

Comparação de Redes Neurais Convolucionais Reais e Quaterniônicas para Tarefas de Classificação de Imagens

Sara A. S. Dias, Vinicius R. Bardelin

¹Instituto de Matemática, Estatística e Computação Científica –
Universidade Estadual de Campinas (UNICAMP)

Resumo. *No campo das redes neurais artificiais, o uso de redes neurais cada vez mais profundas tem sido amplamente explorado. No entanto, a maioria das arquiteturas estudadas é construída utilizando valores reais. Neste artigo, exploramos e comparamos redes convolucionais reais e redes convolucionais quaterniônicas em tarefas de classificação de imagens, utilizando os datasets PathMNIST, CIFAR10 e CIFAR100. O objetivo deste trabalho foi entender como ambas as abordagens se comportam à medida que os problemas de classificação se tornam mais complexos. Os resultados mostraram que as redes de valor quaterniônico superaram as redes de valor real, utilizando cerca de 10% do número de parâmetros.*

1. Introdução

Na crescente área do aprendizado de máquina, mais conhecido como machine learning, as redes neurais têm se destacado como ferramentas poderosas para resolver problemas complexos em diversas disciplinas, incluindo análise de dados, sistemas de recomendação e processamento de imagens, como abordado neste estudo. Inspiradas pelo funcionamento do cérebro humano, as redes neurais são projetadas para reconhecer padrões e aprimorar suas capacidades com a experiência. Em particular, as redes neurais convolucionais se sobressaem no reconhecimento de imagens devido à sua capacidade de processar dados estruturados em forma de grade, como imagens. Utilizando camadas convolucionais, essas redes neurais aprendem a extrair características espaciais relevantes diretamente dos dados de entrada, o que resulta em um desempenho superior em tarefas de visão computacional, onde a precisão na identificação e classificação de objetos é crucial.

A maioria das arquiteturas de redes neurais é baseada em redes neurais de valores reais, como destacado por [Granero et al. 2021]. No entanto, conforme discutido por [Parcollet et al. 2019], os dados do mundo real frequentemente possuem uma natureza multidimensional que exige uma representação adequada capaz de capturar as relações intrínsecas entre os dados. Por exemplo, no contexto do processamento de imagens, cada pixel é composto por três canais de cores: vermelho, verde e azul (os canais RGB). O problema das redes neurais convolucionais de valor real é que elas tratam esses canais de cores como entidades separadas, quando na verdade estão intimamente relacionadas. Esta abordagem pode resultar na perda de informações cruciais sobre a cor, comprometendo a representação eficaz das imagens. Por outro lado, uma abordagem que considere os pixels como entidades únicas emerge como uma proposta mais robusta para o processamento de imagens, especialmente em tarefas que exigem uma compreensão holística das relações multidimensionais.

Nesse sentido, o objetivo deste artigo é comparar o desempenho das redes neurais convolucionais quaterniônicas com o das redes convolucionais tradicionais em tarefas de

classificação de imagens. Para isso, construímos redes convolucionais com arquiteturas similares às das redes LeNet-5 [LeCun et al. 1998] e VGG [Simonyan e Zisserman 2015], aplicando-as a diferentes conjuntos de dados: PathMNIST [Kather et al. 2019], CIFAR10 e CIFAR100 [Krizhevsky 2009]. Nosso foco principal é analisar como essas abordagens se comportam à medida que as tarefas de classificação se tornam progressivamente mais complexas, com imagens menos uniformes entre si e um maior número de classes.

O artigo está organizado de modo que, na Seção 2, apresentamos brevemente a álgebra dos quatérnios, introduzimos a operação de convolução de valor quaterniônico e discutimos sua relevância nas redes neurais convolucionais. Na Seção 3, são apresentados os resultados dos experimentos comparando redes convolucionais tradicionais e redes quaterniônicas. Finalmente, na Seção 4, discutimos as conclusões deste trabalho.

