

# Rede Granular Convolutiva Evolutiva para Classificação de Fluxo de Imagens

Danielle A. Fortunato\* Silvia C. Ferreira\*\* Patricia P. F. Ferraz\*\*  
Rafael A. Santos\*\* Daniel F. Leite\*\*\*

\* *Departamento de Engenharia de Computação e Automação,  
Universidade Estadual de Campinas, UNICAMP, SP  
E-mail: danielle.afortunato@gmail.com*

\*\* *Universidade Federal de Lavras, UFLA, MG*

*E-mail: silvia.ferreira@ufla.br, patricia.ponciano@ufla.br*

\*\*\* *Department of Computer Science, Paderborn University, Germany  
E-mail: daniel.leite@uni-paderborn.de*

---

**Abstract:** Recent advances in machine learning for computer vision and image classification have presented significant challenges, such as interpretability of deep neural network models and ability for continuous learning in dynamic environments. This paper introduces Convolutional Evolving Granular Neural Networks (CEGNN). A CEGNN combines the feature extraction components of a VGG-16 convolutional net with an evolving granular neural network (EGNN) toward advancing the understanding and applications of incremental learning in computer vision, particularly in image recognition and classification. An incremental algorithm is incorporated to the CEGNN to improve model interpretability and continuous learning. Experimental results show that the CEGNN is efficient and competitive in image classification, achieving an accuracy of 78.9% and a precision of 79.0% in a 10-class problem. This opens research avenues in certain applications, such as those dealing with medical images, satellite images, and image-based closed-loop autonomous systems.

**Resumo:** Os recentes avanços em aprendizado de máquina para visão computacional e classificação de imagens têm apresentado desafios importantes, como a interpretabilidade de modelos de redes neurais profundas e a capacidade de aprendizado contínuo em ambientes dinâmicos. Este artigo introduz a Rede Neural Granular Evolutiva Convolutiva (CEGNN). Ela combina os elementos extratores de características de uma rede convolutiva VGG-16 com uma rede neural granular evolutiva (EGNN) visando avançar o entendimento e a aplicação do aprendizado incremental em visão computacional, especialmente no reconhecimento e classificação de imagens. Um algoritmo incremental é incorporado à CEGNN para lidar com a interpretabilidade e o aprendizado contínuo do modelo neural. Os resultados experimentais mostram que a CEGNN é eficiente e competitiva em classificação de imagens, alcançando uma acurácia de 78,9% e uma precisão de 79,0% em um problema de 10 classes. Isto abre frente de pesquisa em certas aplicações, como aquelas que lidam com imagens médicas, de satélite, e sistemas autônomos em malha fechada baseados em imagem.

*Keywords:* Computer vision; Evolving intelligent systems; Deep learning; Granular computing.

*Palavras-chaves:* Visão computacional; Sistemas inteligentes evolutivos; Aprendizado profundo; Computação granular.

---

## 1. INTRODUÇÃO

Recentes avanços em reconhecimento de padrões e análise de imagens têm impulsionado uma variedade de aplicações em diversos setores. Essas tecnologias são fundamentais para melhorar a produtividade e atender às demandas do mercado, oferecendo produtos e serviços de alta qualidade. Métodos baseados em visão computacional e aprendizado de máquina têm sido essenciais em áreas como agricultura, meteorologia, imagens biomédicas, metalografia, robótica móvel e veículos autônomos, entre outras (Li, 2023).

Por exemplo, na agricultura, algoritmos identificam padrões em imagens de plantações, permitindo a detecção precoce de doenças e pragas (Cardoso et al., 2020). Na

meteorologia, a interpretação de imagens de satélite possibilita previsões climáticas mais precisas (Bhardwaj et al., 2019). Em imagens biomédicas, algoritmos auxiliam no diagnóstico e tratamento de condições médicas complexas (Litjens et al., 2017). Na metalografia, a análise de imagens de microestruturas metálicas garante a qualidade e segurança dos produtos (Rodrigues et al., 2017). Em robótica autônoma, sistemas de percepção visual permitem a navegação de robôs móveis a partir da classificação das cenas (Ran et al., 2021) (Shaheen et al., 2022).

Diante dessas aplicações, é essencial continuar desenvolvendo algoritmos para aprimorar a visão computacional para lidar com situações e contextos cada vez mais rea-

lísticos, dinâmicos e complexos. Muitos desses algoritmos são dirigidos por dados e usam técnicas de aprendizado de máquina, incluindo redes neurais profundas. Métodos de otimização como Mini-Batch SGD (Gradiente Descendente Estocástico), AdaGrad e Adam são fundamentais para treinar essas redes em grandes conjuntos de dados (Koushik and Srinivasa, 2021); (Jaraa and Bowenb, 2022). O uso crescente de aprendizado de máquina destaca-se especialmente na análise de grandes volumes de dados (Verne et al., 2022). Dois desafios em particular são aqueles relativos a interpretabilidade e a confiança em modelos profundos (Souza and Dragoni, 2024) (Leite et al., 2024b), e a capacidade de aprendizado online contínuo para capturar novos padrões em ambientes dinâmicos (Mai et al., 2022) (Shaheen et al., 2022) (Leite et al., 2024a).

