Um Classificador Explicável para Transtorno de Estresse Pós-Traumático Utilizando Redes Neurais Convolucionais Tridimensionais

Raphael M. M. Fernandes[1], **Rodrigo J. de Carvalho**[1]
Orlando F. Junior[2], **Liana C. L. Portugal**[2,3] **Taiane C. Ramos**[1]

[1] Instituto de Computação, UFF, Niterói - RJ, Brasil

[2]Lab. de Neurofisiologia do Comportamento, Depto. de Fisiologia e Farmacologia, Instituto Biomédico, UFF, Niterói - RJ, Brasil.

[3]Lab. de Neurofisiologia, Depto. Psicologia, Instituto de Biologia Roberto Alcantara Gomes, Centro Biomédico, UERJ, Rio de Janeiro - RJ, Brasil.

raphael_m@id.uff.br, rod.junger@gmail.com,
orlandofernandesjunior@id.uff.br,
lianalportugal@gmail.com,
taiane_ramos@id.uff.br

Resumo. *O Transtorno de Estresse Pós-Traumático (TEPT) é uma condição psiquiátrica caracterizada por sintomas persistentes de reexperiência, evitação e hiperexcitabilidade em resposta a eventos traumáticos. A identificação precisa de indivíduos com TEPT a partir de dados neurobiológicos permanece um desafio, motivando o uso de abordagens baseadas em aprendizado profundo. Neste estudo, empregamos redes neurais convolucionais tridimensionais (3D-CNNs) para a classificação de TEPT a partir de dados de ressonância magnética funcional (fMRI). Utilizamos uma amostra de 43 participantes (20 com TEPT) expostos a estímulos visuais aversivos e avaliamos o desempenho do modelo por meio de validação cruzada, obtendo uma acurácia média de 86,25%. Além disso, empregamos a técnica de oclusão, baseada no Atlas Harvard-Oxford, para identificar as regiões cerebrais mais relevantes para a classificação. Os resultados destacam o envolvimento de áreas associadas ao processamento visual e emocional, incluindo o giro fusiforme occipital, a divisão superior do córtex occipital lateral e o córtex pré-cúneo.*

1. Introdução

O Transtorno de Estresse Pós-Traumático (TEPT) é uma condição psiquiátrica debilitante que pode se desenvolver após a exposição a eventos traumáticos, como acidentes, agressões ou desastres naturais [Soares et al. 2021]. Estima-se que a prevalência global do TEPT ao longo da vida seja de 3,9% [Koenen et al. 2017], o que reflete o impacto significativo dessa condição a nível mundial. Em eventos traumáticos de larga escala, como a pandemia de COVID-19, os casos de TEPT aumentam substancialmente, afetando milhões de pessoas [Abdalla et al. 2021]. A correta classificação do TEPT é essencial para aprimorar estratégias de diagnóstico e tratamento, permitindo uma abordagem mais precisa e personalizada para os pacientes. Modelos baseados em aprendizado de máquina têm sido explorados para identificar padrões neurobiológicos associados ao transtorno,

utilizando dados de neuroimagem e biomarcadores fisiológicos. A aplicação de classificadores automatizados pode auxiliar na diferenciação entre indivíduos com e sem TEPT, viabilizando a detecção precoce e intervenções mais eficazes.

Além da classificação, a explicabilidade dos modelos é um aspecto fundamental para compreender os mecanismos subjacentes ao TEPT. Métodos como a interpretação de redes neurais e análise de importância de variáveis [Arrieta et al. 2020] permitem identificar as regiões cerebrais mais afetadas pelo transtorno, como a amígdala, o hipocampo e o córtex pré-frontal [Shin et al. 2006]. Essa abordagem não apenas melhora a confiabilidade dos modelos, mas também contribui para avanços no entendimento neurocientífico do TEPT, possibilitando o desenvolvimento de terapias mais direcionadas e eficazes.

A identificação de biomarcadores cerebrais específicos para o TEPT pode contribuir significativamente para aprimorar o diagnóstico e aprofundar a compreensão dos mecanismos neurobiológicos subjacentes à condição. A ressonância magnética funcional (fMRI) tem sido amplamente empregada na investigação de padrões de ativação cerebral em resposta a estímulos emocionais [Hu et al. 2019], permitindo a análise da dinâmica funcional de diferentes regiões cerebrais. Embora diversos estudos indiquem o envolvimento de áreas específicas no processamento de ameaças, a heterogeneidade dos achados entre indivíduos e pesquisas evidencia a necessidade de abordagens mais robustas e generalizáveis para a análise desses dados.

