Conferência Internacional de 2023 sobre Computação e Informática Inteligentes Emergentes (ESCI)
Instituto AISSMS de Tecnologia da Informação, Pune, Índia. 1 a 3 de março de 2023

Previsão de envolvimento do aluno em MOOCs usando Aprendizado profundo

Naeem Ahmad
Departamento do MCA NIT
Raipur, Índia
nahmad.mca@nitrr.ac.in

Zubair Khan
Departamento do MCA NIT
Raipur, Índia
krzubairkhan@gmail.com

Deepak Singh
Departamento de CSE
NIT Raipur, Índia
dsingh.cs@nitrr.ac.in

Resumo—O nível de envolvimento das pessoas em uma determinada tarefa é determinado usando dispositivos de reconhecimento automatizados (por exemplo, sensores fisiológicos e cadeiras com sensores de pressão). Embora essas ferramentas estivessem sendo utilizadas em trabalhos de pesquisa anteriores, elas eram muito caras e intrusivas. Atualmente, o uso de câmeras de vídeo RGB é acessível e também mostrou um efeito significativo na previsão do envolvimento das pessoas nas tarefas. As ferramentas estatísticas estão fornecendo uma base sólida para modelar as técnicas de identificação automática de engajamento que utilizam câmeras de vídeo. Neste artigo, um MobileNetv2 leve é usado para determinar automaticamente o envolvimento do aluno em MOOCs para dispositivos com recursos limitados. Todas as camadas da arquitetura MobileNetV2 foram ajustadas para melhorar o aprendizado e a adaptabilidade. Em vez de 1000 classes como no ImageNet, a camada final é ajustada para 3 classes de saída na etapa de classificação final. O estudo experimental é feito em um conjunto de dados de código aberto criado por sujeitos que assistem a vídeos em cursos online. Os resultados das fases de avaliação mostram que nosso modelo tem desempenho melhor que as outras duas redes pré-treinadas (ResNet50, InceptionV4).

Termos de indexação — Aprendizado profundo, envolvimento do aluno, previsão de envolvimento, MOOCs, aprendizagem por transferência

I. INTRODUÇÃO

As vozes das pessoas, os gestos das mãos e as expressões faciais são as principais habilidades para interagir com outras pessoas. Eles também têm a capacidade de determinar o nível de envolvimento dos seus parceiros durante a interação, o que ajuda na decisão da próxima ação. Muitos esforços têm sido feitos no domínio da visão computacional e seus campos relacionados para copiar essas habilidades através de sistemas inteligentes, uma vez que são aplicáveis em muitas áreas, como interação humano-robô, educação on-line e classificação de vídeos pelos espectadores. Estudos anteriores [19], [24] definiram o envolvimento em diferentes contextos, incluindo interação, educação e tecnologia. Neste trabalho, focamos na identificação automática do nível de engajamento em um ambiente de e-learning. Nossa pesquisa visa identificar o engajamento percebido experimentado por observadores externos, conforme mostrado na Figura 1. O desenvolvimento de sistemas automáticos de reconhecimento de engajamento frequentemente usa esse engajamento percebido como um engajamento alvo [31], [8], [9], [28].

Do estudo anterior [30] em educação, observa-se que o envolvimento percebido é um dos critérios para os professores avaliarem o nível de envolvimento e ajustarem os seus métodos de ensino em conformidade. Agora é necessário estender o uso do envolvimento automático percebido às plataformas de aprendizagem online. Este envolvimento automático percebido pode ser útil para melhorar

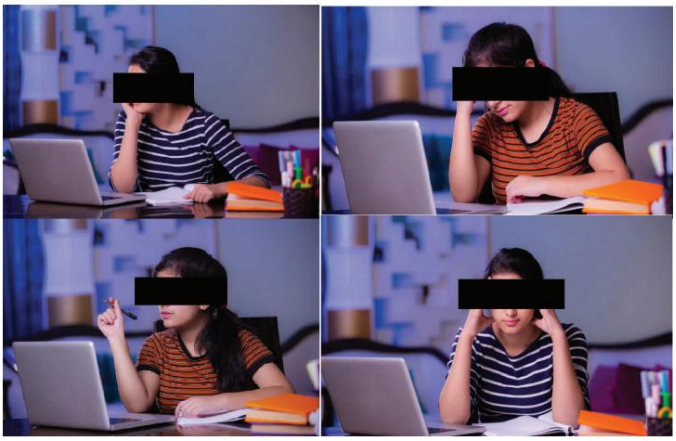

Figura 1. Engajamento percebido em aulas on-line (imagens retiradas de [2])

a metodologia utilizada em plataformas de e-learning. A definição de engajamento no contexto da educação é baseada no comportamento e nas emoções [6], [29], que também é utilizado em nosso estudo para engajamento percebido. Por exemplo, comportamentos como estar ocupado em outro trabalho, menos atento, conversar com outras pessoas, seguir uma tarefa poderiam ser avaliados pelos critérios de engajamento comportamental. Embora as expressões faciais ou o menor interesse dos alunos possam ser julgados por critérios de envolvimento emocional.