2. Métodos

2.1. Álgebra dos Quatérnios

Os quatérnios, propostos por Hamilton, são números hipercomplexos compostos por uma parte real e três partes imaginárias. O conjunto de todos os quatérnios é definido como:

$$\mathbb{H} = \{q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} : q_0, q_1, q_2, q_3 \in \mathbb{R}\}, \quad (1)$$

onde q_0 denota a parte real de q , e q_1, q_2 e q_3 denotam as partes imaginárias do quatérnio. As chamadas unidades hiperimaginárias \mathbf{i} , \mathbf{j} e \mathbf{k} satisfazem a seguinte relação fundamental:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{i}\mathbf{j}\mathbf{k} = -1. \quad (2)$$

Ademais, as operações básicas com quatérnios incluem adição, multiplicação por um escalar, multiplicação de quatérnios e conjugação. A seguir, definimos essas operações:

- Adição: Dados dois quatérnios $p = p_0 + p_1\mathbf{i} + p_2\mathbf{j} + p_3\mathbf{k}$ e $q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$, a adição é realizada componente a componente:

$$p + q = (p_0 + q_0) + (p_1 + q_1)\mathbf{i} + (p_2 + q_2)\mathbf{j} + (p_3 + q_3)\mathbf{k}. \quad (3)$$

- Multiplicação por um escalar: Dado um escalar α e um quatérnio p , a multiplicação é:

$$\alpha \cdot p = \alpha p_0 + \alpha p_1\mathbf{i} + \alpha p_2\mathbf{j} + \alpha p_3\mathbf{k}. \quad (4)$$

- Multiplicação de quatérnios (produto de Hamilton): A multiplicação de dois quatérnios p e q é dada por:

$$p \cdot q = (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)\mathbf{i} \quad (5)$$

$$+ (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)\mathbf{j} + (p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0)\mathbf{k}. \quad (6)$$

- Conjugação: O conjugado de um quatérnio $p = p_0 + p_1\mathbf{i} + p_2\mathbf{j} + p_3\mathbf{k}$ é:

$$p^* = p_0 - p_1\mathbf{i} - p_2\mathbf{j} - p_3\mathbf{k}. \quad (7)$$

Portanto, o modo como o produto de Hamilton está definido faz da álgebra dos quatérnios uma álgebra associativa e não comutativa. Observe que cada coeficiente do quatérnio pq resultante em (5) e (6) é composto das partes reais e imaginárias de p e q .

Uma definição importante é a de quatérnio puro. Um quatérnio $q \in \mathbb{H}$ é chamado quatérnio puro se $\text{Re}\{q\} = 0$, ou seja, se ele pode ser escrito como $q = q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$.

Essas operações definidas acima e a relação fundamental em (2) consitituem um ferramental eficiente para descrever rotações em um espaço tridimensional. Seja \mathbf{w}_θ um versor unitário na sua forma polar:

$$\mathbf{w}_\theta = \cos(\theta) + \sin(\theta)(w_I\mathbf{i} + w_J\mathbf{j} + w_K\mathbf{k}), \quad (8)$$

e q um quatérnio qualquer. Então, o produto de \mathbf{w}_θ por q , $p = \mathbf{w}_\theta q$, aplica uma rotação, com ângulo θ , ao longo de um eixo $w_I\mathbf{i} + w_J\mathbf{j} + w_K\mathbf{k}$. As rotações com os quatérnios serão amplamente utilizadas para performar transformações de escala nas redes convolucionais quaterniônicas. [Altamirano e Gershenson 2023, Parcollet 2018].

2.2. Camada Convolucional Quaterniônica

As camadas convolucionais são as principais estruturas das redes convolucionais, desempenhando um papel fundamental na extração de informações das imagens. A seguir, apresentamos a convolução de valores reais e quaterniônicos, baseando-nos nos estudos de [Senna e Valle 2021, Gaudet e Maida 2018].

Seja \mathbf{I} uma imagem de valor real e \mathbf{F} um filtro de tamanho $(2N + 1) \times (2N + 1)$. A convolução 2D é definida como:

$$(\mathbf{I} * \mathbf{F})^{x,y} = \sum_{i=-N}^N \sum_{j=-N}^N \mathbf{I}^{(x+i,y+i)} \mathbf{F}^{(i,j)} \quad (9)$$