Nos últimos anos, o aprendizado de máquina profundo [LeCun et al. 2015] tem se mostrado uma abordagem promissora para a detecção de padrões em dados de neuroimagem [Abrol et al. 2021, Yin et al. 2022]. Métodos baseados em redes neurais convolucionais profundas (CNN) têm obtido resultados promissores na identificação de padrões complexos em dados cerebrais [Wen et al. 2018]. No entanto, é importante ressaltar que redes neurais profundas requerem grandes volumes de dados para serem treinadas de maneira eficaz, uma vez que sua expressividade e capacidade de generalização dependem diretamente da quantidade e diversidade dos exemplos apresentados durante o aprendizado [Alzubaidi et al. 2023]. Esta característica dificulta o uso de CNNs como método principal em estudos que envolvem a coleta de dados biológicos (como as imagens de fMRI relacionadas a tarefas), pois é comum que um único grupo de pesquisa tenha sua coleta limitada a apenas algumas dezenas de indivíduos.

Neste contexto, o presente estudo emprega técnicas de Aprendizado de Máquina interpretável para identificar as regiões cerebrais que possibilitam distinguir entre indivíduos com TEPT e controles. O principal objetivo é avaliar a viabilidade do uso de uma arquitetura robusta em um conjunto de dados que possui uma pouca quantidade de participantes.

Para mitigar os desafios impostos pelo tamanho reduzido da amostra, empregamos uma estratégia de aumento de dados (data augmentation), buscando ampliar a quantidade de amostras que a rede verá durante o treinamento. É comum que as técnicas de aumento de dados introduzam na amostra dados sintéticos gerados a partir dos dados reais. Porém, nossa abordagem envolve criar combinações das quatro imagens de fMRI que cada participante possui, aumentando a variabilidade das amostras sem prejudicar a avaliação da capacidade de generalização do modelo para dados reais nunca vistos.

Escolhemos utilizar uma arquitetura baseada em redes neurais convolucionais

tridimensionais (3D-CNN) por sua capacidade de preservar e explorar a estrutura volumétrica dos dados de fMRI. Diferentemente de abordagens convencionais, que frequentemente dependem da projeção dos dados tridimensionais em representações bidimensionais ou da extração de estatísticas agregadas, as 3D-CNNs permitem a captura direta das relações espaciais complexas entre diferentes regiões cerebrais [Qureshi et al. 2019]. Essa característica é particularmente relevante para o estudo de fMRI, onde a identificação de padrões sutis e distribuídos na atividade cerebral pode fornecer insights sobre biomarcadores neurais e mecanismos subjacentes da condição [Soares et al. 2021].

Para garantir a interpretabilidade do modelo e compreender quais regiões cerebrais contribuem para a distinção entre indivíduos com TEPT e controles, aplicamos a técnica de oclusão [Zeiler and Fergus 2014]. Essa abordagem é amplamente utilizada em estudos de explicabilidade, pois permite avaliar a relevância de diferentes áreas cerebrais no desempenho da rede neural profunda. A técnica consiste em remover sistematicamente regiões específicas do cérebro e reavaliar a classificação utilizando os dados modificados. Ao comparar o desempenho do modelo antes e depois da remoção de cada região, é possível quantificar seu impacto na acurácia da rede, destacando as áreas mais informativas para distinguir entre pacientes com TEPT e indivíduos sem a condição.

O restante deste artigo está organizado da seguinte forma: Na Seção 2, apresentamos os trabalhos relacionados, destacando abordagens recentes na utilização de redes neurais profundas para análise de neuroimagem. A Seção 3 descreve a metodologia adotada, incluindo detalhes sobre o banco de dados utilizado (Seção 3.1), as etapas de pré-processamento dos dados (Seção 3.2) e a estratégia de aumento de dados empregada para mitigar a restrição amostral (Seção 3.3). Em seguida, apresentamos a arquitetura da rede neural convolucional tridimensional (3D-CNN) utilizada para a classificação do TEPT (Seção 3.4) e a abordagem de explicabilidade baseada na técnica de oclusão (Seção 3.5). Na Seção 4, detalhamos os resultados obtidos, incluindo a performance do modelo na tarefa de classificação e a análise da relevância das regiões cerebrais para a distinção entre pacientes com TEPT e controles. A Seção 5 apresenta a discussão dos achados, contextualizando-os à luz da literatura existente e das limitações do estudo. Por fim, a Seção 6 traz as conclusões do trabalho e direções para pesquisas futuras.

2. Trabalhos Relacionados

A identificação de regiões cerebrais associadas ao TEPT tem sido alvo de diversos estudos utilizando técnicas de neuroimagem e aprendizado de máquina. Nesta seção, discutimos pesquisas relevantes que investigam a ativação cerebral em indivíduos com TEPT e abordagens de predição de sintomas por meio de técnicas computacionais.

