

Inteligência Artificial Explicável Aplicada no Ensino e Aprendizagem de Conceitos de Rede Neural Convolutacional

Title: Explainable Artificial Intelligence Applied to the Teaching and Learning of Convolutional Neural Network Concepts

Título: Inteligencia Artificial Explicable Aplicada a la Enseñanza y el Aprendizaje de Conceptos de Red Neuronal Convolutacional

Marcelo Buscioli Tenorio
Centro Paula Souza (CPS)
ORCID: [0009-0002-1135-9900](https://orcid.org/0009-0002-1135-9900)
marcelo.tenorio@cps.sp.gov.br

Danilo Medeiros Eler
Universidade Estadual Paulista (UNESP)
ORCID: [0000-0002-9493-145X](https://orcid.org/0000-0002-9493-145X)
danilo.eler@unesp.br

Resumo

Este trabalho investiga o potencial da Inteligência Artificial Explicável (XAI) como recurso didático para o ensino de Redes Neurais Convolucionais (CNN) no ensino superior. O estudo teve como objetivo avaliar se técnicas de explicação de modelos de aprendizado profundo, como Mapas de Ativação de Classe e Mapas de Características, podem colaborar com a compreensão conceitual dos estudantes. Foi realizado um estudo de caso comparado com alunos de duas instituições de ensino superior da cidade de Presidente Prudente do estado de São Paulo, Brasil (Fatec e Unesp), totalizando 63 participantes. A pesquisa consistiu em uma aula expositiva dividida em duas partes: na primeira, os conceitos de CNN foram apresentados de forma tradicional; na segunda, foram adicionadas visualizações XAI. Um questionário aplicado ao término de cada parte permitiu mensurar ganhos de aprendizagem imediatos. Os resultados indicaram aumento significativo no número de acertos, especialmente entre os alunos com menor conhecimento prévio, além de alta percepção de contribuição das técnicas XAI para o entendimento do conteúdo (mais de 90% de concordância na Fatec e 73% na Unesp). Conclui-se que os recursos visuais da XAI favorecem o processo de ensino e aprendizagem de CNN, tornando conceitos abstratos mais acessíveis e promovendo maior engajamento dos estudantes.

Palavras-Chave: Inteligência Artificial; CNN; Ensino de Computação; XAI; Educação Superior.

Abstract

This study investigates the potential of Explainable Artificial Intelligence (XAI) as a didactic resource for teaching Convolutional Neural Networks (CNN) in higher education. The main objective was to evaluate whether explanation techniques for deep learning models – such as Class Activation Maps and Feature Maps – can support students' conceptual understanding. A comparative case study was conducted with 63 undergraduate students from two higher education institutions in Presidente Prudente, São Paulo state, Brazil (Fatec and Unesp). The experiment consisted of a lecture divided into two parts: the first presented CNN concepts using traditional visual materials, while the second incorporated XAI-based visualizations. A questionnaire administered at the end of each part allowed us to measure immediate learning gains. Results showed a significant increase in correct responses, particularly among students with less prior knowledge, and a high perception of XAI's contribution to learning (over 90% agreement at Fatec and 73% at Unesp). The findings suggest that XAI visual resources enhance the teaching and learning process of CNN by making abstract concepts more accessible and engaging to students.

Keywords: Artificial Intelligence; CNN; Computer Science Education; XAI; Higher Education.

Resumen

Este estudio investiga el potencial de la Inteligencia Artificial Explicable (XAI) como recurso didáctico para la enseñanza de Redes Neuronales Convolucionales (CNN) en la educación superior. El objetivo principal fue evaluar

si las técnicas de explicación de modelos de aprendizaje profundo, como los Mapas de Activación de Clase y los Mapas de Características, pueden contribuir a la comprensión conceptual de los estudiantes. Se realizó un estudio de caso comparado con 63 estudiantes de dos instituciones de educación superior de la ciudad de Presidente Prudente, estado de São Paulo, Brasil (Fatec y Unesp). El experimento consistió en una clase expositiva dividida en dos partes: la primera presentó los conceptos de las CNN de forma tradicional y la segunda incorporó visualizaciones basadas en XAI. Un cuestionario administrado al final de cada parte nos permitió medir los avances inmediatos en el aprendizaje. Los resultados mostraron un aumento significativo en el número de respuestas correctas, especialmente entre los estudiantes con menor conocimiento previo, y una alta percepción sobre la contribución de las técnicas XAI al aprendizaje (más del 90% de acuerdo en la Fatec y 73% en la Unesp). Se concluye que los recursos visuales de XAI favorecen el proceso de enseñanza y aprendizaje de las CNN, haciendo que los conceptos abstractos sean más accesibles y fomentando una mayor participación de los estudiantes.

Palabras clave: Inteligencia Artificial; CNN; Enseñanza de la Computación; XAI; Educación Superior.

1 Introdução

Os anos de 1940 e 1956 marcam o início das pesquisas sobre Inteligência Artificial (IA), foi nessa época que pesquisadores começaram a explorar o conceito de máquinas inteligentes (Turing, 1950), (McCarthy et al., 1955). As pesquisas na área atravessaram décadas e chegaram nos anos 2000, entretanto, até meados de 2013 as aplicações realizavam tarefas consideradas discriminativas, classificando ou prevendo algo, por exemplo, sistemas de recomendação, detecção de fraudes e diagnóstico médico. A IA que começou a criar algo, chamada de generativa teve seu marco principal nos anos de 2014 e 2017 (Goodfellow, et al., 2014), (Vaswani, et al., 2017). Mas a adoção massiva de aplicações deste tipo por todo mundo ocorreu no ano de 2022. Aplicações computacionais com as quais o ser humano conversa, não mais por palavras-chave, mas por linguagem natural escrita ou falada, para solicitar a realização de tarefas como geração de textos e imagens (Chatterji et al., 2025). Importante destacar que as aplicações aprimoraram a comunicação do humano com a máquina e vice-versa. O pedido é feito pelo humano falando naturalmente e a máquina apresenta respostas prontas, e não mais locais onde possivelmente estariam as respostas.

É neste contexto, que o termo IA se popularizou, em qualquer área da vida existe um grande interesse por aplicações que realizam as mais diversas tarefas possíveis, desde a receita para produção de cachorro-quente para uma festa entre amigos até a geração de programas complexos de computador (Dis et al., 2023). Observando essa demanda, muitos cursos sobre IA estão sendo ofertados, com o foco para iniciantes e profissionais em geral, ou ainda com o foco em Aprendizado de Máquina (ML) e ciência de dados para profissionais específicos (Research, 2025). Entretanto, é importante destacar que no eixo da tecnologia da informação, o ensino sobre IA no nível superior já ocorre muito antes de 2022. É esse público, enquanto estagiários ou profissionais que participaram, participam e participarão no desenvolvimento das aplicações de IA.

Um dos modelos de IA muito utilizado em diversas aplicações é a Rede Neural Convolutacional (CNN). As CNNs são utilizadas em classificação de imagens, detecção de objetos, segmentação de imagens, reconhecimento facial e de gestos, análise de imagens médicas, análise de vídeos, veículos autônomos e robótica. A IA se expandiu para a IA Explicável (XAI), área que apresenta técnicas de explicação de modelos de ML. As técnicas utilizam principalmente recursos visuais gráficos e numéricos para contribuir com o entendimento dos modelos. Geralmente são utilizadas para explicar os resultados apresentados para especialistas como médico, advogado e engenheiro, mas pode inclusive beneficiar o contexto educacional (Saarela & Podgorelec, 2024).

O processo de ensino e aprendizagem de CNN atualmente realizado é impactado por diversas limitações: complexidade do funcionamento interno do modelo; abordagens focadas na prática de programação; ferramentas visuais escassas; falta de estudos aplicando técnicas de XAI (Sanusi & Oyelere, 2020). O objetivo deste trabalho é experimentar e avaliar a proposta de um complemento metodológico em aulas expositivas para o ensino de CNN em cursos de nível superior. A hipótese é que as explicações provenientes das técnicas XAI colaboram com o processo de ensino e aprendizagem do conteúdo em questão. O trabalho tem como principais contribuições: proposta de um método educacional baseado em XAI; produção de um material adaptado para ensino de conceitos de CNN; um estudo de caso comparado para validar a proposta de acordo com o perfil do participante.

O artigo está organizado nas seções de trabalhos relacionados, metodologia utilizada, resultados obtidos, discussões, conclusões com propostas de trabalhos futuros, agradecimentos e referências.

2 Trabalhos Relacionados

Os trabalhos apresentados a seguir foram escolhidos observando os seguintes contextos: ensino tradicional de ML, interesse no uso de XAI na área da educação, técnicas recomendadas para explicação de CNN, ferramentas disponíveis para ensino de CNN, utilização da XAI para ensinar ML e estudos empíricos sobre a aplicação de XAI no processo de ensino e aprendizagem de ML.

O trabalho de Martins et al. (2023) aborda a necessidade de popularizar as competências de ML na educação básica. O objetivo do estudo é analisar a eficácia das metodologias ativas e do suporte tecnológico utilizados no curso ML4ALL. A metodologia adotada consiste em um estudo de caso múltiplo, envolvendo um total de 87 estudantes dos anos finais do ensino fundamental e médio. O curso, desenhado com base em metodologias ativas, propõe o ensino da aplicação de conceitos básicos de ML por meio de atividades práticas. Os resultados da avaliação demonstraram que os estudantes foram capazes de construir uma compreensão conceitual de ML e desenvolver um modelo de classificação de imagem, evidenciando a eficácia da abordagem pedagógica combinada com o suporte tecnológico.

A XAI tem recebido crescente atenção como meio de tornar modelos de ML mais transparentes e confiáveis. Revisões recentes destacam um aumento no interesse pela utilização de XAI em contextos educacionais, visando apoiar a aprendizagem, promover a confiança e oferecer transparência pedagógica. No entanto, essas revisões também enfatizam que a XAI na educação ainda é um campo emergente, com lacunas conceituais e metodológicas importantes, especialmente a escassez de estudos empíricos controlados que avaliem o impacto pedagógico das explicações sobre a aprendizagem dos estudantes (Khosravi, et al., 2022), (Türkmen, 2024).

No domínio técnico da XAI aplicada à visão computacional em Selvaraju et al (2017), os Mapas de Ativação de Classe (CAMs) e variantes baseadas em gradiente, como o Grad-CAM, tornaram-se ferramentas padrão para produzir explicações visuais de predições realizadas por CNNs. Em particular, o Grad-CAM gera mapas de ativação localizados ao ponderar mapas de características com gradientes provenientes da última camada convolucional, produzindo visualizações discriminativas por classe que se mostraram úteis tanto para depuração de modelos quanto para avaliações de interpretabilidade com usuários. Essas características tornam o Grad-CAM e métodos correlatos candidatos naturais para uso pedagógico, com o objetivo de mostrar como as CNNs focam em regiões específicas da entrada.

Complementando as visualizações do tipo CAM em Azam et al (2023), a inspeção de Mapas de Características intermediários fornece uma visão sobre a hierarquia de representações aprendida pelas CNNs: camadas iniciais tendem a codificar características de baixo nível (como bordas e texturas), enquanto camadas mais profundas capturam partes de objetos e conceitos semânticos mais abstratos. Estudos empíricos que analisam mapas de características demonstram seu potencial para ilustrar conceitos fundamentais de CNNs – como extração hierárquica de características, campos receptivos e abstração progressiva da informação –, o que as torna especialmente adequadas como artefatos instrucionais para o ensino conceitual (sem prática de implementação) do funcionamento interno das CNNs.

