

Reconhecimento de Emoções em Sistemas de E-learning

Oussama El Hammoui, Fatimaezzahra Benmarrakchi, Nihal Ouherrou, Jamal El Kafi, Ali El Hore Departamento de
Ciência da Computação
Faculdade de Ciências
Universidade Chouaib Doukkali,
El Jadida, Marrocos
elhammoui.oussama@gmail.com, fz.benmarrakchi@gmail.com,
n.ouherrou@gmail.com, jekafi@gmail.com,
aelhor@gmail.com

Resumo - Os avanços tecnológicos nos sistemas de e-learning oferecem novas oportunidades para os alunos reforçarem o desenvolvimento acadêmico e melhorarem a acessibilidade à educação. O reconhecimento de expressões faciais é uma área cada vez mais importante em sistemas inteligentes de e-learning. No entanto, apoiar o lado emocional dos alunos durante as tarefas de aprendizagem é desafiador e exige uma consciência das emoções dos alunos. Para o efeito, é apresentada uma breve revisão de modo a fundamentar o importante papel do reconhecimento de expressões faciais em sistemas de e-learning. Seu papel na educação especial também é investigado. Este artigo centra-se no reconhecimento de expressões faciais utilizando redes neurais convolucionais e sua aplicação em sistemas de e-learning, introduzindo um novo sistema composto por três etapas principais: pré-processamento, extração de características e classificação. Testamos o sistema proposto com alunos de 8 a 12 anos, em um jogo educativo. Os resultados mostraram que o sistema proposto alcançou resultados de última geração.

Palavras-chave— Reconhecimento de Expressões Faciais; Inteligência artificial; Aprendizagem Profunda; Redes Neurais Convolucionais; Educação; sistema de e-learning.

I. INTRODUÇÃO

As emoções têm um papel importante na aprendizagem e no desempenho dos alunos. As emoções controlam a atenção do aluno, afetam a sua motivação para aprender e influenciam a sua autorregulação da aprendizagem. Conforme afirmado por [1], a aprendizagem autorregulada e a motivação medeiam os efeitos das emoções no desempenho acadêmico. Especialmente, as emoções positivas afetam positivamente o desempenho acadêmico quando são mediadas pela aprendizagem e motivação autorreguladas.

Nos últimos dez anos, algoritmos de aprendizagem profunda dominaram o campo da Inteligência Artificial (IA), um dos subcampos é a visão computacional que teve grandes sucessos graças às Redes Neurais Convolucionais (CNN) [2], que são uma das arquiteturas de aprendizagem profunda que têm sido amplamente utilizados para tarefas de reconhecimento/classificação de imagens. Recentemente, pesquisadores demonstraram um interesse crescente na expressão facial reconhecimento usando CNN [3] [4] devido ao seu potencial

aplicações como e-learning [5][6], monitoramento e segurança, saúde e entretenimento.

Embora alguma investigação tenha sido realizada sobre o reconhecimento de expressões faciais, há poucas investigações empíricas sobre a sua aplicação em sistemas de e-learning, e ainda faltam estratégias adequadas para abordar a presença de emoções na aprendizagem. Portanto, o foco principal deste trabalho é propor uma nova abordagem que analise emoções que possam ser integradas em sistemas de e-learning.

II. TRABALHOS RELACIONADOS

A. Emoção e aprendizagem

Estudos recentes [6][7][8][9] demonstraram que as emoções têm efeitos importantes nos processos de aprendizagem dos alunos. A evidência disponível parece sugerir que as emoções positivas influenciam a aprendizagem, afetando a atenção, a motivação e a auto-regulação da aprendizagem dos alunos. Considerando que as emoções negativas também influenciam a aprendizagem, afetando o desempenho e o desempenho dos alunos. Por exemplo, a vergonha e a ansiedade diminuem o interesse e a motivação. Emoções diferentes podem invocar diferentes cognições não relacionadas com as tarefas, e podem diferir no que diz respeito à intensidade e persistência de tais cognições.

Assim, as emoções positivas e negativas têm um papel muito importante no armazenamento e recuperação de informação na aprendizagem. Conforme mencionado por [9], a ativação das emoções promove a ativação de material associativo na memória de longo prazo, o que facilita a recuperação.