Dessa forma, este trabalho propõe a CEGNN (*Convolutional Evolving Granular Neural Network*), uma nova metodologia para classificação de imagens. A CEGNN combina elementos de uma rede neural bem conhecida, denominada VGG-16, com uma rede neural granular evolutiva (EGNN) (Leite et al., 2013) (Leite et al., 2024b). Essa abordagem permite uma adaptação dinâmica às mudanças no ambiente, contribuindo para uma classificação mais precisa e eficiente de imagens em contexto não-estacionário. Ademais, o *topo* da rede neural CEGNN proposta – isto é, a parte do modelo responsável pela estimação de classe em si – além de ser estruturalmente adaptativo, o que permite a incorporação de novos padrões ao longo do tempo, possui uma estrutura granular, isto é, transparente, ou que pode ser lida como regras linguísticas.

## 2. MATERIAIS E MÉTODOS

### 2.1 Rede Convolutiva VGG-16

A rede neural convolutiva VGG-16 representa uma evolução importante na visão computacional, sendo uma extensão da AlexNet, desenvolvida por Krizhevsky et al. (2017). Incorporando unidades de ativação ReLU e treinamento em GPU, a VGG-16 alcançou sucesso na competição ImageNet, destacando-se por sua profundidade e precisão (Vargas et al., 2016). A VGG-16, por sua vez, incorpora esses princípios, apresentando entradas de  $224 \times 224 \times 3$  e 13 camadas, que alternam entre operações de convolução e pooling.

A VGG-16 utiliza filtros pequenos ( $3 \times 3$ ) em cada camada convolutiva, contrariando abordagens anteriores. Essa escolha permitiu reduzir a quantidade de parâmetros, mantendo a eficácia na extração de características (Rodrigues et al., 2021). A arquitetura é composta por cinco blocos convolutivos, seguidos por três camadas totalmente conectadas e uma camada softmax de saída. Essa configuração padrão é treinada com milhares de imagens do ImageNet, um conjunto de dados extenso que desempenha um papel crucial no treinamento de modelos de redes neurais (Rezende et al., 2020).

O ImageNet, com mais de 15 milhões de imagens classificadas, é essencial para o treinamento e avaliação de modelos de visão computacional. O conjunto de dados é utilizado anualmente no ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), promovendo avanços na área (Krizhevsky et al., 2017). A Figura 1 apresenta o

diagrama da estrutura da VGG-16, a qual é constituída por cinco camadas convolutivas, três camadas totalmente conectadas e uma camada de saída softmax.

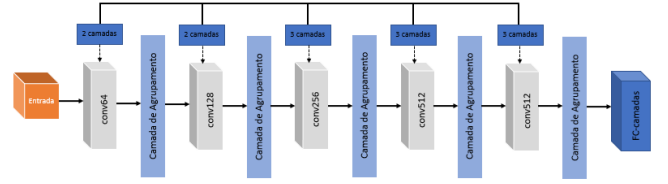


Figura 1. Diagrama da estrutura da rede convolutiva denominada VGG-16

### 2.2 Análise de Componentes Principais

A análise de componentes principais (PCAs, *principal component analysis*) é uma técnica crucial da estatística multivariada, que visa transformar um conjunto de variáveis originais em um conjunto de variáveis de mesma dimensão, chamadas de componentes principais. Essa técnica é fundamental para a redução da dimensionalidade dos dados, mantendo o máximo possível de informação contida nos mesmos. Por meio da PCA, busca-se redistribuir a variação observada nos dados originais em um conjunto de eixos ortogonais não correlacionados, permitindo a representação dos dados em um espaço de dimensão inferior. Suas aplicações abrangem desde a geração de índices até o agrupamento de indivíduos, proporcionando uma visão mais clara e simplificada das relações entre as variáveis.

A determinação dos componentes principais é realizada através da resolução da equação característica da matriz de covariância ou correlação, resultando em autovalores e autovetores que representam a variância e a direção dos componentes, respectivamente. Os escores dos componentes principais são utilizados para análise e interpretação dos dados em dimensões reduzidas, facilitando a compreensão dos padrões subjacentes aos dados. Em suma, a PCA é uma ferramenta poderosa para simplificar a análise de conjuntos de dados complexos, proporcionando uma representação mais eficiente e compreensível dos mesmos (Varela, 2008). A equação (1) é utilizada para determinar os valores dos componentes principais para cada observação:

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ip}x_p \quad (1)$$

em que  $y_i$  representa o  $i$ -ésimo componente principal;  $a_{ij}$  são os elementos dos autovetores e  $x_j$  são as variáveis originais. Essa equação mostra como cada componente principal é calculado como uma combinação linear das variáveis originais, ponderadas pelos coeficientes do autovetor correspondente.