O envolvimento cognitivo é difícil de reconhecer com base apenas na observação, pois não está diretamente associado ao envolvimento percebido, especialmente no caso dos alunos [6]. De acordo com o estudo em [6], o engajamento tem aspectos de múltiplos lados.
Portanto, é necessário focar em dois outros tipos de envolvimento, incluindo comportamental e emocional. Aparentemente, esses três fatores estão dinamicamente associados entre si em um indivíduo. De acordo com o resultado deste estudo, o engajamento percebido está relacionado diretamente ao comportamento e às emoções, e indiretamente ao engajamento cognitivo.

Vários esforços de pesquisa têm sido feitos para construir sistemas automáticos sistemas de reconhecimento de engajamento em estudos anteriores [7]. No entanto, as técnicas baseadas em câmeras de vídeo RGB são mais utilizadas por serem menos dispendiosas e não intrusivas [31], [9]. A seguir estão as etapas do processo de identificação de engajamento mais comumente usado com segmentos de vídeo: primeiro, os quadros de vídeo são utilizados para produzir um recurso de baixo nível em um período de tempo especificado. Em segundo lugar, a agregação de baixos

978-1-6654-7524-2/23/$31,00 ©2023 IEEE

recursos de nível obtidos na etapa anterior são usados para extrair um recurso de alto nível. Finalmente, esses recursos de alto nível são usados na tomada de decisões. No entanto, o referido processo que usa aprendizado de recursos baseado em dados não é popular devido à insuficiência de dados para abordagens de aprendizado de máquina baseado em dados.

Recentemente, o aprendizado profundo ganhou gradualmente mais popularidade na superação de problemas de aprendizado de máquina convencional em imagens faciais para determinar o nível de envolvimento dos alunos. No entanto, esta popularidade da aprendizagem profunda não é possível sem o uso de recursos computacionais suficientes e uma enorme quantidade de dados. Vê-se que esses três fatores nem sempre estão disponíveis na utilização de dispositivos de ponta devido ao baixo poder computacional e menos memória. Portanto, um modelo leve de aprendizado profundo é necessário para superar essas limitações [21]. Os métodos convencionais baseados em CNNs não conseguem atingir alta precisão, pois requerem grandes conjuntos de dados. Para superar essas limitações, muitas improvisações de modelos CNN foram propostas [16]. Na última década, os esforços de pesquisa se concentraram no projeto de arquitetura leve sem comprometer o desempenho de imagens faciais [14]. Esses tipos de modelos leves funcionariam em dispositivos com recursos limitados, o que seria muito útil para os acadêmicos.

Uma solução possível seria um modelo leve de CNNs que classificasse imagens faciais [32].

Um dos problemas comuns com os dados é o desequilíbrio de classes, que requer uma abordagem de classificação eficaz.
Embora vários estudos tenham apresentado o aprendizado profundo improvisado, um trabalho muito limitado é feito para o problema de dados desequilibrados no aprendizado profundo. Essas técnicas existentes podem ser amplamente classificadas com base no funcionamento de algoritmos e dados, e em seu conjunto. Neste artigo, um modelo leve e ajustado é proposto para lidar com o desequilíbrio de classes do conjunto de dados. Aqui, o aumento de dados é aplicado no conjunto de dados desequilibrados classificados e de treinamento usando o modelo MobileNetv2 pré-treinado. O principal objetivo deste artigo é propor um método de reconhecimento automatizado baseado em aprendizagem profunda para determinar o nível de envolvimento dos alunos usando um conjunto de dados de imagens faciais.

O restante dos artigos está organizado da seguinte forma: Discutimos trabalhos relacionados que já foram realizados na Seção II na área de detecção de engajamento, engajamento percebido e outras pesquisas relacionadas. A metodologia proposta para reconhecimento do trabalho está descrita na Seção III. Os resultados obtidos são discutidos na Seção IV. Por fim, concluímos nosso trabalho na Seção V.

