SEVEN
publicações acadêmicas

COMBINAÇÃO DE DEEP LEARNING E MACHINE LEARNING PARA DIAGNÓSTICO AUXILIADO POR INTELIGÊNCIA ARTIFICIAL

https://doi.org/10.56238/sevened2025.008-005

Fábio Lofredo Cesar
Licenciatura em Física
Universidade Estadual de Campinas (UNICAMP)
E-mail:fabiolofredo@gmail.com

Hygor Santiago Lara
Engenheiro Mecânico, Doutor em Mecânica dos Sólidos e Projetos Mecânicos
Universidade Estatual de Campinas (UNICAMP)
ORCID: https://orcid.org/0000-0002-4835-5498
CV Lattes: http://lattes.cnpq.br/8858059011756333
E-mail:hsantiagolara@gmail.com

RESUMO
Este trabalho propõe um sistema híbrido que combina a ResNet (Residual Network - amplamente reconhecida por seu impacto no aprendizado profundo, sendo um marco na área de visão computacional) com Árvores Extremamente Aleatórias (Extra Trees) para classificar imagens de raio-X de tórax e auxiliar na detecção de doenças. A abordagem utiliza a técnica de Transfer Learning, onde a ResNet, previamente treinada, é empregada para extrair características relevantes das imagens. Em seguida, o algoritmo Extra Trees realiza a classificação com base nessas características. Na etapa inicial, utilizando apenas a ResNet combinada com uma pequena rede neural, obtivemos uma acurácia de 95,40% na validação e 79,33% nos testes. Com a implementação do sistema híbrido, os resultados foram significativamente aprimorados, alcançando 96,90% de acurácia na validação e 89,98% nos testes, representando uma melhoria expressiva de aproximadamente 10 pontos percentuais nos testes. Esses resultados destacam o potencial do sistema híbrido em aplicações, demonstrando como a combinação de técnicas avançadas de aprendizado profundo e aprendizado de máquina pode contribuir significativamente para a melhoria da precisão.

Palavras-chave: Raio-x. CNN. Transfer learning. Resnet.

1 INTRODUÇÃO

Visão computacional é um campo que envolve várias áreas projetando alguma interpretação da visão humana computacionalmente. O campo busca resolver desafios de áreas como reconhecimento e detecção de objetos ou faces. Detecção de movimento, análise tridimensional a partir de imagens bidimensionais, reconstrução de cenas e restauração de imagens. E isso traz o interesse de pesquisadores e empresas (Kovaleski, 2018).

Redes neurais convolucionais utilizam operações de convolução em camadas. Os filtros, nessas redes, contribuem para extrair características da imagem, o que torna úteis para o reconhecimento de objetos e faces (Kovaleski, 2018).

Uma das motivações deste trabalho é utilizar imagens de raio-x para a detecção de doenças, algo que pode ser muito útil na área médica, agindo como uma ferramenta de diagnóstico para os médicos.

1.1 HISTÓRIA

No contexto histórico das Redes neurais convolucionais (CNN - Convolutional Neural Network) e Redes totalmente convolucionais (FCN - Fully Convolutional Networks) pode ser divididas em:

Entre 1989 e 1999, ocorreu a origem das Redes Neurais Convolucionais (CNNs). Durante esse período, as redes eram capazes de aprender automaticamente os padrões dos filtros e reconhecer variações rotacionais nas imagens, marcando o início de uma nova abordagem no processamento de imagens. No início dos anos 2000, no entanto, houve uma estagnação no desenvolvimento das CNNs, com poucos avanços relevantes na área. Entre 2006 e 2011, as CNNs passaram por um renascimento graças à introdução de técnicas como greedy layer-wise unsupervised training e o max pooling, além de melhorias no processamento proporcionadas pela evolução dos hardwares disponíveis (Cunha, 2020).

O período entre 2012 e 2014 foi marcado pela ascensão das CNNs, com avanços significativos em seu desempenho, impulsionando sua popularidade e aplicabilidade em diversas áreas. Em 2014, ocorreu a descoberta das Redes Neurais Convolucionais Totalmente Conectadas (FCN). Essa inovação consistiu em substituir a camada totalmente conectada por uma nova camada de convolução, permitindo que as CNNs realizassem a segmentação semântica de imagens de maneira mais eficiente e precisa (Cunha, 2020).