Além disso, a pesquisa [10] mostrou que os alunos experimentam uma grande variedade de emoções auto-referenciadas, relacionadas a tarefas e sociais em ambientes acadêmicos, como prazer em aprender, esperança, orgulho, admiração, ansiedade, vergonha, desesperança e tédio. Além disso, resultados de pesquisas anteriores sobre consciência emocional e dificuldades de aprendizagem mostraram que os alunos com dificuldades de aprendizagem têm dificuldades emocionais, como depressão e ansiedade. Na verdade, a vida emocional e seus problemas são muito negligenciados [11]. Além disso, a grande maioria das crianças com dificuldades de aprendizagem tem

algum problema emocional associado à dificuldade de aprendizagem [12]. Embora extensas pesquisas [13] [14] [15] [16] tenham sido realizadas em alunos com dificuldades de aprendizagem e sistemas de e-learning, não foram encontrados estudos que investiguem o estado emocional dos alunos-alvo.

Com base nas evidências atualmente disponíveis, parece justo sugerir que um ambiente de aprendizagem que tenha em conta a consciência emocional dos alunos poderá afetar positivamente a motivação e o desempenho dos alunos. Portanto, os sistemas de e-learning devem integrar sistemas de reconhecimento de emoções para apoiar os alunos durante as tarefas de aprendizagem.

B. Técnicas de reconhecimento de expressões faciais

Em 1971, Ekman e Friesen afirmaram que existem 6 emoções faciais globais: raiva, nojo, medo, felicidade, tristeza e surpresa [17], essas emoções contêm sinais repletos de informações que podem ser úteis para determinar o estado mental de alguém. Cada pessoa tem uma maneira de mostrar as emoções faciais, e com a mesma pessoa a expressão de uma emoção também varia, o que torna um problema difícil para os métodos de aprendizado de máquina. Outro problema é que os dados geralmente não estão aptos para processamento, neste caso significa que muitos fatores devem ser levados em consideração, como iluminação, pose da cabeça, detecção de rosto e outros problemas.

A atenção mais recente se concentrou na expressão facial reconhecimento. Um exemplo é Chih-Hung Wu [18], que propôs um sistema de reconhecimento de expressão facial baseado em SVM e árvores de decisão que usa o software de reconhecimento facial Luxund-FaceSDK para detectar rostos do banco de dados de expressões faciais JAFFE. Outro exemplo é Chen et al.[5], que também propôs um sistema baseado em SVM, mas usou Active Shape Models para detectar rosto e localização com a ferramenta VOSM e wavelets Gabor para adquirir informações de aparência facial. Ou Shan et al. [19], que apresentaram um método que utiliza LBP para representação facial. Os pesquisadores usaram SVM para o reconhecimento de expressão com recursos de LBP reforçado. No entanto, o problema do LBP é que ele precisa de mais faces adicionadas ao conjunto de treinamento para poder detectar faces em condições de pouca luz [20].

Além disso, existe uma variedade de pesquisas nas áreas de inclusão da consciência emocional em sistemas de e-learning. Entre os estudos há um artigo apresentado por [6], que afirma que os Sistemas Tutoriais Inteligentes (STI) são privilegiados com capacidades de feedback afetivo, são capazes de enviar sinais afetivos ou cognitivos apropriados aos alunos, em resposta à detecção do seu estado afetivo, garantindo dessa forma, a sua segurança emocional e o seu envolvimento ou persistência na experiência de aprendizagem. Trabalhos semelhantes, uma estrutura integrada de reconhecimento de emoções para ITS introduzida por [7], que elabora dois tipos de recursos de sistemas proativos e reativos, mapeando seus diferentes componentes, dependências e inter-relacionamentos, a fim de capturar e estruturar a variedade rica e criativa de maneiras pelas quais o ITS pode apoiar emoções positivas.