Bastos et al. (2022) foi responsável pela coleta de dados utilizada no presente trabalho e, investigaram, por meio de fMRI, o engajamento de pacientes com TEPT em pistas de segurança ao visualizar imagens aversivas. O estudo analisou a atividade cerebral de 20 pacientes com TEPT e 23 controles enquanto observavam imagens de mutilação e neutras, apresentadas em dois contextos: um *real* e um *safe*. Os dados foram analisados utilizando a Análise de Variância de Medidas Repetidas (Repeated-Measures ANOVA), e usaram como variáveis o fator entre grupos (TEPT e controle) e os fatores dentro dos grupos (*real* vs *safe*). Os resultados mostraram uma interação significativa entre grupo, contexto e valência ($F(1, 41) = 13.33, p = 0.001$), evidenciando que, os participantes

controle engajam mais em pistas de segurança, portanto, apresentam menos reatividade no contexto *safe* ($p = 0.303$). Em contraste, os pacientes com TEPT mantiveram uma alta reatividade cerebral, independentemente do contexto ($p < 0.001$), sugerindo uma dificuldade em modular respostas emocionais diante de pistas de segurança.

Em um estudo subsequente, Portugal et al. (2023) aplicaram técnicas de aprendizado de máquina no mesmo banco de dados para predizer os valor da Escala de *Checklist* de Estresse Pós-Traumático (PCL-5), separando os dados entre os contextos *real* e *safe*. Utilizando um modelo de regressão por processo gaussiano, os autores obtiveram uma correlação entre o PCL real e predito de $0,59$ no contexto *real* e de $0,01$ no contexto *safe*. Os resultados, apesar de serem estatiscamente significativos no contexto *real*, apresentam baixa correlação. O estudo não encontrou resultados significativos para o contexto *safe*.

Por fim, a revisão sistemática realizada por Jia et al. (2024) destaca a escassez de estudos que utilizam modelos de aprendizado de máquina para a classificação de TEPT. Além disso, nenhum desses trabalhos emprega *deep learning* como ferramenta de classificação, limitando-se ao uso de redes neurais para redução de dimensionalidade. A maioria desses estudos também conta com mais de 80 participantes, o que pode ser um desafio para novos estudos utilizando dados de fMRI devido ao alto custo de coleta e ao desconforto que esse processo pode causar. Nenhum dos trabalhos apresentados na revisão fazem uso de técnicas de explicabilidade, o que dificulta a compreensão das regiões cerebrais envolvidas no transtorno. Assim, torna-se fundamental desenvolver abordagens que permitam a classificação com quantidades menores de dados e maior interpretabilidade dos resultados.

O objetivo deste trabalho é reanalisar os dados coletados por Bastos et al. (2022) para verificar se é possível obter uma boa classificação de sujeitos entre pacientes com TEPT ou controle utilizando um modelo de *Deep Learning* com base em exames de fMRI em uma amostra reduzida. Também verificaremos a viabilidade de obter informações sobre as regiões cerebrais envolvidas nos processos biológicos do TEPT através de um método de explicabilidade para redes neurais.

3. Métodos

3.1. Banco de Dados

Este estudo utilizou um banco de dados privado, coletado pela Universidade Federal do Rio de Janeiro (UFRJ) em parceria com o Instituto D'or (IDOR) [Bastos et al. 2022]. A coleta dos dados foi aprovada pelo comitê de ética da UFRJ (número do processo 1.749.604). O banco de dados conta com 52 participantes: 23 indivíduos diagnosticados com TEPT e 29 indivíduos do grupo controle, que vivenciaram eventos traumáticos, mas não desenvolveram TEPT. Durante o experimento, os participantes foram expostos a dois conjuntos de imagens: um representando cenas reais de partes do corpo humano mutiladas (mutilado) e outro mostrando partes do corpo em situações cotidianas típicas (neutro) [Bastos et al. 2022]. As imagens foram apresentadas em dois contextos distintos: o contexto *real* e o contexto *safe*. No contexto *real*, os participantes foram informados de que as imagens exibidas correspondiam a registros autênticos de mutilações, o que gerava uma percepção de ameaça e maior impacto emocional. Já no contexto *safe*, os participantes foram instruídos de que as imagens eram simuladas, reduzindo a sensação de perigo e permitindo uma resposta emocional diferenciada. Neste trabalho, utilizamos apenas as

imagens cerebrais referentes ao indivíduo visualizando imagens mutiladas em ambos os contextos (*real* e *safe*). Cada indivíduo foi apresentado à imagens mutiladas e neutras em ambos os contextos em duas rodadas (*run 1* e *run 2*).

3.2. Pré-processamento

O pré-processamento foi realizado para padronizar os cérebros e facilitar a entrada dos dados no modelo. Além disso, participantes que apresentaram movimentos de cabeça excessivos durante a coleta foram removidos, para garantir a precisão e a consistência dos dados. O procedimento resultou na remoção de 8 participantes, sendo 3 do grupo de TEPT e 5 do grupo controle. Para mais detalhes sobre os procedimentos realizados durante o pré-processamento consultar Bastos et al. (2022).