Chiaburu et al. (2024) aborda o potencial didático de explicações (XAI) em tarefas visuais complexas (classificação de espécies), os autores realizaram um experimento com 1200 participantes, comparando desempenho com e sem explicações, e testando efeitos de aprendizado duradouros. Os principais achados relatam que as explicações não melhoraram significativamente a precisão além da previsão simples, mas reduziram a incerteza. Usuários tenderam a replicar previsões erradas do modelo. Não foram encontrados efeitos de aprendizado duradouros.

O trabalho de Hao et al (2024) afirma que os avanços metodológicos recentes na geração de CAMs – como técnicas de redução de ruído, fusão e seleção, exemplificadas por abordagens

como o UnionCAM – têm como objetivo produzir mapas de ativação mais limpos e informativos, o que pode aumentar a utilidade pedagógica dessas visualizações ao reduzir ruídos e destacar regiões salientes de forma mais consistente. Incorporar tais refinamentos pode resultar em materiais instrucionais mais claros e mitigar interpretações equivocadas que podem surgir quando mapas de ativação ruidosos são apresentados a estudantes.

Do ponto de vista prático e de reprodutibilidade, diversos repositórios e ferramentas de código aberto reúnem Grad-CAM, Guided Backpropagation, Occlusion Maps e variantes, reduzindo a barreira para que docentes gerem visualizações explicativas de alta qualidade a partir de modelos pré-treinados. Esses recursos permitem a produção rápida de exemplos visuais para uso em sala de aula, sem exigir que estudantes treinem ou implementem modelos por conta própria (Gildenblat, 2025).

Apesar da disponibilidade de técnicas e ferramentas XAI, o trabalho de Naveed et al (2024) relata que um tema recorrente na literatura de avaliação de XAI é a escassez relativa de estudos empíricos robustos e centrados no usuário que meçam como diferentes estilos de explicação afetam a compreensão e os resultados de aprendizagem. Revisões sobre avaliação empírica de XAI recomendam desenhos de estudo rigorosos (pré/pós-teste, métricas de compreensão validadas e amostras com poder estatístico adequado) e fornecem diretrizes para alinhar tipos de explicação com medidas de avaliação – práticas que são essenciais para estudos pedagógicos que buscam quantificar ganhos de aprendizagem obtidos a partir de visualizações XAI.

Em síntese, a literatura aponta duas lacunas que motivaram o presente trabalho: (1) embora as técnicas de XAI para CNNs estejam maduras e suportadas por ferramentas práticas, há pouca evidência empírica sobre sua eficácia como intervenções educacionais para o ensino de conceitos de CNNs a estudantes de graduação; e (2) as melhores práticas da literatura de avaliação de XAI sugerem o uso de desenhos experimentais rigorosos (pré/pós-teste, instrumentos validados e análise quantitativa) para fundamentar alegações sobre ganhos de aprendizagem. Este estudo busca preencher essas lacunas ao aplicar visualizações do tipo Mapa de Ativação e Mapa de Características como artefatos didáticos conceituais para estudantes de graduação da área de tecnologia da informação, avaliando seu efeito sobre a compreensão conceitual por meio de questionários aplicados antes e depois da intervenção.

3 Metodologia

A condução do experimento foi da seguinte forma: 1 - Exposição inicial (sem as técnicas XAI); 2 - Aplicação do pré-questionário; 3 - Exposição complementar (com as técnicas XAI); 4 - Aplicação do pós-questionário. O objetivo da exposição inicial é de nivelamento através do ensino tradicional e o objetivo da exposição complementar é de aprimoramento através de uma inovação no método de ensino. O experimento foi uma aula expositiva teórica dividida em duas partes, os itens 1 e 2 compõem a primeira parte da aula e os itens 3 e 4 a segunda parte. Tanto na exposição inicial quanto na complementar, a comunicação foi visual (imagens e textos projetados em tela) e verbal entre o professor e os alunos. Os alunos podiam tirar dúvidas durante a exposição, mas não podiam fazer anotações para consulta quando fossem responder os questionários. Com exceção da primeira pergunta do pré-questionário e das duas últimas perguntas do pós-questionário, as demais oito perguntas foram as mesmas nos dois questionários. A duração do experimento foi em média 100 minutos.

Com exceção da primeira pergunta sobre conhecimento prévio do pré-questionário, das últimas perguntas sobre a percepção do aluno em relação ao recurso utilizado e da pergunta em aberto (discursiva) do pós-questionário, para elaboração das oito demais perguntas que compõem os dois questionários, levou-se em consideração o conceito que a pergunta aborda, qual conteúdo

da aula apoia a resposta, se a XAI colabora ou não com a resposta e se a resposta seria por memorização ou não.

Este experimento, conduzido pelo número CAAE 91521025.6.0000.5402, via sistema do Governo Brasileiro, foi elaborado e aprovado com todos os cuidados recomendados por um comitê de ética, são eles: termo de consentimento livre e esclarecido; termo de responsabilidade com os dados coletados; declaração de autorização de ambas as instituições e termo de compromisso de cumprimento de resolução, norma e regimento pertinentes.

O experimento foi realizado em duas instituições de ensino superior na cidade de Presidente Prudente do estado de São Paulo, a Faculdade de Tecnologia (Fatec) e a Universidade Estadual Paulista (Unesp). As turmas escolhidas no 2º semestre de 2025 foram, Inteligência Artificial com 40 alunos do curso de Tecnologia em Análise e Desenvolvimento de Sistemas da Fatec e Inteligência Artificial com 30 alunos do curso de Bacharelado em Ciência da Computação da Unesp. A amostra foi composta por 37 participantes do curso da Fatec e 26 participantes do curso da Unesp.

O tema da aula foi Redes Neurais Convolucionais – um modelo caixa-preta de aprendizado de máquina para classificação de imagens. Na exposição inicial os tópicos abordados foram:

A) Introdução

O aprendizado de máquina tem como objetivo principal disponibilizar recursos para que a máquina realize as mesmas atividades humanas. Na vertente de aprendizagem profunda estão os modelos de redes neurais considerados de caixa-preta pois possuem alta complexidade para acompanhamento de seu funcionamento interno.

B) Arquitetura de uma CNN

A figura 1 ilustra as camadas de uma rede neural convolucional.

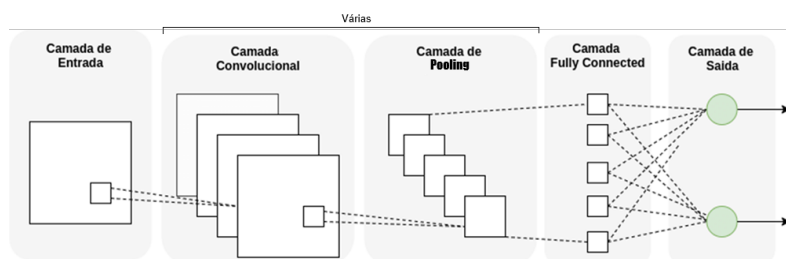


Figura 1: Arquitetura de uma CNN.

C) Camada de Entrada

Conforme ilustra a figura 2, nesta camada as imagens são representadas como matrizes de pixels, cada pixel tem um valor que representa a intensidade da cor e os valores são normalizados para facilitar o processamento.

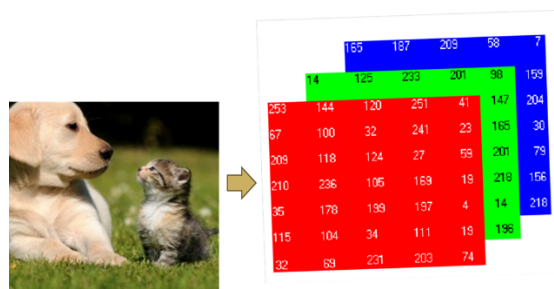


Figura 2: Representação de imagens.

D) Camada de Convolução

Conforme ilustra a figura 3, nesta camada são detectados os padrões como bordas, texturas e formas, a função de ativação utilizada é a ReLU. São utilizados filtros para extrair as características, cada filtro é uma matriz pequena que percorre a imagem aplicando operações matemáticas. A saída é um mapa de características destacando informações relevantes.

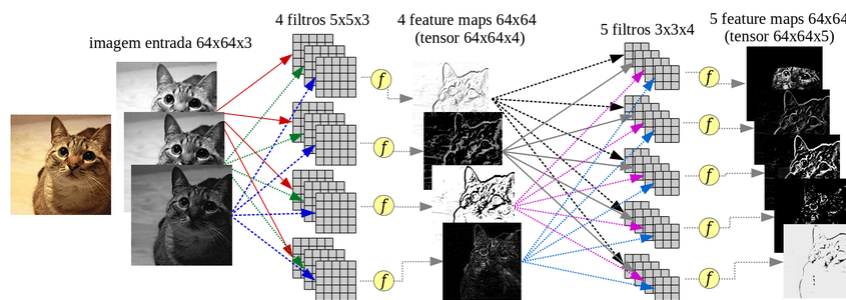


Figura 3: Aplicação de filtros.

E) Camada de Pooling

Conforme ilustra a figura 4, nesta camada ocorre a redução da dimensionalidade da imagem sem perder informações essenciais. Esse passo impacta no tempo de processamento, na eficiência e na visualização da informação.



Figura 4: Redução da dimensionalidade.

F) Camada de Saída

Conforme ilustra a figura 5, esta camada é conhecida por totalmente conectada, possui uma rede neural artificial multicamada, a função de ativação utilizada é a Softmax e a saída é uma probabilidade associada a cada classe.

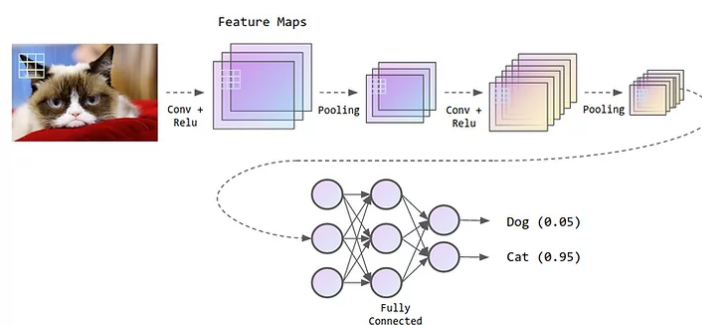


Figura 5: Classificação de uma imagem.

Em seguida, os participantes responderam anonimamente o pré-questionário online composto por nove perguntas.

Na exposição complementar, os participantes receberam explicações através dos recursos provenientes das técnicas da XAI, mais especificamente o Mapa de Ativação de Classe e o Mapa

de Características, pois de acordo com a literatura descrita na seção dois, são técnicas recomendadas para entendimento do funcionamento de CNN. Foram utilizadas ferramentas online de interação humano computador, apenas por parte do professor. As ferramentas utilizam a rede GoogLeNet que possui camadas de convolução chamadas mixed (Olah, et al., 2018), (Carter et al., 2019). Os tópicos abordados foram:

A) Camadas Ocultas

Conforme ilustra a figura 6, os números que para o ser humano é algo bastante abstrato são transformados em imagens que oferecem significados, como contorno de orelha, focinho de cachorro, cabeça de gato, pernas peludas e grama.