II. TRABALHOS RELACIONADOS

Vários estudos de pesquisadores em interação humano-qualquer (pode ser humano para HHI, computador para HCI e recursos para (HRI)) tecnologia investigaram o significado do engajamento [24].
Em [17], os autores explicaram que o engajamento é um estado mental que inclui o ponto de engajamento, o período de engajamento sustentado, o ponto de desligamento e o ponto de reengajamento. Embora não exista uma definição universalmente acordada de envolvimento, os autores concentraram-se no tema da manutenção e continuação da participação entre as partes envolvidas. Os estudos de investigação sobre o envolvimento em HRI e HCI podem ser divididos em dois grupos distintos. O primeiro ponto de vista é descobrir como tornar o robô mais interessante para interagir com o participante. O outro ponto de vista é a identificação automática do envolvimento humano durante a interação [15].

Nossa pesquisa está mais alinhada com o segundo ponto de vista. Nossa pesquisa se concentrou no desenvolvimento de um sistema que pudesse detectar automaticamente o interesse e a participação dos alunos em um ambiente de aprendizagem. O nível de participação dos alunos é descrito em diversos contextos pela comunidade de pesquisadores educacionais. Na literatura [6], o estudo apresentou 3 tipos de engajamento compreendendo cognitivo, emocional e comportamental. Vários estudiosos propuseram outras formas de envolvimento, como comportamental, intelectual, cognitivo e psicológico [1]. Estudos anteriores [9], [13] sobre detecção automática de engajamento focaram no engajamento percebido. Como o feedback dos alunos sobre o conteúdo entregue por um instrutor é baseado no envolvimento percebido, nosso estudo de detecção automática de envolvimento também utiliza o envolvimento percebido.

Vários sistemas de detecção automática de engate foram propostos usando diferentes tipos de sinais de entrada [31], [4], [18], [12]. Dados de fala, fisiológicos, visuais, contextuais e multimodais são exemplos de sinais de entrada. Focamos nas literaturas mais relevantes para nossa pesquisa. Os autores em [31] desenvolveram um sistema de detecção automática para categorizar os níveis de envolvimento dos alunos enquanto trabalhavam em um quebra-cabeça educacional baseado em computador. Com o uso de uma câmera e uma cadeira sensível à pressão, eles conseguiram analisar expressões faciais e movimentos de cabeça. A pesquisa deles foi a primeira tentativa de medir o interesse de um indivíduo.

Um algoritmo foi proposto para determinar o nível de interesse em situações de reunião. O sistema usou uma técnica de reconhecimento baseada no modelo oculto de Markov para identificar o subconjunto de alto interesse do grupo com base em entradas audiovisuais.
Sua pesquisa mostrou que a informação auditiva era mais importante que a visual [20]. Em 2015, os autores apresentaram um estudo sobre o impacto da variação do tamanho dos grupos, comparando modelos de categorização treinados em dados coletados de indivíduos e de grupos no HRI [15]. A pesquisa afirma que modelos treinados com dados sobre interações de grupo podem ser usados em uma variedade maior de cenários. Eles usaram recursos recuperados manual e automaticamente de dados de áudio, visuais e contextuais para detectar o desligamento. Os autores em [28] introduziram um mecanismo para fornecer feedback automaticamente aos professores que ajudam a determinar o nível de participação dos alunos.

Os pontos de olhar dos alunos foram determinados usando dados de suas cabeças em movimento. Eles usaram dados do olhar dos alunos como uma proxy para o seu nível de participação nas aulas.

Os autores [31] desenvolveram um sistema que analisa uma sequência de imagens para determinar o nível de interesse de um aluno com base apenas em suas expressões faciais. Eles demonstraram a viabilidade de usar dois níveis de envolvimento e um videoclipe de 10 segundos é a duração mais discriminativa. Também foram desenvolvidos alguns métodos para acumular características ou sugestões dos quadros de uma imagem fornecida em um período de tempo específico. Ao modelar