1.2 OBJETIVOS

Existe uma busca por arquiteturas mais eficientes, e esse artigo pretende ajudar a mostrar as eficiências de duas técnicas para predição de imagens.

Objetivo deste artigo é criar um sistema híbrido de CNN(Resnet) com árvores extremamente aleatórias, usando transfer learning e mensurar sua acurácia.

1.3 METODOLOGIA: REDAÇÃO E CÓDIGO

Certas partes deste artigo e no código foram escritas com o auxílio da inteligência artificial ChatGPT, com o objetivo de melhorar a clareza textual. Nesse processo, incorporamos tanto trechos de autoria própria quanto fragmentos de artigos originais, sempre garantindo a devida citação dos autores originais nas passagens utilizadas. Posteriormente, submetemos essas seções a revisões para garantir a precisão e a integridade do conteúdo.

2 DISCUSSÃO E ANÁLISE BIBLIOGRÁFICA

Segundo a análise de (Rodrigues, 2018), o qual é empregado um rede neural convolucional para mostrar a viabilidade da técnica para leitura de caracteres das placas de licenciamento veiculares, é obtido resultados de inferência na ordem de 89,24%. Estão entre as possibilidades de aplicações, controlar o tráfego rodoviário, identificar carros em estacionamentos, ou verificar infratores das leis de trânsito.

De acordo com (Cruz, 2019), consoante com o estudo de reconhecimento de emoções por expressão facial utilizando redes neurais convolucionais, são diversas as suas aplicações possíveis como interação humano-computador, psiquiatria e cuidados médicos, deficiência visual, interação humano-robô e personagens virtuais e animação.

Segundo (Alvear-Sandoval et al., 2019) em que é aplicado técnicas de melhorias na aplicação do CNN e nos Stacked Denoising Auto-Encoder classifiers . E em seguida é aplicado Stacked Denoising Auto-Encoder classifiers na saída de uma CNN obtendo melhores resultados. Conclui-se que combinar as técnicas de diferentes naturezas pode adquirir melhores resultados.

Em Ahlawat & Choudhary, (2020) e em Niu & Suen (2012) realizam uma abordagem na qual se utiliza um sistema híbrido de CNN com Support Vector Machine (SVM), o CNN funciona extraindo características e o SVM como um classificador binário, obtendo-se assim uma acurácia de 99,28% no banco de dígitos do MNIST no Ahlawat & Choudhary (2020) e 99,81% sem rejeição e 94,40% com rejeição de 5,60% no Niu & Suen (2012).

3 TEORIA

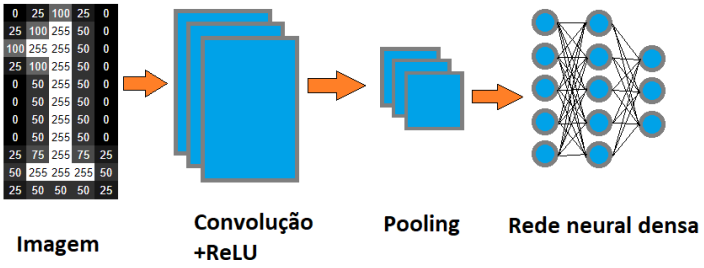
3.1 PROCESSOS DA CNN

Uma Convolutional Neural Network (CNN) é composta por três elementos fundamentais: a camada de convolução, a camada de pooling e a rede neural densa. A camada de convolução extrai características da entrada usando filtros de tamanho reduzido, que convoluem os dados em largura,

altura e profundidade. Durante o treinamento, os filtros se ajustam para identificar características comuns nos dados, como arestas e cores, evoluindo para estruturas mais complexas. A camada de pooling reduz o tamanho dos dados após a convolução, permitindo que a rede aprenda diferentes representações dos dados e evite o overfitting. A rede neural densa, normalmente no final da arquitetura, utiliza as características extraídas para classificar a saída. O equilíbrio entre recursos e desempenho é essencial na definição da arquitetura (Rodrigues, 2018).

A Figura 1 ilustra um processo completo de uma Rede Neural Convolucional (CNN). Os pixels, representados por valores quantitativos, passam inicialmente por camadas de convolução ativadas pela função ReLU. Em seguida, é aplicado um processo de pooling para redução dimensional e extração de características mais relevantes. Por fim, as informações processadas são encaminhadas para uma rede neural densa, que realiza a etapa final de classificação ou regressão.