III. NOSSO MÉTODO PROPOSTO

No presente estudo, desenvolvemos um sistema de reconhecimento de expressões faciais baseado em CNN, com o objetivo de integrá-lo em sistema de e-learning. Para tanto, utilizamos dois publicamente

bancos de dados disponíveis CK+ [21] [22] e KDEF [23], para realização dos experimentos. Adicionamos uma etapa de pré-processamento de imagens usando OpenCV para detectar e recortar rostos. A segunda etapa é a extração de características utilizando a rede neural de convolução proposta e a etapa final é a classificação das imagens faciais utilizando uma rede totalmente conectada. A figura 1 ilustra as três etapas principais do nosso sistema de reconhecimento de expressões faciais proposto.

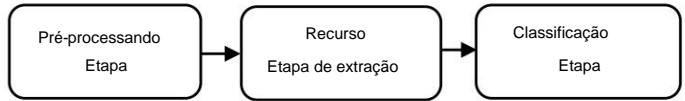


Figura 1. As três etapas do sistema de reconhecimento de expressões faciais proposto

A. Conjunto de dados

O modelo é treinado e testado em uma combinação de imagens dos bancos de dados CK+ e KDEF. A base de dados CK+ é amplamente utilizada para reconhecimento de expressões faciais, inclui 327 sequências de vídeo representadas por 123 participantes com idades entre De 18 a 50 anos, 69% mulheres, 81% euro-americanos, 13% afro-americanos e 6% de outros grupos, cada exibição começou e terminou com rosto neutro. O KDEF é um conjunto de 4.900 imagens de expressões faciais humanas, contém 70 sujeitos, 35 mulheres e 35 homens, com idades entre 20 e 30 anos, exibindo sete expressões emocionais diferentes. O banco de dados combinado levou apenas imagens de pose de cabeça reta do banco de dados KDEF e todas as imagens disponíveis do CK +, foi dividido em 683 amostras de treinamento, 81 amostras de validação e 81 amostras de teste, conforme mostrado na figura 2.

Também utilizamos o banco de dados de expressões faciais femininas japonesas (JAFFE) para testar o modelo proposto, o banco de dados contém 213 imagens de 7 expressões faciais (6 expressões faciais básicas + 1 neutra) representadas por 10 sujeitos conforme mostrado na figura 3.

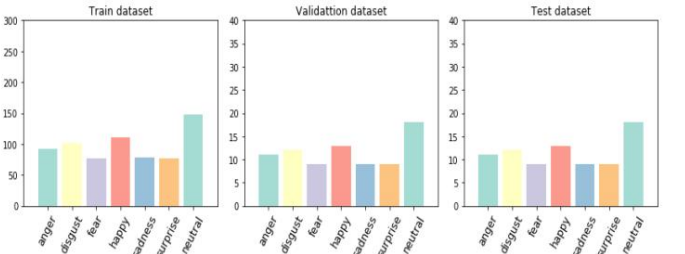


Figura 2. O banco de dados personalizado (combinação de KDEF e CK+).

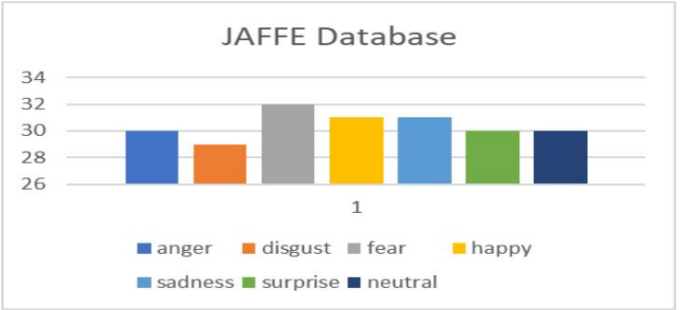


Figura 3. Banco de dados de expressões faciais femininas japonesas (JAFFE).