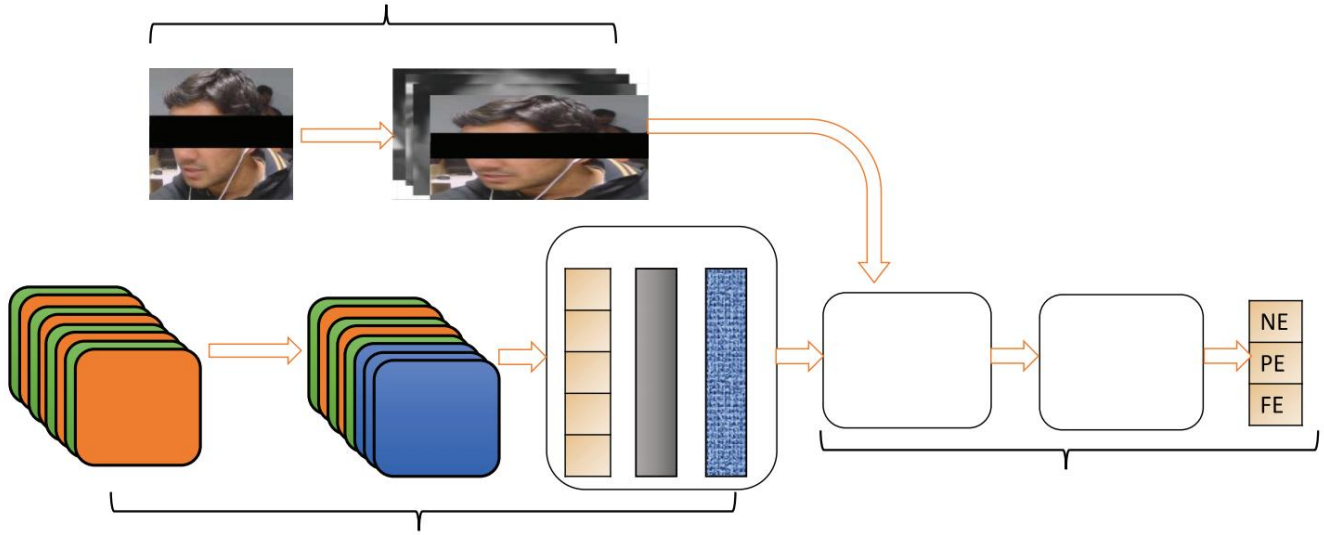

Figura 2. Fluxo de trabalho de multiclassificação baseado em aprendizagem por transferência

dinâmica temporal, funcionais estatísticos são normalmente utilizados como uma estratégia de agregação [8], [21], [27], [23]. Os autores em [27] propuseram um modelo que utiliza a análise das posturas humanas e do movimento corporal para prever os níveis de envolvimento das crianças. Em seu modelo, os recursos da série temporal são agregados e transformados em meta-recursos usando funções mínimas, máximas, médias e de histograma normalizadas.

Para classificar vídeos em uma estrutura de aprendizagem profunda, os autores desenvolveram uma abordagem que combina dados de CNNs de quadro único [26]. Para a fusão de informações de tempo, eles analisaram CNNs com múltiplas conexões no domínio temporal e criaram três modelos de fusão diferentes, incluindo Late, Early e Slow. Os modelos propostos diferem na forma como combinam essas informações no domínio do tempo. Os autores em [11] demonstraram que os resultados do modelo Slow Fusion são superiores aos do modelo Late Fusion e aos do modelo Early Fusion. Várias redes neurais aprendem recursos espaciais e temporais usando kernels convolucionais 3D [10]. Nosso método está relacionado ao modelo MobileNetV2 [22]. Usando uma camada convolucional 3D, este modelo coleta cada mapa de características espaciais das imagens de entrada. Observa-se que o congelamento da rede pré-treinada em camadas baixas melhora mais a precisão do que o ajuste fino [5]. Com esta pesquisa de literatura, conclui-se que é necessário um modelo leve de CNNs para classificar a imagem facial e prever o nível de envolvimento do aluno em dispositivos com recursos limitados, o que seria muito útil para os acadêmicos.

III. METODOLOGIA PROPOSTA

As CNNs são consideradas os melhores modelos para processamento e análise de dados de imagem, pois demonstraram excelente desempenho em segmentação de imagens, classificação de imagens e muitas outras aplicações. CNNs é uma rede feed-forward tendo um arranjo hierárquico de vários níveis de muitas camadas. A configuração básica das CNNs consiste em camadas alternativas de convolução e pooling acopladas a camadas totalmente conectadas no ponto final. As coleções de kernels de convolução são feitas em cada camada que realiza múltiplas transformações. Às vezes, a camada de pooling média global é usada para restaurar a camada totalmente conectada [3], [14]. Os modelos de aprendizagem profunda baseados em CNNs e dispositivos móveis são amplamente utilizados para diversas aplicações industriais. Os dispositivos móveis têm recursos e poder computacional limitados, o que requer modelos leves de aprendizagem profunda. Determinar as áreas locais em uma imagem usando extração por convolução é a operação principal das CNNs. No entanto, esta operação sofre de dois problemas principais: (1) falta de extração direta de características globais e (2) computação complexa devido ao aumento nos parâmetros do kernel de convolução. Diferentes métodos de convolução foram propostos para resolver esses problemas. Convolução dilatada, convolução de grupo, convolução separável em profundidade e convolução deformável são métodos muito famosos [32]. Aqui, é aplicado um modelo de redes pré-treinadas MobileNetV2 com aprendizagem por transferência, que utiliza convoluções separáveis em profundidade.