### 2.3 Rede Neural Granular Evolutiva EGNN

O aprendizado em EGNN segue um princípio comum que, geralmente, envolve dois estágios. Primeiro, grânulos de informação – intervalos ou conjuntos fuzzy – são construídos a partir de uma base de representação numérica. Em seguida, a aprendizagem – construção e refinamento – da

rede neural é baseada nos grânulos de informação ao invés de ser baseada nos dados originais. Assim, a rede EGNN não é exposta a todos os dados de treinamento, muito mais numerosos que os grânulos formados. Por exemplo, quando amostras de dados não transportam novas informações, os exemplos são incorporados pelos grânulos já conhecidos (Fortunato et al., 2020), (Leite, 2019).

Considere que o fluxo de dados  $(\mathbf{x}, y)^{[h]}$ ,  $h = 1, \dots$ , é medido a partir de uma função desconhecida  $f$ . Entradas  $x_j$  e saída  $y$  são dados numéricos e uma classe. A Figura 2 mostra um modelo EGNN de quatro camadas. A camada de entrada recebe  $x^{[h]}$ . A camada granular é um conjunto de grânulos  $G_j^i$ ,  $i = 1, \dots, c$ , estratificado a partir dos dados de entrada, formando uma partição fuzzy da  $j$ -ésima entrada domínio. Um grânulo  $G^i = G_1^i \times \dots \times G_n^i$  é uma relação fuzzy, ou seja, um conjunto fuzzy multidimensional em  $X_1 \times \dots \times X_n$ . Assim,  $G^i$  tem a função de pertinência  $G^i(x) = \min(G_1^i(x_1), \dots, G_n^i(x_n))$  em  $X_1 \times \dots \times X_n$  (Decker et al., 2020).

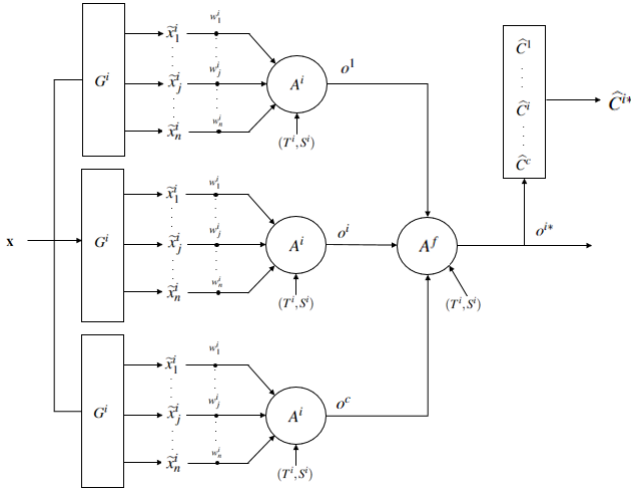


Figura 2. Rede neural granular evolutiva (EGNN) para classificação de dados numéricos, como aqueles produzidos no espaço latente de CNNs

Para a atualização dos pesos da rede neural,  $\omega_j^i \in [0, 1]$  é proporcional à importância do  $j$ -ésimo atributo de  $G_j^i$  à saída da rede neural. Quando um novo grânulo  $G^{c+1}$  é gerado, os pesos são definidos como  $\omega_j^{c+1} = 1, \forall j$ . O  $\omega_j^i$  atualizado é associado ao grânulo mais ativo  $G^i$ ,  $i = \text{argmax}(o^1, \dots, o^c)$ , por:

$$\omega_j^i(\text{novo}) = \omega_j^i(\text{velho}) - \beta^i \tilde{x}_j^i \in | \quad (2)$$

no qual  $\tilde{x}_j^i$  é a similaridade entre o elemento  $x_j$  do vetor de entrada e o grânulo  $G_j^i$ ;  $\beta^i$  depende do número de acertos ( $\text{acerto}^i$ ) e erros ( $\text{erro}^i$ ) locais de classificações conforme

$$\beta^i = \frac{\text{erro}^i}{\text{acerto}^i + \text{erro}^i}, \quad (3)$$

e

$$\in^{[h]} = C^{[h]} - \hat{C}^{[h]} \quad (4)$$

em que  $\in^{[h]}$  é o erro de estimação atual,  $\hat{C}^{[h]}$  é a classe estimada e  $C^{[h]}$  é a classe real.

Ademais os neurônios de agregação  $A^i$ ,  $i = 1, \dots, c$ , e  $A^f$ , conforme apresentados na Figura 2, podem implementar quaisquer funções de agregação (Leite and Skrzjanc, 2019) (Souza and Dragoni, 2024). Este artigo adota a T-norma  $\min$  para  $A^i \forall i$ , e a S-norma  $\max$  para  $A^f$  (Beliakov et al., 2008). É importante notar que inicialmente  $c = 0$ , ou seja, a estrutura da rede não existe antecipadamente. Toda a estrutura evolui conforme novos padrões e classes surgem no fluxo de dados  $x^{[h]}$ ,  $h = 1, \dots$ .

A construção incremental da estrutura, conexões e neurônios de agregação da rede EGNN – especializada para a tarefa de classificação conforme mostrado na Figura 2 – tão bem como os mecanismos de mesclagem, deslize e remoção de grânulos de informação ao longo do tempo serão omitidos deste artigo. O algoritmo pode ser encontrado em (Leite, 2019). Novos refinamentos do algoritmo podem ser encontrados em (Leite et al., 2024b).