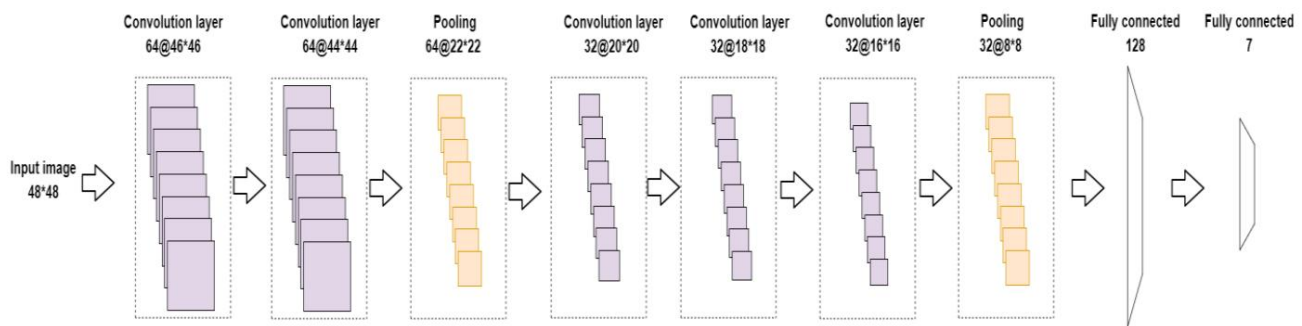


Figura 4. Estrutura da Rede Neural Convolucional (CNN)

B. Pré-processamento

Esta fase consiste em quatro etapas:

- Primeiro as imagens são convertidas em tons de cinza porque informações de cores não nos ajudam a identificar informações importantes bordas ou outros recursos.
- Detecção de rosto usando Haar Cascades em OpenCV, é um método eficaz de detecção de objetos proposto por Paul Viola e Michael Jones [24]. Este método é baseado em uma abordagem de aprendizado de máquina onde uma função em cascata é treinada em muitos aspectos positivos (imagens de rostos) e imagens negativas (imagens sem rostos), é então utilizado para detectar rostos. Existem três categorias de haar recursos, recursos de borda, recursos lineares e recursos de quatro retângulos mostrados na figura 5, cada recurso é composto por um retângulo preto e um retângulo branco, subtraindo a soma de todos os pixels no branco retângulo da soma de todos os pixels no preto retângulo, obtemos um único valor do recurso, o objetivo dessas características é indicar a existência de certos caracteres na imagem, finalmente para distinguir a parte não facial e a parte facial.
- Recortar os rostos detectados.
- Redimensionar as imagens para um tamanho fixo de (48,48). Figura 6 demonstra a etapa de pré-processamento.

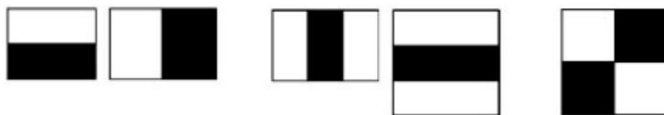


Figura 5. Modelos de recursos do tipo Haar.

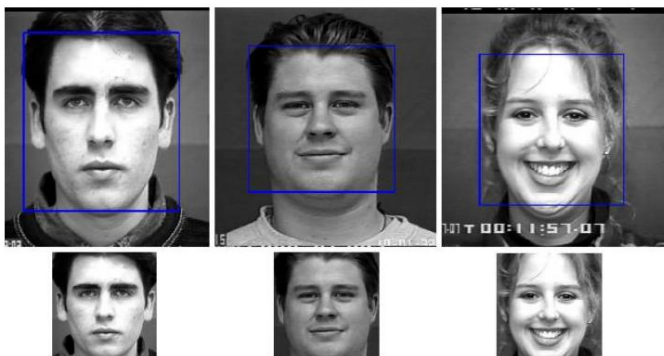


Figura 6. Demonstração da etapa de pré-processamento.

C. Extração e classificação de características.

As imagens da etapa de pré-processamento são enviadas para a CNN na figura 4. A principal vantagem da CNN é que ela aceita imagens 2D diretamente como dados de entrada. Portanto, tem uma vantagem na área de visão computacional. A arquitetura proposta é composta por cinco camadas de convolução, duas camadas de pooling e duas camadas totalmente conectadas.

A camada de convolução é usada para extrair recursos como bordas, cantos e formas para gerar mapas de recursos. A saída de convolução é calculada da seguinte forma.

$$\text{Largura de saída} = (W - Fw + 2P) / Sw + 1$$

$$\text{Altura de saída} = (H - Fh + 2P) / Sh + 1$$

Onde (W, H) é o tamanho da entrada, (Fw, Fh) é o tamanho do filtro, P é o preenchimento de zero e (Sw, Sh) é o avanço da convolução.

