras que fornecem uma terminologia comum para comunicação, a compreensão das inter-relações e a identificação de lacunas de conhecimento (VEGAS et al., 2009). Algumas propostas de classificação sobre o estilo de videoaulas já foram apresentadas na literatura.

Os trabalhos de (HANSCH et al., 2015) e (CROOK; SCHOFIELD, 2017) apresentaram propostas de classificação de videoaulas como resultados secundários de seus trabalhos. Em (HANSCH et al., 2015), os autores buscaram analisar a padronização do processo de produção de vídeos e sua eficácia como uma ferramenta pedagógica. Nesse processo, foram catalogados 18 estilos. Esse resultado foi baseado em uma revisão da literatura, observação de cursos online e em 12 entrevistas semiestruturadas com profissionais da área de produção de vídeo educacional. Já o trabalho de (CROOK; SCHOFIELD, 2017) buscou identificar como as variações de estilos podem implicar em diferentes experiências para os alunos. Foram identificados 16 estilos através de uma amostra de 200 vídeos de diferentes disciplinas e de diferentes ambientes virtuais de aprendizagem.

Já os artigos de (SANTOS ESPINO et al., 2016) e (CHORIANOPOULOS, 2018) objetivaram organizar os estilos de videoaulas em duas dimensões. Em (SANTOS ESPINO et al., 2016), os autores realizaram uma classificação de acordo com o estratégia de comunicação em duas categorias opostas: centrada no palestrante ou centrada no quadro. Foram identificados 7 estilos mais frequentes por meio de um mapeamento do catálogo de estilos apresentados em (HANSCH et al., 2015). Por outro lado, o trabalho de (CHORIANOPOULOS, 2018) generalizou as duas categorias apresentadas em (SANTOS ESPINO et al., 2016) por considerar que elas não são conflitantes. Tais categorias foram apresentadas em duas dimensões com noções mais amplas e mais comuns nas ciências da aprendizagem: incorporação humana e mídia instrucional, ambas possuem os mesmos limites do digital ao físico, por exemplo, de animação humana à corpo humano inteiro. Foram apresentados 13 estilos descobertos através da revisão da literatura e pela examinação de vídeos educacionais. Porém, esses estilos não foram nomeados, mas

representados por símbolos. Já em (KÖSE et al., 2021), os autores tiveram como principais referências os artigos de (HANSCH et al., 2015; CROOK; SCHOFIELD, 2017) quanto a estilos de videoaula. Também, buscaram ampliar a classificação das videoaulas para além de seu estilo, identificaram novas características como interação (capacidade de parar, avançar ou retroceder o vídeo); conexão (requisito de acesso à internet para reprodução do vídeo); sequência (transições entre as partes do vídeo são estruturadas); componentes (imagem, imagem em movimento, áudio e texto); formato de imagem (2D, 3D); instante (vídeo ao vivo); conteúdo (o assunto do vídeo tem partes individuais e independentes ou não).

2.3.2 Classificação automática de estilos de videoaulas

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 o trabalho de (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. Em (LEE et al., 2017), os autores identificaram e melhoraram a qualidade do *frame* que melhor representa o conteúdo. O artigo de (DAVILA; ZANIBBI, 2018) reconhece fórmulas matemáticas presentes na lousa. Os trabalhos de (KOTA et al., 2021; URALA 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 em (CIUREZ et al., 2019), os autores apresentaram 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 automática quanto aos estilos de videoaulas, concentrando-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 *Convolutional Neural Networks* (CNN), incluindo VGG16, InceptionV3 e ResNet50, onde cada *frame* é classificado em um dos estilos.

2.3.3 Relevância deste estudo

Os trabalhos apresentados na Subseção **2.3.1** tentaram definir diferentes estilos de videoaula ou os organizar em algumas dimensões. Porém, não houve uma definição consistente de cada estilo, apresentando uma descrição detalhada e os aspectos que os caracterizam. A falta de uma definição suficiente para os estilos impede uma avaliação precisa dos resultados das pesquisas. Ainda, são apresentados diferentes nomenclaturas para um mesmo estilo, sendo possível, em alguns casos, o agrupamentos de classes de diferentes autores. A criação de uma classificação detalhada, utilizando uma linguagem comum e compreensível que facilite a definição dos estilos, tornaria possível a correta caracterização dos objetos de estudos de futuras investigações. O presente trabalho pretende preencher essa lacuna ao levantar os aspectos que caracterizam cada estilo e identificar novos estilos com base em uma Revisão Sistemática da Literatura (RSL). Então, propor uma classificação unificada, contribuindo assim para a formação de um entendimento comum de definição de estilos de videoaulas com base em estudos científicos existentes.

Já a Subseção **2.3.2** aborda que diversos artigos focam na extração de características de videoaulas porém pouco deles focam exatamente na classificação automática de estilos de videoaulas. Assim, o diferencial deste trabalho, em relação às propostas existentes na literatura, está na utilização de diferentes características visuais das videoaulas, 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. O resultado desse pesquisa poderá ajudar a fornecer recomendações personalizadas de cursos para um aluno com base no estilo da videoaula, uma vez que diferentes estilos têm diferentes propósitos.

3 UMA PROPOSTA DE CLASSIFICAÇÃO UNIFICADA DE ESTILOS DE VIDEOAULAS

O estilo de um vídeo instrucional pode afetar o envolvimento do aluno e os resultados de aprendizagem. Na literatura, existem diferentes propostas de classificação de estilos de videoaulas mas com divergências entre si. Consolidar os estilos de produção identificável permite entender as possibilidades de produção de materiais e facilita a comunicação entre pesquisadores ou produtores de conteúdo. Este capítulo apresenta um levantamento desses trabalhos a partir de uma Revisão Sistemática da Literatura (RSL) a fim de identificar os estilos de videoaulas e quais aspectos os caracterizam, então propor uma classificação unificada desses estilos. Além disso, apresenta uma investigação de como sistemas de classificação automática têm sido usados para identificar estilos de vídeos educacionais.

Esse capítulo está organizado da seguinte forma: a metodologia utilizada na Revisão Sistemática da Literatura e a análise das questões de pesquisa são descritas na Seção 3.1 e 3.2, respectivamente. Já a Seção 3.3 apresenta o modelo de classificação de estilos de videoaula proposto. Por fim, a Seção 3.4 apresenta as considerações finais do capítulo.

3.1 METODOLOGIA

A RSL foi utilizada para garantir uma cobertura adequada dos trabalhos existentes, abrangendo diferentes visões e revelando pontos de consenso e discordâncias de nomenclaturas, bem como mostrando as lacunas e questões em aberto que permanecem. A Revisão Sistemática da Literatura (RSL) foi embasada no método proposto por KITCHENHAM; CHARTERS (2007), o qual permite a identificação, avaliação e análise de estudos significativos. As questões de pesquisa identificadas neste estudo foram:

1. Quais aspectos têm sido utilizados para caracterizar cada estilo de videoaula?
2. Como trabalhos que utilizam IA para classificar automaticamente estilos de videoaulas têm caracterizado estes estilos?

Foram realizadas buscas por trabalhos apresentados em periódicos ou anais de eventos online de relevância nos repositórios: ACM Digital Library, ERIC Institute of Education Sciences, IEEE Digital Library e Scopus. A *string* usada em todos esses repositórios digitais foi ((“*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*”)).

Os seguintes critérios de inclusão foram utilizados: (1) O artigo utiliza, discute ou apresenta estilos de videoaulas. (2) O artigo está em idioma português ou inglês. Já os critérios de exclusão definidos foram: (1) Não foi possível ter acesso ao artigo de

maneira gratuita. (2) Trabalhos repetidos que apareçam em mais de uma base de busca, considerando apenas o resultado mais recente.

A *string* de busca foi executada nos repositórios de pesquisa considerando dados atualizados até o dia 31 de maio de 2022 e resultando em um total de 1.438 artigos encontrados. A Tabela 1 apresenta a quantidade de artigos retornados por cada repositório de pesquisa.

Tabela 1: Quantidade de artigos retornados por repositório de pesquisa

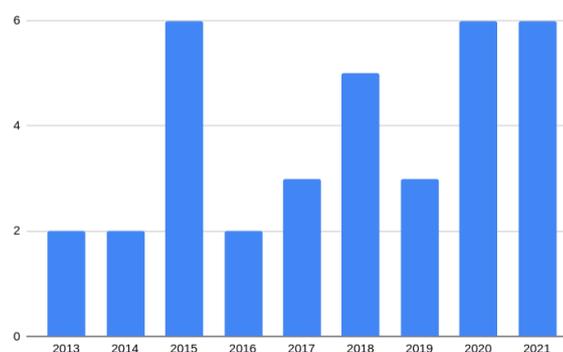
Repositório	Link de Acesso	Quantidade Artigos Retornados
ACM Digital Library	https://dl.acm.org/	338
ERIC	https://eric.ed.gov	196
Scopus	http://www.scopus.com	814
IEEE Digital Library	http://ieeexplore.ieee.org	90

Fonte: Elaborada pela autora (2022).

Na primeira etapa, foi possível eliminar trabalhos que não atendiam ao escopo da pesquisa com a leitura dos títulos dos trabalhos e a análise dos resumos com objetivo de apontar se os trabalhos continham de fato contribuições para a pesquisa. Na segunda etapa, os estudos foram selecionados e/ou rejeitados a partir da leitura completa do artigo e confronto com critérios de inclusão e exclusão. Assim, foram selecionados 35 trabalhos, onde 24 auxiliaram na resposta da primeira questão de pesquisa e 11 contribuíram para resposta da segunda questão de pesquisa. Esses artigos foram utilizados para o preenchimento das lacunas identificadas nos trabalhos relacionados.

A Figura 1 apresenta a distribuição dos 35 artigos utilizados para responder as questões da pesquisa por ano de publicação. Com ela, é possível verificar que os estilos de videoaula vêm sendo mais estudados nos últimos anos.

Figura 1 - Quantidade de artigos por ano de publicação



Fonte: Figura produzida pela autora (2022).