Figura 1. Exemplo de um processo de CNN

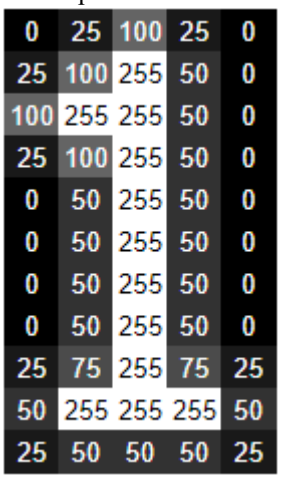


Fonte: Autoria própria.

3.1 ENTRADA DA IMAGEM

A capacidade dos seres humanos de interpretar imagens é uma habilidade intrínseca, mas os computadores dependem de uma representação numérica para processar imagens. Nas máquinas, as imagens são traduzidas em matrizes de pixels, onde cada pixel é representado por um número que varia de 0 a 255, refletindo a intensidade de cor. Essas matrizes organizam os pixels, permitindo que os computadores processem e analisem imagens de forma algorítmica. Esse processo de conversão é fundamental para que as máquinas possam compreender e tomar decisões com base em informações visuais, desempenhando um papel crucial em campos como visão computacional e aprendizado de máquina. Imagens monocromáticas possuem 1 canal, já coloridas com RGB possuem 3 canais (Cunha, 2020).

Figura 2. Imagem monocromática representando a variação de intensidade de 0 a 255.



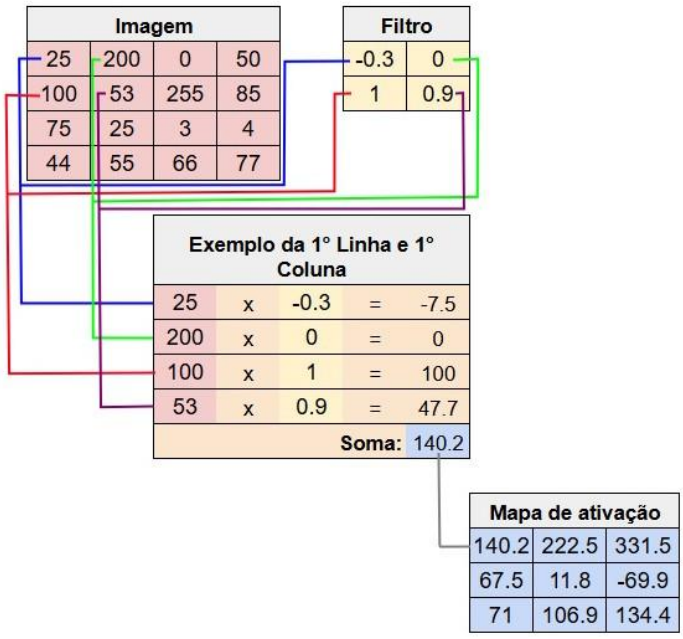
0	25	100	25	0
25	100	255	50	0
100	255	255	50	0
25	100	255	50	0
0	50	255	50	0
0	50	255	50	0
0	50	255	50	0
0	50	255	50	0
25	75	255	75	25
50	255	255	255	50
25	50	50	50	25

Fonte: Autoria própria.

3.2 CONVOLUÇÃO

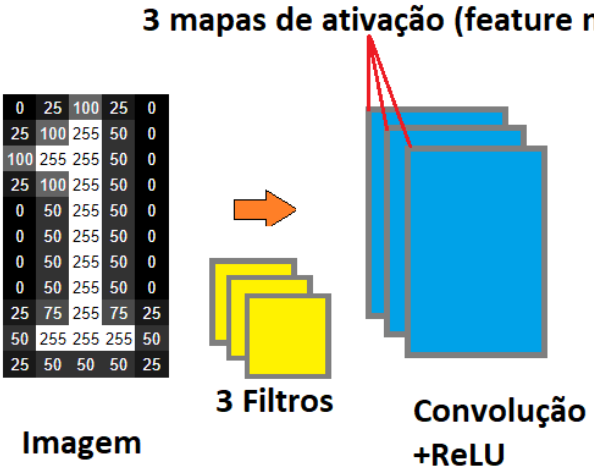
A camada de convolução é o componente central de uma CNN, onde a convolução é uma operação matemática que envolve o deslizamento de uma função sobre outra para processar sub-regiões específicas da imagem. Isso difere da abordagem tradicional, onde os neurônios estão conectados a todos os dados de entrada. A vantagem da CNN reside na eficiência do processamento de características locais e na redução de parâmetros, tornando-a ideal para tarefas de visão computacional, como reconhecimento de padrões e classificação de imagens (Cruz, 2019), ou seja, o é feito um deslizamento de um filtro na imagem produzindo matematicamente um ou mais mapas de ativação (feature map)(Cunha, 2020).