A. MobileNetV2

MobileNetV2 é uma rede pré-treinada que pertence à classe dos modelos leves. Ele usa uma estrutura residual invertida, gargalos lineares e filtros de convolução leves e separáveis em profundidade. Esses filtros operam nas camadas intermediárias e contribuem para a não linearidade. Toda essa arquitetura foi construída especificamente para dispositivos móveis com recursos e poder computacional limitados [22]. Os principais blocos de construção desta arquitetura são de dois tipos: (1) bloco residual com passo e (2) redução com passo 2. Compreende camadas incluindo Conv 1x1 com ReLU6, separáveis em profundidade

convolução e Conv 1x1 sem qualquer não linearidade. Conseqüentemente, a arquitetura geral do MobileNetV2 contém 53 camadas.

B. Paradigma da aprendizagem por

transferência A aprendizagem por transferência desempenha um papel vital na melhoria do desempenho dos alunos-alvo, transferindo o conhecimento recebido de fontes diferentes e relacionadas. É necessário quando os dados ficam facilmente desatualizados. Isso significa que os dados rotulados coletados em uma janela de tempo específica não seguem a mesma distribuição na janela de tempo futura. Por exemplo, é possível que a utilização de um modelo treinado num conjunto de dados de imagens de um tipo de pessoa num período de tempo específico degrade o desempenho de uma estimativa de expressão quando aplicada a diferentes tipos de pessoas em diferentes períodos de tempo. A aprendizagem por transferência utiliza redes pré-treinadas e ajustadas em uma nova tarefa para um melhor aprendizado [25]. A Figura 2 descreve o processo de classificação dos alunos com base em seu nível de envolvimento usando aprendizagem por transferência e rede MobileNetV2 pré-treinada.

Isso é feito seguindo as seguintes etapas: 1)

Criar imagens faciais a partir dos vídeos de entrada 2) O
aumento de imagem é aplicado no conjunto de dados, pois contém ambos os tipos de imagens: imagens coloridas e imagens em preto e branco.

3) Os recursos de baixo nível são gerados a partir das camadas inferiores da rede pré-treinada, incluindo MobileNetV2, ResNet-50 e IceptionV4.

4) Recursos de baixo nível são passados para as camadas superiores para extrair recursos de alto nível das imagens aumentadas.
5) O conjunto de dados de imagens faciais é categorizado em 3 classes para imagens de treinamento.
6) As redes treinadas são ajustadas nos dados de validação e testadas em modelos ajustados para as imagens de teste.

Neste trabalho, todas as camadas do MobileNetV2 são ajustadas para realizar a tarefa de classificação. Para realizar a análise comparativa, são utilizadas mais duas redes pré-treinadas (ResNet-50, IceptionV4) onde cada camada é ajustada usando o aprendizado de transferência para realizar a tarefa de classificação.

4. RESULTADOS E DISCUSSÃO

Nesta seção, os resultados obtidos são apresentados para análise comparativa. Esta comparação é feita entre o modelo leve MobileNetv2 e dois outros modelos, incluindo ResNet-50, IceptionV4. Para mostrar a comparação, usamos dados de código aberto baixados de https://github.com/e-drishti/wacv2016. Esses dados são capturados para que os sujeitos assistam a vídeos em cursos on-line e sejam rotulados por meio de crowdsourcing com diversas opções de rótulos de engajamento. Como o conjunto de dados contém três classes (Não engajado, Parcialmente engajado, Muito engajado) e o tamanho da imagem de entrada é 100x100x3, uma ampla variedade de configurações experimentais foi considerada. Na fase de treinamento, 70% das imagens de entrada foram utilizadas, enquanto 30% das imagens de entrada foram utilizadas para teste. Os parâmetros de treinamento são um tamanho de lote de 16, valor gama de 0,1 para um solucionador Adam, uma taxa de aprendizado de 0,001 e um tamanho de passo de 7. O modelo é treinado para 25 épocas. O modelo é treinado usando uma perda de entropia cruzada. Todas as camadas da arquitetura MobileNetV2 foram ajustados. Em vez de 1.000 classes como ImageNet, a camada densa final é ajustada para gerar três classes.

A Deep Learning Toolbox do MATLAB foi usada para todos os experimentos associados a este estudo. Três dessas redes pré-treinadas são treinadas e testadas no conjunto de dados de e-learning.
Esses modelos são ajustados modificando a saída da última camada para três classes (muito engajado, nominalmente engajado e não engajado) para realizar a tarefa de classificação. Os resultados mostram que o modelo MobileNetv2 leve é útil para dispositivos com recursos limitados, em vez de CNNs pesadas e profundas para tarefas de classificação sem comprometer o desempenho. A precisão média do modelo proposto é 74,55% melhor do que outras duas redes pré-treinadas fornecidas na Tabela I. A precisão da validação do treinamento e os gráficos de perda do modelo proposto são mostrados nas Figuras 3 e 4, respectivamente.