3.2 ANÁLISE DAS QUESTÕES DE PESQUISA

A seguir, são respondidas as duas questões de pesquisa dessa revisão sistemática com a apresentação dos aspectos que caracterizam cada estilo e dos aspectos utilizados nas classificações automáticas de estilos.

3.2.1 Q1 - Quais aspectos têm sido utilizados para caracterizar cada estilo de videoaula?

Com a leitura dos artigos, foi possível verificar que são utilizados diferentes termos para identificar um mesmo estilo de videoaula. Além disso, diferentes dimensões de um estilo são apresentadas em uma mesma lista, sem hierarquia, as tornando extremamente longas e possibilitando a realização de alguns agrupamentos. Por exemplo, no trabalho de (HANSCH et al., 2015), os estilos *Talking Head*, *Webcam Capture*, *On Location* e *Green Screen* possuem muito em comum, focando no instrutor e diferindo apenas na configuração da imagem de fundo do vídeo. Já em (CROOK; SCHOFIELD, 2017), os estilos *Fixed Frame Outside*, *Mobile Frame Outside*, *Fixed but Overlapping*, *Mobile Frame and Overlapping*, *Presence in Split Screen* e *Presence in Picture* focam no instrutor e nos slides, diferindo somente na posição de um em relação ao outro.

Foi identificado também que o nome de classe *Talking Head* foi utilizado para descrever estilos diferentes. Enquanto em (ILIOUDI et al., 2013), foi utilizado para representar *Classroom Lecture*, em (CHOE et al., 2019) foi utilizado para descrever *Presentation Style*. Porém, a grande maioria dos artigos o descreve como sendo o *close-up* de um narrador solitário.

Os dados detalhados da pesquisa estão descritos na Tabela 2 com os artigos recuperados, o detalhamento dos estilos de videoaulas e os nomes de todos os estilos descritos.

Tabela 2: Estilos de Videoaulas dos Artigos Analisados

Estilos Mapeados	Correspondências com Artigos da RSL
<i>Talking Head</i>	<i>Talking Head</i> (SANTOS ESPINO et al., 2016; OZAN; OZARSLAN, 2016)
	<i>Presence In Full Screen</i> ; <i>Presence overlapped by content</i> (CROOK; SCHOFIELD, 2017)
	<i>Talking Head</i> ; <i>Webcam Capture</i> ; <i>On Location</i> ; <i>Green Screen</i> ; <i>Text Overlay</i> (HANSCH et al., 2015)
	<i>Recording Using a Webcam</i> (INMAN; MYERS, 2018)

Continua na próxima página...

Tabela 2 – continuação

Estilos Mapeados	Correspondências com Artigos da RSL
	<i>Studio; Office Desk</i> (GUO et al., 2014)
	<i>Visible Narrator</i> (ARRUABARRENA et al., 2021)
	<i>Talking Head; Text Overlay</i> (CHEN, 2017)
Voice Over Slides	<i>Presentation Slides with Voice-Over</i> (HANSCH et al., 2015; CHEN, 2017)
	<i>Slides</i> (SANTOS ESPINO et al., 2016; GUO et al., 2014)
	<i>Voice Over Slides</i> (CROOK; SCHOFIELD, 2017)
	<i>Voice Over Presentation</i> (CHEN; WU, 2015)
	<i>PowerPoint Presentation With Teacher Narration Only</i> (HEW; LO, 2020)
	<i>Narration-over-PowerPoint</i> (CHEN; THOMAS, 2020)
	<i>Slides On/Off</i> (CHOE et al., 2019)
	<i>Voice-over Slideshow</i> (INMAN; MYERS, 2018)
	<i>Voiceover</i> (ROSENTHAL; WALKER, 2020; KOKOÇ et al., 2020)
	<i>Textual Presentation</i> (ARRUABARRENA et al., 2021)
Presentation Style	<i>PPT-Style</i> (WANG et al., 2019)
	<i>Picture-in-Picture</i> (HANSCH et al., 2015; KOKOÇ et al., 2020; SANTOS et al., 2020; CHEN; WU, 2015; CHEN, 2017)
	<i>Fixed Frame Outside; Mobile Frame Outside; Fixed but Overlapping; Mobile Frame and Overlapping; Presence in Split Screen; Presence in Picture</i> (CROOK; SCHOFIELD, 2017)
	<i>SussexDL</i> (GILARDI et al., 2015)
	<i>Picture-in-Picture; Live Composite</i> (ROSENTHAL; WALKER, 2020)
	<i>Presentation Style</i> (OZAN; OZARSLAN, 2016)
	<i>Talking Head, Weatherman, Pen Tablet</i> (CHOE et al., 2019)
	<i>Picture-in-picture; Overlay mode</i> (BHAT et al., 2015)
	<i>PowerPoint Presentation With Teacher Narration Only</i> (HEW; LO, 2020)
	<i>Picture-in-Picture; Chroma Key</i> (CAO, 2018)
Khan-Style Tablet Capture	<i>Khan-Style Tablet Capture</i> (HANSCH et al., 2015; CHEN, 2017)
	<i>Khan-style Freehand Writing Video</i> (HEW; LO, 2020)
	<i>Khan style</i> (ILIOUDI et al., 2013; GUO et al., 2014; WANG et al., 2019)
	<i>Khan Academy</i> (KÖSE et al., 2021)

Continua na próxima página...

Tabela 2 – continuação

Estilos Mapeados	Correspondências com Artigos da RSL
<i>Hand-drawn</i>	<i>Hand-drawn</i> (CHEN; THOMAS, 2020)
	<i>Udacity Style Tablet Capture</i> (HANSCH et al., 2015; CHEN, 2017)
	<i>Kahn whiteboard</i> (CROOK; SCHOFIELD, 2017)
	<i>Handwriting on a Whiteboard</i> (SANTOS et al., 2020)
<i>Actual Paper / Whiteboard</i>	<i>Actual Paper/Whiteboard</i> (HANSCH et al., 2015; CHEN, 2017)
	<i>Presence Active on Whiteboard</i> (CROOK; SCHOFIELD, 2017)
	<i>Chalk-and-talk</i> (SANTOS et al., 2020)
<i>Screencast</i>	<i>Screencast</i> (HANSCH et al., 2015; SANTOS ESPINO et al., 2016; INMAN; MYERS, 2018; KOKOÇ et al., 2020; SANTOS et al., 2020; GILARDI et al., 2015; CHEN, 2017)
	<i>Voice Over Screencast</i> (CROOK; SCHOFIELD, 2017)
	<i>Code</i> (GUO et al., 2014)
<i>Enhanced Screencast</i>	<i>Enhanced Screencast</i> (GILARDI et al., 2015)
<i>Classroom Lecture</i>	<i>Classroom Lecture</i> (HANSCH et al., 2015; HEW; LO, 2020; CHEN, 2017)
	<i>Presence in Lecture</i> (CROOK; SCHOFIELD, 2017)
	<i>Live Lecture</i> (SANTOS ESPINO et al., 2016)
	<i>Lecture Theatre Recording</i> (GILARDI et al., 2015)
	<i>Lecture Capture</i> (CHEN; WU, 2015; INMAN; MYERS, 2018; ROSENTHAL; WALKER, 2020; LACKMANN et al., 2021)
	<i>Classic Classroom</i> (CHOE et al., 2019)
	<i>Talking Head Style</i> (ILIOUDI et al., 2013)
	<i>Classroom</i> (GUO et al., 2014)
	<i>Classroom environment videos</i> (KÖSE et al., 2021)
<i>Conversation</i>	<i>Interview, Seminar, Conversation, Live Video</i> (HANSCH et al., 2015)
	<i>Presence in Interview, Presence in Discourse</i> (CROOK; SCHOFIELD, 2017)
	<i>In-person lecture</i> (GILARDI et al., 2015)
	<i>Interview</i> (SANTOS ESPINO et al., 2016; OZAN; OZARSLAN, 2016; CHOE et al., 2019)
	<i>Interview, conversation</i> (KÖSE et al., 2021)
<i>Demonstration</i>	<i>Demonstration</i> (HANSCH et al., 2015; MAYER et al., 2020)
	<i>Hands-on Demonstration of a Procedure</i> (SANTOS et al., 2020)

Continua na próxima página...