### 3. METODOLOGIA

#### 3.1 Conjuntos de Dados

Para avaliar o desempenho do modelo proposto na classificação de fluxo de imagens, dois conjuntos de dados distintos foram empregados. O conjunto de dados *Cats-vs-Dogs* ou Gato versus Cachorro foi obtido do *Kaggle Repository* e é dedicado à classificação de imagens de dois tipos de animais: gatos e cachorros. Estes dados foram disponibilizados em 2018 pelo repositório Kaggle. As principais características desse conjunto são as seguintes: 5011 imagens de gatos e 5017 imagens de cachorros. As imagens neste conjunto apresentavam diferentes tamanhos, variando de  $(327 \times 500)$ ,  $(499 \times 375)$ , entre outros. Na Figura 3 são apresentados alguns exemplos de imagens do conjunto de dados



Figura 3. Imagens do conjunto de dados Cats-vs-Dogs

O conjunto de dados *Common Objects Day and Night* (CODaN) foi selecionado para avaliação mais abrangente, considerando um ambiente de classificação de imagens em 10 classes no domínio diurno. As classes incluídas no conjunto de dados CODaN são: bicicleta, carro, motocicleta, ônibus, barco, gato, cachorro, garrafa, xícara e cadeira. As principais características desse conjunto incluem 10.000 imagens de treinamento para a CNN e 2.500 imagens de teste, com 250 imagens disponíveis para cada classe. É importante ressaltar que para a avaliação foram utilizadas exclusivamente imagens do domínio diurno do conjunto de dados CODaN. Na Figura 4 são apresentados alguns exemplos de imagens do conjunto de dados CODaN, utilizado com dez classes.



Figura 4. Imagens do conjunto de dados CODaN

### 3.2 Proposta para Classificação de Imagens

A proposta para classificação de imagens apresenta um sistema composto por quatro etapas fundamentais como é representado no diagrama da Figura 5.

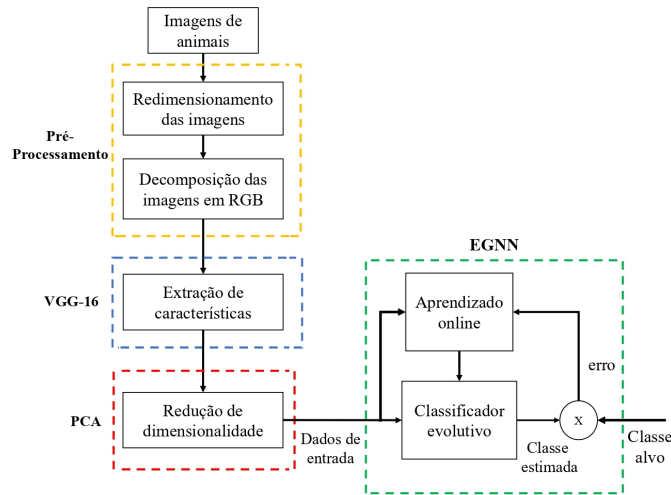


Figura 5. Diagrama convolucional granular evolutivo

Primeiramente, as imagens são pré-processadas, passando por redimensionamento e decomposição dos canais RGB, permitindo a extração de características discriminativas. Em seguida, a rede neural convolucional VGG-16 é empregada para extrair características visuais complexas das imagens. Posteriormente, a Análise de Componentes Principais (PCA) é aplicada para reduzir a dimensionalidade e preservar informações essenciais para a classificação. Por fim, a rede neuro-fuzzy granular evolutiva (EGNN) é utilizada para a classificação das imagens com base nas características extraídas, combinando aprendizado de máquina incremental e interpretabilidade de grânulos, resultando em um modelo adaptável e eficiente para suporte à tomada de decisão.

### 3.3 Modelagem Proposta para Análise de Fluxo de Imagens

A Figura 6 ilustra o fluxograma do funcionamento da arquitetura convolucional incremental proposta, destacando a substituição das camadas estáticas e totalmente conectadas do topo da rede VGG-16 – camadas efetivamente responsáveis pela classificação – pela rede EGNN equipada com algoritmo incremental. A arquitetura CEGNN apresenta capacidade de incorporar novos padrões e refinar grânulos de informação ao longo da vida.

As camadas convolucionais da rede VGG-16 extraem características das imagens, enquanto as camadas subsequentes utilizam as características latentes mais importantes em termos da maximização da variabilidade dos dados para classificar os objetos presentes nas imagens.