V. CONCLUSÃO

Este trabalho apresenta um sistema automatizado de reconhecimento do envolvimento dos alunos em sessões online utilizando redes leves pré-treinadas, útil para dispositivos com recursos limitados. Aqui, uma multiclasse (muito engajado, nominalmente engajado, não engajado) é realizada usando um banco de dados de e-learning crowdsource. É desenvolvido um modelo leve baseado em aprendizagem por transferência, obtido pelo ajuste fino das camadas do MobileNetv2. Os resultados obtidos são comparados com outras duas redes pré-treinadas ResNet-50 e IceptionV4. O MobileNetv2 leve tem desempenho superior com precisão média de 74,55%.

Os resultados mostram que o modelo MobileNetv2 leve é útil para dispositivos com recursos limitados, em vez de CNNs pesadas e profundas para tarefas de classificação sem comprometer o desempenho. No futuro, este trabalho pode ser estendido a quaisquer métodos avançados de aprendizagem profunda para dispositivos de ponta, o que pode ser útil para acadêmicos e instrutores em aulas online.

RECONHECIMENTO

Este trabalho faz parte de um grande projeto de alunos do MCA apoiado pelo Departamento de Aplicações de Computador, Instituto Nacional de Tecnologia (NIT) Raipur, Índia.

REFERÊNCIAS

[1] ME Alvarez e AJ Frey, “Promovendo o sucesso acadêmico por meio envolvimento dos alunos”, pp. 1–2, 2012.
[2] S. Anand, “As aulas online estão desgastando as crianças?” no India Today Insight, 23 de novembro de 2020.
[3] RK Barbhuiya, N. Ahmad e W. Akram, “Aplicação de redes neurais convolucionais no diagnóstico de câncer”, em Computational Intelligence in Oncology, Springer, 2022, pp.
[4] M. Dewan, M. Murshed e F. Lin, “Detecção de engajamento na aprendizagem online: uma revisão”, Smart Learning Environments, vol. 6, não. 1, pp. 1–20, 2019.
[5] C. Feichtenhofer, “X3d: Expansão de arquiteturas para reconhecimento de vídeo eficiente”, em Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp.
[6] JA Fredricks, PC Blumenfeld e AH Paris, “Engajamento escolar: Potencial do conceito, estado da evidência”, Revisão da pesquisa educacional, vol. 74, não. 1, pp. 59–109, 2004.
[7] A. Ghandeharioun, D. McDuff, M. Czerwinski e K. Rowan, “Emma: um chatbot de bem-estar com consciência de emoção”, em 2019, 8ª Conferência Internacional sobre Computação Afetiva e Interação Inteligente (ACII). IEEE, 2019, pp.

TABELA I

DIFERENTES REDES PRÉ-TREINADAS USADAS PARA O ESTUDO

Dobras	**Precisão pré-treinada** Inception-	**Sensibilidade**	**Especificidade**	**Precisão**
Dobra-1	V4 0,7159 Resnet 50 0,7186	0,7351	0,8579	0,6563
	Mobilenetv2 0,7371 Proposto	0,7505	0,8546	0,6622
	0,7451 Inception-V4 0,7005	0,7395	0,8866	0,6860
	Resnet 50 0,7125 Mobilenetv2	0,7457	0,8936	0,6920
Dobra-2	0,7344 Proposto 0,7424	0,7469	0,8514	0,6498
	Inception-V4 0,7086 Resnet	0,7528	0,8618	0,6623
	50 0,7129 Mobilenet v2	0,7468	0,8793	0,6798
	0,7318 Proposta 0,7461	0,7498	0,8854	0,6891
Dobra-3	Inception-V4 0,7107 Resnet	0,7359	0,8466	0,6534
	50 0,7175 Mobilenetv2 0,7387	0,7338	0,8693	0,6644
	Proposta 0,7462 Inception-	0,7504	0,8679	0,6801
	V4 0,7184 Resnet 50 0,7156	0,7617	0,8734	0,6967
Dobra-4	Mobilenetv2 0,7389 Proposto	0,7440	0,8523	0,6562
	0,7477	0,7404	0,8615	0,6597
		0,7355	0,8665	0,6747
		0,7487	0,8746	0,6882
Dobra-5		0,7339	0,8713	0,6674
		0 0,7525	0. 8607	0,6625
		0,7762	0,8667	0,6836
		0,7887	0,8773	0,6976