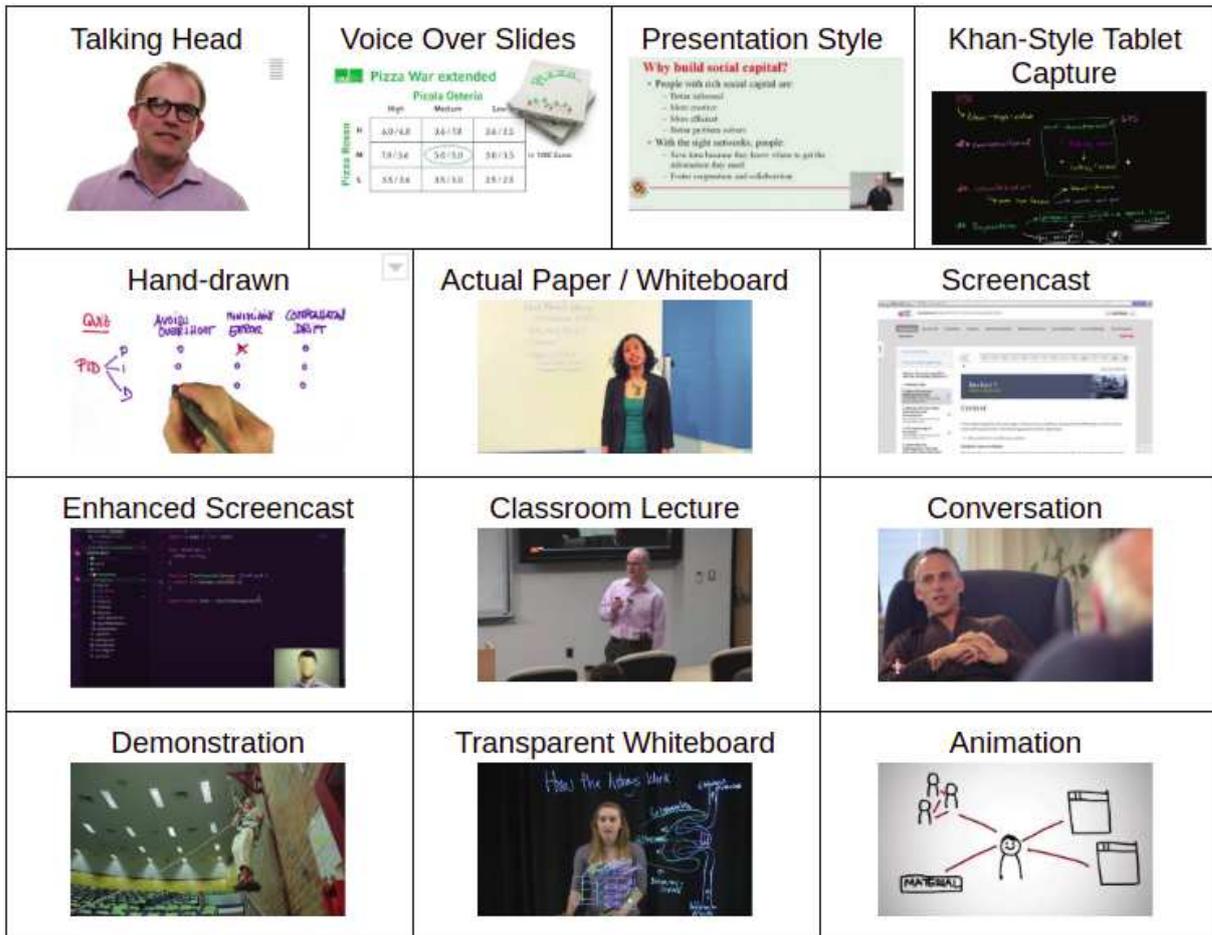
Tabela 2 – continuação

Estilos Mapeados	Correspondências com Artigos da RSL
	<i>Demo</i> (CHOE et al., 2019)
	<i>Demonstration videos</i> (KÖSE et al., 2021)
<i>Transparent Whiteboard</i>	<i>Transparent Whiteboard</i> (MAYER et al., 2020; STULL et al., 2018)
	<i>Virtual Whiteboard</i> (SANTOS ESPINO et al., 2016)
	<i>Learning Glass</i> (CHOE et al., 2019)
<i>Animation</i>	<i>Animation</i> (HANSCH et al., 2015; CHEN, 2017)
	<i>Infographic</i> (LACKMANN et al., 2021)

Fonte: Elaborada pela autora (2022).

Foram realizados dois tipos de agrupamentos: agrupamento de termos diferentes que representam um mesmo estilo ou agrupamento de subdimensões de um mesmo estilo. A fim de alcançar esses agrupamentos, foi criada uma tabela com as descrições de cada estilo por artigo. A partir da leitura de novas descrições, essas eram comparadas com as anteriores a fim de identificar se não representavam um mesmo estilo. Por último, novas leituras da tabela foram realizadas a fim de identificar estilos com muitas características em comum. Para escolha do nome de cada agrupamento, foi priorizada a denominação proposta por (HANSCH et al., 2015), o qual apresenta os estilos de forma mais completa, utilizado o termo mais frequente ou, em último caso, cunhado um novo termo que melhor representa o agrupamento. A Figura 2 apresenta exemplos de cada um dos agrupamentos, os quais são assim descritos:

- ***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; 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). O narrador deve aparecer, esta é a característica marcante deste estilo (ARRUABARRENA et al., 2021). Alguns agrupamentos foram realizados nos trabalhos de (HANSCH et al., 2015) e (GUO et al., 2014) pelas classes diferirem apenas na configuração da cena ou fundo.
- ***Voice Over Slides***: consiste de slides em PowerPoint ou em outro formato de apresentação visível em tela inteira com a narração do instrutor (INMAN; MYERS,

Figura 2 - *Frames* de exemplos de cada estilo de videoaulas

Fonte: Figuras recuperadas dos artigos (HANSCH et al., 2015; MAYER et al., 2020) ou de videoaulas no Youtube.

2018). Uma subclasse desse estilo é *Writing Over Slides*, a qual inclui também a escrita do instrutor.

- ***Presentation Style***: 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; WALKER, 2020) ou com o instrutor sendo fisicamente sobreposto ao slide no momento da gravação (*slides* e apresentador aparecem na tela como uma única unidade) (GILARDI et al., 2015). Algumas classes de (CROOK; SCHOFIELD, 2017; ROSENTHAL; WALKER, 2020) foram agrupadas por apresentarem diferenças somente na posição do instrutor em relação ao *slide*.
- ***Khan-Style Tablet Capture***: é um vídeo 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).

- ***Hand-drawn***: inclui atos manuais de escrita onde a mão do instrutor se torna visível enquanto ele realiza as anotações (CHEN; THOMAS, 2020). No estilo proposto pela Udacity, a mão do instrutor é apresentada de forma transparente para que a escrita não fique obscurecida (HANSCH et al., 2015). Assim, consideramos o estilo Udacity uma subclasse de *Hand-drawn*.
- ***Actual Paper / Whiteboard***: é uma apresentação de monólogo feita enquanto o instrutor se move na frente do conteúdo em um quadro branco e age sobre ele (CROOK; SCHOFIELD, 2017). O instrutor olha diretamente para a câmera, passando a impressão de que se fala diretamente com o espectador.
- ***Screencast***: é a gravação visual da saída da tela de uma sessão de computador. Geralmente, inclui uma narração de voz com uma descrição das ações que estão sendo realizadas (SANTOS ESPINO et al., 2016).
- ***Enhanced Screencast***: a apresentação é gravada como um *screencast* com a adição do instrutor. Isso geralmente ocorre com o rosto ou a metade superior do instrutor mostrado em um canto ou na lateral dos *slides* (GILARDI et al., 2015).
- ***Classroom Lecture***: É a gravação de uma palestra em sala de aula ou em uma conferência com audiência da sala estando visível ou implícita (SANTOS ESPINO et al., 2016). Envolve a gravação de vídeo de uma palestra física (ROSENTHAL; WALKER, 2020). Sempre mantém a percepção de estar sendo gravado em uma única tomada e o instrutor fala diretamente com a audiência.
- ***Conversation***: É uma gravação de um diálogo com um palestrante ou especialista no domínio que responde perguntas ou discute sobre um tópico (OZAN; OZARSLAN, 2016). Alguns agrupamentos foram realizados em (HANSCH et al., 2015; CROOK; SCHOFIELD, 2017) por todos esses estilos estarem centrados em diálogos.
- ***Demonstration***: é a captura do instrutor realizando experimentos orquestrados para ilustrar alguma conceito (CHOE et al., 2019). Permite que o espectador veja um experimento em ação, em vez de ver apenas alguém falando sobre ele (HANSCH et al., 2015).
- ***Transparent Whiteboard***: o instrutor se posiciona atrás de um grande painel de vidro voltado para a câmera enquanto escreve ou desenha no vidro. A câmera inverte a escrita e o desenho do instrutor para que seja legível para o espectador (STULL et al., 2018).
- ***Animation***: refere-se a imagens em movimento geradas por computador mostrando associações entre as figuras desenhadas (MAYER; MORENO, 2002).

3.2.2 Q2 - Como trabalhos que utilizam IA para classificar automaticamente estilos de videoaulas têm caracterizado estes estilos?

Durante a RSL, diversos artigos recuperados tinham como foco a classificação automática de videoaulas, porém poucos deles estavam de alguma forma relacionados com o foco dessa pesquisa: a classificação de estilos da videoaula. Contudo, percebe-se que a tarefa de classificação de videoaulas tem sido objeto de estudo de diversos laboratórios e os trabalhos apresentam diferenças metodológicas que iniciam na definição dos tipos de videoaulas que serão alvo de cada abordagem. Com isso, os resultados desses trabalhos muitas vezes são difíceis de comparar uns com os outros.

Os trabalhos recuperados com base em características de estilos de videoaula se diferenciam em diversos aspectos. Em relação ao foco desta revisão, somente o trabalho realizado por (ARYAL et al., 2018) classificou automaticamente quanto os estilos de videoaulas. Assim, percebe-se que há ainda muito espaço para aplicação da teoria (classificações teóricas) na prática (sistemas reconhedores de estilos).

O trabalho de (ARYAL et al., 2018) 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 *Convolutional Neural Networks* (CNN), incluindo VGG16, InceptionV3 e ResNet50, onde cada *frame* é classificado em um dos estilos. Já o trabalho de (RAWAT et al., 2014) identificou conceitos visuais que podem auxiliar na identificação dos estilos, como professor falando, professor escrevendo no quadro-negro e professor explicando slide. Foram identificados 3 aspectos básicos que são professor, quadro-negro e slide, identificados por meio de recursos quanto as cores das imagens.

Os demais artigos recuperados, apesar de não citarem diretamente um estilo, utilizam-o para extrair automaticamente diferentes características das videoaulas. Quanto ao estilo *Voice Over Slides*, o artigo de (ALI et al., 2021) apresentou melhorias para a indexação de videoaulas e o artigo de (BALASUBRAMANIAN et al., 2015) mostrou uma maneira de resumir os conteúdos apresentados. Com relação ao estilo *Writing Over Slides*, os autores de (KAO et al., 2013) desenvolveram um método para detecção das anotações realizadas nas apresentações dos slides.

Quanto ao estilo *Classroom Lecture*, alguns artigos focaram no professor. Por exemplo, o trabalho de (XU et al., 2019) apresenta uma metodologia para extração de conteúdo de videoaula usando as ações de escrever ou apagar do professor que potencialmente indicam características diretamente relacionadas ao conteúdo da palestra. Já em (YOUSAF et al., 2015), os autores expõem uma abordagem para a avaliação do desempenho e comportamento do professor na sala de aula usando reconhecimento de face e estimativa de pose do professor. Outros artigos concentram-se na extração do conteúdo escrito na lousa (LEE et al., 2017; DAVILA; ZANIBBI, 2018; KOTA et al., 2021; DAVILA

et al., 2021).