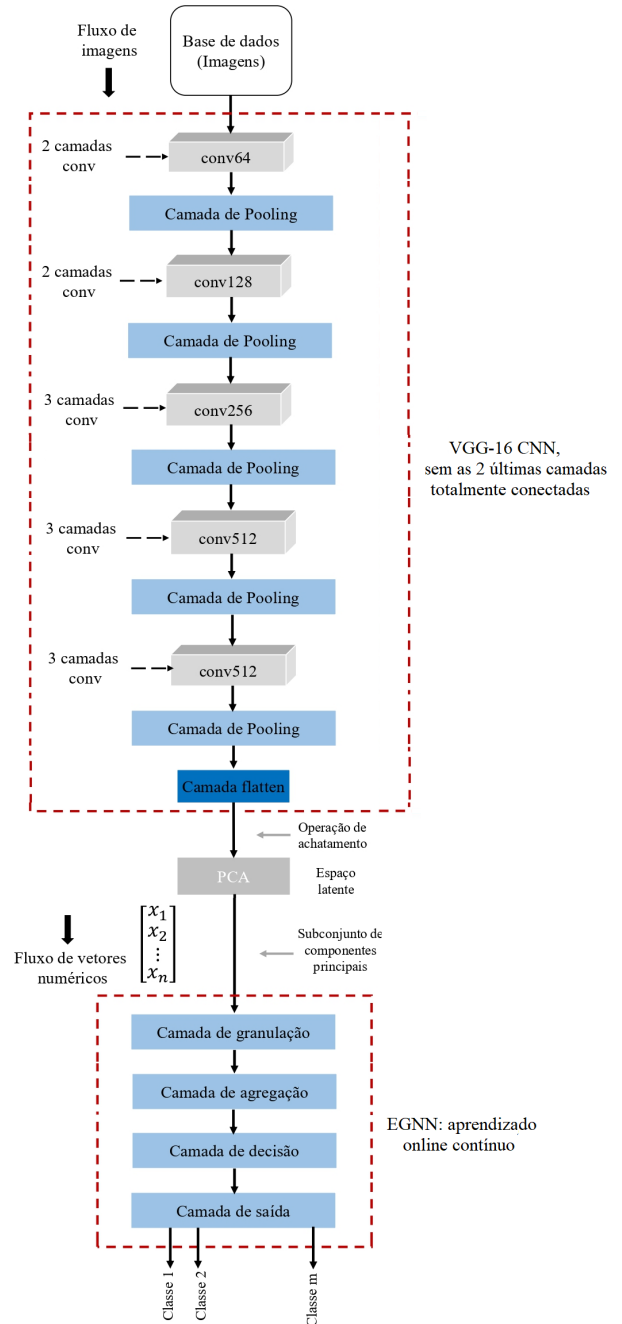


Figura 6. Diagrama da estrutura do modelo CEGNN, que combina a rede VGG-16 com a rede EGNN equipada com aprendizado incremental

Em muitas arquiteturas de redes neurais convolucionais, as últimas camadas com pesos ajustáveis em modo offline são totalmente conectadas. Este trabalho substituiu estas por camadas de granulação dinâmica e agregação para captura online de novas informações. O topo EGNN conta com um número de saídas igual ao número de classes no novo conjunto de dados, conforme estes são observados a partir do fluxo.

É relevante notar que todo esse processo de substituição e adaptação ocorreu em um ambiente computacional significativamente avançado. O hardware utilizado, caracterizado por um Inspiron 24 All in One de 13ª geração Intel® Core™ i7-1355U, Windows 11 Home, Intel® Iris® Xe

Graphics, 16 GB DDR4, 512 GB SSD, desempenhou um papel fundamental na eficiência do processamento. Essas características do computador garantiram a capacidade de lidar com a complexidade das redes acopladas e a adaptabilidade do modelo, proporcionando uma implementação eficaz e rápida dessa metodologia.

O modelo VGG-16 foi adaptado para preservar suas camadas iniciais pré-treinadas, mas substituindo a última camada totalmente conectada pela arquitetura de uma EGNN. Os pesos do modelo VGG-16, inicialmente obtidos a partir da classificação no conjunto de dados ImageNet, foram mantidos utilizando *transfer learning*, com apenas as camadas finais sendo modificadas para se adequarem ao novo contexto.

#### 4. RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados os resultados obtidos a partir da avaliação do modelo treinado utilizando o conjunto de dados *Cats-vs-Dogs* e CODaN. Foram consideradas métricas essenciais, como acurácia, precisão, *Recall* e F1-Score, além da matriz de confusão, para analisar o desempenho do modelo. Também foram examinados alguns cenários, como a média de regras geradas durante o processo de classificação. Essas regras representam padrões identificados pelo algoritmo para distinguir as classes, e sua média oferece uma medida da complexidade do modelo. Um maior número de regras pode sugerir uma complexidade elevada ou um ajuste excessivo do modelo aos dados. Os resultados relacionados ao conjunto de dados de duas classes, com foco na classificação entre Gato e Cachorro são apresentados na Tabela 1.

Tabela 1. CEGNN: Desempenho de classificação de imagens para duas classes *Cats-vs-Dogs*

Classificador	CEGNN			
	5	10	15	20
<b>Características</b>				
<i>A<sub>cc</sub></i> (%)	91,23	89,03	88,70	86,94
<b>Média de regras</b>	66,07	136,03	172,72	195,05
<b>Tempo (s)</b>	115,80	116,20	123,20	511,60
<b>Precisão (%)</b>	91,00	88,00	92,00	89,00
<b>Recall (%)</b>	91,00	88,00	92,00	89,00
<b>F1 (%)</b>	90,00	89,00	89,00	86,00