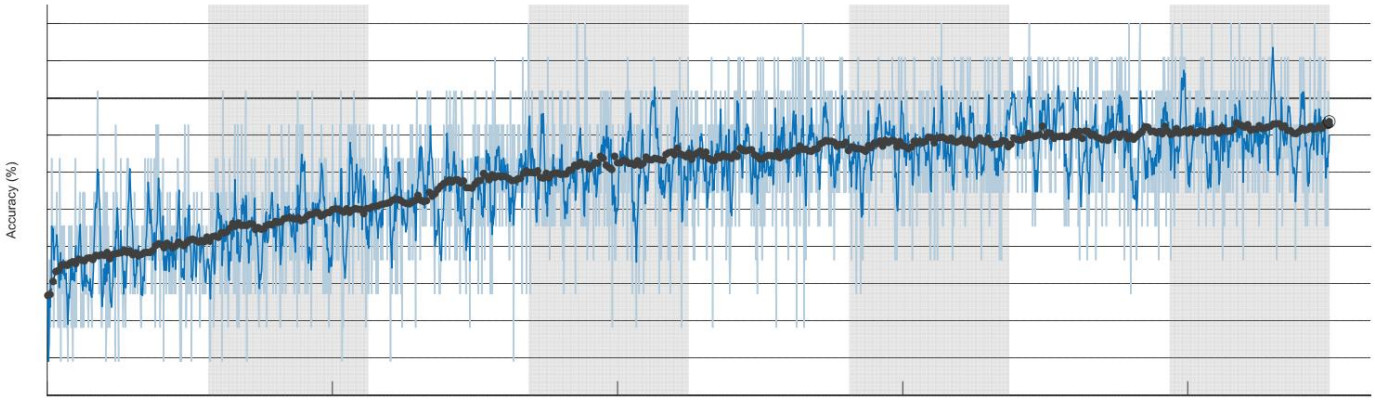

Figura 3. Gráfico de precisão do modelo proposto

[8] J. Hernandez, Z. Liu, G. Hulten, D. DeBarr, K. Krum e Z. Zhang, "Medindo o nível de envolvimento dos telespectadores", em 2013 10º IEEE conferência internacional e workshops sobre rosto e gestos automáticos reconhecimento (GF). IEEE, 2013, pp.

[9] M. Jang, D.-H. Lee, J. Kim e Y. Cho, "Identificando os principais sinais em interações privadas entre alunos e professores para educação aprimorada por robôs", em 2013 IEEE RO-MAN. IEEE, 2013, pp.

[10] C. Jing, P. Wei, H. Sun e N. Zheng, "Redes neurais espaçotemporais para reconhecimento de ação com base em perda conjunta", Computação Neural e Aplicações, vol. 32, pp.

[11] A. Kamel, B. Sheng, P. Yang, P. Li, R. Shen e DD Feng, "Redes neurais convolucionais profundas para reconhecimento de ações humanas usando mapas de profundidade e posturas", IEEE Transactions on Systems, Man, and Cibernética: Sistemas, vol. 49, não. 9, pp.

[12] A. Khan, JP Li, N. Ahmad, S. Sethi, AU Haq, SH Patel e S. Rahim, "Prever tendências emergentes nas mídias sociais modelando-as como redes bipartidas temporais", IEEE Access, vol. 8, pp. 2020.

[13] A. Khan, JP Li, A. u. Haq, S. Nazir, N. Ahmad, N. Varish, A. Malik, e SH Patel, "Modelo de processo de decisão de observador parcial para guindaste-robô ação", Programação Científica, vol. 2020, pp. 1–14, 2020.

[14] A. Khan, A. Sohail, U. Zahoora e AS Qureshi, "Uma pesquisa do arquiteturas recentes de redes neurais convolucionais profundas", Artificial revisão de inteligência, vol. 53, não. 8, pp.

[15] I. Leite, M. McCoy, D. Ullman, N. Salomons e B. Scassellati, "Comparar modelos de desligamento nas interações individuais e grupais", em 2015 10ª Conferência Internacional ACM/IEEE sobre Humano-Robô Interação (HRI). IEEE, 2015, pp.

[16] Z. Li, F. Liu, W. Yang, S. Peng e J. Zhou, "Uma pesquisa de convolucionais redes neurais: análise, aplicações e perspectivas", transações IEEE em redes neurais e sistemas de aprendizagem, 2021.

[17] HL O'Brien e EG Toms, "O que é o envolvimento do usuário? um conceitual estrutura para definir o envolvimento do usuário com a tecnologia", Journal of a sociedade americana de ciência e tecnologia da informação, vol. 59, não. 6, pp. 938–955, 2008.

[18] C. Oertel, G. Castellano, M. Chetouani, J. Nasir, M. Obaid, C. Pelachaud e C. Peters, "Engajamento na interação humano-agente: Uma visão geral", Frontiers in Robotics and AI, vol. 7, pág. 92, 2020.