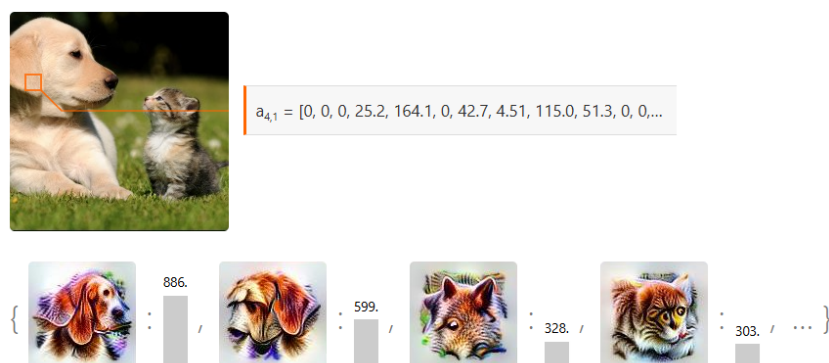


Figura 6: Significado dos números (Olah, et al., 2018).

B) A Visão da Rede

Conforme ilustra a figura 7, é preciso fazer otimizações por conta da grande quantidade de neurônios existentes na rede.

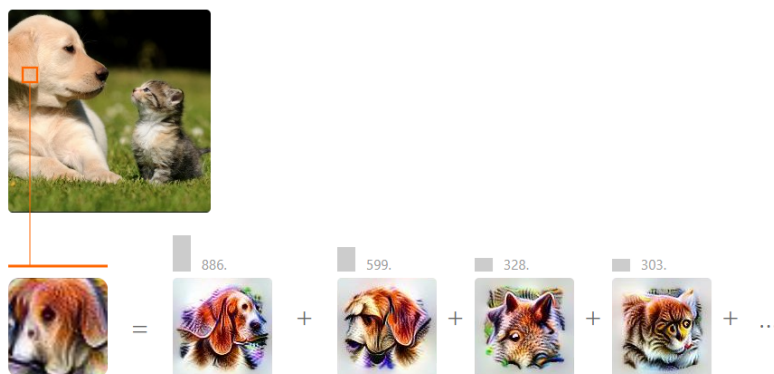


Figura 7: Vetor de ativação (Olah, et al., 2018).

Conforme ilustra a figura 8, é possível ver não apenas o que a rede detecta em cada posição, mas também o que a rede entende da imagem de entrada como um todo.

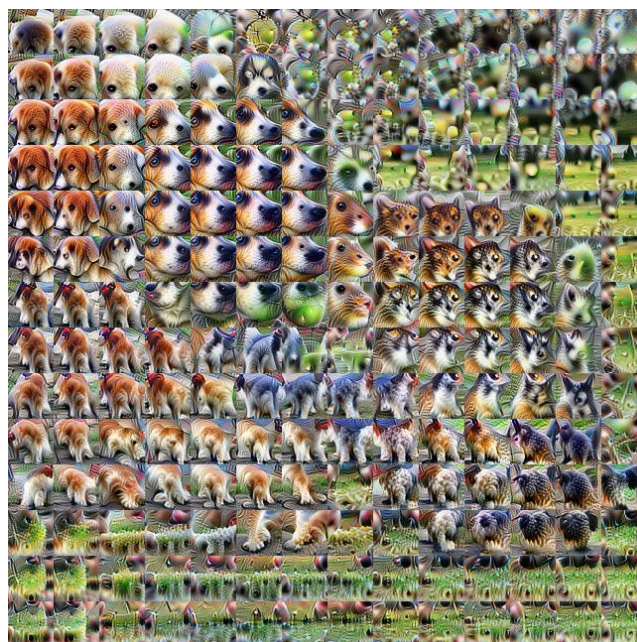


Figura 8: Entendimento da imagem (Olah, et al., 2018).

Conforme ilustra a figura 9, é possível observar como a compreensão da rede evolui entre as camadas, desde a detecção de bordas em camadas iniciais até formas e partes de objetos mais sofisticados nas últimas.

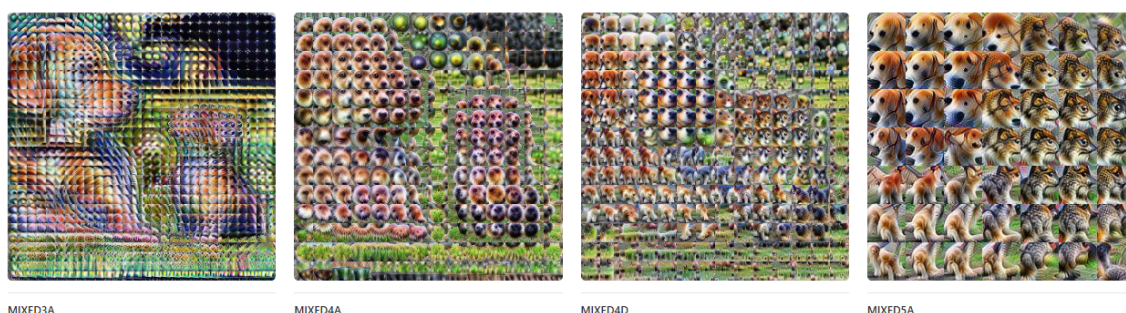


Figura 9: Evolução da compreensão da rede (Olah, et al., 2018).

Essas visualizações, no entanto, omitem uma informação crucial, a magnitude das ativações. Conforme ilustra a figura 10, ao dimensionar a área de cada célula pela magnitude do vetor de ativação, é possível indicar a intensidade com que a rede detectou as características.

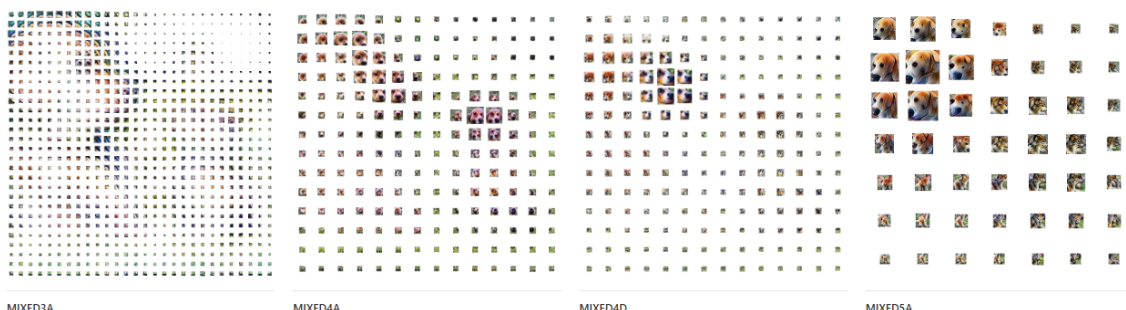


Figura 10: Intensidade das características (Olah, et al., 2018).

C) Montagem dos Conceitos

A visualização de características nos ajuda a responder o que a rede detecta, mas não responde como a rede monta essas peças individuais para chegar a decisões posteriores ou por que essas decisões foram tomadas. A interface mais comum para atribuição é chamada de mapa

de saliência – um mapa de calor simples que destaca os pixels da imagem de entrada que mais influenciaram a classificação. Conforme ilustra a figura 11, é possível observar através da combinação das características e saliências das camadas ocultas de convolução, quais as regiões mais importantes da imagem de entrada.

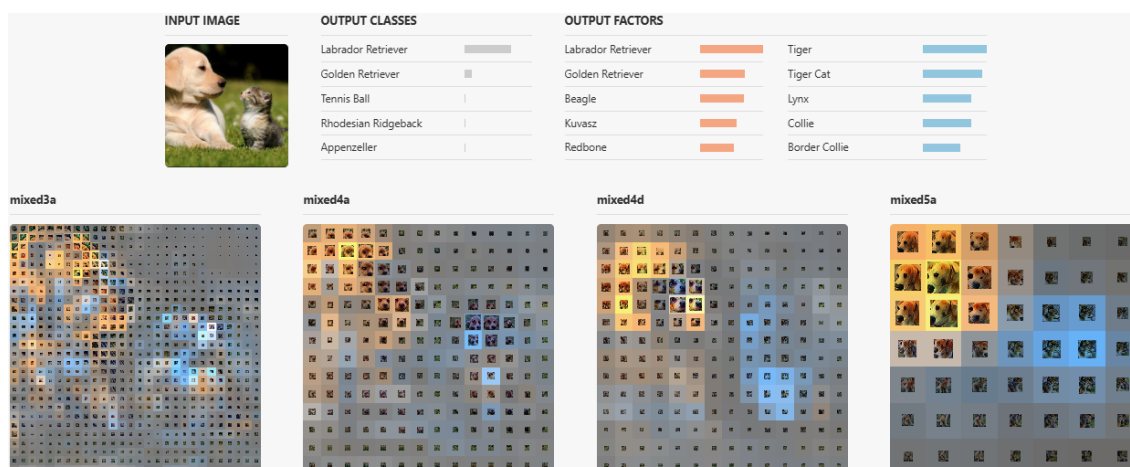


Figura 11: A influência das camadas de convolução (Olah, et al., 2018).

A figura 12 complementa a visualização através do foco numa determinada célula da grade da camada mixed5a.

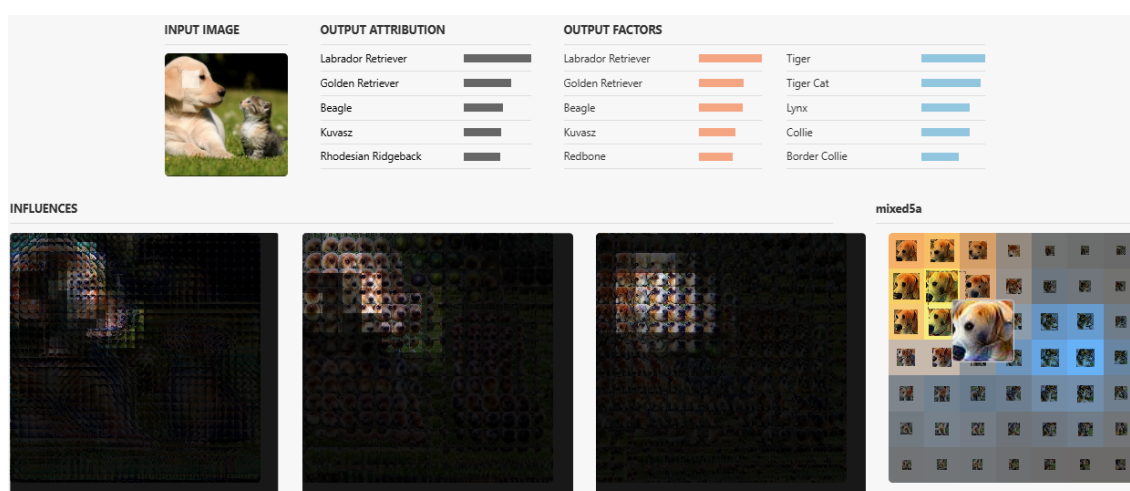


Figura 12: Observando a camada mixed5a (Olah, et al., 2018).

D) Tornando em Escala Humana

Os números expressivos de uma CNN, como a quantidade de camadas de convolução, a quantidade de neurônios e a quantidade de filtros dificultam a visualização da informação. Conforme ilustra a figura 13, para superar essa dificuldade foi utilizada a abordagem de Olah, et al. (2018) que utiliza o conceito de grupos relevantes de neurônios.