A camada de pooling é uma técnica de redução de dimensão usada para reduzir o tamanho e formar uma nova camada. Aceita uma entrada de $W1 \times H1 \times D1$ e produz $W2 \times H2 \times D2$, onde D é a profundidade. É calculado da seguinte forma.

$$W2 = (W1 - Fw) / Sw + 1$$

$$H2 = (H1 - Fh) / Sh + 1$$

$$D2 = D1$$

Conseqüentemente, essas camadas são aplicadas várias vezes até que o sistema extraia recursos suficientes, então a saída 2D da camada final é rasterizada em um vetor 1D que será alimentado no classificador de rede de camada totalmente conectada tradicional, na camada final uma função Softmax usada para multiclassificação.

4. EXPERIMENTOS E RESULTADOS

As etapas de pré-processamento e extração de recursos foram realizadas usando bibliotecas python como OpenCV e TensorFlow em GPU Nvidia Tesla K80 (12GB), dividimos o conjunto de dados em 90:10

para (treinamento, validação) e testes para avaliar a eficácia do sistema. Neste estudo, utilizamos todas as emoções do banco de dados CK+, exceto o desprezo, e geramos as imagens neutras retirando a primeira face de cada exibição. Todas as imagens foram pré-processadas e alimentadas na CNN para extração de características, uma etapa final para classificação utilizou duas camadas totalmente conectadas para multiclassificação. O sistema proposto possui uma precisão de teste de 97,53% conforme mostra a figura 7.

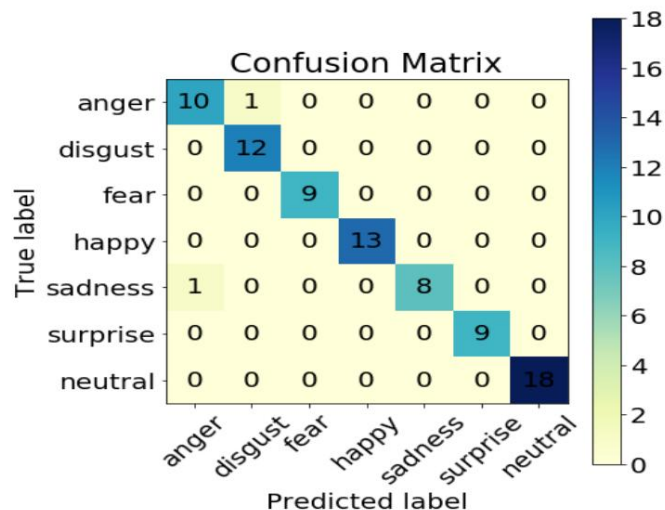


Figura 7. Matriz de confusão para dados de teste do conjunto de dados combinado.

Para testar o sistema proposto em um novo conjunto de dados, utilizamos todas as imagens do banco de dados JAFFE para testar o sistema proposto e ele possui uma precisão de 97,18% conforme mostrado na figura 8.

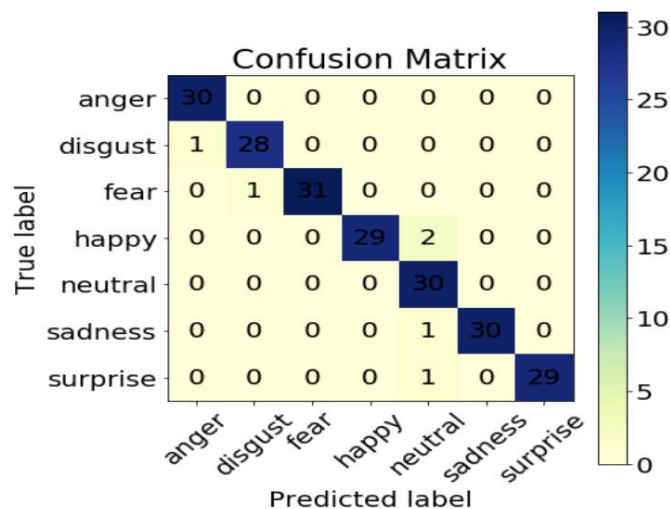


Figura 8. Matriz de confusão para conjunto de dados JAFFE.