Figura 3. Uma imagem sendo convolucionada por um filtro gerando um mapa de ativação. Mostrando também exemplos dos cálculos na geração da matrix.



Fonte: Autoria própria.

Figura 4. Exemplo da geração de 3 mapas de ativação usando 3 filtros.



Fonte: Autoria própria.

3.3 FUNÇÃO DE ATIVAÇÃO

Após as camadas de convolução em uma rede neural convencional, é comum inserir uma camada não-linear, conhecida como camada de ativação. Essa camada é crucial para introduzir a não-linearidade em um sistema que, até aquele ponto, realizou operações predominantemente lineares, como multiplicação e soma (Cunha, 2020).

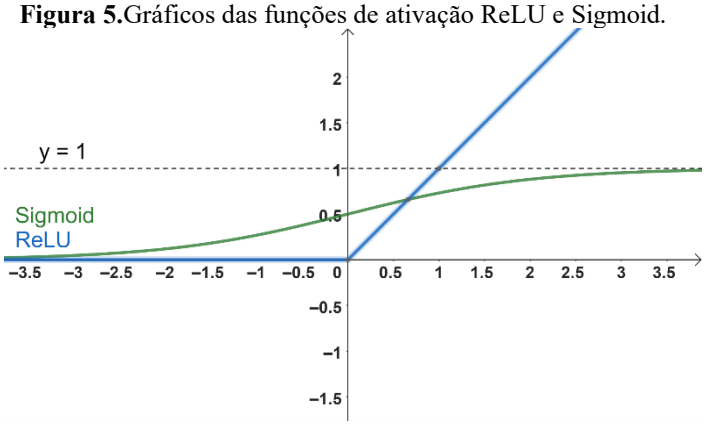
Pesquisas recentes demonstraram que a função ReLU (Rectified Linear Units) oferece benefícios superiores, permitindo um treinamento mais rápido e eficiente da rede sem comprometer a precisão. A camada ReLU aplica a função f(x) = max(0, x) a todos os valores do volume de entrada, substituindo os valores negativos por zero, aumentando a capacidade de modelagem não-linear sem afetar negativamente as ativações da camada de convolução. A função ReLU é (Cunha, 2020):

$$f(x) = max(0, x) \text{ , (1)}$$

A função Sigmoid, semelhante aos neurônios biológicos, restringe seus valores a um intervalo entre 0 (não ativação) e 1 (ativação), refletindo o comportamento de ativação ou inativação dos neurônios diante das entradas recebidas. A função Sigmoid é dada por (Rodrigues, 2018):

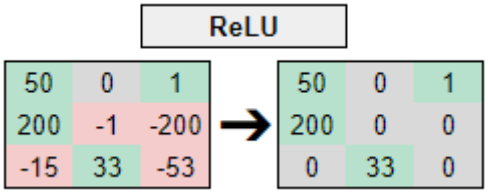
$$\sigma(x) = \frac{1}{1+e^{-x}} \text{ , (2)}$$

Figura 5. Gráficos das funções de ativação ReLU e Sigmoid.



Fonte: Autoria própria.

Figura 6. Exemplo da função ReLU em uma matrix.



Fonte: Autoria própria.

Usualmente usam a função Softmax no final de uma CNN para classificar, é um modelo generalizado de uma regressão logística e pode estimar probabilidades para múltiplas classes (Cruz, 2019). A probabilidade para uma classe i, dado uma entrada x é obtida na fórmula:

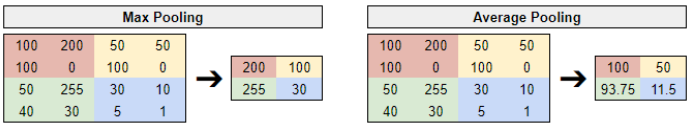
$$p(classe = i|x) = \frac{e^{y_i}}{\sum_{j=1}^{K} e^{y_j}}, (3)$$

Sendo K o número total de classes (Kovaleski, 2018).