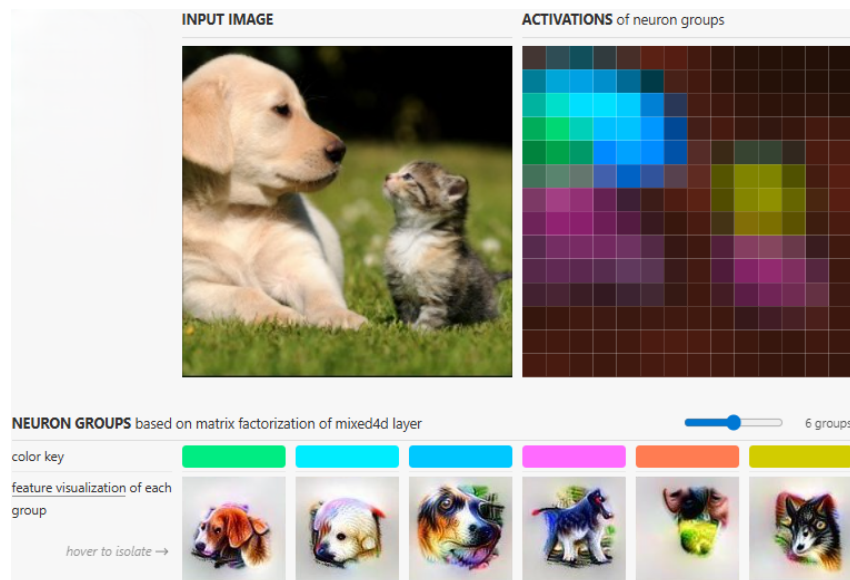


Figura 13: Grupos relevantes de neurônios (Olah, et al., 2018).

E) Agregando Várias Imagens

As grades de ativação mostram como a rede vê uma única imagem, para entender como ela reage a muitas imagens é preciso olhar para a grade de ativação individual de cada imagem. Entretanto, o cérebro humano não é bom em comparar muitos exemplos. Conforme ilustra a figura 14, a abordagem de Carter et al. (2019), chamada de Atlas de Ativação foi utilizada nesta situação, ela coleta a ativação de diversas imagens, reduz a dimensão e com o critério de semelhança desenha uma grade contendo em cada célula a característica extraída da média das ativações. A célula tem as medidas de acordo com a densidade da ativação média.

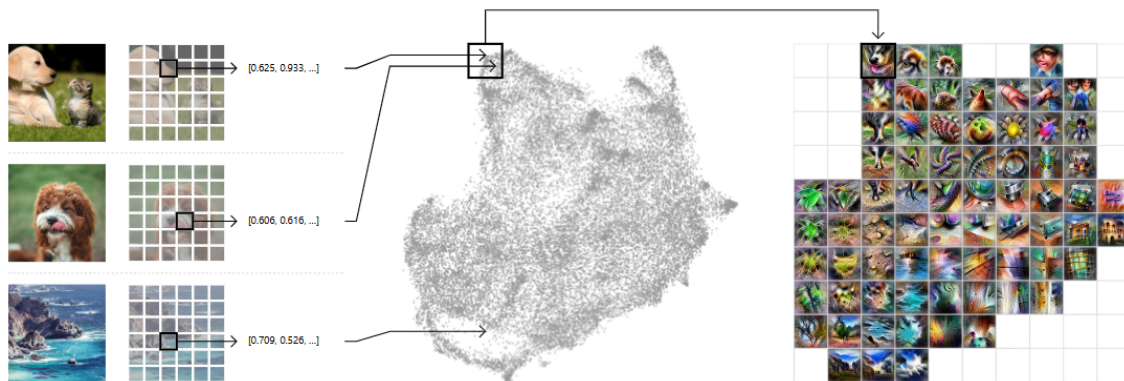


Figura 14: Geração do atlas de ativação (Carter et al., 2019).

Conforme ilustra a figura 15, no canto superior esquerdo do atlas, é possível ver imagens que parecem com cabeças de animais. À medida que avança para baixo, é possível ver diferentes tipos de pelo e as costas de animais de quatro patas. Abaixo disso, encontra-se diferentes pernas e pés de animais apoiados em diferentes tipos de solo. Abaixo dos pés, é possível ver terrenos e pisos isolados. Esses fundos arenosos e rochosos se misturam lentamente à água.



Figura 15: Atlas de ativação da camada mixed4c (Carter et al., 2019).

F) Olhando para Várias Camadas

Conforme ilustra a figura 16, é possível notar que a camada inicial mixed3b é muito inespecífica em comparação com as outras. As características que emergem são de padrões e manchas de cor. Na camada intermediária mixed4c, as características definitivamente se assemelham a folhas, mas podem ser qualquer tipo de planta. Na camada final mixed5b é possível ver folhagem com texturas específicas do repolho e curvadas em bolas arredondadas.

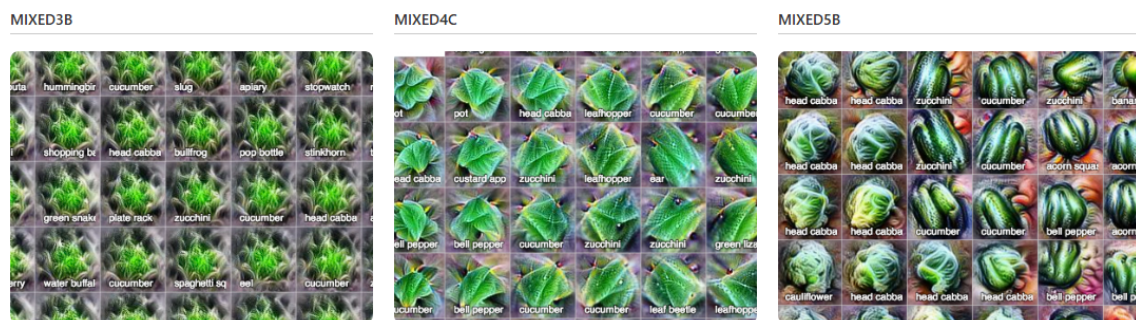


Figura 16: Evolução das imagens (Carter et al., 2019).

G) Observando uma Classificação

Conforme ilustra a figura 17, os círculos numerados apresentam as regiões do atlas da camada mixed4d que a rede utilizou para classificar barco de bombeiro e bonde. Observe no círculo 1 que tanto o bonde quanto o barco de bombeiro contêm ativações para o que parecem ser janelas. No círculo 2 observe que em ambas as classes contêm ativações para aparelhos semelhantes a guindastes, embora sejam menos proeminentes (opacidade) do que as ativações de janelas. No círculo 3, as ativações de barco de bombeiro têm atribuições muito mais fortes para a água em comparação com as ativações do bonde, onde praticamente não há evidências positivas. No círculo 4, as ativações para bonde têm fortes atribuições de edifícios do que para barco de bombeiro.

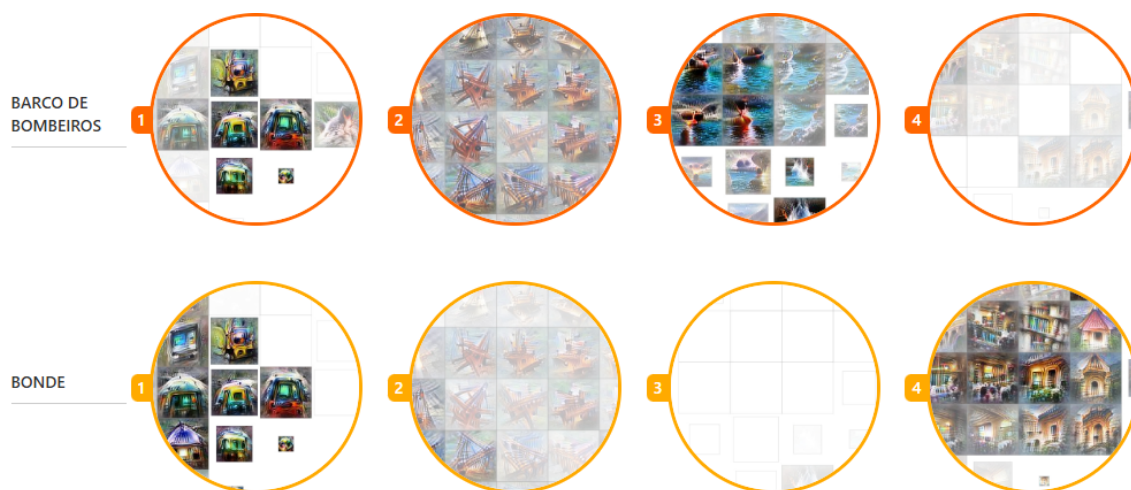


Figura 17: Ativações no atlas (Carter et al., 2019).

Concluindo, olhando para alguns exemplos de imagens de entrada, é possível ver como fundos de edifícios ou de água é uma maneira fácil de diferenciar entre um barco de bombeiro e um bonde.

Em seguida, os participantes responderam anonimamente o pós-questionário online composto por dez perguntas.

4 Resultados e Discussões

A seguir são apresentados os resultados do estudo empírico deste trabalho, que descrevem e comparam os dados de forma quantitativa e qualitativa.

4.1 Resultados da Fatec

A figura 18 apresenta o resultado da primeira pergunta do pré-questionário, a respeito do nível de conhecimento sobre CNN. Quase 46% dos participantes não possuíam conhecimento sobre CNN e 43,2% possuíam pouco conhecimento, um total de quase 90%.

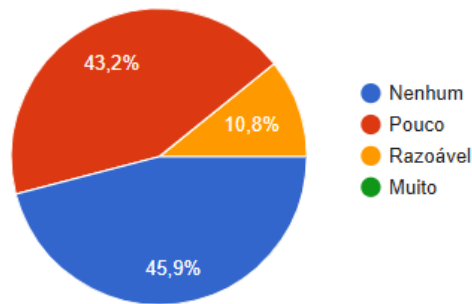


Figura 18: Antes desta aula, você já possuía algum conhecimento sobre rede neural convolucional?

A seguir são apresentados os resultados das oito perguntas feitas aos participantes no pré e pós-questionário. O gráfico à esquerda é do pré-questionário, o gráfico à direita é do pós-questionário e em seguida a legenda.

A figura 19 apresenta o resultado da pergunta número um, relacionada a arquitetura básica de uma CNN e envolve a habilidade de memorização por parte do estudante. Observa-se que o índice de respostas corretas (vermelho) aumentou 5,7% e o índice de incerteza (roxo) reduziu, porém, algumas respostas erradas ainda aparecem (verde e azul).

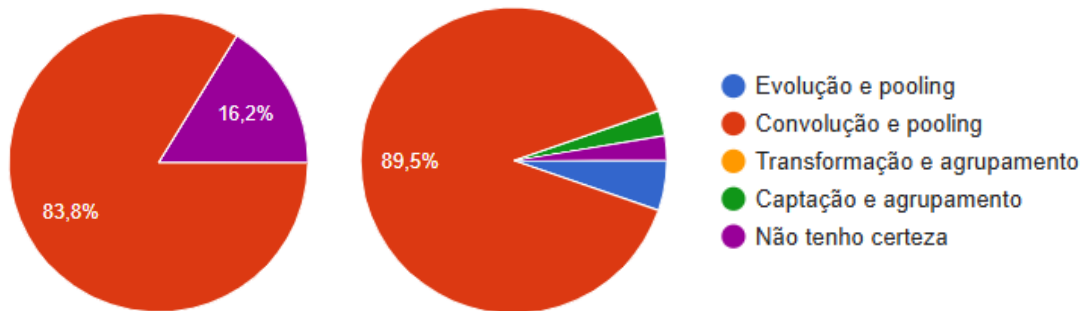


Figura 19: 1) A arquitetura de uma rede neural convolucional é composta por uma camada de entrada, várias camadas ocultas e uma camada de saída. Quais são os nomes das camadas ocultas?