O sistema teve um bom desempenho em comparação com Shan et al.[25] seu método atingiu 76,4742% no banco de dados JAFFE, Mayya et al.[26] alcançaram uma precisão de reconhecimento de 98,12% no JAFFE, utilizaram uma rede SVM na extremidade do sistema para fazer a classificação.

Testamos o nosso sistema proposto num jogo educativo, com 4 alunos com idades compreendidas entre os 8 e os 12 anos (2 participantes com dificuldades de aprendizagem), selecionados do ensino básico e Serviço de Fonoaudiologia-Centro de Saúde, El Jadida

Marrocos. A escolha dos nossos utilizadores-alvo teve a tendência de mostrar o importante papel do reconhecimento de emoções no e-learning para alunos com e sem dificuldades de aprendizagem.

Para realizar esta experiência integramos o nosso sistema proposto num jogo educativo desenvolvido por [8] que proporciona um ambiente de aprendizagem que promove a aprendizagem e ajuda os alunos com as suas dificuldades de aprendizagem, melhorando algumas das suas competências elementares, como a leitura e a escrita. Os participantes foram convidados a jogar enquanto o sistema analisa sua emoção em tempo real, ao final do jogo o sistema salva os resultados.

Usamos Plotly, uma estrutura online que fornece soluções online para cientistas de dados criarem gráficos, análises e estatísticas. As figuras a seguir (9, 10, 11 e 12) mostram uma série temporal de todos os resultados do sistema, cada um representa probabilidades emocionais do aluno durante a aprendizagem.

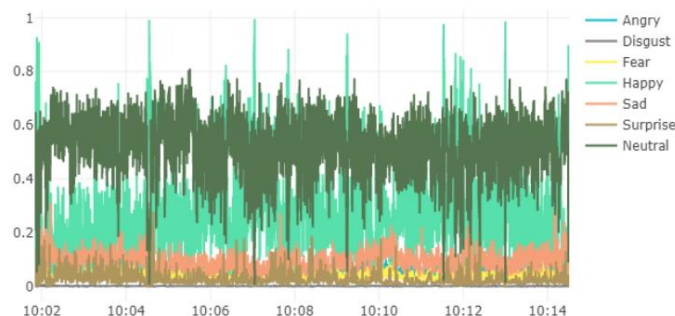


Figura 9. Probabilidades emocionais do participante 1.

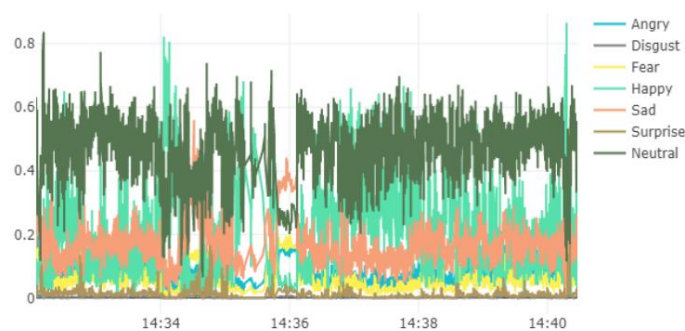


Figura 10. Probabilidades emocionais do participante 2.



Figura 11. Probabilidades emocionais do participante 3 (com dificuldades de aprendizagem)

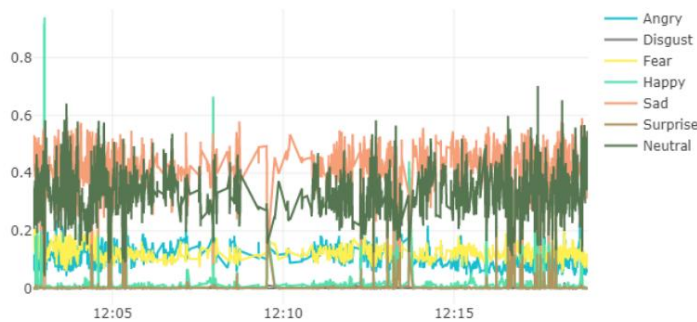


Figura 12: Probabilidades emocionais do participante 4 (com dificuldades de aprendizagem).