Por meio dessa questão de pesquisa, foi possível observar duas lacunas na literatura. A primeira é que os artigos que trabalham com a identificação de características em um determinado estilo de videoaula não fazem uso do estudo teórico e das nomenclaturas definidas por autores que estudam estilos de vídeos educacionais. Em sua maioria, estes trabalhos apresentam apenas descrições pouco detalhadas para caracterizar o estilo que será escopo do seu sistema. Isso dificulta a compilação dos resultados encontrados entre os diferentes artigos. A segunda lacuna diz respeito a falta de estudos na área de IA aplicada à Educação para identificar automaticamente alguns estilos de vídeos educacionais. Somente os estilos *Talking Head*, *Voice Over Slides* e *Screencast* foram identificados automaticamente no trabalho de (ARYAL et al., 2018). A apresentação de soluções para identificação automática de estilos de vídeos educacionais contribuiria para a área de Tutores Inteligentes e Sistemas Adaptativos para recomendação de conteúdo mais aderentes às características de cada aluno e do curso.

3.3 MODELO DE CLASSIFICAÇÃO DE ESTILOS DE VIDEOAULAS

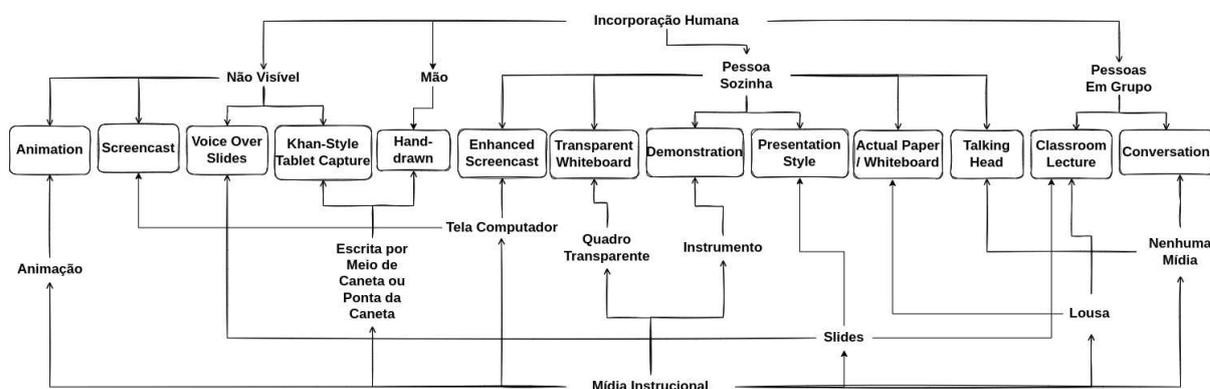
Tendo em vista a importância da definição dos estilos das videoaulas, é proposto um modelo de classificação que visa unificar as definições da literatura. Conforme pontuado por (CROOK; SCHOFIELD, 2017), uma videoaula também pode conter mais de um estilo. Para a definição do modelo, utilizamos como base a taxonomia proposta por (CHORIANOPOULOS, 2018), onde o autor 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. Contudo, o presente trabalho difere do apresentado por (CHORIANOPOULOS, 2018) principalmente por nomear cada estilo de videoaula, incluir novos estilos identificados na literatura e por mapear os estilos com os aspectos que os definem. A classificação foi baseada em itens visuais presentes nas videoaulas e nos estilos encontrados nesta RSL.

A Figura 3 apresenta os aspectos que caracterizam cada estilo de videoaula separados pelos dois fatores¹: Incorporação Humana e Mídia Instrucional. Nessa classificação, a Incorporação Humana pode estar não visível; ser uma pessoa sozinha; pessoas em grupo ou uma mão. Já as Mídias Instrucionais podem referir-se a animação; a escrita por meio de caneta ou ponta da caneta; tela do computador; slides; instrumento; quadro transparente; lousa ou não possuir nenhuma mídia. Em comparação com os atributos apresentados por (CHORIANOPOULOS, 2018), quanto a Incorporação Humana mapeamos os valores

¹ Pode ser visualizada em melhor resolução no link <https://sites.google.com/view/classificacao-estilo-videoaula/>

de pessoas e audiência para pessoas em grupo; instrutor e cabeça falante para pessoa sozinha. Já quanto às Mídias Instrucionais adicionamos a tela de computador e o quadro transparente por pertencerem a novos estilos identificados.

Figura 3 - Modelo de Classificação dos Estilos de Videoaulas



Fonte: Figura produzida pela autora (2022).

Para fins de demonstração do modelo proposto, foram realizadas classificações de videoaulas recuperadas aleatoriamente do repositório público da Universidade de São Paulo². Foram recuperadas 4 videoaulas com estilos diferentes: A³, B⁴, C⁵ e D⁶. A Tabela 3 apresenta exemplos da utilização do modelo. Tendo uma captura de tela da videoaula, é possível atribuir um valor para cada um dos fatores, Incorporação Humana e Mídia Instrucional, dentre os valores apresentados pelo modelo. Com essas informações em mãos, é possível utilizar o modelo para definir qual será o estilo da videoaula.

3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Diferentes palavras são utilizadas na literatura para nomear um mesmo estilo de videoaula o que torna difícil a interpretação e unificação dos resultados encontrados. Nesse capítulo, foi realizada uma RSL a fim de identificar os estilos de videoaulas e quais aspectos os caracterizam. Assim, foi possível propor uma classificação unificada que possibilita a comunicação por meio de uma terminologia comum. Levantou-se também quais aspectos são utilizados na classificação automática das videoaulas quanto ao estilo. Por fim, foi proposto um modelo de classificação de estilos de videoaulas que unifica a definição de cada estilo.

² <https://eaulas.usp.br/portal/search.action>

³ <https://eaulas.usp.br/portal/video.action?idItem=3662>

⁴ <https://eaulas.usp.br/portal/video.action?idItem=1477>

⁵ <https://eaulas.usp.br/portal/video.action?idItem=12412>

⁶ <https://eaulas.usp.br/portal/video.action?idItem=1287>

Tabela 3: Exemplos de Classificações de Videoaulas

	Captura de Tela	Incorporação Humana	Mídia Instrucional	Estilo da Videoaula
A		Mão	Escrita por meio de caneta ou ponta de caneta	<i>Hand-drawn</i>
B		Pessoa Sozinha	Instrumento	<i>Demonstration</i>
C		Pessoa Sozinha	Slides	<i>Presentation Style</i>
D		Pessoas em Grupo	Nenhuma Mídia	<i>Conversation</i>

Fonte: Elaborada pela autora (2022).

Por meio desse estudo, foram identificadas lacunas nesse campo de pesquisa, tais como, existem poucos trabalhos que fazem a classificação automática quanto à estilos de videoaulas e esses não utilizam uma nomenclatura unificada. A classificação unificada e o modelo de classificação que foram apresentados nessa dissertação possibilitam a distinção de cada estilo, facilitando a identificação automática dos mesmos. Com base nisso, o próximo capítulo apresenta uma proposta de classificação automática de estilos de videoaulas, podendo ser utilizado para investigar os estilos de videoaulas disponíveis para os alunos em repositórios de objetos de aprendizado.

4 CLASSIFICAÇÃO AUTOMÁTICA DE ESTILOS DE VIDEOAULAS

Muitos estudos foram realizados para investigar os efeitos dos estilos das videoaulas no envolvimento do aluno e no resultado de aprendizagem. Tais estudos sugerem que os estilos de videoaula afetam o desempenho acadêmico e que os alunos aprendem melhor através de um determinado estilo de videoaula em comparação a outros. Contudo, poucos estudos buscaram classificar automaticamente esses estilos. Diversos estudos tiveram como foco a extração automática de características das videoaulas, porém poucos deles estão relacionados diretamente a classificação automática quanto aos estilos da videoaula. Assim, neste capítulo, propõe-se uma abordagem para classificação automática de 4 estilos de videoaulas utilizando o modelo de classificação proposto na Seção 3.3. Para isso, utilizou-se da extração de características visuais desses estilos quanto aos fatores incorporação humana e mídia instrucional que foram descritos no modelo de classificação proposto, e diferentes métodos de classificação foram empregados para a avaliação do método. 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. 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.

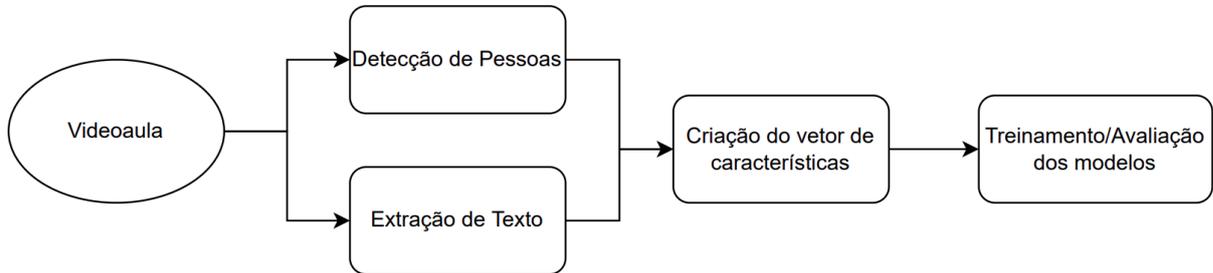
Este capítulo está organizado da seguinte forma: o método utilizado e a análise dos resultados são descritos na Seção 4.1 e 4.2, respectivamente. Por fim, a Seção 4.4 apresenta as considerações finais.