3.3. Aumento de dados

Ao fim do pré-processamento, são obtidas 4 imagens do cérebro por paciente, sendo duas referentes ao contexto *real* e duas referente ao contexto *safe* (run 1 e run 2). Os dados são compostos por 20 participantes do grupo com TEPT e 23 participantes do grupo de controle. Para aumentar a variabilidade do conjunto de dados, foi realizado um processo de aumento de dados (*data augmentation*). Cada paciente possui duas imagens correspondentes ao contexto *real* e duas ao contexto *safe*, que foram combinadas, gerando quatro pares distintos para cada participante. Esse procedimento duplicou a quantidade de dados disponíveis para treinamento, preservando as características originais das imagens e ampliando a diversidade do conjunto. A Figura 1 ilustra o processo de aumento de dados realizado. O dataset final é composto de 172 pares de imagens, sendo uma do contexto *real* e uma do contexto *safe*.

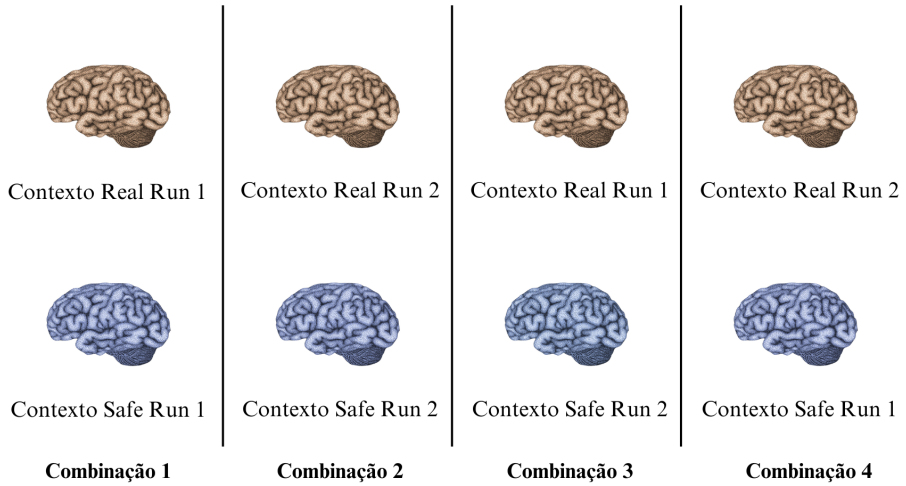

Figura 1. Pares de imagens cerebrais para um participante. Cada participante vê as imagens nos contexto *real* e *safe* duas vezes (*run 1* e *run 2*), e as imagens são combinadas para aumentar a quantidade de dados disponíveis no treinamento de forma que cada par contenha uma imagem do contexto *real* e uma imagem do contexto *safe*.

3.4. Arquitetura da Rede Neural 3D-CNN

A rede neural desenvolvida possui duas entradas tridimensionais, uma para o contexto *real* e outra para o contexto *safe*, ambas com dimensões (53, 63, 52, 1). Cada entrada é seguida por camadas convolucionais, de *max pooling* e de *batch normalization*. Essas camadas são individuais para cada entrada, e os pesos são treinados de forma independente.

Cada entrada passa por quatro camadas de convolução 3D (*Conv3D*), com 3, 6, 12 e 24 filtros, respectivamente, e um *kernel* de 3 x 3 x 3. Após a aplicação das primeiras duas camadas convolucionais, é realizada uma operação de *MaxPooling3D* com um *kernel* 2 x 2 x 2 para redução dimensional. O mesmo processo de convolução e *MaxPooling3D* é repetido para as camadas seguintes. Em seguida, é aplicada uma normalização em batch (*BatchNormalization*) e, ao final da etapa convolucional, a saída é convertida em um vetor unidimensional por meio de uma operação de *Flattening*.

Ao final, as saídas das duas redes convolucionais são concatenadas e passam por duas camadas densas (*Dense*) com 256 e 128 unidades, respectivamente, ambas seguidas por camadas de *Dropout* com taxa de 0,7. O uso do *Dropout* de 0,7 foi essencial para evitar *overfitting* devido à baixa quantidade de dados na amostra. Quando executamos com um *Dropout* de 0,5, a acurácia média foi de 79,32%. A camada final do modelo é composta por uma unidade de saída com função de ativação *sigmoid*, responsável pela classificação binária utilizando a função de perda de entropia cruzada binária (*binary crossentropy*). A figura 2 mostra em detalhes a arquitetura utilizada.