A figura 20 apresenta o resultado da pergunta número dois, relacionada ao formato de entrada dos dados na rede e que envolve a habilidade de memorização por parte do estudante. Observa-se que um índice de 19% dos participantes migrou da incerteza (laranja) para a resposta correta (azul), restando ainda um índice de 10,5% de inseguros.

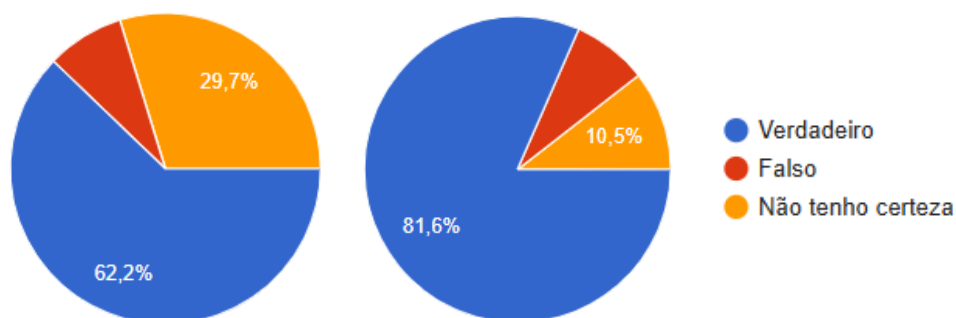


Figura 20: 2) Na entrada da rede, a imagem que tem altura (A) e largura (L) é transformada em uma matriz de pixels que representam a intensidade das cores. A dimensão da matriz depende da quantidade de canais (C) da imagem. Nas imagens coloridas (RGB) a matriz tem dimensão $A \times L \times 3$.

A figura 21 apresenta o resultado da pergunta número três, relacionada aos filtros da rede. Após o uso dos recursos visuais da XAI observa-se um aumento de 16,5% de acerto (laranja) e

uma redução na incerteza (roxo) e respostas erradas (verde, azul e vermelho), demonstrando contribuição com o processo de ensino e aprendizagem.

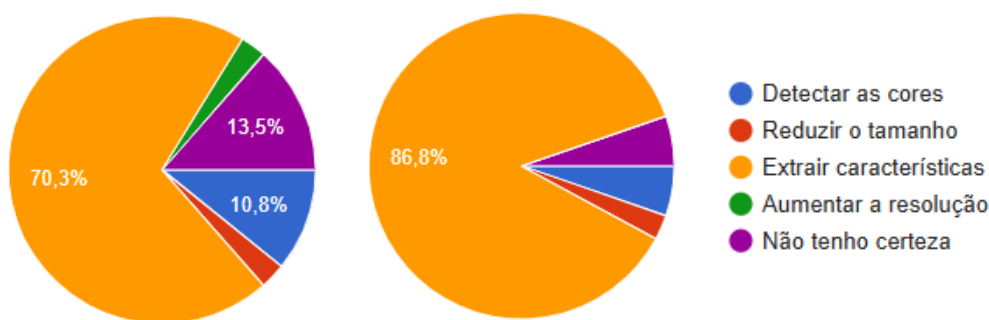


Figura 21: 3) Qual a função dos filtros (cálculos aplicados aos pixels da imagem) que são utilizados nas camadas de convolução?

A figura 22 apresenta o resultado da pergunta número quatro, relacionada à evolução de aprendizado da rede (de contornos e texturas às formas), a resposta correta é falso (vermelho), entretanto, aproximadamente 60% dos participantes responderam verdadeiro (azul) antes e depois da intervenção com os recursos visuais da XAI. No pós-questionário, aproximadamente 30% acertaram (vermelho) e aproximadamente 10% permaneceram inseguros (laranja). Este resultado não era esperado, a explicação utilizada apresentou a evolução de aprendizado da rede.

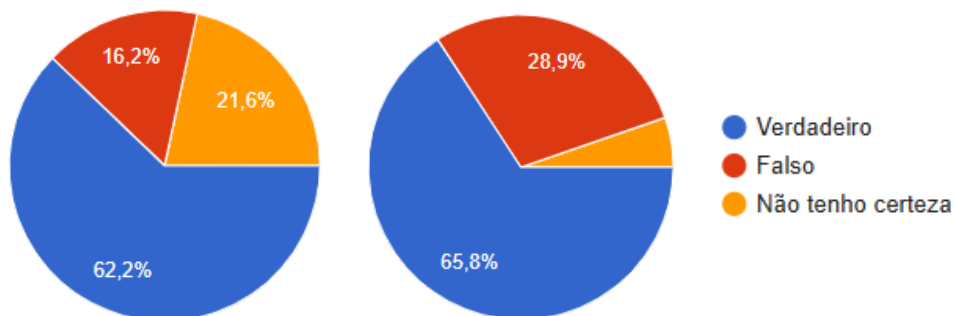


Figura 22: 4) Existem várias camadas de convolução. As camadas de convolução iniciais detectam formas mais complexas (por exemplo, face do cachorro), as camadas intermediárias detectam contornos e as camadas finais detectam texturas (por exemplo, pelo do cachorro).

A figura 23 apresenta o resultado da pergunta número cinco, relacionada ao relacionamento entre classes e padrões aprendidos pela rede (função de ativação). A intervenção com os recursos da XAI podem colaborar de alguma forma, mas a habilidade de memorização é mais relevante nesta pergunta. Tanto no pré quanto no pós-questionário observa-se um percentual de 60% de acerto (verde), apesar de erro (vermelho) e incerteza (roxo) estarem presentes. Importante destacar que no final da primeira parte da aula, a função Softmax é citada na camada de saída, isto pode ter influenciado as repostas erradas nesta pergunta.

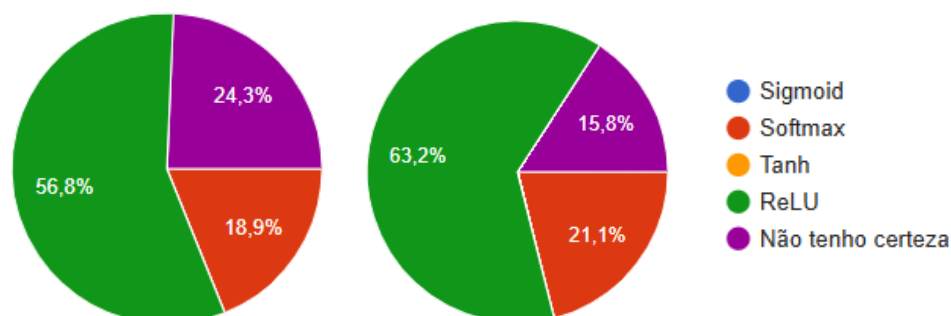


Figura 23: 5) Com o uso de funções de ativação não lineares, a rede pode aprender contornos e texturas de imagens. Qual é a função de ativação mais comum nas camadas de convolução?

A figura 24 apresenta o resultado da pergunta número seis, relacionada às partes da imagem que mais influenciam a decisão da rede. Nesta pergunta era esperado uma boa contribuição da XAI no percentual de respostas corretas e foi o que ocorreu, do pré para o pós-questionário houve um aumento de 25% de acertos (azul), exatamente o percentual reduzido dos que não tinham certeza.

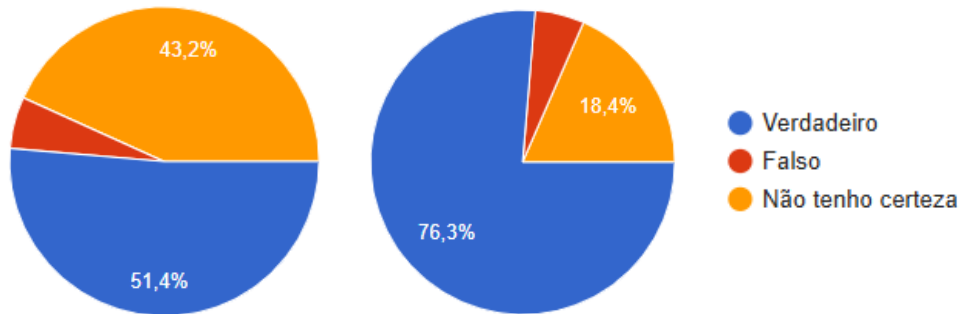


Figura 24: 6) As camadas ocultas (convolução) extraem as características da imagem. A maneira que a rede vê e entende a imagem é através dessas características. Portanto, para explicar a classificação final da rede, é importante observar a relação que a saída (classe) tem com as regiões (pixel) da imagem nas camadas ocultas.

A figura 25 apresenta o resultado da pergunta número sete, relacionada à observação se a rede está focando em regiões relevantes da imagem. A base teórica para responder esta pergunta é apresentada na primeira parte da aula, entretanto, observa-se melhora nos acertos (verde) após a segunda parte da aula. Talvez porque a segunda parte da aula reforça o objetivo da camada de convolução na extração de características, favorecendo lembrar qual é o objetivo da camada de pooling.

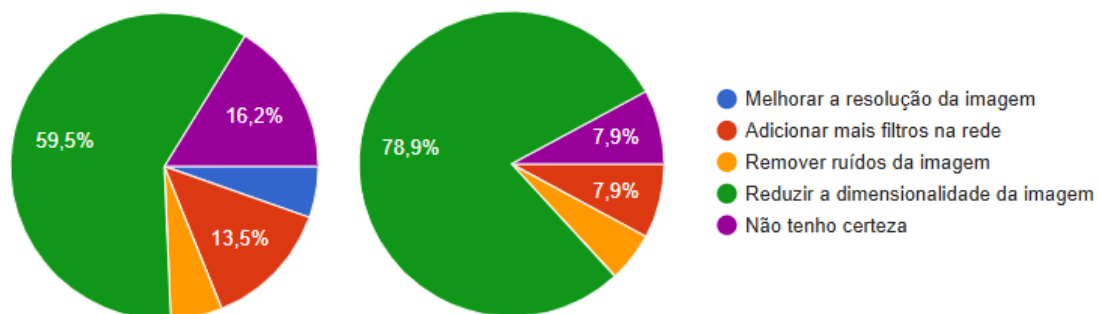


Figura 25: 7) As camadas de pooling proporcionam melhorias no tempo de processamento, na eficiência e na visualização da informação. Qual é o principal objetivo das camadas de pooling?

A figura 26 apresenta o resultado da pergunta número oito, relacionada ao relacionamento entre classes e padrões aprendidos pela rede (camada de saída). Este assunto é tratado na primeira parte da aula e pouco comentado na segunda parte. O percentual de acerto (vermelho) foi em torno de 50%, entretanto, chama a atenção o percentual de erro (azul) que aumentou no pós-questionário.