Notavelmente, o desempenho mais expressivo foi alcançado ao incorporar um conjunto de novas variáveis, nomeadamente os cinco primeiros componentes principais do conjunto de dados original. Nessa configuração, a performance total alcançou uma taxa de acerto de 91,23% para as classes consideradas com um tempo de processamento de 115,8 segundos para todas as amostras, utilizando granularidades iniciais pequenas, como  $\rho = 0,5$ , um total de 1000 iterações, sendo uma iteração por instância de entrada de teste e aproximadamente 86 regras. A precisão do modelo foi avaliada em 91,00%, indicando sua habilidade em evitar classificar erroneamente instâncias negativas como positivas. Além disso, o *Recall* atingiu 91,00%, evidenciando a capacidade do algoritmo em identificar corretamente todas as instâncias positivas. Um aspecto crucial foi o valor do F1-Score, que atingiu 90,00%. Este índice reflete um equilíbrio efetivo entre precisão e *Recall*, sublinhando o desempenho do modelo em termos de acurácia e capacidade de identificação correta.

A Figura 7 exhibe a matriz de confusão para CEGNN, destacando a contagem de classificações precisas em comparação com aquelas incorretas para as imagens de Gato e Cachorro. Para a análise seguinte, a matriz de confusão possui duas classes distintas, sendo a classe (0) referente a imagens de cachorros e a classe (1) a imagens de gatos. Observou-se que, entre as 5017 amostras de imagens contendo cachorros, 322 foram erroneamente classificadas como imagens contendo gatos, enquanto 4695 foram corretamente classificadas. Ao contrário, entre 5011 amostras contendo imagens de gato, 557 foram erroneamente classificadas como contendo cachorros, e 4454 foram corretamente classificadas.

		Classe estimada		
		0	1	
Classe de saída	0	4695	557	89,40%
	1	322	4454	93,30%
		93,60%	88,90%	<b>91,20%</b>

Figura 7. Matriz de confusão para a classificação de duas classes

Os resultados relacionados a classificação do conjunto de dados de 10 classes são apresentados na Tabela 2.

Tabela 2. CEGNN: Desempenho de classificação de imagens para dez classes CODaN

Classificador	CEGNN			
	5	10	15	20
<b>Características</b>				
<i>A<sub>cc</sub></i> (%)	62,08	78,88	73,80	71,25
<b>Média de regras</b>	37,59	37,75	39,31	82,37
<b>Tempo (s)</b>	25,98	24,51	26,85	71,25
<b>Precisão (%)</b>	62,00	79,00	76,00	72,00
<b>Recall (%)</b>	62,00	78,00	73,00	71,00
<b>F1 (%)</b>	61,00	79,00	74,00	71,00

Observou-se que um aprimoramento de desempenho foi alcançado ao se utilizar 10 componentes principais na análise. Nessa configuração expandida, a acurácia global atingiu a marca de 78,88% com um tempo de processamento de 24,51 segundos para todas as amostras, empregando granularidade pequenas, como  $\rho = 0,5$ , um total de 800 iterações e aproximadamente 46 regras. A precisão do modelo foi calculada em 79,00%, indicando eficácia em evitar classificações equivocadas de instâncias negativas como positivas. Além disso, a taxa de *Recall* atingiu 78,00%, demonstrando a capacidade do algoritmo em identificar corretamente todas as instâncias positivas. O índice F1-Score atingiu 79,00%, indicando um equilíbrio sólido entre precisão e *Recall*.

Quando se está lidando com um conjunto de dados de 10 classes, como é o caso mencionado, a complexidade e a diversidade das características presentes nas imagens podem ser significativamente maiores do que em um conjunto de dados com apenas 2 classes. Isso ocorre porque, ao lidar com mais classes, há uma ampla gama de objetos, formas, texturas e contextos visuais que precisam

ser distintamente reconhecidos e classificados pelo modelo. Em contrapartida, com um número maior de PCAs, o modelo pode capturar mais variações nos dados do que com apenas 10 PCAs. Isso pode resultar em uma representação mais detalhada das características das imagens no espaço de características. No entanto, se a complexidade adicional não estiver alinhada com a estrutura dos dados ou se introduzir ruído desnecessário, isso pode afetar negativamente o desempenho do modelo. Nem sempre um número maior de componentes principais significa melhor desempenho. A seleção inadequada de PCAs pode levar a uma representação inadequada dos dados ou à inclusão de características irrelevantes, o que pode prejudicar o desempenho do modelo.

A Figura 8 exibe a matriz de confusão para CEGNN, destacando a contagem de classificações precisas em comparação com aquelas incorretas para as imagens de dez classes de animais e objetos. Na análise subsequente, a matriz de confusão contém dez classes distintas, onde a classe (0) refere-se a imagens de bicicletas, a classe (1) a imagens de barcos, a classe (2) a imagens de garrafas, a classe (3) a imagens de ônibus, a classe (4) a imagens de carros, a classe (5) a imagens de gatos, a classe (6) a imagens de cadeiras, a classe (7) a imagens de xícaras, a classe (8) a imagens de cachorros e a classe (9) a imagens de motocicletas.