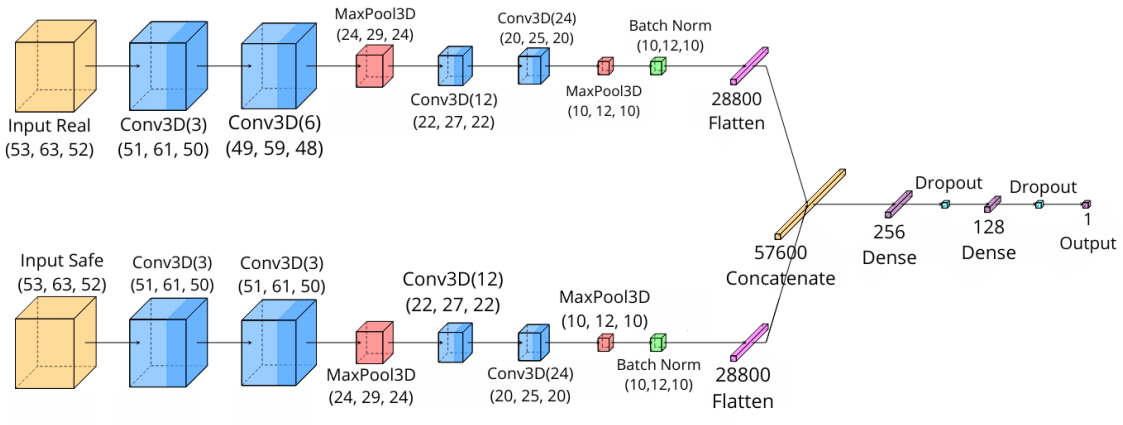

Figura 2. A arquitetura possui uma entrada para imagem real e uma para imagem *safe*. eles são processadas separadamente durante as convoluções e concatenadas para gerar uma classificação de TEPT ou controle.

Para treinamento e teste da rede, foi realizada a validação cruzada estratificada com k = 5, para diminuir o risco de uma escolha específica de divisão de dados influenciar os resultados do modelo. Os dados foram separados por fold de forma que todos os dados de uma pessoa estivessem no mesmo fold, eliminando assim a possibilidade de contaminação de dados. O modelo foi treinado por 100 épocas com um *batch size* de 32, utilizando a função de perda *binary crossentropy* e o otimizador *Adam* com uma taxa de aprendizado de 0,0001.

3.5. Explicabilidade

Este estudo aplicou a técnica de oclusão para avaliar a relevância das regiões cerebrais na classificação de pacientes com TEPT. A oclusão consiste em remover diferentes regiões do cérebro durante o teste, sem retreinamento da rede neural. Isso permite que as previsões reflitam apenas o impacto da remoção de cada região, sem interferência nos ajustes do modelo.

A técnica foi implementada com base no Atlas Harvard-Oxford [Jenkinson et al. 2012], substituindo os valores dos *voxels* das regiões removidas por um valor mínimo da imagem. Cada imagem com uma região oclusa foi novamente classificada pela rede e a diferença entre a acurácia original e a obtida após a oclusão da região foi registrada. Esse processo foi repetido para cada um dos cinco *folds* da validação cruzada, e as variações na acurácia foram usadas para quantificar a importância de cada região cerebral na distinção entre pacientes com TEPT e controles.

4. Resultados

A Figura 3 apresenta a evolução de acurácia durante a etapa de treinamento ao longo das 100 épocas de treinamento para cada um dos cinco *folds*. Já para os dados de teste, a figura 4 mostra que o modelo apresentou boa capacidade de generalização, atingindo em média 86,25% de acurácia (desvio padrão de 7,55%) no conjunto de teste, com valores específicos de 81,25%, 78,12%, 100%, 87,5% e 84,37% em sua última época. O F1-Score também se mostra bem consistente, apresentando um valor médio de 88,64%. Embora se note alguma variação entre os *folds*, o desvio padrão relativamente baixo indica consistência nos resultados, considerando o tamanho reduzido do conjunto de dados.

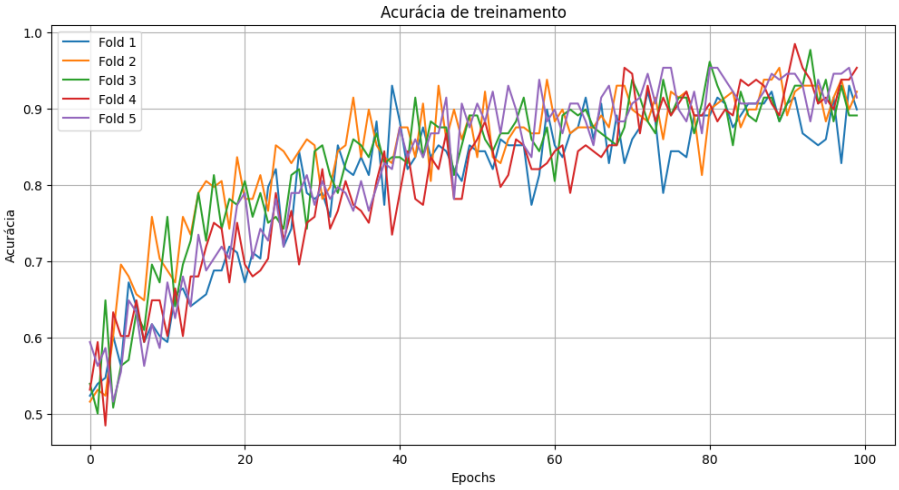

Figura 3. Acurácia do grupo de treinamento ao longo das épocas para os 5 *folds*.