Assim, a convolução quaterniônica será definida de forma análoga à equação (9), porém aplicada a uma imagem de valor quaterniônico $\mathbf{I} = (\mathbf{I}_0, \mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3)$ e um filtro também de valor quaterniônico $\mathbf{F} = (\mathbf{F}_0, \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$:

$$\mathbf{I} * \mathbf{F} = [\mathbf{I}_0 \quad \mathbf{I}_1 \quad \mathbf{I}_2 \quad \mathbf{I}_3] * \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 \\ -\mathbf{F}_1 & \mathbf{F}_0 & -\mathbf{F}_3 & \mathbf{F}_2 \\ -\mathbf{F}_2 & \mathbf{F}_3 & \mathbf{F}_0 & -\mathbf{F}_1 \\ -\mathbf{F}_3 & -\mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_0 \end{bmatrix} \quad (10)$$

Em particular, na equação (10), embora \mathbf{I} seja uma imagem de valor quaterniônico, $\mathbf{I}_0, \mathbf{I}_1, \mathbf{I}_2$ e \mathbf{I}_3 são imagens reais e $\mathbf{F}_0, \mathbf{F}_1, \mathbf{F}_2$ e \mathbf{F}_3 também são filtros de valor real. Ou seja, o produto é computado utilizando convoluções de valores reais.

2.3. O Uso dos Quatérnios em Classificação de Imagens

Como destacado por [Parcollet et al. 2020], é fundamental encontrar uma representação adequada de dados para codificar efetivamente as relações presentes entre os pixels de uma imagem. Uma abordagem eficaz para essa representação é considerar cada pixel como uma entidade única, composta por três elementos (os canais RGB, por exemplo) fortemente inter-relacionados, ao invés de representar os pixels como grupos de elementos unidimensionais que *poderiam* estar relacionados, que é a abordagem adotada pelas redes convolucionais tradicionais (de valor real).

A álgebra dos quatérnios oferece uma maneira natural de expressar e processar características de entradas multidimensionais como entidades únicas, já que possui quatro dimensões. Os pixels são os elementos básicos empregados durante o processamento de imagens e geralmente são definidos usando a codificação RGB (vermelho, verde e azul). Os quatérnios representam uma forma eficiente de codificar cada pixel como um elemento completo (uma entidade única), compondo pixels com valores usando um quatérnio puro como:

$$q = 0 + Ri + Gj + Bk. \quad (11)$$

Alternativamente, alguns trabalhos utilizam a codificação HSV (Hue, Saturation, Value), baseada na matiz, saturação e brilho, constituindo um modelo mais intuitivo para a representação de cores. Uma abordagem comparando a codificação RGB e HSV no contexto de redes convolucionais quaterniônicas pode ser vista em [Granero et al. 2021] e [Vieira et al. 2024]. Em particular, neste trabalho utilizaremos a implementação RGB proposta por [Vieira et al. 2024].

Para imagens em tons de cinza, uma abordagem proposta por [Beijing et al. 2021], mapeia os valores em tons de cinza na parte real do quatérnio:

$$q = \text{valores-cinza} + 0i + 0j + 0k. \quad (12)$$

[Parcollet et al. 2020] apresentam várias outras abordagens de representação de pixels no campo de processamento de imagens. Além disso, [Zhu et al. 2018] comparam redes reais e quaterniônicas, apontando que, enquanto a convolução com valores reais é capaz de aplicar apenas transformações de escala nas imagens de entrada, a convolução quaterniônica escalona e rotaciona as imagens no espaço de cores, proporcionando uma representação mais estrutural da informação de cor de uma imagem.

3. Experimentos Computacionais

Nesta seção, comparamos redes convolucionais tradicionais com redes de valor quaterniônico em tarefas de classificação de imagens. Em particular, buscamos entender o comportamento de ambas as abordagens à medida que as tarefas de classificação se tornam mais complexas, ou seja, conforme o número de classes aumenta e as imagens se tornam menos uniformes entre si.

Os conjuntos de dados utilizados para os experimentos foram: PathMNIST [Kather et al 2019], que possui 9 classes; CIFAR10 e CIFAR100 [Krizhevsky 2009], com 10 e 100 classes, respectivamente. Nas subseções a seguir, uma breve descrição sobre cada um dos datasets é apresentada.

Para cada conjunto de dados, adotamos arquiteturas similares às das redes LeNet-5 e VGG (em suas versões com valores reais e de quatérnios) para monitorar o desempenho enquanto construíamos redes convolucionais cada vez mais profundas. No treinamento, padronizamos o uso de 100 épocas e um batch size de 64 em todas as redes deste estudo. Utilizamos a função de ativação ReLU para as camadas convolucionais e Softmax para a camada de saída.