4.1 MATERIAIS E MÉTODO

O processo de classificação de estilos de videoaula apresentado neste trabalho é realizado por meio da identificação dos fatores Incorporação Humana e Mídia Instrucional descritos no modelo de classificação apresentado na Figura 3. A ideia geral é verificar quais atributos que representam esses fatores estão presentes na videoaula e, em seguida, realizar a classificação. O fator Incorporação Humana possui atributos relacionados a presença de pessoas na videoaula, assim para extrair características relacionadas a esses atributos, foi utilizado um sistema de detecção de objetos para identificar pessoas nas videoaulas. Já o fator Mídia Instrucional possui atributos relacionados aos tipos de mídias presentes na videoaula, para sua extração foi utilizada uma ferramenta de *Optical Character Recognition* (OCR) para reconhecer e extrair o texto embutido nas imagens, já que nos estilos selecionados os atributos quanto a Mídia Instrucional se referem a presença de textos no vídeo, como slides e escrita por meio de caneta ou ponta de caneta. Com base nesses dados, foram extraídas algumas características das videoaulas e então essas características foram utilizadas por modelos de classificação. A Figura 4 mostra a arquitetura em alto

nível dessa solução.

Figura 4 - Arquitetura em Alto Nível da Solução

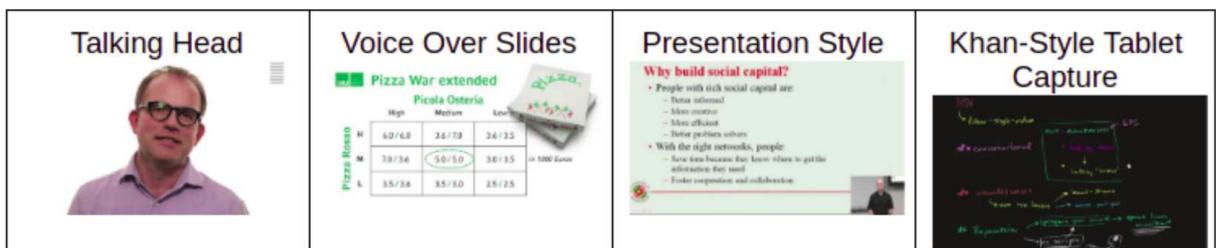


Fonte: Figura produzida pela autora (2022).

4.1.1 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*. O trabalho de (HANSCH et al., 2015) cita *Talking Head* e *Khan-style* como os estilos mais comumente usados. O autor de (WANG et al., 2019) cita o estilo *Voice Over Slides*. Já o artigo (OZAN; OZARSLAN, 2016) além do estilo *Voice Over Slides* cita também a popularidade do estilo *Presentation Style*. Além disso, esses estilos foram alguns dos quais mais apareceram nos 24 artigos recuperados pela RSL que foram utilizados para responder a primeira questão de pesquisa descrita na subseção 3.2.1. Dos 24 artigos recuperados, *Talking Head* aparece em 8, *Voice Over Slides* em 14, *Presentation Style* em 13 e *Khan-style* em 7. A Figura 5 apresenta exemplos desses estilos.

Figura 5 - Exemplos dos estilos de videoaula



Fonte: Figuras recuperadas do artigo (HANSCH et al., 2015)

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; 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; 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; 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).

4.1.2 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 Academy¹, Ted Talks², VideoLectures.NET³ e University of Oxford⁴ por estarem licenciados sob uma licença *Creative Commons* e permitem uso não comercial. 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. Esse grupo contém 20 videoaulas no estilo *Classroom Lecture* e 15 no estilo *Conversation*. A classe de Outros, então, representa uma classe de confusão para o modelo, a qual auxilia na verificação da qualidade da solução 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 4 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 disponível⁵ para outros pesquisadores.

4.1.3 Extração de características

Para a extração das características da videoaula, utilizou-se como base o modelo de classificação proposto no capítulo anterior, o qual apresenta os principais fatores

¹ <http://ko.mujica.org/>

² <https://www.ted.com/talks>

³ <http://videlectures.net/>

⁴ <https://podcasts.ox.ac.uk/>

⁵ <https://drive.google.com/drive/folders/168SRUrT00vtZYKkQndd5dzuGquWvDeSc?usp=sharing>

Tabela 4: 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	-	-	-	-
Total	35	35	35	35	35

Fonte: Elaborada pela autora (2022).

(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 presente capítulo utiliza desses aspectos apresentados para realizar a classificação automática.

Por meio da utilização do modelo de classificação apresentado na Figura 3 do capítulo anterior, é possível verificar que quanto ao fator Incorporação Humana, os estilos *Talking Head* e *Presentation Style* possuem o atributo “pessoa sozinha”, já nos estilos *Khan-Style* e *Voice Over Slides*, esse atributo é “não visível”. Quanto ao fator Mídia Instrucional, os estilos *Voice Over Slides* e *Presentation Style* possuem o atributo “slides”, o estilo *Khan-Style* contém o atributo “escrita por meio de caneta ou ponta de caneta”, enquanto que o estilo *Talking Head* apresenta “nenhuma mídia” nesse fator. Assim, buscou-se por características que incorporem esses dois fatores. O fator Incorporação Humana foi identificado através da presença de pessoas e o fator Mídia Instrucional pela presença de textos uma vez que essa é uma característica de “slides” e “escrita por meio de caneta ou ponta de caneta”.

Primeiramente, foi identificada 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 os 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 *Optical Character Recognition* (OCR) para 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.

A 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 $\frac{\sum_{i=1}^F P^{(Novas)}}{F}$.

Uma vez que o algoritmo de OCR apresentou erros ao reconhecer as palavras escritas à mão, não identificando todas as letras das palavras, foi adicionada a característica 'Média de palavras inexistentes' com o intuito de contabilizar essa característica do algoritmo.

4.1.4 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. Orange é uma ferramenta de código aberto para visualização de dados, aprendizado de máquina e mineração de dados (ARASU et al., 2020).

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 5 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 utilizada a validação cruzada dividindo-se o conjunto total de dados em 10 subconjuntos.

4.2 ANÁLISE DOS RESULTADOS

A Tabela 6 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

Tabela 5: 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 Nº Iterações	SGD / Adam / L-BFGS-B 100 / 150 / 200	L-BFGS-B 150

Fonte: Elaborada pela autora (2022).

que as *features* utilizadas são relevantes para a classificação, uma vez que os modelos tiveram alta acurácia e alta área sob a curva ROC.

Tabela 6: Resultados das avaliações dos diferentes modelos

Modelo	Acurácia	F1	Precisão	Recall	Área ROC
<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
AdaBoast	0,857	0,857	0,859	0,857	0,911

Fonte: Elaborada pela autora (2022).

O modelo de classificação que apresentou maior acurácia foi o *Logistic Regression* com 92%. A Figura 6 apresenta a matriz de confusão desse classificador. As matrizes de confusão de todos os classificadores podem ser consultadas no pênndice A. A partir da análise da matriz de confusão do modelo *Logistic Regression*, 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, ambas 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.

Figura 6 - Matriz de confusão do algoritmo *Logistic Regression*

	Khan-Style	Outros	Presentation Style	Talking Head	Voice Over Slides	Σ
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
Σ	35	33	34	37	36	175

Fonte: Figura produzida pela autora (2022).

Com base na observação dos erros do algoritmo é possível identificar algumas melhorias na identificação das *features* de classificação. A análise da presença de pessoas foi utilizada para a extração das *features* “Tempo de aparição de uma pessoa” e “Tempo de aparição de mais de uma pessoa”, contudo alguns aperfeiçoamentos podem ser realizados nessa análise para a correta definição dessas *features*. Algumas videoaulas apresentaram fotos dos professores no slide de capa ou em videoaulas sobre medicina, foto de pessoas eram utilizadas para a explicação do conteúdo. Essas fotos foram consideradas como a imagem do professor. Assim, passar a analisar se a pessoa realiza algum tipo de movimento pode

garantir a correta definição das *features* sobre o tempo de aparição de pessoas na videoaula, uma vez que as videoaulas podem conter imagens estáticas e isso não configura 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*, assim foi considerado a aparição de somente uma pessoa. Além disso, algumas videoaulas do estilo *Presentation Style* não tiveram a identificação do professor por ele estar em um quadro muito pequeno do vídeo. Logo, poderia-se realizar novos testes com novas configurações do algoritmo de identificação de pessoas utilizado ou com outros algoritmos de reconhecimento de objetos, pois nem todos os casos onde existiam pessoas nas videoaulas foram identificados.

Já a técnica de identificação de textos foi utilizada para a definição das *features* “Tempo de aparição de slides”, “Quantidade de slides”, “Média de palavras inexistentes” e “Média de palavras novas por *frame*”. Em algumas videoaulas do estilo *Classroom Lecture* o algoritmo não foi capaz de identificar o conteúdo escrito na lousa, uma possibilidade seria passar a identificar a própria lousa. Ainda utilizar outras técnicas como, por exemplo, as descritas no artigo de (DAVILA; ZANIBBI, 2018) que reconhece fórmulas matemáticas, pois em alguns *frames* as fórmulas não foram identificadas.

4.3 ANÁLISE DA IMPORTÂNCIA DAS *FEATURES*

A fim de determinar se as previsões realizadas pelo modelo *Logistic Regression* são razoáveis, utilizou-se a avaliação qualitativa do modelo através da importância de cada *feature* para cada classe. Foram gerados gráficos que indicam quais *features* tiveram o maior impacto em cada previsão gerada pela classificação. A Figura 7 apresenta 5 gráficos, onde cada um deles expõe as *features* mais importantes para a definição de um determinado estilo analisado.

Os gráficos da Figura 7 podem ser analisados da seguinte forma: os valores que têm alto impacto na previsão da classe analisada estão à direita e aqueles que votam contra a classe analisada estão à esquerda. A cor do ponto representa o valor da *feature* (vermelho para valores altos e azul para valores baixos). Por exemplo, uma *feature* que tenha mais pontos vermelhos à esquerda significa que valores altos para essa *feature* contribuem contra o valor alvo.