3.4 POOLING

A camada de pooling desempenha um papel crucial em Redes Neurais Convolucionais (CNNs), focando na subamostragem para reduzir o tamanho da imagem durante o processamento. Isso ajuda a diminuir a carga computacional, o número de parâmetros e, consequentemente, a prevenir o overfitting e economizar memória. A operação de pooling é inspirada no funcionamento do córtex visual, onde campos de recepção locais representam sub-regiões do campo visual de neurônios específicos. Reduzir o tamanho da imagem torna a CNN mais robusta às variações de posição dos objetos e elimina ruídos menores. Existem dois tipos principais de pooling: o max pooling, que seleciona o valor máximo em um campo de recepção, e o average pooling, que calcula a média dos valores próximos. O max pooling é a técnica mais comumente usada nas CNNs (Cruz, 2019).

Figura 7. Exemplos de Pooling para reduzir o tamanho da imagem. No Max Pooling mantém o maior número e no Average Pooling faz-se a média dos valores.

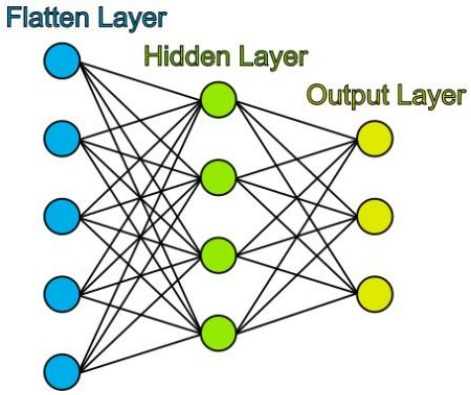


Fonte: Autoria própria.

3.5 REDE NEURAL DENSA

A camada totalmente conectada em uma Rede Neural Convolucional (CNN) desempenha um papel essencial na decisão de classificação da imagem como um todo. Sua função é unir todas as características e atributos extraídos da imagem pelas camadas anteriores, ou seja, as camadas de convolução e pooling. Essa camada interconecta todos os neurônios de maneira tradicional, com a saída da camada anterior sendo achatada em um único vetor antes de entrar na camada totalmente conectada (Cunha, 2020).

Figura 8. Rede neural densa.



Fonte: Autoria própria.

3.6 TREINAMENTO

O treinamento envolveu a divisão da base de dados em batches para processamento e atualização dos pesos da CNN após cada batch. Depois de rodar todos os batches é calculada uma nova época (epoch). O modelo ideal foi definido com base na menor taxa de função de perda (Loss Function ou loss) na base de validação. A taxa de aprendizado (Learning Rate) é ajustada dinamicamente pelo otimizador (optimizer), começando alta e diminuindo ao longo do treinamento, funcionando como um acelerador, obtendo assim uma acurácia(Accuracy) satisfatória (Cruz, 2019).

3.7 TRANSFER LEARNING

Transfer learning se baseia em transferir um aprendizado, já aprendido, para aprender algo em algum domínio diferente, porém semelhante. Em Zhuang et al. (2021) cita os exemplos intuitivos de transfer learning, como quem aprende violino, pode aprender piano mais rapidamente, assim como quem aprende a andar de bicicleta, pode aprender a andar de moto mais rapidamente. Mas é preciso

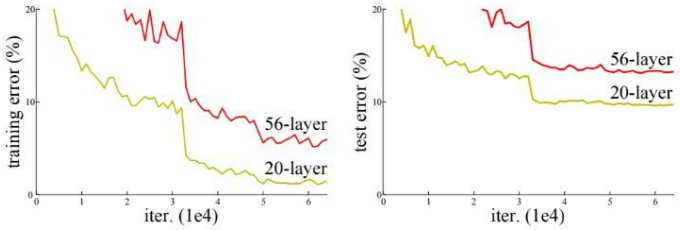
tomar cuidado, pois se os domínios tiverem pouca coisa em comum, o aprendizado não será eficiente, pois quem aprende a andar de bicicleta pode não ajudar a aprender a tocar piano.

Assim, no contexto prático de inteligência artificial e deste trabalho, é aproveitado algo já treinado, para aprender em outro domínio semelhante, levando assim menos tempo de treino e precisando de menos dados. O modelo já treinado utilizado neste trabalho será o Resnet, será adicionada camadas que serão treinadas com as imagens de raio-X, otimizando tempo e requerendo menos imagens.