Com base nos gráficos acima, notamos que o sistema detectou diferentes emoções durante o aprendizado, as figuras 11 e 12 mostram que em determinados momentos o sistema não detectou as emoções, isso porque os alunos não estavam olhando diretamente para a câmera. Podemos notar que a emoção triste nas figuras 11 e 12 é maior em comparação com as figuras 9 e 10 onde a emoção feliz é maior, provavelmente devido às suas dificuldades de aprendizagem. Portanto, o reconhecimento de emoções é importante em sistemas de e-learning para ajudar e apoiar os alunos na aprendizagem.

V. CONCLUSÃO

Neste artigo propusemos um sistema de expressão facial reconhecimento baseado na CNN, o teste foi bem-sucedido e o sistema foi capaz de detectar rostos e classificar emoções com uma precisão de 97,53% nos dados de teste e uma precisão de 97,18% no conjunto de dados JAFFE. E foi realizada uma experiência com alunos com idades entre 8 e 12 anos para testar o sistema proposto. Os resultados mostram que as emoções foram detectadas e que o sistema alcançou resultados de última geração.

Recomendam-se, portanto, estudos futuros sobre o tema atual, especialmente, para investigar o estado emocional dos alunos em sistemas de e-learning.

Nesta fase, são necessários testes iterativos com utilizadores-alvo para recolher feedback sobre os nossos designs propostos e garantir o design resultante com o objetivo de identificar as alternativas mais promissoras e combiná-las melhor com os sistemas de e-learning.

RECONHECIMENTOS

Este trabalho foi apoiado financeiramente por uma Bolsa de Excelência concedida a Oussama El Hammoui (11UAE2017) e a Fatimaezzahra Benmarrakchi (2UCD2015) pelo Centro Nacional de Pesquisa Científica e Técnica (CNRST)-Ministro da Educação Nacional, Ensino Superior, Formação de Pessoal e Investigação Científica, Marrocos.

Os autores gostariam de agradecer ao presidente e à equipe do Centro de Saúde do Serviço de Fonoaudiologia, El Jadida Marrocos e também aos estudantes que participaram deste estudo.

Os autores gostariam também de agradecer à fonoaudióloga Ilham ELhousni pelas valiosas sugestões e recomendações e outras coisas do Groupe Scolaire l'Ange Bleu -El Jadida pela sua cooperação.