As áreas cerebrais cuja oclusão resultou nas maiores quedas na acurácia do classificador estão na Tabela 1 e na Figura 5, sendo as cinco principais: giro fusiforme - parte occipital (21,87%), córtex occipital lateral (14,37%), córtex pré-cúneo (12,50%), giro do cíngulado posterior (11,25%) e giro pré-central (9,37%).

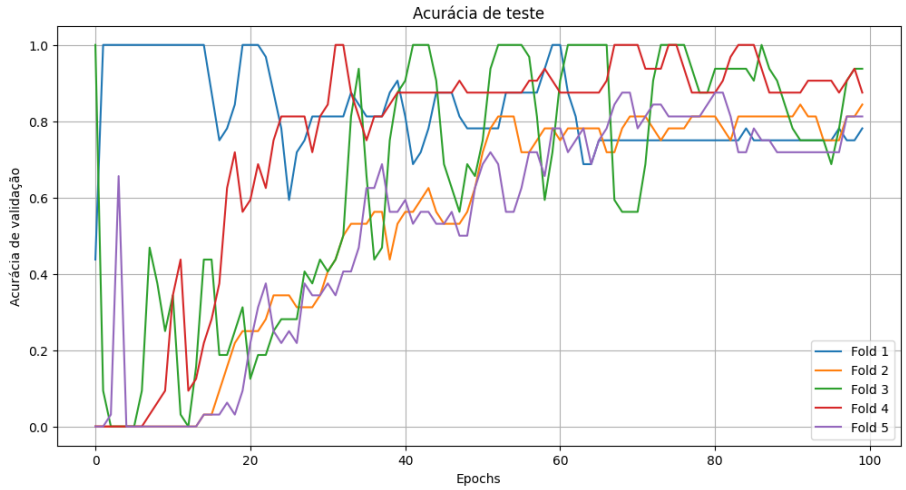

Figura 4. Acurácia no grupo de teste ao longo das épocas para os 5 *folds*.

Região do cérebro	Piora na classificação quando oclusa
Giro Fusiforme, parte occipital	21,87%
Córtex Occipital Lateral, parte superior	14,37%
Córtex Pré-cúneo	12,50%
Giro Cingulatado Posterior	11,25%
Giro pré-central	9,37%
Giro Lingual	9,37%
Giro Fusiforme, occipitotemporal	9,37%
Córtex Insular	7,50%
Giro Frontal Médio	7,50%
Giro frontal inferior, pars opercularis	6,87%

Tabela 1. Principais regiões cerebrais, classificadas pela piora média nos 5 folds (%) na classificação quando oclusas.

5. Discussão

Nosso objetivo neste estudo era verificar se, usando um método de *deep learning* poderíamos classificar dados de fMRI entre pacientes com TEPT e controles, usando um conjunto de dados com poucas amostras. Obtivemos bons resultados de acurácia, indicando que foi possível fazer a classificação. Além disso, concluímos que a técnica de aumento de dados colaborou para que o modelo pudesse ser treinado, mesmo com um baixo número de amostras. Importante destacar que os folds da validação cruzada foram separados a nível de participante, para que não houvesse contaminação do conjunto teste.

O desempenho do modelo ao longo das 100 épocas de treinamento mostra uma consistência na capacidade de generalização, com uma acurácia média de 86,25% e desvio padrão de 7,55% no conjunto de teste. Esses resultados indicam que a arquitetura proposta de uma rede neural convolucional tridimensional (3D-CNN) foi eficaz na extração de padrões relevantes para a distinção entre os grupos. O uso de técnicas de regularização, como *Dropout* e *Batch Normalization*, foi essencial para mitigar o *overfitting*, mesmo com um conjunto de dados relativamente pequeno.

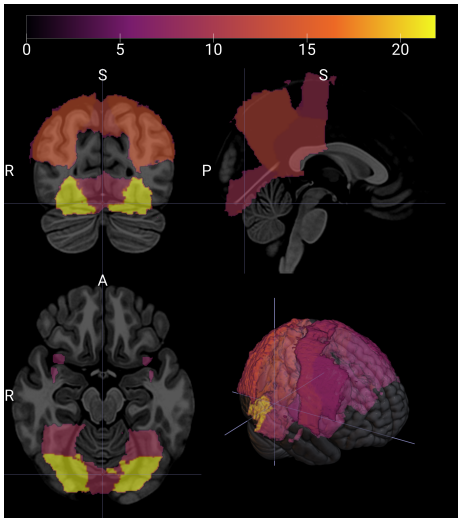

Figura 5. Regiões cerebrais mais relevantes para a classificação de pacientes com TEPT de acordo com a técnica de oclusão.