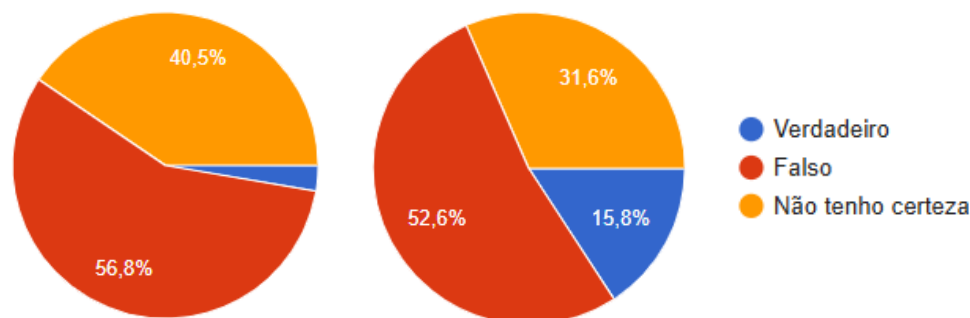


Figura 26: 8) Os dados passam pela camada de saída que é uma rede neural tradicional de múltiplas camadas, mas a probabilidade de pertencimento a cada classe ainda não é calculada nesta camada.

A figura 27 apresenta o resultado da penúltima pergunta do pós-questionário que possui uma escala de 1 para “Pouco” a 5 para “Muito”. Somando os que responderam 4 e 5, um total de 92% reconhecem que os recursos da XAI contribuem com a aprendizagem.

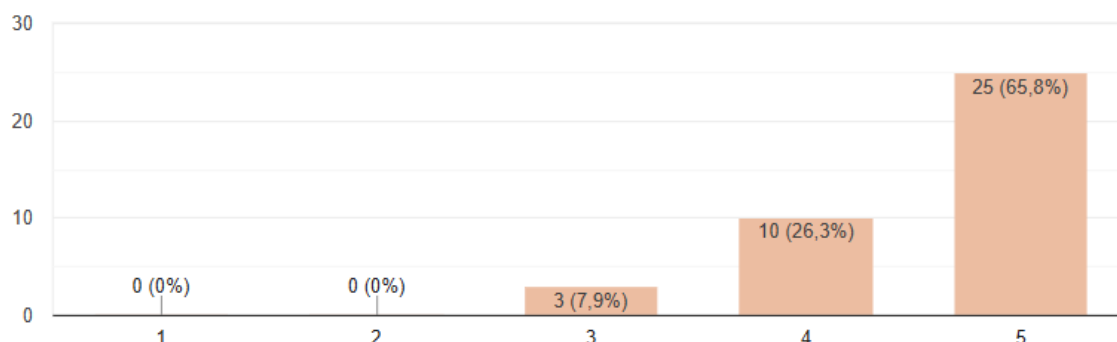


Figura 27: Nesta parte da aula foi utilizado técnicas de Inteligência Artificial Explicável (mapa de características e mapa de ativação de classe). Ver o funcionamento interno da rede através de recursos visuais contribuiu com sua aprendizagem?

Na última pergunta do pós-questionário, a única opcional, o participante podia escrever o que gostou e dar sugestões de melhorias. A seguir são descritos alguns dos principais pontos positivos e negativos encontrados:

a) Didática e clareza da explicação: Muitos participantes destacaram que o conteúdo foi didático, claro e fácil de compreender, mesmo para quem não tinha familiaridade com o tema. Destaque para o comentário “O trabalho está muito didático e fácil de compreender.”.

b) Forma da exposição: As demonstrações visuais e os exemplos interativos foram apontados como o ponto alto da aula. Os alunos apreciaram ver o funcionamento interno da rede, como ela processa e classifica imagens. Muitos mencionaram que ver os resultados ajudou a tornar o conteúdo mais concreto. Destaque para os comentários “As imagens mostradas ajudaram a visualizar o funcionamento.” e “Gostei da explicação com o site interativo.”.

c) Interesse e impacto do conteúdo: Diversos comentários indicam encantamento e curiosidade sobre o tema. Destaque para o comentário “Nunca tinha visto algo assim, contribuiu bastante para meu conhecimento.”.

d) Tempo insuficiente: Alguns participantes sentiram que o tempo para cada parte da aula foi curto. Destaque para o comentário “Na segunda aula, senti falta de explicação teórica.”.

e) Necessidade de mais exemplos: Alguns participantes sugerem mais exemplos com variações de imagens para reforçar os conceitos. Destaque para o comentário “Poderia usar mais exemplos para melhor comparação e visualização.”.

4.2 Resultados da Unesp

Conforme observado na figura 28, o resultado da primeira pergunta do pré-questionário apresenta que 23% não possuíam conhecimento em CNN, 50% possuíam pouco conhecimento e quase 27% possuíam conhecimento razoável. Dividindo em dois grupos, baixo e alto conhecimento, apesar de um percentual total considerado de baixo conhecimento (73%), há um percentual significativo (27%) de conhecimento alto.

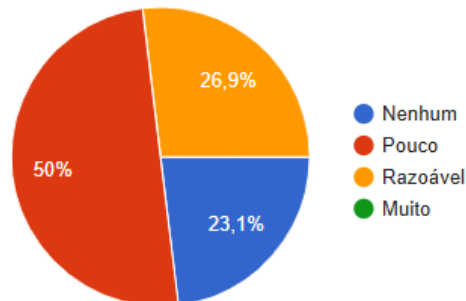


Figura 28: Antes desta aula, você já possuía algum conhecimento sobre rede neural convolucional?

A seguir são apresentados os resultados das oito perguntas feitas aos participantes no pré e pós-questionário. O gráfico à esquerda é do pré-questionário, o gráfico à direita é do pós-questionário e em seguida a legenda.

A figura 29 apresenta o resultado da pergunta número um, referente ao nome das camadas ocultas. No pré-questionário 88,5% acertaram (vermelho) a resposta e no pós 92,3% acertaram. Chama a atenção a redução do percentual de incerteza (roxo) e os baixos índices de respostas erradas (laranja e azul).

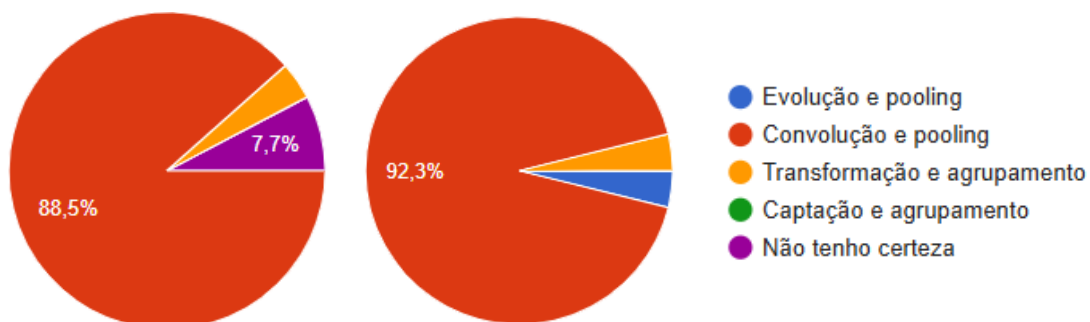


Figura 29: 1) A arquitetura de uma rede neural convolucional é composta por uma camada de entrada, várias camadas ocultas e uma camada de saída. Quais são os nomes das camadas ocultas?

A figura 30 apresenta o resultado da pergunta número dois, referente ao formato dos dados de entrada. Observa-se uma baixa variação dos percentuais entre o pré e o pós-questionário, destacando um alto índice de respostas corretas (azul), importante destacar que este assunto é tratado no início da primeira parte da aula.

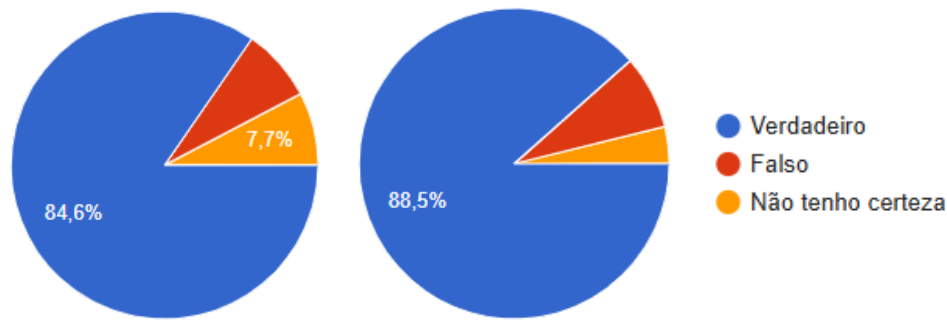


Figura 30: 2) Na entrada da rede, a imagem que tem altura (A) e largura (L) é transformada em uma matriz de pixels que representam a intensidade das cores. A dimensão da matriz depende da quantidade de canais (C) da imagem. Nas imagens coloridas (RGB) a matriz tem dimensão $A \times L \times 3$.

A figura 31 apresenta o resultado da pergunta número três, referente ao objetivo dos filtros presentes nas camadas de convolução. Apesar de uma resposta errada (vermelho) no pré-questionário equivalente a 3,8%, no geral houve um índice de 100% de acerto.

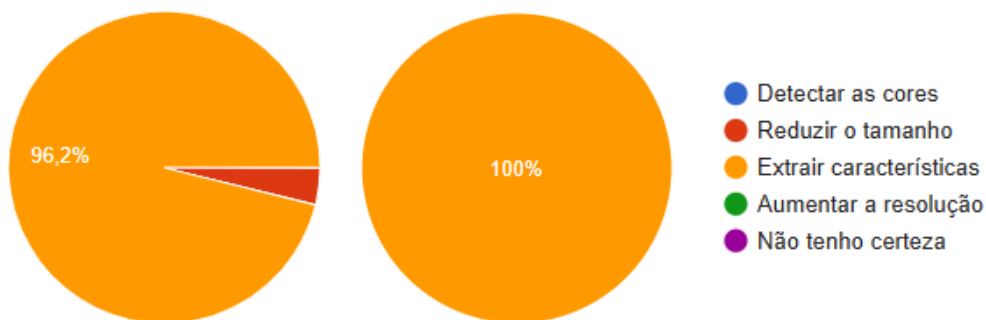


Figura 31: 3) Qual a função dos filtros (cálculos aplicados aos pixels da imagem) que são utilizados nas camadas de convolução?

A figura 32 apresenta o resultado da pergunta número quatro, referente a evolução do aprendizado da rede entre as camadas. Com o uso da XAI observa-se uma pequena melhora no percentual de acertos (vermelho). Importante destacar que a pergunta utiliza uma sequência errada na evolução do aprendizado da rede, dificultando a resposta.

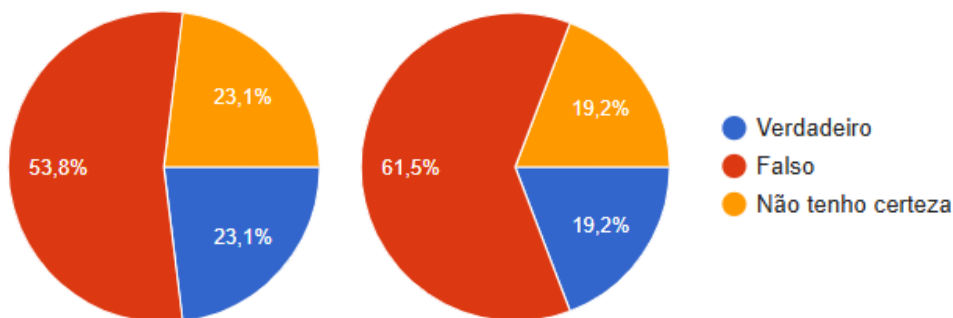


Figura 32: 4) Existem várias camadas de convolução. As camadas de convolução iniciais detectam formas mais complexas (por exemplo, face do cachorro), as camadas intermediárias detectam contornos e as camadas finais detectam texturas (por exemplo, pelo do cachorro).