Assim, por meio da Figura 7a, é possível verificar que no estilo *Talking Head*, valores menores para as duas *features* mais importantes “Tempo de Aparição de slides” e “Tempo Aparição Mais de Uma Pessoa” contribuem mais para a definição desse estilo, uma vez que esse estilo não possui slides e somente uma pessoa aparece na videoaula. Já no estilo *Voice Over Slides* apresentado na Figura 7b, as principais *features* foram “Tempo Aparição Uma Pessoa” com contribuição maior de valores menores já que esse estilo não possui a presença do professor e “Tempo de Aparição de slides” com contribuição maior de valores maiores

visto que esse estilo é composto por slides. Enquanto que no estilo *Presentation Style* (Figura 7c), valores maiores para as *features* “Tempo Aparição Uma Pessoa” e “Tempo de Aparição de slides” cooperam para essa definição, visto que esse estilo é composto por slides e pela presença do professor. Por sua vez, no estilo *Khan-Style Tablet Capture*, apresentado na Figura 7d, valores menores para a variável “Tempo Aparição Uma Pessoa” foi mais importante já que esse estilo não apresenta a imagem do instrutor, acredita-se não ter tido uma clara definição dos valores “Tempo de Aparição de slides” pois em algumas videoaulas o algoritmo de OCR não foi capaz de detectar as palavras por completo. Já na Figura 7e que representa a classe de confusão Outros, valores maiores para a variável “Tempo Aparição Mais de Uma Pessoa” foi o que mais contribuiu, dado que o repositório desse estilo contém muitas videoaulas no estilo *Classroom Lecture* e *Conversation* onde é uma característica a videoaula possuir mais de uma pessoa, o mesmo ocorre para a *feature* “Tempo Aparição Uma Pessoa”.

Com base na análise dos gráficos, é possível concluir que o modelo *Logistic Regression* realizou previsões pertinentes sobre os estilos das videoaulas analisadas.

4.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

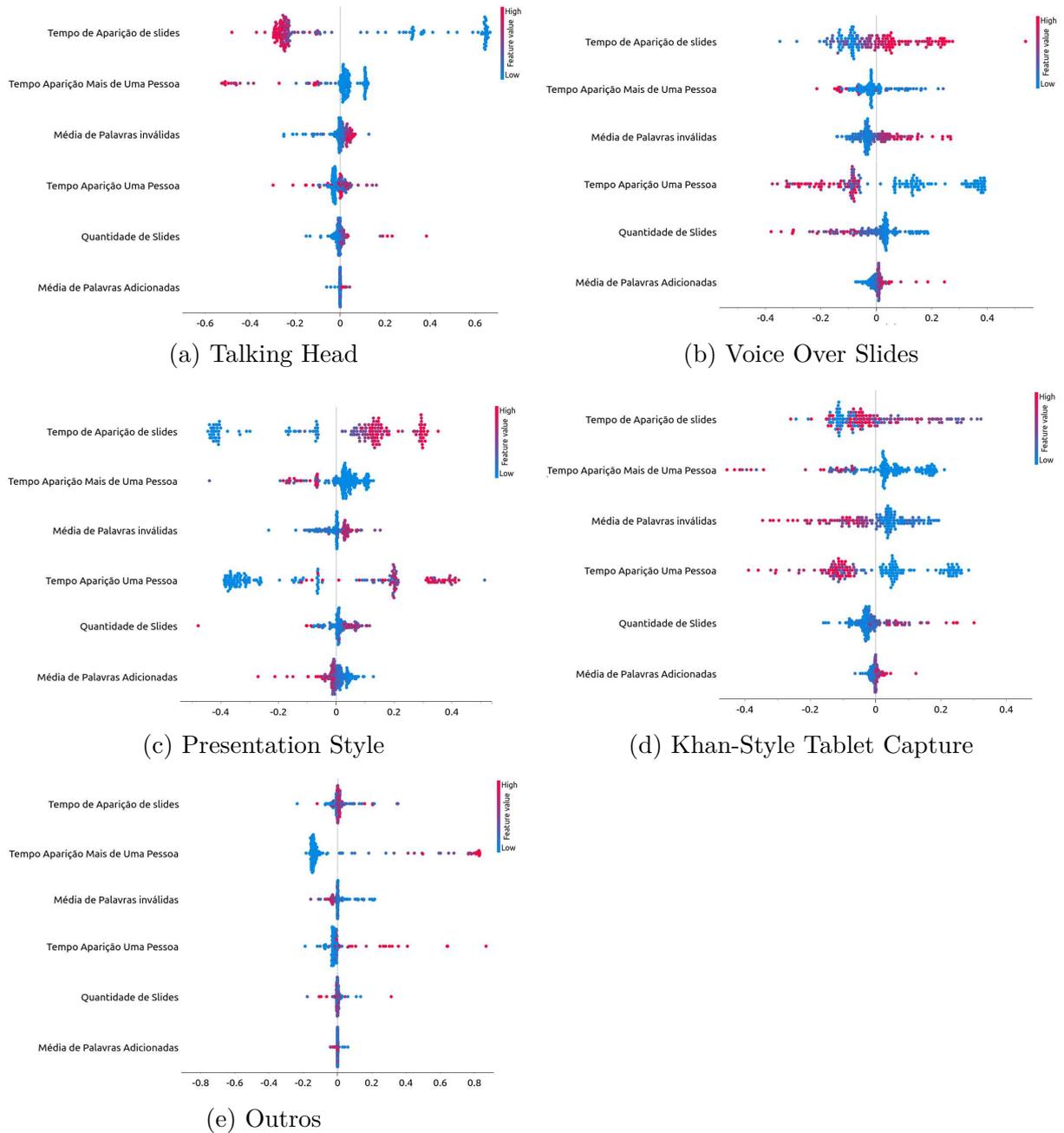
Pesquisadores argumentam que a escolha do estilo de videoaula correto é crucial para o envolvimento dos alunos e resultados de aprendizagem, além de influenciar nos custos de produção do vídeo (LACKMANN et al., 2021). Esse capítulo, apresenta uma proposta de classificação de estilos de videoaulas. A caracterização de estilos e sua identificação automática trazem benefícios para sistemas de recomendação educacionais uma vez que possibilita a entrega de materiais mais adequados aos objetivos do estudante.

Este trabalho mostra que as *features* adotadas na classificação automática são capazes de diferenciar bem os estilos. Essas *features* foram extraídas tendo-se como base a utilização do modelo de classificação proposto no capítulo anterior que utiliza os fatores Incorporação Humana e Mídia Instrucional para a classificação das videoaulas. Além disso, a classificação pôde ser realizada por meio da utilização de *features* simples e fáceis de serem coletadas.

Essa proposta de classificação automática tem boa capacidade de generalização, porém para alguns dos estilos que não foram tratados aqui, novas *features* deverão ser investigadas no futuro. Como, por exemplo, o estilo *Demonstration* possui o fator Incorporação Humana como “pessoa sozinha” e o fator Mídia Instrucional como “instrumento”, o modelo de classificação proposto aqui não extrai *features* relacionadas a apresentação de “instrumento”, o que deve ser coberto por novas *features* em trabalhos posteriores.

Neste estudo foram utilizadas somente videoaulas na língua inglesa. Para sua utilização em outros idiomas como português ou espanhol, não basta apenas recuperar novas videoaulas nesses idiomas, serão necessárias também algumas adaptações durante a

Figura 7 - Importância das *features* no modelo *Logistic Regression*



Fonte: Figura produzida pela autora (2022).

extração das *features*. Por exemplo, será necessário definir o novo idioma como parâmetro do algoritmo de *Optical Character Recognition* (OCR) para a correta identificação do texto e utilizar também um novo corpus definido para o idioma para identificação das palavras inexistentes. Além disso, a aplicação do modelo foi realizada por meio da utilização do Orange e para a sua implantação em produção ainda é necessário esforço de construção de códigos mais eficientes.

5 CONSIDERAÇÕES FINAIS

O conteúdo multimídia e o aprendizado baseado em vídeo têm um papel central no mundo pós-pandemia. Assim, fornecer novas interfaces e serviços avançados que explorem ainda mais seu potencial torna-se de suma importância (TORRE et al., 2022). Embora as videoaulas sejam frequentemente utilizadas em diversas áreas abrangendo uma ampla gama de estudos e aplicações, a falta de uma abordagem comum para a definição e classificação de videoaulas resulta na utilização de diversos modelos diferentes para esses fins (KÖSE et al., 2021). Existem na literatura diferentes classificações para os estilos de videoaula utilizadas na educação, na realidade, nenhuma foi aceita por unanimidade. Os nomes dos estilos geralmente não coincidem, mesmo que descrevam o mesmo estilo de vídeo (ARRUABARRENA et al., 2021). Além disso, estudos recentes indicam que o estilo da videoaula influencia no envolvimento do aluno, engajamento, retenção dos alunos e alguns estilos parecem ter um impacto maior na aprendizagem do que outros (LACKMANN et al., 2021). Com base nisso, torna-se importante a caracterização dos estilos das videoaulas e sua identificação automática permitindo que sistemas de recomendação possam selecionar materiais compatíveis aos objetivos de aprendizagem dos estudantes.

Esse trabalho apresenta uma RSL a fim de identificar os estilos de videoaulas e quais aspectos os caracterizam. Assim, foi possível propor uma classificação unificada que possibilita a comunicação por meio de uma terminologia comum. Levantou-se também quais aspectos são utilizados na classificação automática das videoaulas quanto ao estilo. Com esse levantamento, foi possível identificar duas lacunas na literatura: (i) os artigos que trabalham com essa classificação automática não utilizam as classes de estilo já propostas na literatura, geralmente definem um novo nome para a classe ou descrevem subjetivamente a classe, o que dificulta a compilação de resultados; e (ii) nem todos os estilos de videoaulas possuem estudos para identificá-los automaticamente, o que traria avanço em soluções de ambiente de personalização da educação. Por fim, foi proposto um modelo de classificação de estilos de videoaulas que unifica a definição de cada estilo.