	0	1	2	3	4	5	6	7	8	9	
0	205		3		2	1	17	1	1	20	88,34%
1	2	201	5	4	4	6	17	3	1	7	94,83%
2	1	3	184		1	5	18	33		5	72,44%
3		2	2	209	21	2	13			1	89,60%
4			12	12	201	8	7	3	1	6	68,49%
5		1	2		4	192	10	5	35	1	70,80%
6	7	2	15	2	11	9	189	8	4	3	71,59%
7	2	3	21		9	6	19	182		8	65,96%
8	1		1		1	40	6	2	197	2	78,81%
9	16		6		6		9	1		212	78,81%
	85,08%	91,00%	73,75%	84,73%	75,97%	71,76%	75,29%	69,62%	78,00%	83,26%	<b>78,88%</b>
	0	1	2	3	4	5	6	7	8	9	

Figura 8. Matriz de confusão para a classificação de dez classes

A maior confusão do modelo ao classificar as amostras foi em relação a classe (7) referente as imagens de xícaras, onde das 250 amostras de imagens de xícaras, 182 foram adequadamente classificadas. Entretanto, foram identificados equívocos na predição, onde 1 amostra foi incorretamente associada a imagens de bicicletas, 3 amostras a imagens de barcos, 33 amostras a imagens de garrafas, 3 amostras a imagens de carros, 5 amostras a imagens de gatos, 8 amostras a imagens de cadeiras, 2 amostras a imagens de cachorros e 1 amostra a imagens de motocicletas. Essa análise detalhada destacou áreas específicas de desafio em relação à classe de imagens de garrafas, proporcionando uma compreensão mais aprofundada do desempenho do modelo em cada categoria. A análise detalhada enfoca especialmente a classe de imagens de garrafas, evidenciando que a forma mais alongada das xícaras pode ter contribuído para confusões com garrafas. A presença dessas confusões permitiu reflexões sobre possíveis melhorias e otimizações no processo de classificação, contribuindo para o refinamento do modelo em questão.

A escolha da arquitetura VGG-16 neste estudo se baseia na sua habilidade em mapear suavemente o espaço de entrada para o espaço de saída, resultando em uma superfície de mapeamento geralmente suave. Apesar de possuir 138 milhões de parâmetros, a VGG-16 demonstra notável capacidade de generalização, refletida na suavidade de sua superfície de mapeamento, o que é crucial para a interpretabilidade. No entanto, maximizar a variância no espaço transformado de menor dimensão enfrenta desafios com a aplicação do PCA, pois requer acesso à base de dados offline completa, o que é inviável em um contexto de fluxo de dados. Para contornar essa limitação, foi adotada uma abordagem incremental combinada com PCA, permitindo a análise contínua dos dados. A interpretabilidade do modelo foi aprimorada ao fundamentá-lo em partições do espaço latente, o que permite compreender como o modelo chega a uma decisão, associando cada região do espaço de entrada a uma classe ou categoria específica. Essa abordagem facilita a interpretação das características relevantes para a classificação das imagens, proporcionando transparência e interpretabilidade na construção resultante.

## 5. CONCLUSÃO

Este artigo apresentou uma nova abordagem, denominada CEGNN, para a classificação de fluxos não-estacionários de imagens em tempo real, combinando o extrator da rede neural convolucional VGG-16 com a flexibilidade estrutural e paramétrica da rede granular incremental EGNN. A aplicação da análise de componentes principais no ponto de conexão entre as redes (espaço latente) contribuiu significativamente para otimizar o desempenho do modelo, permitindo uma representação mais compacta das características latentes das imagens no classificador granular. Essa combinação de técnicas resultou em um modelo robusto e eficaz, capaz de lidar com a complexidade e variabilidade dos dados de forma relativamente precisa. Destacou-se a escolha estratégica do extrator VGG-16, cuja superfície de mapeamento entrada-saída mais suave proporcionou melhor generalização e interpretabilidade. Os resultados demonstraram uma acurácia de 78,88% e uma precisão de 79,00% na tarefa de classificação para dez classes CODaN, refletindo a capacidade do modelo de classificar corretamente as amostras positivas e negativas. A abordagem CEGNN mostra-se promissora em diversas aplicações onde os dados não podem ser armazenados em memória para serem analisados posteriormente em vista de seu grande volume e/ou de sua dinâmica. Fluxos de dados têm a sua essência capturada em um único passo de treinamento incremental por meio do modelo CEGNN.

Em pesquisa futura será explorada e avaliada a interpretabilidade das variáveis produzidas no espaço latente de CEGNN. Um índice que quantifica a interpretabilidade do classificador será incorporado como termo de regularização na função de perda que dirige o treinamento incremental. Ademais, aplicações considerando fluxo de imagens médicas e imagens capturadas por robôs móveis em diferentes contextos serão consideradas.

## AGRADECIMENTOS

O primeiro autor agradece a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por meio do au-

xílio 2024/04031-2 e a Universidade Estadual de Campinas (FEEC-UNICAMP). O último autor agradece ao ministério de ciência e cultura, *Ministry of Culture and Science of the State of North Rhine-Westphalia*, pelo suporte NW21-059D (projeto SAIL) e a Paderborn University.