3.8 RESNET

Resnet é uma rede neural residual treinada, que ganhou em primeiro lugar na competição de ILSVRC 2015. A rede tem uma saída para 1000 classes, tendo sido treinada com 1,28 milhões de imagens (He et al., 2015).

Figura 9. Erro de treinamento na esquerda e erro de teste na direita no CIFAR-10 com 20 e 56 camadas. A rede mais profunda tem maior erro de treinamento e de teste.



Fonte: He et al., 2015.

Segundo o artigo da figura 9, redes neurais com muitas camadas não necessariamente darão um resultado melhor na acurácia.

Figura 10. Um exemplo de um bloco do aprendizado por resíduo no Resnet.

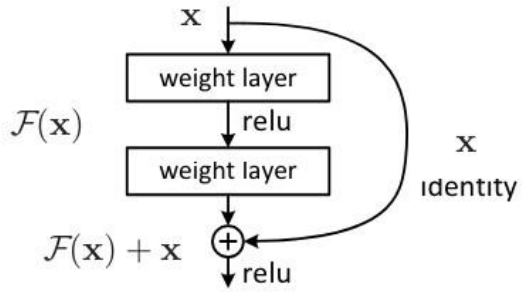


Figure 2. Residual learning: a building block.

Fonte: He et al., 2015.

Na figura 10 acima, o exemplo do funcionamento de uma parte da rede Resnet, na qual apresenta “conexões de atalhos”, elas pulam uma ou mais camadas. as conexões de atalho executam mapeamento de identidade e suas saídas são adicionadas as saídas das camadas empilhadas.

3.9 ÁRVORES EXTREMAMENTE ALEATÓRIAS

Árvores extremamente aleatórias, ou Extra Tree, é um algoritmo de Machine Learning para predição. Com ele, é possível treinar o algoritmo com dados seguindo uma lógica similar às árvores de decisão. Segundo Geurts et al. (2006), a principal diferença em relação a outros algoritmos de árvores, traduzindo para o português, é:

> O algoritmo Extra-Trees constrói um conjunto de árvores de decisão ou de regressão não podadas de acordo com o procedimento top-down clássico. Suas duas principais diferenças com outros métodos de conjunto baseados em árvore são que ele divide os nós escolhendo pontos de corte totalmente aleatórios e que ele usa toda a amostra de aprendizagem (em vez de uma réplica de bootstrap) para fazer crescer as árvores.

4 METODOLOGIA

4.1 PRIMEIRA ETAPA

Primeiramente foi utilizado o banco de dados de imagens de raio-x do tórax do site, nele pegamos as imagens de treino e teste com e sem pneumonia. Adotamos também a arquitetura de resnet_v2. Adicionamos uma camada de 64 neurônios com função de ativação Relu e 20% de dropout, outra camada de 16 neurônios também com Relu e dropout de 20% e no seu final adicionamos 2 neurônios com função de ativação Softmax para a classificação de saudável ou doente. Como na Tabela 1:

Tabela 1. Arquitetura sendo o keras_layer o resnet_v2. Adicionalmente uma camada de 64, outra de 16 e a última de 2 neurônios.

Camada	Nome	Saídas de neurônios
1	Resnet	1001
2	Densa64 + Dropout	64
3	Densa16 + Dropout	16
4	Densa2	2

Fonte: Autoria própria.

A arquitetura foi treinada usando transfer learning, ou seja, congelando o resnet e treinando os neurônios seguintes.

4.2 SEGUNDA ETAPA

Na segunda etapa, foi tirado as camadas de 2 neurônios e o dropout da camada de 16 neurônios, ficando como na Tabela 2. Após isso, a arquitetura de árvores extremamente aleatórias foi treinada no resultado da camada dos 16 neurônios de forma supervisionada. Foi obtido o resultado de saudável ou doente na saída do modelo das árvores extremamente aleatórias.

Tabela 2. Arquitetura sendo o keras layer o resnet v2. Adicionalmente uma camada de 64, outra de 16.

Camada	Nome	Saídas de neurônios
1	Resnet	1001
2	Densa64 + Dropout	64
3	Densa16	16

Fonte: Autoria própria.

5 RESULTADO E DISCUSSÕES

Na primeira etapa, do resnet com saída de 2 neurônios, foi obtida uma acurácia para validação de 95,4% e para teste de 79,33%, mostrando nitidamente o overfitting.