A arquitetura da rede LeNet-5 utilizada consiste em três camadas convolucionais, duas camadas de average pooling e duas camadas densas. Adicionalmente, a técnica de dropout é aplicada após as camadas de pooling, e um batch normalization foi usado após a primeira camada densa, com o objetivo de reduzir o risco de sobreajuste. Para uma visualização detalhada desta arquitetura, veja a Figura 1.

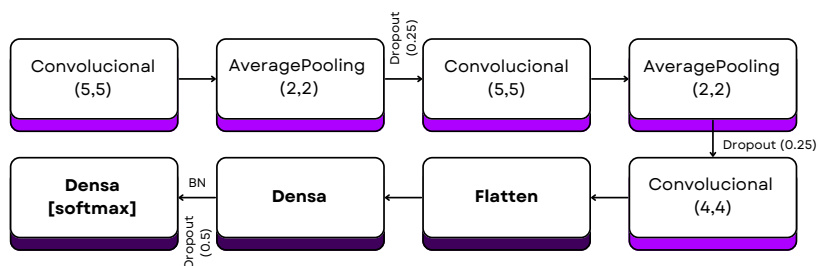


Figura 1. Estrutura: Rede tipo LeNet-5.

A arquitetura da rede VGG utilizada consiste em oito camadas convolucionais, quatro camadas de max pooling e duas camadas densas. Além disso, nessa rede também utilizamos dropout e batch normalization. Na Figura 2, exibimos um esquema da rede do tipo VGG construída.

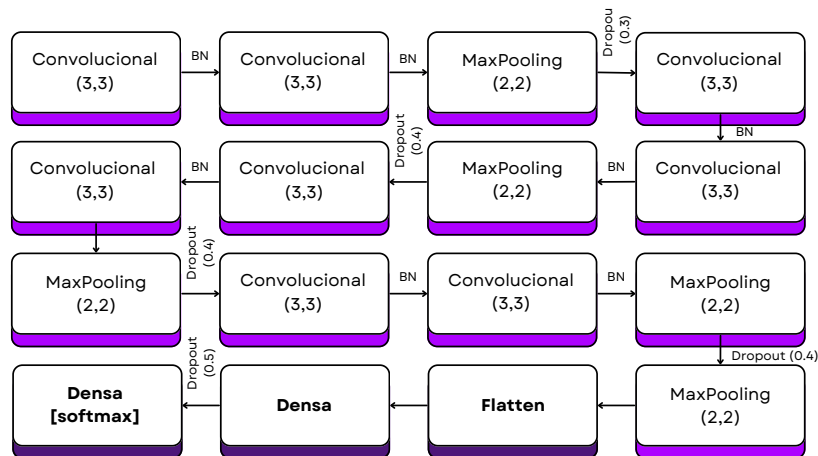


Figura 2. Estrutura: Rede tipo VGG.

Com efeito, assim como em [Granero et al. 2021], durante a construção de cada uma das redes nas suas versões quaterniônicas, dividimos o número de filtros de todas as camadas reais por quatro, para manter as versões quaterniônicas dentro do mesmo orçamento de parâmetros. Essa abordagem garante que todas as redes construídas tenham uma estrutura semelhante, permitindo uma comparação direta de seus desempenhos.

Por fim, na apresentação dos resultados, denotamos as redes convolucionais reais com o prefixo \mathbb{R} e as redes convolucionais quaterniônicas com o prefixo \mathbb{Q} .

3.1. PathMNIST (MEDMNIST)

O MEDMNIST [Yang et al. 2021] é uma compilação de conjuntos de dados que abrangem vários tipos de imagens médicas. Para os nossos experimentos, utilizamos o dataset PathMNIST, um dos 18 conjuntos de dados disponíveis no MEDMNIST, que contém imagens coloridas de dimensão 28×28 de histologia relacionadas a patologias do cólon. O PathMNIST é um dataset com 9 classes e 107.180 amostras, das quais 89.996 são usadas para treinamento, 10.004 para validação e 7.180 para teste. Na Figura 3, apresentamos algumas amostras do PathMNIST.