Tendo como base o modelo de classificação proposto foi realizada uma avaliação desse modelo por meio da utilização de diferentes algoritmos classificadores. Focou-se em 4 estilos diferentes *Talking Head*, *Voice Over Slides*, *Presentation Style* e *Khan-style*. Uma classe Outros foi adicionada como uma classe de confusão para o modelo. Embora outros modelos de classificação possam ser avaliados ou novos experimentos com outros hiperparâmetros possam ser realizados, a abordagem apresentada mostra que as *features* extraídas dos vídeos são capazes de classificar de forma acurada o conjunto de estilos apresentados neste estudo. Os resultados demonstram que a abordagem permite identificar de forma satisfatória os estilos de videoaulas selecionados, sendo essa identificação realizada por meio de *features* simples e fáceis de serem extraídas.

5.1 CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

- Uma RSL sobre os estilos de videoaulas. Esse estudo apresenta uma visão geral de quais aspectos têm sido utilizados para caracterizar cada estilo de videoaula e uma discussão de como trabalhos recentes que têm tentado aplicar técnicas de IA para classificar automaticamente vídeos educacionais definem os estilos de videoaulas muitas vezes distantes das definidas na literatura teórica da área.
- Uma proposta de modelo de classificação de estilos de videoaulas baseado em características visuais das videoaulas que objetiva unificar as definições da literatura e mapear os estilos com base nos aspectos que os definem. Essa proposta de classificação oferece contribuições práticas aos pesquisadores ao permitir que classifiquem vídeos de acordo com os aspectos elencados nos dois diferentes fatores (Incorporação Humana e Mídia Instrucional), o que facilitará e agilizará o acesso aos tipos de vídeo.
- Uma abordagem para classificação automática de estilos de videoaulas com base em *features* simples e fáceis de serem extraídas, o que permite que novas aplicações em Informática na Educação possam explorar o uso desses vídeos nas suas soluções uma vez que estudos apontam que os alunos aprendam melhor através de um determinado estilo de videoaula em comparação com outros (LACKMANN et al., 2021; ROSENTHAL; WALKER, 2020; NG; PRZYBYŁEK, 2021; RAHIM; SHAMSUDIN, 2019a)
- Um novo *dataset* com um conjunto de vídeos do mundo real do domínio da educação contendo os estilos *Khan-Style*, *Talking Head*, *Voice over Slides*, *presentation Style* e Outros disponível para os pesquisadores que permite replicar e avançar em pesquisar na área de estilos de videoaulas. Até onde é de conhecimento da autora, este é o primeiro *dataset* anotado com estilos de videoaulas.
- A publicação de três artigos durante os estudos para essa dissertação, os quais mostram a trajetória para essa dissertação, mesmo que não sendo todos relacionados com o tema aqui discutido. Estes artigos discutem a geração de sequências curriculares adaptativas por meio da utilização de uma abordagem multiobjetivo AQUINO et al. (2020), o engajamento dos alunos por meio da detecção de comunidades AQUINO et al. (2020) e a abordagem para classificação automática de estilos de videoaula AQUINO et al. (2022) descrita no Capítulo 4.

5.2 LIMITAÇÕES

As limitações desse trabalho podem ser divididas em três partes: (i) condução da RSL, (ii) usabilidade do modelo de classificação, e (iii) Execução da classificação

automática.

Condução da RSL. A RSL foi pautada em um protocolo robusto, porém passível de exclusão de algum trabalho relevante por não corresponder a um dos critérios de seleção, seja por não conter os termos de busca nas palavras-chave, títulos ou resumos ou por decisões subjetivas ocorridas durante o processo de leitura para inclusão do artigo. Além disso, este trabalho depende da representatividade dos repositórios selecionados para responder às questões de pesquisa.

Usabilidade do modelo de classificação. O modelo não foi avaliado por especialistas da área de Tecnologias Educacionais. Possíveis deficiências no modelo de classificação proposto podem ser eliminadas por meio da examinação do trabalho pela comunidade científica ou por meio de uma validação manual do modelo classificação, através de um procedimento de classificação de um número significativo de videoaulas com base no modelo proposto.

Execução da classificação automática. O desempenho dos algoritmos de classificação está diretamente relacionado a uma escolha adequada de parâmetros de projeto. Contudo, esses parâmetros foram selecionados de maneira empírica. 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. Além disso, uma mesma videoaula pode conter mais de um estilo, porém essa classificação foi realizada utilizando videoaulas que possuem somente um estilo. Por fim, o *dataset* criado não possui videoaulas sobre todas as áreas de conhecimento e algumas dessas áreas estão com um número desbalanceado de videoaulas representando cada estilo, o que pode causar algum viés na classificação.

5.3 TRABALHOS FUTUROS

Este trabalho abre um leque de possibilidades para pesquisas futuras. O modelo de classificação proposto pode ser melhor avaliado em análises futuras por especialistas da área de Tecnologias Educacionais, garantindo assim sua compreensão sem ambiguidades. Além disso, incluir uma validação manual do modelo, através de um procedimento de classificação de um número significativo de videoaulas.

Existe uma variedade de estilos, os quais não foram incluídos na classificação automática deste estudo e que ainda precisam ser avaliados. Novas *features* possivelmente serão necessárias para permitir a classificação desses 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. Por fim, uma videoaula pode conter mais de um estilo, porém esses casos não foram tratados no presente trabalho.

REFERÊNCIAS

- ABREU, R. et al. Detecção automática de estilos de aprendizagem por meio de técnicas de clusterização e classificação. **XXXI Simpósio Brasileiro de Informática na Educação**, Porto Alegre: Sociedade Brasileira de Computação, p. 1022–1031, 2020.
- ALI, M. M. et al. Segmenting lecture video into partitions by analyzing the contents of video. **International Conference on Data Analytics for Business and Industry (ICDABI)**, p. 191–196, 2021.
- AQUINO, Bernadete; BARRÉRE, Eduardo; DE SOUZA, Jairo Francisco. Classificação Automática de Estilos de Videoaulas. **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 33. , 2022, Manaus. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, p. 956-967, 2022.
- AQUINO, Bernadete; SOUZA, Jairo; GONÇALVES, Luciana Brugiolo; SOARES, Stênio São Rosário Furtado. Geração de Sequências Curriculares Adaptativas baseada no Perfil dos Alunos e Materiais Didáticos utilizando o Algoritmo NSGA-III. **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 31. , 2020, Online. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, p. 902-911, 2020.
- AQUINO, Bernadete; STROELE, Victor; SOUZA, Jairo. Análise do Engajamento dos Alunos em Ambientes Virtuais de Aprendizagem para detecção de comunidade. **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 31. , 2020, Online. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020 . p. 952-961.
- ARASU, B. Senthil et al. A machine learning-based approach to enhancing social media marketing. **Computers & Electrical Engineering**, v. 86, p. 106723, 2020.
- ARRUABARRENA, R. et al. A novel taxonomy of student-generated video styles. **International Journal of Educational Technology in Higher Education**, v. 18, p. 68, 2021.
- ARYAL, S. et al. 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)**, p. 1–5, 2018.
- BALASUBRAMANIAN, V. et al. A multimodal approach for extracting content descriptive metadata from lecture videos. **Journal of Intelligent Information Systems**, v.46, p. 121–145, 2015.
- BARRÉRE, E. et al. Framework para segmentação temporal de vídeos educacionais. **XXXI Simpósio Brasileiro de Informática na Educação**, Porto Alegre: Sociedade Brasileira de Computação, p. 972–981, 2020.
- BHAT, S. et al. Seeing the instructor in two video styles: Preferences and patterns. **8th International Conference on Educational Data Mining (EDM)**, Madri, 2015.
- BORDES, S. et al. Towards the optimal use of video recordings to support the flipped classroom in medical school basic sciences education. **Medical education online**, v. 26, p. 1841406, 2021.

- CAO, J. et al. Effect of instructor inclusion type and course complexity on students' learning in lecturer video. **International Conference on Education Technology and Computers (ICETC '18)**. p. 181–188, 2018.
- CHEN, Chwen Jen et al. MOOC Videos-Derived Emotions. **Journal of Telecommunication, Electronic and Computer Engineering**, p. 137-140, 2017.
- CHEN, H. T.; THOMAS, M. Effects of lecture video styles on engagement and learning. **ETRD**, v. 68, p. 2147–2164, 2020.
- CHEN, C.-M.; WU, C.-H. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. **Computers & Education**, v. 80, p. 108–121, 2015.
- CHOE, R. et al. Student satisfaction and learning outcomes in asynchronous online lecture videos. **CBE Life Sciences Education**, v. 18, 2019.
- CHORIANOPOULOS, K. A taxonomy of asynchronous instructional video styles. **International Review of Research in Open and Distributed Learning**, v. 19(1), 2018.
- CIUREZ, M. A. et al. Automatic categorization of educational videos according to learning styles. **International Conference on Software, Telecommunications and Computer Networks (SoftCOM)**, p. 1–6, 2019.
- COLLIOT, T.; JAMET, É. Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study. **Education Tech Research Dev** **66**, p. 1415–1433, 2018.
- CROOK, C.; SCHOFIELD, L. The video lecture. **The Internet and Higher Education**, v. 34, p. 56–64, 2017.
- DAVILA, K. et al. Fcn-lecturenet: Extractive summarization of whiteboard and chalkboard lecture videos. **IEEE Access**, v.9, p. 104469–104484, 2021.
- DAVILA, K.; ZANIBBI, R. Visual search engine for handwritten and typeset math in lecture videos and latex notes. **16th International Conference on Frontiers in Handwriting Recognition (ICFHR)**, p. 50–55, 2018.
- DE OLIVEIRA, E. S. et al. Identificação Automática de Estilos de Aprendizagem: Uma Revisão Sistemática da Literatura. **XXVI Workshop sobre Educação em Computação**, Porto Alegre: Sociedade Brasileira de Computação, 2018.
- 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**, v.18, p. 28, 2021.
- GILARDI, M. et al. The effects of video lecture delivery formats on student engagement. **Science and Information Conference**, p. 791–796, 2015.
- GUO, P. et al. How video production affects student engagement: An empirical study of mooc videos. **First ACM Conference on Learning @ Scale Conference**, p. 41–50, 2014.

HANSCH, A. et al. Video and online learning: Critical reflections and findings from the field. **SSRN eLibrary**, 2015.