REFERÊNCIAS

- [1] C. Mega, L. Ronconi e R. De Beni, "O que torna um bom aluno? Como as emoções, a aprendizagem autorregulada e a motivação contribuem para o desempenho acadêmico.", *J. Educ. Psicol.*, vol. 106, não. 1, pág. 121, 2014.
- [2] A. Krizhevsky, I. Sutskever e GE Hinton, "Classificação ImageNet com Redes Neurais Convolucionais Profundas", em *Avanços em Sistemas de Processamento de Informação Neural 25*, F. Pereira, CJC Burges, L. Bottou e KQ Weinberger, Eds. Curran Associates, Inc., 2012, pp.
- [3] Z. Meng, P. Liu, J. Cai, S. Han e Y. Tong, "Rede Neural Convolucional com Consciência de Identidade para Reconhecimento de Expressão Facial", em 2017, 12ª Conferência Internacional IEEE sobre Reconhecimento Automático de Gestos Faciais (FG 2017), 2017, pp.
- [4] Z. Yu e C. Zhang, "Image Based Static Facial Expression Recognition with Multiple Deep Network Learning", em *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Nova York, NY, EUA, 2015, pp. -442.
- [5] L. Chen, C. Zhou e L. Shen, "Reconhecimento de Expressão Facial Baseado em SVM em E-learning", *IERI Procedia*, vol. 2, não. Suplemento C, pp. 781-787, janeiro de 2012.
- [6] M. Feidakis, T. Daradoumis e S. Caballe, "Dotando sistemas de e-Learning com consciência emocional", em 2011, Terceira Conferência Internacional sobre Redes Inteligentes e Sistemas Colaborativos, 2011, pp.
- [7] JM Harley, SP Lajoie, C. Frasson e NC Hall, "An Integrated Emotion-Aware Framework for Intelligent Tutoring Systems", em *Inteligência Artificial na Educação*, 2015, pp.
- [8] F. Benmarrakchi, JE Kafi, A. Elhore e S. Haie, "Explorando o uso das TIC no apoio aos estilos de aprendizagem preferidos dos alunos disléxicos: Uma avaliação preliminar", *Educ. Inf. Tecnologia*, pp. 1-19, outubro de 2016.
- [9] GH Bower, "Psicólogo Americano de Humor e Memória 36: 129-148", *CrossRef Google Sch.*, 1981.
- [10] R. Pekrun, T. Goetz, W. Titz e RP Perry, "Emoções Acadêmicas na Aprendizagem e Realização Autorreguladas dos Alunos: Um Programa de Pesquisa Qualitativa e Quantitativa", *Educ. Psicol.*, vol. 37, não. 2, pp. 91-105, junho de 2002.
- [11] AR Arthur, "A vida emocional das pessoas com dificuldades de aprendizagem", *Ir. J. Aprenda. Desabil.*, vol. 31, não. 1, pp. 25-30, março de 2003.
- [12] JC Abrams, "Sobre Dificuldades de Aprendizagem: Considerações Afetivas", *J. Ler. Escreva. Aprender. Desabil. Internacional*, vol. 2, não. 3, pp. 189-196, janeiro de 1986.
- [13] F. Benmarrakchi, JE Kafi e A. Elhore, "Tecnologia de comunicação para usuários com dificuldades específicas de aprendizagem", *Procedia Comput. Ciência*, vol. 110, não. Suplemento C, pp. 258-265, janeiro de 2017.
- [14] FE Benmarrakchi, JE Kafi e A. Elhore, "Abordagem de modelagem de usuário para alunos disléxicos em ambientes virtuais de aprendizagem", *Int. J. Cloud Appl. Computação. IJCAC*, vol. 7, não. 2, pp. 1-9, abril de 2017.
- [15] F. Benmarrakchi, JE Kafi e AE Hore, "Uma maneira de aprendizagem diferente para alunos com dificuldades de aprendizagem específicas", *Int. J. Computação. Tecnologia*, vol. 14, não. 10, pp. 6157-6162, outubro de 2015.
- [16] F. Benmarrakchi, JE Kafi e A. Elhore, "Apoiando as preferências de estilo de aprendizagem do disléxico em ambiente de aprendizagem virtual adaptativo", em 2016 Conferência Internacional sobre Engenharia MIS (ICEMIS), 2016, pp.
- [17] P. Ekman e WV Friesen, "Constantes entre culturas no rosto e na emoção.", *J. Pers. Soc. Psicol.*, vol. 17, não. 2, pág. 124, 1971.
- [18] CH Wu, "Nova tecnologia para o desenvolvimento de reconhecimento de expressões faciais em e-learning", em 2016, Conferência Internacional de Portland sobre Gestão de Engenharia e Tecnologia (PICMET), 2016, pp.
- [19] C. Shan, S. Gong e PW McOwan, "Reconhecimento de expressão facial baseado em padrões binários locais: um estudo abrangente", *Image Vis. Computação*, vol. 27, não. 6, pp. 803-816, maio de 2009.
- [20] J. Chang-yeon, Detecção de rosto usando recursos LBP. 2008.
- [21] T. Kanade, JF Cohn e Y. Tian, "Banco de dados abrangente para análise de expressões faciais", em *Proceedings Fourth IEEE International*

Conferência sobre Reconhecimento Automático de Rosto e Gestos (Cat. No. PR00580), 2000, pp.

- [22] P. Lucey, JF Cohn, T. Kanade, J. Saragih, Z. Ambadar e I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): Um conjunto de dados completo para unidade de ação e expressão especificada por emoção", em 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp.

- [23] "Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM do Departamento de Neurociência Clínica, seção de Psicologia, Karolinska Institutet, ISBN 91-630-7164-9."

- [24] P. Viola e M. Jones, "Detecção rápida de objetos usando uma cascata reforçada de recursos simples", em Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, vol. 1, pág. I-511-I-518 vol.1.

- [25] K. Shan, J. Guo, W. You, D. Lu e R. Bie, "Reconhecimento automático de expressão facial baseado em uma estrutura de rede neural convolucional profunda", 2017, pp.

- [26] V. Mayya, RM Pai e MM Manohara Pai, "Reconhecimento automático de expressão facial usando DCNN", Procedia Comput. Ciência, vol. 93, pp.