Os trabalhos de Harricharan et al. (2020), Suo et al. (2022) e Zilcha-Mano et al. (2020) apresentaram acurácias de 80,4%, 73,8% e 76,7% respectivamente, para a mesma tarefa de classificação utilizando modelos tradicionais de aprendizado de máquina. Esses trabalhos obtiveram uma menor acurácia em comparação com o nosso resultado, mesmo contando com um número maior de participantes na amostra. Acreditamos que esse resultado se deve à técnica de aprendizado profundo empregada neste trabalho, que permitiu um aprendizado mais eficiente do modelo. Yang et al. (2021) e Zhu et al. (2021) obtiveram acurácias de 71,2% e 80,0% respectivamente, apesar de utilizarem redes neurais em suas arquiteturas. No entanto, nesses estudos, as redes foram empregadas apenas para redução de dimensionalidade, mas a classificação também foi feita empregando técnicas de aprendizado supervisionado tradicionais. Acreditamos que o uso de CNNs no presente trabalho tenha contribuído para um melhor desempenho do modelo, dada sua maior capacidade de extração de características relevantes para a classificação de imagens.

A técnica de oclusão identificou as regiões cerebrais mais relevantes para a classificação do modelo. Essas regiões coincidem com aquelas frequentemente relatadas na literatura sobre TEPT. [Harricharan et al. 2016, Chao et al. 2012] Especificamente, o córtex do pré-cúneo e o giro do cíngulo posterior desempenham um papel central na integração de informações autorreferenciais e na memória autobiográfica, sugerindo uma possível relação com o TEPT [Summerfield et al. 2009].

Uma limitação importante deste estudo é a utilização de um atlas bilateral, que agrupa regiões de ambos os hemisférios cerebrais em uma única representação. Essa abordagem impede a análise específica das alterações funcionais ou estruturais em cada hemisfério separadamente, limitando a compreensão de possíveis assimetrias neurais associadas ao TEPT.

Em trabalhos futuros, pretendemos investigar a importância do contexto para a classificação de TEPT ou controle. Pela hipótese proposta em Bastos et al. (2022), os participantes com TEPT apresentariam menos engajamento nas pistas de segurança e por isso, as imagens de contexto *safe* apresentariam maior diferença na ativação cerebral entre

os grupos e, por tanto, maior importância para a rede. Esta investigação poderá contribuir para corroborar as conclusões obtidas no trabalho de Bastos et al. (2022).

6. Conclusão

Neste estudo, investigamos a aplicação de 3D-CNNs para a identificação de regiões cerebrais associadas ao TEPT a partir de dados de fMRI. Nosso principal objetivo foi interpretar as contribuições das diferentes regiões cerebrais no TEPT.

Para isso, treinamos um modelo 3D-CNN com validação cruzada em cinco *folds*, atingindo uma acurácia média de 86,25%. A análise de interpretabilidade foi conduzida por meio da técnica de oclusão, na qual removemos sistematicamente diferentes regiões cerebrais do Atlas Harvard-Oxford para avaliar seu impacto na performance do classificador. Os resultados indicaram que as áreas mais relevantes para a predição do TEPT incluíram o Giro Fusiforme Occipital, o Córtex Occipital Lateral, o Pré-cúneo e o Giro do Cíngulo Posterior, regiões conhecidas por seu papel no processamento visual e na integração de informações emocionais. Além disso, nossos achados reforçam a importância de abordagens baseadas em aprendizado profundo para o avanço da neurociência clínica.

7. Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) E-26/210.759/2024 (295437).

Referências

Abdalla, S. M., Ettman, C. K., Cohen, G. H., and Galea, S. (2021). Mental health consequences of covid-19: a nationally representative cross-sectional study of pandemic-related stressors and anxiety disorders in the usa. *BMJ Open*, 11(8).

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., and Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature Communications*, 12(1):353.

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-Dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., et al. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, 10(1):46.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.

Bastos, A., Silva, L., Oliveira, J., Oliveira, L., Pereira, M., Figueira, I., Mendlowicz, M., Berger, W., Luz, M., Campos, B., Marques-Portella, C., Moll, J., Bramati, I., Volchan, E., and Erthal, F. (2022). Beyond fear: patients with posttraumatic stress disorder fail to engage in safety cues. *Journal of Affective Disorders Reports*, 10:100380.