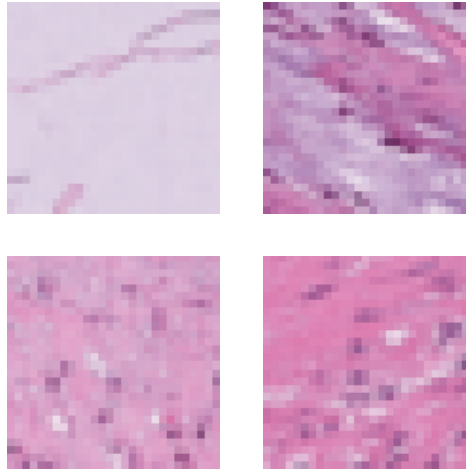


Figura 3. Amostras do PathMNIST.

Comparando o PathMNIST com os conjuntos de dados CIFAR10 e CIFAR100, observamos que ele é o mais simples para tarefas de classificação, tanto pela quantidade de classes quanto pela uniformidade das amostras.

Os resultados de todas as redes utilizadas para o PathMNIST estão dispostos na Tabela 1. Com efeito, a acurácia reportada é referente ao conjunto de validação.

Podemos observar a diferença significativa dos parâmetros entre as redes. Além disso, as acurácias das redes neurais reais ficaram abaixo das redes quaterniônicas, mesmo considerando a diferença do número de parâmetros.

Tabela 1. Nº de Parâmetros e Acurácia - PathMNIST

Arquitetura	Parâmetros	Acurácia
R-LeNet-5	221.257	0.81
R-VGG	4.960.329	0.97
Q-LeNet-5	56.073	0.88
Q-VGG	624.009	0.98

3.2. CIFAR10

O CIFAR10 [Krizhevsky 2009] contém imagens coloridas de dimensão 32×32 , distribuídas em 10 classes, possuindo 60.000 amostras, das quais 50.000 são usadas para treinamento e 10.000 para teste. As classes incluem aviões, automóveis, pássaros, gatos, cervos, cães, sapos, cavalos, navios e caminhões. Na Figura 4, apresentamos algumas amostras dos CIFAR10.

Como as classes do CIFAR10 são bastante distintas entre si, ao compararmos suas amostras com as do PathMNIST, notamos que a tarefa de classificação para o CIFAR10 é consideravelmente mais complexa do que para as imagens de histologia.

Os resultados de todas as redes utilizadas para o CIFAR10 estão dispostos na Tabela 2.

**Figura 4. Amostras do CIFAR10.****Tabela 2. Nº de Parâmetros e Acurácia - CIFAR10**

Arquitetura	Parâmetros	Acurácia
R-LeNet-5	319.306	0.78
R-VGG	5.743.434	0.87
Q-LeNet-5	96.330	0.77
Q-VGG	722.570	0.87

3.3. CIFAR100

O CIFAR100 [Krizhevsky 2009] contém imagens coloridas de dimensão 32×32 , distribuídas em 100 classes, possuindo 60.000 amostras, das quais 50.000 são usadas para treinamento e 10.000 para teste. As classes do CIFAR100 incluem uma variedade de objetos dez vezes maior em comparação com o CIFAR10, abrangendo peixes, flores, insetos, frutas, eletrodomésticos, veículos, entre outros. Cada classe principal é dividida em 20 subclasses, facilitando a categorização em diferentes níveis. Na Figura 5, são apresentadas algumas amostras do CIFAR100.

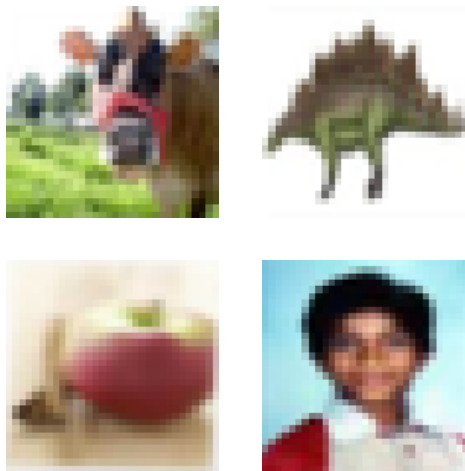


Figura 5. Amostras do CIFAR100.

Comparado aos outros dois datasets utilizados nos experimentos, o CIFAR100 é significativamente mais desafiador para tarefas de classificação devido ao seu elevado número de classes.