A figura 33 apresenta o resultado da pergunta número cinco, referente ao nome da função de ativação das camadas de convolução. Quase 77% acertaram (verde) a resposta tanto antes quanto depois das técnicas XAI.

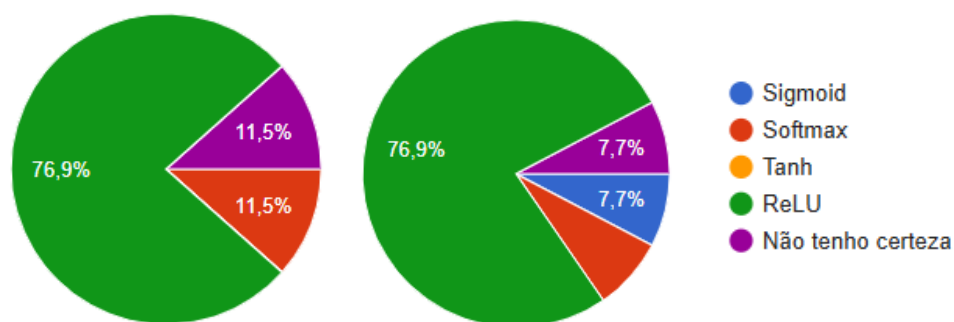


Figura 33: 5) Com o uso de funções de ativação não lineares, a rede pode aprender contornos e texturas de imagens. Qual é a função de ativação mais comum nas camadas de convolução?

A figura 34 apresenta o resultado da pergunta número seis, referente à maneira que a rede vê e entende a imagem. No pré-questionário o percentual de acerto (azul) foi alto (84,6%), porém, no pós-questionário o percentual de acerto foi de 100%. Evidenciando a colaboração das técnicas XAI com o processo de ensino e aprendizagem de CNN.

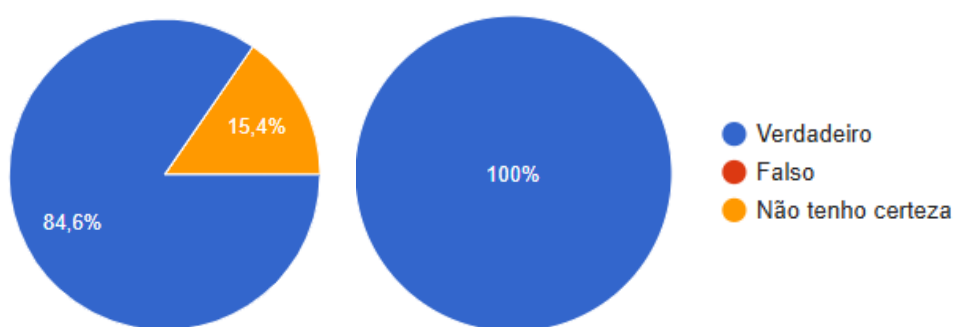


Figura 34: 6) As camadas ocultas (convolução) extraem as características da imagem. A maneira que a rede vê e entende a imagem é através dessas características. Portanto, para explicar a classificação final da rede, é importante observar a relação que a saída (classe) tem com as regiões (pixel) da imagem nas camadas ocultas.

A figura 35 apresenta o resultado da pergunta número sete, referente às camadas de pooling. Houve uma pequena melhora, em torno de 7%, no pós-questionário. Este resultado não é decorrente do uso das técnicas XAI pois na segunda parte da aula não é comentado sobre a camada de pooling.

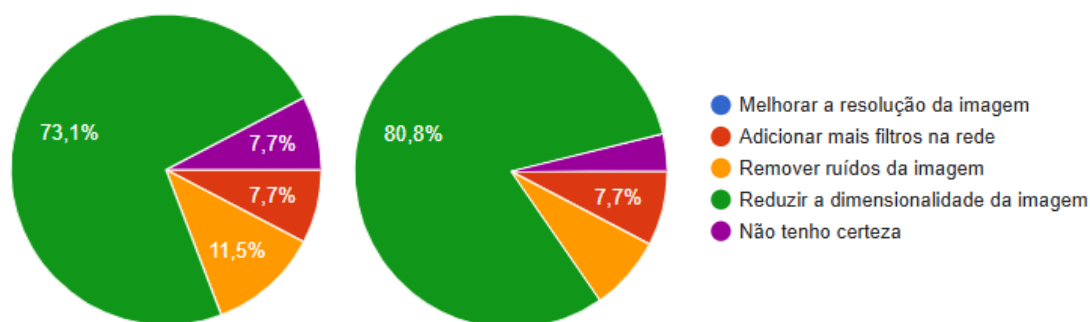


Figura 35: 7) As camadas de pooling proporcionam melhorias no tempo de processamento, na eficiência e na visualização da informação. Qual é o principal objetivo das camadas de pooling?

A figura 36 apresenta o resultado da pergunta número oito, referente a camada de saída. A pergunta faz uma afirmação errada, a resposta é falso (vermelho). A resposta desta pergunta é tratada apenas na primeira parte da aula, onde o percentual de acerto foi de 50%, com 34,6% de incerteza e 15,4% de resposta errada (azul). Após a segunda parte da aula, o percentual de acerto reduziu para 46,2%, a incerteza reduziu para 23,1%, entretanto, o percentual de erro aumentou para 30,8%. Importante destacar que as técnicas XAI utilizadas não contribuem com a resposta desta pergunta.

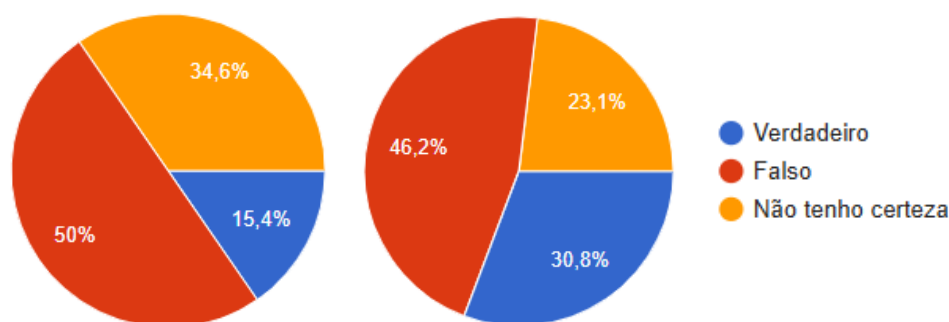


Figura 36: 8) Os dados passam pela camada de saída que é uma rede neural tradicional de múltiplas camadas, mas a probabilidade de pertencimento a cada classe ainda não é calculada nesta camada.

A figura 37 apresenta o resultado da penúltima pergunta do pós-questionário, referente a percepção do aluno em relação ao uso da XAI no contexto de ensino e aprendizagem. 73% concordam (soma das respostas 4 e 5) que os recursos visuais utilizados contribuíram com a aprendizagem, 23% mantiveram-se neutro (resposta 3) e apenas um aluno acredita que a contribuição foi pequena (resposta 2). Nenhum aluno respondeu que não houve contribuição.

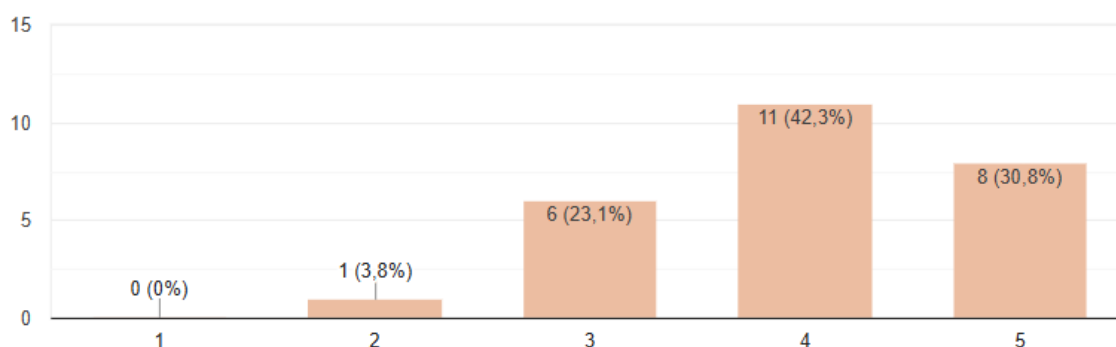


Figura 37: Nesta parte da aula foi utilizado técnicas de Inteligência Artificial Explicável (mapa de características e mapa de ativação de classe). Ver o funcionamento interno da rede através de recursos visuais contribuiu com sua aprendizagem?

Referente à última pergunta (optativa) do pós-questionário para responder de forma discursiva, seguem os principais pontos positivos e negativos encontrados:

a) Recursos visuais eficazes: A maior parte dos participantes elogiou a qualidade das representações visuais, especialmente os mapas de características e a demonstração do funcionamento da rede. Destaque para os comentários “Visualizar as features extraídas pelas camadas me ajudou a entender melhor.” e “Bem explicado e as partes visuais mostradas foram bem interessantes!”.

b) Compreensão aprimorada: Alguns comentários mostram que, mesmo com limitações, os alunos saíram com melhor entendimento dos conceitos da CNN. Destaque para o comentário “Gostei bastante da aula, deu pra entender melhor os conceitos de redes convolucionais.”.

c) Potencial didático da metodologia: Os participantes reconheceram que a técnica de ensino tem alto potencial pedagógico, especialmente para tornar visível o que normalmente é abstrato. Destaque para os comentários “Gostei da dinâmica de demonstração visual dos conceitos.” e “A técnica pode funcionar bem em uma sequência de aulas.”.

d) Falta de base mais sólida: Vários participantes sentiram falta de explicações mais detalhadas sobre os fundamentos das redes neurais simples antes de chegar nas convolucionais. Destaque para os comentários “Talvez explicar mais a base das redes neurais mais básicas antes das convolucionais.” e “Gostaria de uma explicação mais a fundo da função de ativação.”.

e) Mais interatividade: Muitos comentários indicam que os alunos queriam participar mais ativamente, eles próprios interagindo com as ferramentas. Destaque para os comentários “Talvez

deixar mais interativo.” e “Acredito que seria mais interessante os próprios alunos interagirem com o site.”.

f) Tempo e carga de conteúdo: Alguns participantes sentiram que o conteúdo era denso demais para uma única aula, o que dificultou a compreensão. Destaque para o comentário “Achei difícil relacionar essas imagens e a geração delas às redes em uma única aula experimental.”.

4.3 Comparações entre os Resultados

Em comparação com os estudantes da Fatec, a tabela 1 mostra que os estudantes da Unesp possuem maior conhecimento em CNN. Importante destacar que em ambas as instituições, a disciplina de IA é oferecida no final do curso, isto é, muitos já tiveram alguma vivência com a área, seja em disciplinas anteriores, eventos, iniciação científica ou trabalho de conclusão de curso.

Tabela 1: 1) Conhecimento em CNN.

Instituição	Nenhum	Pouco	Razoável	Muito
Fatec	45,9	43,2	10,8	0,0
Unesp	23,1	50,0	26,9	0,0

A tabela 2 mostra que apesar dos percentuais de acertos serem maiores tanto no pré quanto no pós na Unesp, o ganho de aprendizagem foi mais significativo na Fatec. Apenas na pergunta oito houve uma diminuição nos acertos após a utilização da XAI, nas demais perguntas houve aumentos em ambas as instituições.