Na segunda etapa, do resnet, com a rede neural mais a árvore extremamente aleatória, foi obtida uma acurácia para validação de 96,90% e para teste de 89,98%, melhorando os resultados do teste.

Isso mostra que houve um aumento de aproximadamente 10 pontos percentuais para o teste, melhorando significativamente a acurácia do modelo.

A eficiência do resultado obtido foi semelhante à da técnica de leitura de caracteres das placas de licenciamento veiculares (Rodrigues, 2018), que alcançou 89,24%. No entanto, é importante ressaltar a diferença entre as realidades desses dois casos: enquanto um se trata de dígitos em placas de veículos, o outro envolve a análises de raios-x de tórax.

Leão et al. (2020), que investigaram a aplicação de redes convolucionais para detecção de Covid-19 em imagens de raio-X, em sua análise de com duas e três classes e com e sem transfer learning, obteve acurácias entre 82,11% e 87,91%. Mostrando que a segunda etapa, feita neste trabalho, obteve resultados superiores.

Em Castro et al. (2023), que exploraram o uso de Curvas Principais como uma técnica para otimizar a triagem de pacientes com tuberculose, usando também raiox-X, obteve acurácias entre 84% e 89%. Mostrando resultados semelhantes obtidos neste trabalho.

6 CONCLUSÃO

O processo de transfer learning no resnet com as árvores extremamente aleatórias teve um desempenho melhor do que somente com transfer learning para as imagens de raio-x do tórax, com aproximadamente 10 pontos percentuais de diferença.

Os resultados finais de acurácia deste trabalho são comparáveis aos da literatura, mas a melhora da acurácia do primeiro ao segundo modelo, mostra uma ferramenta potencial para ser utilizada em conjunto com outros modelos.

Para trabalhos futuros é possível pensar em aplicar este sistema em outros dados e contextos, é possível também tentar novas arquiteturas baseadas no sistema híbrido, com potencialidade de aumentar a acurácia final. Também é possível utilizar “Finning tuning”, que consiste em treinar também a rede Resnet, para tentar obter melhores resultados.

AGRADECIMENTOS

Agradeço ao professor Hygor S. Lara por todo apoio e dedicação ao projeto, sem o qual não seria possível realizar esse trabalho. Também agradeço ao pessoal do GEIA (Grupo de Estudos em Inteligência Artificial).

REFERÊNCIAS

AHLAWAT, S.; CHOUDHARY, A. Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. Procedia Computer Science, v. 167, p. 2554-2560, 2020.

ALVEAR-SANDOVAL, R. F.; SANCHO-GÓMEZ, J. L.; FIGUEIRAS-VIDAL, A. R. On improving CNNs performance: The case of MNIST. Information Fusion, v. 52, p. 106-109, 2019.

CASTRO, D. H. H. de et al. Utilização de Curvas Principais na triagem de pacientes com tuberculose. In: XVI Brazilian Conference on Computational Intelligence (CBIC 2023), Salvador, BA, 8-11 de outubro, 2023.

CRUZ, A. A. Uma abordagem para reconhecimento de emoção por expressão facial baseada em redes neurais de convolução. 2019. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal do Amazonas, Manaus, 2019.

CUNHA, L. C. Redes Neurais Convolucionais e Segmentação de Imagens - Uma Revisão Bibliográfica. 2020. Trabalho de Conclusão de Curso (Graduação em Engenharia de Controle e Automação) — Universidade Federal de Ouro Preto, Ouro Preto, 2020.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. Machine Learning, v. 63, p. 3-42, 2006. DOI: 10.1007/s10994-006-6226-1.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385, 2015.

KOVALESKI, P. A. Implementação de Redes Neurais Profundas para Reconhecimento de Ações em Vídeo. 2018. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

LEÃO, P. P. de S. et al. Detecção de Covid-19 em imagens de raio-X utilizando redes convolucionais. J. Health Inform., Número Especial SBIS, p. 393-398, dez. 2020.

NIU, X.-X.; SUEN, C. Y. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recognition, v. 45, n. 4, p. 1318-1325, 2012.

RODRIGUES, D. A. Deep Learning e Redes Neurais Convolucionais: Reconhecimento Automático de Caracteres em Placas de Licenciamento Automotivo. 2018. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) — Universidade Federal da Paraíba, João Pessoa, 2018.

ZHUANG, F. et al. A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, v. 109, n. 1, p. 43-76, 2021. DOI: 10.1109/JPROC.2020.3004555.