Os resultados de todas as redes utilizadas para o CIFAR100 estão dispostos na Tabela 3.

Tabela 3. Nº de Parâmetros e Acurácia - CIFAR100

Arquitetura	Parâmetros	Acurácia
\mathbb{R} -LeNet-5	342.948	0.40
\mathbb{R} -VGG	5.793.444	0.56
\mathbb{Q} -LeNet-5	119.460	0.44
\mathbb{Q} -VGG	551.652	0.57

4. Conclusão

Neste trabalho, comparamos os desempenhos de redes neurais convolucionais de valor real e de valor quaterniônico em tarefas de classificação. Utilizamos a codificação RGB e duas arquiteturas de rede convolucional, LeNet-5 e VGG, a fim de monitorar o

desempenho de ambas as abordagens à medida que usávamos redes mais profundas e as tarefas de classificação se tornavam mais complexas, com um maior número de classes e imagens menos uniformes.

Nossos resultados mostram que as redes convolucionais quaterniônicas superam as redes de valor real para imagens mais uniformes, utilizando apenas cerca de 12.5% do número de parâmetros e alcançando uma acurácia de 98%. Para imagens mais complexas, como as do CIFAR-10, percebemos que a rede quaterniônica, mesmo usando 12.5% dos parâmetros da rede real, alcançou o mesmo nível de acurácia. Para o problema mais complexo deste trabalho, o CIFAR-100, a rede quaterniônica apresentou uma acurácia 0.01 maior que a rede VGG real, usando aproximadamente 9.5% dos parâmetros da rede real.

Esses resultados indicam que as redes quaterniônicas performaram melhor do que as redes reais, especialmente em termos de eficiência no uso de parâmetros, mantendo ou até melhorando a acurácia em diferentes níveis de complexidade das imagens.

Referências

- Altamirano-Gomez, G., and Gershenson, C. (2023). Quaternion Convolutional Neural Networks: Current Advances and Future Directions. arXiv preprint arXiv:2307.08663.
- Beijing, C., Xingwang, J., Ye, G., & Jinwei, W. (2021). A Quaternion Two-Stream R-CNN Network for Pixel-Level Color Image Splicing Localization. *Chinese J. Electron.*, 30, 1069-1079. <https://doi.org/10.1049/cje.2021.08.004>
- Gaudet, C. J., & Maida, A. S. (2018). Deep quaternion networks. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
- Granero, M. A., Hernández, C. X., & Valle, M. E. (2021, November). Quaternion-valued convolutional neural network applied for acute lymphoblastic leukemia diagnosis. In *Brazilian Conference on Intelligent Systems* (pp. 280-293). Cham: Springer International Publishing.
- Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C. A., Gaiser, T., Marx, A., Valous, N. A., Ferber, D., Jansen, L., Reyes-Aldasoro, C. C., Zörnig, I., Jäger, D., Brenner, H., Chang-Claude, J., Hoffmeister, M., & Halama, N. (2019). Predicting survival from colorectal cancer histology using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1), e1002730.
- Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. Technical report.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Parcollet, T., Zhang, Y., Morchid, M., Trabelsi, C., De Mori, R., & Bengio, Y. (2018). Quaternion Convolutional Neural Networks for End-to-End Speech Recognition. *Proc. Interspeech 2018*, 22-26.
- Parcollet, T., Ravanelli, M., Morchid, M., Linarès, G., Trabelsi, C., & Bengio, Y. (2018). Quaternion recurrent neural networks.

Parcollet, T., Morchid, M., & Linar'es, G. (2020). A survey of quaternion neural networks. *Artificial Intelligence Review*, 53(4), 2957–2982.

Senna, F., & Valle, M. (2021). Tessarine and Quaternion-Valued Deep Neural Networks for Image Classification. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional* (pp. 350-361). Porto Alegre: SBC. doi:10.5753/eniac.2021.18266

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Vieira, G., Valle, M. E., & Lopes, W. (2024). Clifford Convolutional Neural Networks for Lymphoblast Image Classification. *Advanced Computational Applications of Geometric Algebra*.

Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., ..., & Ni, B. (2023). Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1), 41.

Zhu, X., Xu, Y., Xu, H., & Chen, C. (2018). Quaternion convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 631-647).