Tabela 2: Percentual de acerto.

Nº Pergunta) Conceito	Fatec			Unesp		
	Pré	Pós	Diferença	Pré	Pós	Diferença
1) Arquitetura	83,8	89,5	5,7	88,5	92,3	3,8
2) Entrada	62,2	81,6	19,4	84,6	88,5	3,9
3) Convolução – filtros	70,3	86,8	16,5	96,2	100,0	3,8
4) Convolução – evolução	16,2	28,9	12,7	53,8	61,5	7,7
5) Convolução – ativação	56,8	63,2	6,4	76,9	76,9	0,0
6) Convolução – relação	51,4	76,3	24,9	84,6	100,0	15,4
7) Pooling	59,5	78,9	19,4	73,1	80,8	7,7
8) Saída	56,8	52,6	-4,2	50,0	46,2	-3,8

Observando a média na relação candidato/vaga (tabela 3) do vestibular de ambas as instituições nos últimos três anos, é possível concluir que o processo seletivo é mais concorrido na Unesp, ou seja, os estudantes da Unesp passam por uma seleção mais rigorosa. Também é importante destacar as diferenças entre os cursos, o curso da Fatec¹ tem o foco mais prático e aplicado com duração de 3 anos, o curso da Unesp² tem o foco mais teórico e científico com duração de 4 anos. Estas variáveis devem ser levadas em consideração para reflexões a respeito dos resultados obtidos nesta pesquisa, tendo em vista que o percentual de acertos da Unesp foi maior que o da Fatec.

¹ <https://fatecpp.cps.sp.gov.br/cursos-fatec/analise-e-desenvolvimento-de-sistemas/>

² <https://www.fct.unesp.br/#!/graduacao/ciencia-da-computacao/>

Tabela 3: Concorrência no vestibular.

Instituição	Candidato/Vaga
Fatec	7
Unesp	15

Fonte: Fat³ e Vunesp⁴.

O resultado apresentado na tabela 4, referente a penúltima pergunta do pós-questionário, reforça o resultado observado na tabela 1, onde mostra que os participantes da Unesp possuem mais conhecimento em CNN do que os participantes da Fatec. Desta forma, a percepção do aluno em relação a contribuição da XAI na aprendizagem dele, foi mais relevante na Fatec do que na Unesp.

Tabela 4: Contribuição com a aprendizagem.

Instituição	Pouco (1 + 2)	Razoável (3)	Muito (4 + 5)
Fatec	0,0	7,9	92,1
Unesp	3,8	23,1	73,1

Comparando as respostas discursivas da última pergunta, é possível observar que em ambas as turmas, a recepção foi majoritariamente positiva. Os participantes reconheceram o valor didático da proposta, destacando a clareza das explicações, a qualidade das demonstrações e a relevância do tema.

A visualização foi unanimemente elogiada nas duas turmas. Os participantes destacaram o impacto da interpretação visual das camadas na compreensão do conteúdo. Essa estratégia foi vista como a principal responsável por tornar o tema palpável e interessante. O aspecto visual e demonstrativo é o ponto de convergência entre ambas as turmas, sendo considerado o componente mais eficaz da metodologia.

Ambas as turmas concentraram as críticas no tempo limitado. A turma da Unesp enfatizou o desejo por maior interatividade, sugerindo que os alunos tenham contato direto com a ferramenta. Ambas as turmas expressaram a necessidade de explicações adicionais sobre fundamentos – como funções de ativação e redes neurais básicas.

5 Conclusões e Trabalhos Futuros

A XAI é uma recente linha de pesquisa que tem sido aplicada em áreas como saúde, finanças, indústria, segurança, ambiental, logística e comércio eletrônico, entretanto, áreas como direito, assistência social e educação carecem de aplicações (Kalasampath, et al., 2025). A motivação deste trabalho foi avaliar se as técnicas de explicação de modelos de IA, principalmente para modelos caixa-preta, podem colaborar com o processo de ensino e aprendizagem de Rede Neural Convolucional. Um estudo de caso comparado foi realizado com estudantes de duas instituições de ensino superior da cidade de Presidente Prudente do estado de São Paulo. As técnicas de explicação foram aplicadas durante uma aula expositiva sobre os conceitos de CNN.

A aula que compôs o experimento, foi dividida em duas partes de aproximadamente 50 minutos cada. Durante a primeira parte os alunos tiveram um ensino de nivelamento através de projeções textuais e de imagens como de costume educacional, em seguida responderam um questionário. Na segunda parte da aula os alunos receberam um complemento visual resultado das

³ <https://vestibular.fatec.sp.gov.br>

⁴ <https://vestibular.unesp.br/>

técnicas de explicação da XAI e por fim responderam outro questionário. Em ambas as partes a comunicação verbal também foi utilizada.

Os resultados encontrados são relevantes pois demonstram que as técnicas utilizadas, mapas de características e mapas de ativação de classe, foram importantes para o sucesso do processo de ensino e aprendizagem, os números apresentam ganhos educacionais. É importante destacar que a avaliação foi de ganho de aprendizagem imediato e não duradouro. Um dos principais achados está relacionado com a efetividade da metodologia de acordo com o perfil do estudante, ou seja, houve maior contribuição com os que possuíam menor conhecimento sobre o assunto. De forma descritiva, o uso de recursos visuais da XAI foi reconhecido pelos participantes da pesquisa como potencial colaborador com o sucesso do processo de ensino e aprendizagem.

Esta metodologia experimental possui a dependência da internet, pois as ferramentas utilizadas são online. Trabalhos futuros podem providenciar métodos offline, permitindo que os recursos sejam aplicados mesmo que o site ou a internet estejam fora do ar. Tendo em vista a densidade dos assuntos, a questão do tempo de aula (duração) também é um fator limitador. Ainda como trabalho futuro, sugere-se avaliar qual a contribuição da explicabilidade de IA no ensino de modelos como o Transformer para processamento de texto em linguagem natural e modelos de árvores de decisão com dados tabulares, seja modelos que fazem uso de árvores em paralelo ou sequencial, pois também são considerados caixa-preta. Sugere-se ainda realizar uma comparação entre metodologias de ensino com grupo controle (sem XAI) e grupo tratamento (com XAI).

Agradecimentos

Agradeço à Deus pelo sustento e à minha família (minha esposa Patrícia e meus filhos Enzo, Sofia e Lucas) por todo apoio, agradeço ao meu orientador (Danilo Eler) pelo direcionamento de todo processo, agradeço à Fatec de Prudente pelo afastamento de metade de minhas aulas para dedicação à pesquisa, agradeço aos responsáveis pelo laboratório LaPESA da Unesp de Prudente onde realizo os estudos, agradeço aos diretores e coordenadores que disponibilizaram as instituições de ensino, agradeço aos professores (Adriane Cavichioli e Danillo Pereira) das disciplinas de IA que permitiram a coleta dos dados, agradeço aos estudantes participantes da pesquisa, e por fim agradeço ao comitê de ética da Unesp de Prudente pela avaliação e aprovação desta presente pesquisa.

Referências

- Azam, S., Montaha, S., Fahim, K. U., Rafid, A. K., Mukta, M. S., & Jonkman, M. (2023). Using Feature Maps to Unpack the CNN ‘Black Box’ Theory with Two Medical Datasets of Different Modality. *Intelligent Systems with Applications*, 18, p. 200233. <https://doi.org/10.1016/j.iswa.2023.200233> [GS Search]
- Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation Atlas. *Distill*. [link] [GS Search]
- Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How People Use ChatGPT (Working Paper No. 34255). National Bureau of Economic Research. <https://doi.org/10.3386/w34255> [GS Search]
- Chiaburu, T., Haußer, F., & Bießmann, F. (2024). Confident Teacher, Confident Student? A Novel User Study Design for Investigating the Didactic Potential of Explanations and their Impact on Uncertainty. *arXiv*. [link] [GS Search]

- Dis, E. A., Bollen, J., Zuidema, W., Rooij, R. v., & Bockting, C. L. (2023). ChatGPT: Five Priorities for Research. *Nature*, 614(7947), pp. 224-226. <https://doi.org/10.1038/d41586-023-00288-7> [GS Search]
- Gildenblat, J. (2025). Advanced AI Explainability for PyTorch. Acesso em 26 de 05 de 2026. [\[link\]](#) [GS Search]
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative Adversarial Nets. 28th International Conference on Neural Information Processing Systems, 2, pp. 2672–2680. Cambridge. [GS Search]
- Hao, H., Rui, W., Hao, L., & Huai, Y. (2024). UnionCAM: Enhancing CNN Interpretability Through Denoising, Weighted Fusion, and Selective High-Quality Class Activation Mapping. *Frontiers in Neurorobotics*, 18. <https://doi.org/10.3389/fnbot.2024.1338149> [GS Search]
- Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Maruthamuthu, A. (2025). A Literature Review on Applications of Explainable Artificial Intelligence (XAI). *IEEE Access*, 13, pp. 41111-41140. <https://doi.org/10.1109/ACCESS.2025.3546681> [GS Search]
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., . . . Gašević, D. (2022). Explainable Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 3, p. 100074. <https://doi.org/10.1016/j.caeai.2022.100074> [GS Search]
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. [\[link\]](#) [GS Search]
- Martins, R. M., von Wangenheim, C. G., Rauber, M. F., & Hauck, J. C. R. (2023). Eficácia das Metodologias Ativas e do Suporte Tecnológico no Ensino de Machine Learning na Educação Básica. *Anais do III Simpósio Brasileiro de Educação em Computação (EDUCOMP)* (pp. 154–162). Sociedade Brasileira de Computação. <https://doi.org/10.5753/educomp.2023.228162> [GS Search]
- Naveed, S., Stevens, G., & Robin-Kern, D. (2024). An Overview of the Empirical Evaluation of Explainable AI (XAI): A Comprehensive Guideline for User-Centered Evaluation in XAI. *Applied Sciences*, 14(23), p. 11288. <https://doi.org/10.3390/app142311288> [GS Search]
- Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The Building Blocks of Interpretability. *Distill*. [\[link\]](#) [GS Search]
- Research, G. V. (2025). AI In Education Market Size, Share & Trends Analysis Report. San Francisco: Grand View Research. [\[link\]](#)
- Saarela, M., & Podgorelec, V. (2024). Recent Applications of Explainable AI (XAI): A Systematic Literature Review. *Applied Sciences*, 14(19). <https://doi.org/10.3390/app14198884> [GS Search]
- Sanusi, I. T., & Oyelere, S. S. (2020). Pedagogies of Machine Learning in K-12 Context. *Proceedings of the IEEE Frontiers in Education Conference (FIE)*, (pp. 1-8). <https://doi.org/10.1109/FIE44824.2020.9274129> [GS Search]
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual Explanations from Deep Networks via Gradient-Based Localization. *IEEE International Conference on Computer Vision*, (pp. 618-626). [\[link\]](#) [GS Search]
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), pp. 433-460. <https://doi.org/10.1093/mind/LIX.236.433> [GS Search]

- Türkmen, G. (2024). The Review of Studies on Explainable Artificial Intelligence in Educational Research. *Journal of Educational Computing Research*, 63(2), pp. 277-310. <https://doi.org/10.1177/07356331241310915> [GS Search]
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is All you Need. 31st International Conference on Neural Information Processing Systems, (pp. 6000–6010). Red Hook. [\[link\]](#) [GS Search]