Chao, L. L., Lenoci, M., and Neylan, T. C. (2012). Effects of post-traumatic stress disorder on occipital lobe function and structure. *Neuroreport*, 23(7):412–419.

Harricharan, S., Nicholson, A. A., Thome, J., Densmore, M., McKinnon, M. C., Théberge, J., Frewen, P. A., Neufeld, R. W., and Lanius, R. A. (2020). Ptsd and its dissociative subtype through the lens of the insula: Anterior and posterior insula resting-state functional connectivity and its predictive validity using machine learning. *Psychophysiology*, 57(1):e13472.

Harricharan, S., Rabellino, D., Frewen, P. A., Densmore, M., Théberge, J., McKinnon, M. C., Schore, A. N., and Lanius, R. A. (2016). fmri functional connectivity of the periaqueductal gray in ptsd and its dissociative subtype. *Brain and behavior*, 6(12):e00579.

Hu, J., Kuang, Y., Liao, B., Cao, L., Dong, S., and Li, P. (2019). A multichannel 2d convolutional neural network model for task-evoked fmri data classification. *Computational Intelligence and Neuroscience*, 2019(1):5065214.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Jia, Y., Yang, B., Yang, Y., Zheng, W., Wang, L., Huang, C., Lu, J., and Chen, N. (2024). Application of machine learning techniques in the diagnostic approach of ptsd using mri neuroimaging data: A systematic review. *Heliyon*.

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., Karam, E. G., Ruscio, A. M., Benjet, C., Scott, K., Atwoli, L., Petukhova, M., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bunting, B., Ciutan, M., de Girolamo, G., Degenhardt, L., Gureje, O., Haro, J. M., Huang, Y., Kawakami, N., Lee, S., Navarro-Mateu, F., Pennell, B.-E., Piazza, M., Sampson, N., Ten Have, M., Torres, Y., Viana, M. C., Williams, D., Xavier, M., and Kessler, R. C. (2017). Posttraumatic stress disorder in the world mental health surveys. *Psychological Medicine*, 47(13):2260–2274. Epub 2017 Apr 7.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Portugal, L. C. L., Ramos, T. C., Fernandes, O., Bastos, A. F., Campos, B., Mendlowicz, M. V., da Luz, M., Portella, C., Berger, W., Volchan, E., David, I. A., Erthal, F., Pereira, M. G., and de Oliveira, L. (2023). Machine learning applied to fmri patterns of brain activation in response to mutilation pictures predicts ptsd symptoms. *BMC Psychiatry*, 23(1):719.

Qureshi, M. N. I., Oh, J., and Lee, B. (2019). 3d-cnn based discrimination of schizophrenia using resting-state fmri. *Artificial Intelligence in Medicine*, 98:10–17.

Shin, L. M., Rauch, S. L., and Pitman, R. K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in ptsd. *Annals of the New York Academy of Sciences*, 1071(1):67–79.

Soares, D. C. S., dos Santos, L. A., and Donadon, M. F. (2021). Transtorno de estresse pós-traumático e prejuízos cognitivos, intervenções e tratamentos: uma revisão de literatura. *Revista Eixo*, 10(2):15–24.

Summerfield, J. J., Hassabis, D., and Maguire, E. A. (2009). Cortical midline involvement in autobiographical memory. *Neuroimage*, 44(3):1188–1200.

Suo, X., Lei, D., Li, W., Sun, H., Qin, K., Yang, J., Li, L., Kemp, G. J., and Gong, Q. (2022). Psychoradiological abnormalities in treatment-naive noncomorbid patients with posttraumatic stress disorder. *Depression and Anxiety*, 39(1):83–91.

Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., and Han, W. (2018). Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: A brief overview and our opinion. *Frontiers in Neuroinformatics*, 12:23.

Yang, J., Lei, D., Qin, K., Pinaya, W. H., Suo, X., Li, W., Li, L., Kemp, G. J., and Gong, Q. (2021). Using deep learning to classify pediatric posttraumatic stress disorder at the individual level. *BMC psychiatry*, 21:1–10.

Yin, W., Li, L., and Wu, F.-X. (2022). Deep learning for brain disorder diagnosis based on fmri images. *Neurocomputing*, 469:332–345.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham. Springer International Publishing.

Zhu, Z., Lei, D., Qin, K., Suo, X., Li, W., Li, L., DelBello, M. P., Sweeney, J. A., and Gong, Q. (2021). Combining deep learning and graph-theoretic brain features to detect posttraumatic stress disorder at the individual level. *Diagnostics*, 11(8):1416.

Zilcha-Mano, S., Zhu, X., Suarez-Jimenez, B., Pickover, A., Tal, S., Such, S., Marohasy, C., Chrisanthopoulos, M., Salzman, C., Lazarov, A., et al. (2020). Diagnostic and predictive neuroimaging biomarkers for posttraumatic stress disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(7):688–696.