HEW, K. F.; LO, C. K. Comparing video styles and study strategies during video-recorded lectures: effects on secondary school mathematics students' preference and learning. **Interactive Learning Environments**, v. 28(7), p. 847–864, 2020.

ILIOUDI, C. et al. Investigating differences among the commonly used video lecture styles. **WAVE Workshop on Analytics on Video-based Learning**, 2013.

INMAN, J.; MYERS, S. Now streaming: Strategies that improve video lectures. **IDEA Center**, Inc, 2018.

JAYOMA, J. M. et al. 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)**, p. 1–6, 2020.

KAO, J.-L. et al. Detecting handwritten annotation by synchronization of lecture slides and videos. **Computer Engineering and Applied Computing**, v. 29, 2013.

KARTHIKRAJ, V. et al. Attention Span Detection for Online Lectures. **International Conference on Advances in Computing and Communications (ICACC)**, p. 1-6, 2021.

KITCHENHAM, B. A.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, **Keele University and Durham University Joint Report**, v.2, 2007.

KOKOÇ, M. et al. Effects of sustained attention and video lecture types on learning performances. **Educational Technology Research and Development**, v.68, p. 3015-3039, 2020.

KOTA, B. U. et al. Automated whiteboard lecture video summarization by content region detection and representation. **25th International Conference on Pattern Recognition (ICPR)**, p. 10704–10711, 2021.

KÖSE, E. et al. Classification of instructional videos. **Technology, Knowledge and Learning**, v. 26, p. 1079-1109, 2021.

LACKMANN, S. et al. The influence of video format on engagement and performance in online learning. **Brain Sciences**, v. 11, p. 128, 2021.

LEE, G. C. et al. Robust handwriting extraction and lecture video summarization. **Multimedia Tools and Applications**, p. 357-360, 2017.

LIN, J. et al. Automatic knowledge discovery in lecturing videos via deep representation. **IEEE Access**, v. 7, p. 33957–33963, 2019.

LU, X. et al. 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)**, p. 513-516, 2021.

- MAYER, R. E. et al. Five ways to increase the effectiveness of instructional video. **Educational Technology Research and Development**, v. 68, 2020.
- MAYER, RICHARD; MORENO, ROXANA. Animation as an Aid to Multimedia Learning. **Educ Psychol Rev.** v.14, p 87-99, 2002.
- NG, YEN YING; PRZYBYŁEK, ADAM. Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. **IEEE Access**, v. 9, p. 36485-36499, 2021.
- OZAN, OZLEM; OZARSLAN, YASIN. Video lecture watching behaviors of learners in online courses. **Educational Media International**, v. 53, p. 1-15, 2016.
- RAHIM, MUHAMAD IZZAT; SHAMSUDIN, SARIMAH. Video Lecture Styles in MOOCs by Malaysian Polytechnics. **3rd International Conference on Education and Multimedia Technology, Association for Computing Machinery**, New York, NY, USA, p. 64–68, 2019.
- RAHIM, M. I.; SHAMSUDIN, S. Categorisation of Video Lecture Designs in Mooc for Technical and Vocational Education and Training Educators. **Journal of Technical Education and Training**, [S. l.], v. 11, n. 4, 2019.
- RAWAT, Y. et al. Mode of teaching based segmentation and annotation of video lectures. **12th International Workshop on Content-Based Multimedia Indexing (CBMI)**, p. 1-4, 2014.
- RODRIGUES, A. et al. Inconsistências no processo de recomendação de objetos de aprendizagem utilizando teoria de estilos de aprendizagem. **XXXI Simpósio Brasileiro de Informática na Educação**, Porto Alegre: Sociedade Brasileira de Computação, p. 1303–1312, 2020.
- ROSENTHAL, SONNY; WALKER, ZACHARY. Experiencing Live Composite Video Lectures: Comparisons with Traditional Lectures and Common Video Lecture Methods. **International Journal for the Scholarship of Teaching and Learning**, v. 14, 2020.
- SABLIC, M. et al. Video-Based Learning (VBL)—Past, Present and Future: an Overview of the Research Published from 2008 to 2019. **Technology, Knowledge and Learning**, p. 1-17, 2020.
- SANTOS, E. et al. Video for teaching: classroom use, instructor self-production and teachers' preferences in presentation format. **Technology Pedagogy and Education**, v. 29, 2020.
- SANTOS ESPINO et al. Speakers and boards: A survey of instructional video styles in MOOCs. **Technical Communication**, v.63, p. 101-115, 2016.
- SHANMUKHAA, G. S. et al. Construction of knowledge graphs for video lectures. **6th International Conference on Advanced Computing and Communication Systems (ICACCS)**, p. 127–131, 2020.
- SONIA, S. et al. Automatic question-answer generation from video lecture using neural machine translation. **8th International Conference on Signal Processing and Integrated Networks (SPIN)**, p. 661–665, 2021.

STULL, T. et al. Using transparent whiteboards to boost learning from online STEM lectures. **Computers & Education**, v. 120, p. 146-159, 2018.

TORRE, I. et al. Video augmentation to support video-based learning. **International Conference on Advanced Visual Interfaces (AVI 2022)**, p. 1–5, 2022.

URALA, B. et al. Automated detection of handwritten whiteboard content in lecture videos for summarization. **16th International Conference on Frontiers in Handwriting Recognition (ICFHR)**, p. 19–24, 2018.

VEGAS, S. et al. Maturing Software Engineering Knowledge through Classifications: A Case Study on Unit Testing Techniques. **IEEE Transactions on Software Engineering**, v.35, p. 551-565, 2009.

WANG, Y. et al. Research on Learners' Eye Movements for Online Video Courses. **14th International Conference on Computer Science & Education (ICCSE)**, p. 661-666, 2019.

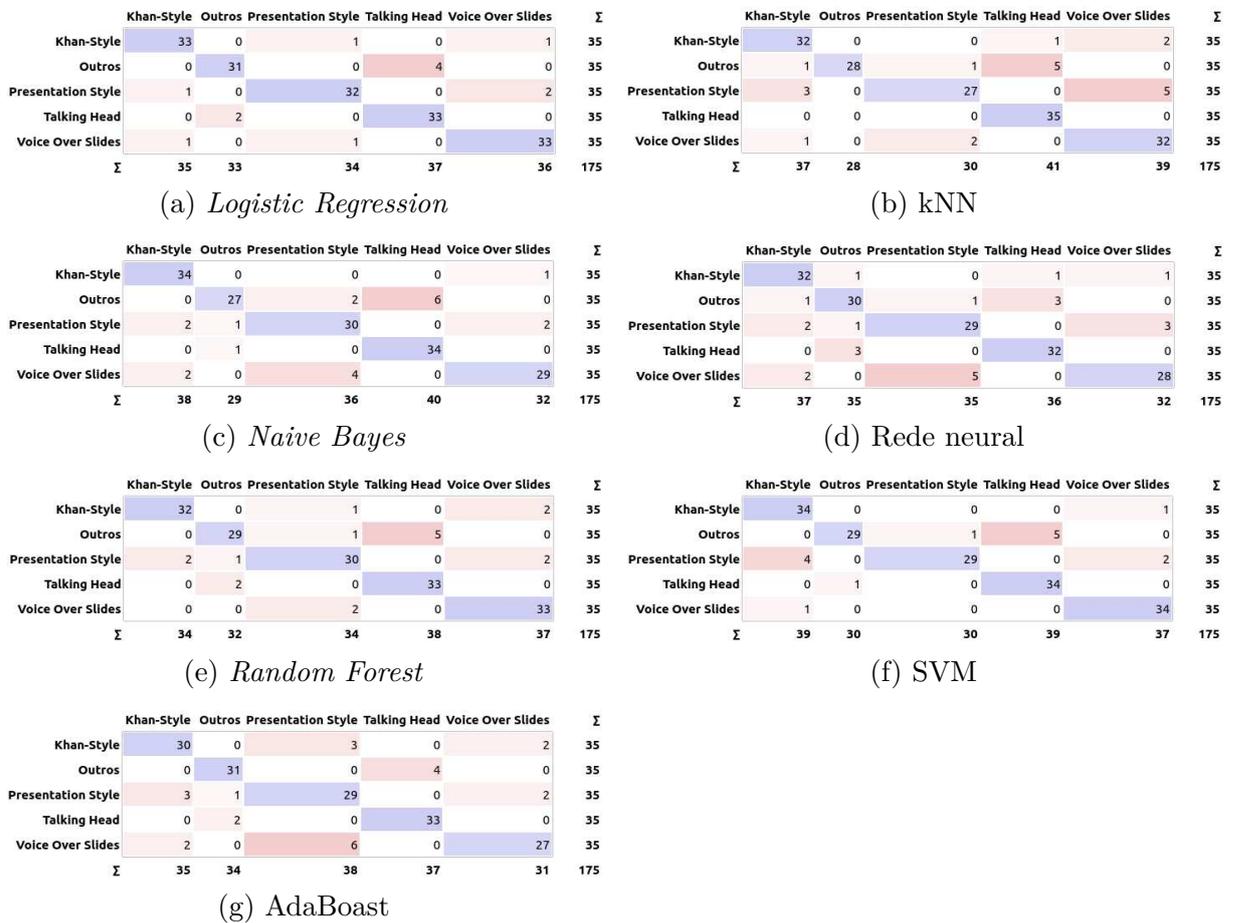
XU, F. et al. Content Extraction from Lecture Video via Speaker Action Classification Based on Pose Information. **International Conference on Document Analysis and Recognition (ICDAR)**, p. 1047-1054, 2019.

YILMAZ, A. et al. Detection and breed classification of cattle using yolo v4 algorithm. International Conference on INnovations in Intelligent SysTems and Applications (INISTA), p. 1–4, 2021.

YOUSAF, M. H. et al. A novel vision based approach for instructor's performance and behavior analysis. **International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15)**, p. 1–6, 2015.

APÊNDICE A – Matriz de Confusão dos Modelos de Classificação

Figura 8 - Matriz de confusão dos modelos de classificação



Fonte: Figura produzida pela autora (2022).