## REFERÊNCIAS

- Beliakov, G., Pradera, A., and Calvo, T. (2008). *Aggregation Functions: A Guide for Practitioners*. Springer.
- Bhardwaj, A., Sam, L., Martín-Torres, F.J., Zorzano, M., and Luque, J.A.R. (2019). Uav imaging of a martian brine analogue environment in a fluvio-aeolian setting. *Remote Sens*, 11, 1–32.
- Cardoso, J.R.F., F., C.E.A., T., J.E.P., Zerbato, C., Carneiro, F.M., and Estevam, F.N.L. (2020). Development of a robotic structure for acquisition and classification of images (eraci) in sugarcane crops. *Revista Ciência Agrônômica*, 51, 1–15.
- Decker, L., Leite, D., Viola, F., and Bonacorsi, D. (2020). Comparison of evolving granular classifiers applied to anomaly detection for predictive maintenance in computing centers. *IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 1–7.
- Fortunato, D., Santana, M., Gomes, J., and Leite, D. (2020). Modelagem granular neuro-fuzzy evolutiva para classificação de distúrbios em sistemas de distribuição de potência. In *Congresso Brasileiro de Automática (CBA), Porto Alegre*, 1–7.
- Jaraa, J.D.Z. and Bowenb, S. (2022). Learning curve analysis on adam, sgd, and adagrad optimizers on a convolutional neural network model for cancer cells recognition. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 11, 263–283.
- Koushik, S.S. and Srinivasa, K. (2021). Detection of respiratory diseases from chest x rays using nesterov accelerated adaptive moment estimation. *Measurement*, 176, 109–153.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84–90.
- Leite, D., Sharma, A., Demir, C., and Ngomo, A.C. (2024a). Interpretability index based on balanced volumes for transparent models and agnostic explainers. In *IEEE International Conference on Fuzzy Systems (WCCI - FUZZ-IEEE), Yokohama, Japan*, 1–10.
- Leite, D. and Skrzjanc, I. (2019). Ensemble of evolving optimal granular experts, owa aggregation, and time series prediction. *Information Sciences*, 504, 95–112.
- Leite, D.F. (2019). *Comparison of Genetic and Incremental Learning Methods for Neural Network-Based Electrical Machine Fault Detection*. In: *Predictive Maintenance in Dynamic Systems (Lughofer, E., Sayed-Mouchaweh, M. (eds))*. Springer, Cham: Predictive Maintenance in Dynamic Systems.
- Leite, D., Costa, P., and Gomide, F. (2013). Evolving granular neural networks from fuzzy data streams. *Neural Networks*, 38, 1–16.
- Leite, D., Silva, A., Casalino, G., Sharma, A., Fortunato, D., and Ngomo, A.C. (2024b). EGNN-C+: Interpretable evolving granular neural network and application in classification of weakly-supervised eeg data streams. In *IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), Madrid, Spain*, 1–10.
- Li, Z. (2023). Identification for red automobile based on cifar10 using machine learning models. *Journal of Physics: Conference Series*, 2428, 1–7.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., v. Ginneken, B., and Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
- Mai, Z., Li, R., Jeong, J., Quispe, D., Kim, H., and Sanner, S. (2022). Online continual learning in image classification: An empirical survey. *Neurocomputing*, 469, 28–51.
- Ran, T., Yuan, L., and Zhang, J. (2021). Scene perception based visual navigation of mobile robot in indoor environment. *ISA Transactions*, 109, 389–400.
- Rezende, V., Costa, M., A.Santos, and de Oliveira, R.C.L. (2020). Aplicação de redes neurais convolucionais na classificação de doenças de plantas. *Congresso Brasileiro de Inteligência Computacional*, 1–7.
- Rodrigues, D.A., dos Santos, G.P., Fernandes, M.C., dos Santos, J.C., Freitas, F.N.C., and Filho, P.P.R. (2017). Classificação automática do tipo de ferro fundido utilizando reconhecimento de padrões em imagens de microscopia. *Revista Matéria*, 22, 1–16.
- Rodrigues, J., Sousa, A., and Santos, A. (2021). Aplicação de redes neurais convolucionais e processamento digital de imagens para classificação do estado dos olhos e avaliação de sonolência. *Revista Brasileira de Computação Aplicada*, 13, 1–10.
- Shaheen, K., Hanif, M.A., Hasan, O., and Shafique, M. (2022). Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. *J Intell Robot Syst*, 105 (9).
- Souza, P.V. and Dragoni, M. (2024). Efnm-nul0 - a trustworthy knowledge extraction about stress identification through evolving fuzzy neural networks. *Fuzzy Sets and Systems*, 487, 109008.
- Varella, C..A. (2008). Análise multivariada aplicada as ciências agrárias. *Pós-graduação em Agronomia. Universidade Federal do Rio de Janeiro*, 1–12.
- Vargas, A.C.G., Paes, A., and Vasconcelos, C.N. (2016). Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. *Proceedings of the XXIX Conference on Graphics, Patterns and Images*, 1–4.
- Verne, T.S., Silva, F.A., Almeida, L.L., Pereira, D., and Artero, A.O. (2022). Detecção e reconhecimento de plantas de pequeno porte utilizando aprendizagem de máquina”. *Colloquium Exactarum*, 14, 36–45.