[19] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini e I. Poggi, "A modelo de atenção e interesse usando o comportamento do olhar", em International Workshop sobre Agentes Virtuais Inteligentes. Springer, 2005, pp.

[20] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder e T. Langlotz, "O olho na realidade estendida: uma pesquisa sobre interação do olhar e rastreamento ocular na realidade estendida usada na cabeça", ACM Computing Surveys (CSUR), vol. 55, não. 3, pp. 1–39, 2022.

[21] MS Ryoo, B. Rothrock e L. Matthies, "Recursos de movimento agrupados para vídeos em primeira pessoa", em Proceedings of the IEEE Conference on Visão computacional e reconhecimento de padrões, 2015, pp.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov e L.-C. Chen, "Mobilenetv2: Resíduos invertidos e gargalos lineares", em Proceedings

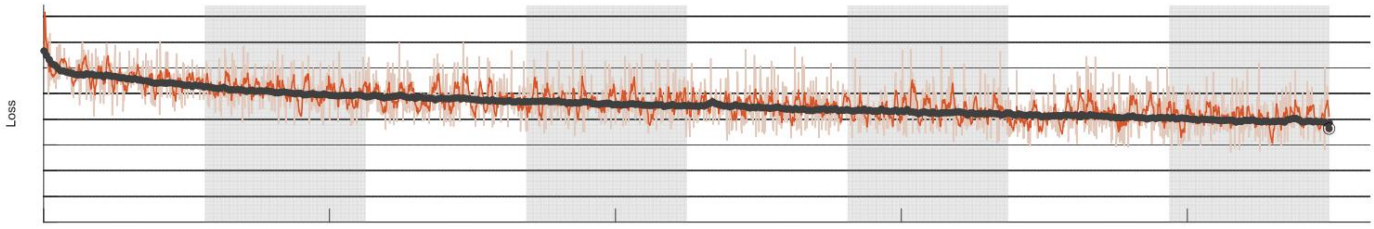

Figura 4. Gráfico de perdas do modelo proposto

da conferência IEEE sobre visão computacional e reconhecimento de padrões, 2018, pp.
[23] S. Senecal, L. Cuel, A. Aristidou e N. Magnenat-Thalmann, “Sistema contínuo de reconhecimento de emoções corporais durante apresentações teatrais,” Animação por Computador e Mundos Virtuais, vol. 27, não. 3-4, pp. 311–320, 2016.

[24] CL Sidner, C. Lee, CD Kidd, N. Lesh e C. Rich, “Explorações em engajamento para humanos e robôs”, Inteligência Artificial, vol. 166, não. 1-2, pp.

[25] D. Singh, A. Shukla e M. Sajwan, “Estrutura de aprendizagem de transferência profunda para a identificação de atividades maliciosas para combater ataques cibernéticos”, Sistemas de Computador de Geração Futura, vol. 125, pp.
[26] B. SravyaPranati, D. Suma, C. ManjuLatha e S. Putheti, “Classificação de vídeo em larga escala com redes neurais convolucionais”, em Tecnologia de Informação e Comunicação para Sistemas Inteligentes: Proceedings of ICTIS 2020, Volume 2. Springer , 2021, pp.
[27] B. Stephens-Fripp, F. Naghdy, D. Stirling e G. Naghdy, “Percepção automática de afeto com base na marcha e postura corporal: uma pesquisa”, International Journal of Social Robotics, vol. 9, não. 5, pp. 617–641, 2017.
[28] O. Sumer, P. Goldberg, S. D'Mello, P. Gerjets, U. Trautwein e E. Kasneci, “Análise de engajamento multimodal de vídeos faciais na sala de aula”, IEEE Transactions on Affective Computing, 2021 .
[29] P. Sunitha, N. Ahmad e RK Barbhuiya, “Impact of covid-19 on education”, in ICCCE 2021,Springer, 2022, pp.
[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke e A. Rabinovich, “Indo mais fundo com convoluções”, em Proceedings of a conferência IEEE sobre visão computacional e reconhecimento de padrões, 2015, pp.
[31] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster e JR Movellan, “As faces do envolvimento: Reconhecimento automático do envolvimento do aluno a partir de expressões faciais”, IEEE Transactions on Affective Computing, vol. 5, não. 1, pp. 86–98, 2014.
[32] Y. Zhou, S. Chen, Y. Wang e W. Huan, “Revisão da pesquisa em redes neurais convolucionais leves”, em 2020 IEEE 5ª Conferência de Tecnologia da Informação e Engenharia Mecatrônica (ITOEC). IEEE, 2020, pp.